CMIFed: A Transportable Hypermedia Authoring System

Lynda Hardman, Guido van Rossum, Jack Jansen, Sjoerd Mullender

Multimedia Kernel Systems Group
Department CST
CWI (Centre for Mathematics and Computer Science)
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Email: Lynda.Hardman@cwi.nl, Guido.van.Rossum@cwi.nl, etc.

INTRODUCTION

Creating multimedia presentations can be a complex and time-consuming process. We believe that a key to reducing the authoring burden is the use of structure-based authoring tools [1], [2]. These allow the explicit manipulation of the structure of a presentation rather than the implicit manipulation of this structure via, for example, a time-line. The CMIF authoring system supports the composition of hypermedia presentations from existing media data objects. CMIF (CWI Multimedia Interchange Format) is a system-independent representation for hypermedia presentations; CMIFed (CMIF editor) is the main tool for creating CMIF presentations [3].

CMIF presentations can contain text, images, video and audio. These are referred to as (media) data objects. The data objects are generally kept in separate files, which may use a number of standard formats (e.g. GIF, TIFF and JPEG images are all acceptable). The author can impose structure by placing data objects in a tree. Rough synchronization (presenting data objects in parallel) is obtained by marking a tree node as parallel. Fine synchronization (the specification of precise delays) can be added in the form of specific synchronization constraints.

Interactivity is added to a presentation by creating hyperlinks, which can be attached to stretches of text or areas of images. The author has considerable freedom in the effect of activating a hyperlink — it may replace all or part of the current context or initiate an activity in a different context (a context can be thought of as part of the screen [4]). It is also possible to run external programs or scripts.

CMIF defines (logical) "channels" to capture the essence of resources — a channel represents a particular screen area or another resource such as an audio channel. A media data object must be bound to a channel in order to be displayed or otherwise presented to the end user. The channel specifies default values for a variety of media-specific attributes such as fonts and colors. Thus, when a presentation style is changed in a channel, this change is propagated throughout the whole presentation.

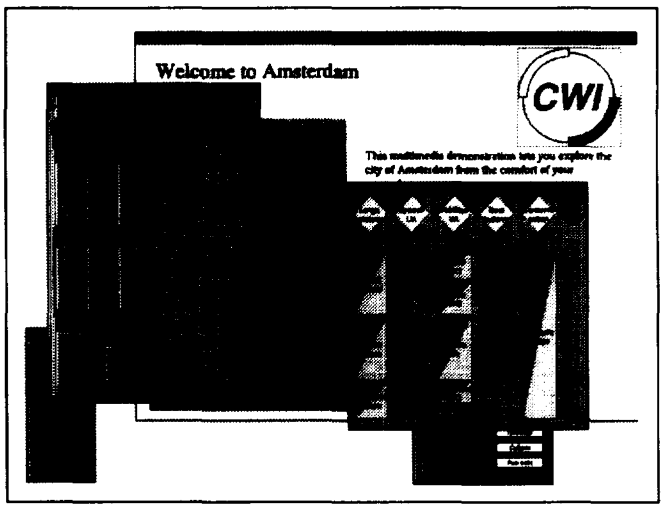
Channels form the key to the transportability of presentations created with CMIFed — essentially, when a presentation has to be mapped to a new environment, only the channels have to be mapped, not all individual media data objects.

In the video we show the basics of structure editing using the two main editing views of CMIFed: the Hierarchy View and the Channel View. The third main view, the Player, is also shown.

THE HIERARCHY VIEW

The Hierarchy View (on the left in the figure below) is the primary authoring view, allowing the author to create multimedia presentations using a top-down or bottom-up approach. The hierarchically structured nodes of the presentation are represented as nested boxes, where children of a node are played either sequentially or in parallel.

Authoring is carried out by creating parallel and sequential structures (composite nodes — displayed in gray) and assigning media data objects as the leaf nodes of this structure (displayed in orange). Although the size of the boxes bears no relation to their duration, time flows from top to bottom — nodes played sequentially are displayed one above the other, while nodes played in parallel are displayed next to each other. Independent contexts (displayed in blue) can be placed anywhere in the tree — their position determines when they are active.



A typical view of the CMIFed environment, showing the hierarchy view and the channel view, with the presentation and control panels in the background.

THE CHANNEL VIEW

The Channel View (bottom right in the figure) shows the logical resource usage of a presentation, including timing relations derived from the structure defined in the Hierarchy View. The media data objects making up the presentation are shown in the Channel View with their precise durations and timing relationships. More complex timing constraints can be specified using synchronization arcs (the arrows in the figure).

A channel enables the author to define high-level presentation characteristics for each media type, so that presentations can be composed without having to specify details for each object: for example, a text channel defines a rectangular area on the screen and a font. Attribute values assigned to a channel can be overridden by individual media data objects.

THE PLAYER

The Player is used to play the finished presentation or to preview it while authoring. (A separate playing program, without editing facilities, is also available.) The author or end user can turn channels on and off, for example allowing the selection of alternative languages.

The Player also allows the author to preview a selection from the Hierarchy or Channel View without having to go through a complete sequence. Finally, the Player can be used to rearrange the lay-out of channels on the screen.

REFERENCES

[1] Structured Multimedia Authoring, by Lynda Hardman, Guido van Rossum and Dick C A Bulterman (ACM Multimedia '93, Anaheim, Aug '93, 283 - 289)

[2] The Amsterdam Hypermedia Model: adding Time and Context to the Dexter Model, by Lynda Hardman, Dick C. A. Bulterman and Guido van Rossum (Communications of the ACM 37 (2), Feb 94, 50-62)

[3] CMIFed: A Presentation Environment for Portable Hypermedia Documents, by Guido van Rossum, Jack Jansen, K.Sjoerd Mullender and Dick C.A. Bulterman (ACM Multimedia '93, Anaheim, Aug '93, 183 - 188)

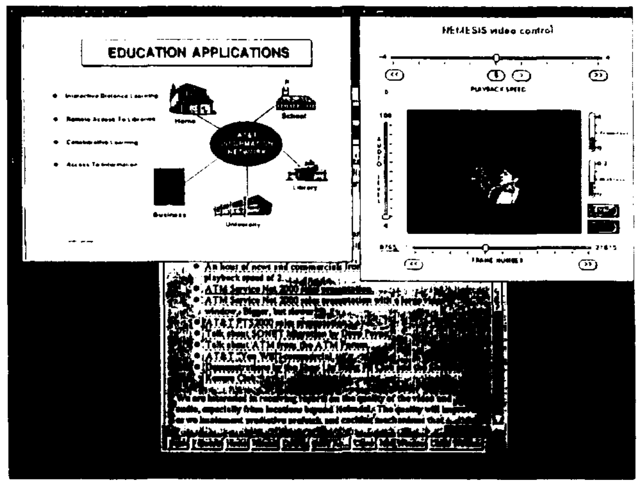
[4] Links in Hypermedia: the Requirement for Context, by Lynda Hardman, Dick C.A. Bulterman and Guido van Rossum (ACM Hypertext '93, Seattle WA, Nov '93, 183-191)

Nemesis: Multimedia Information Delivery

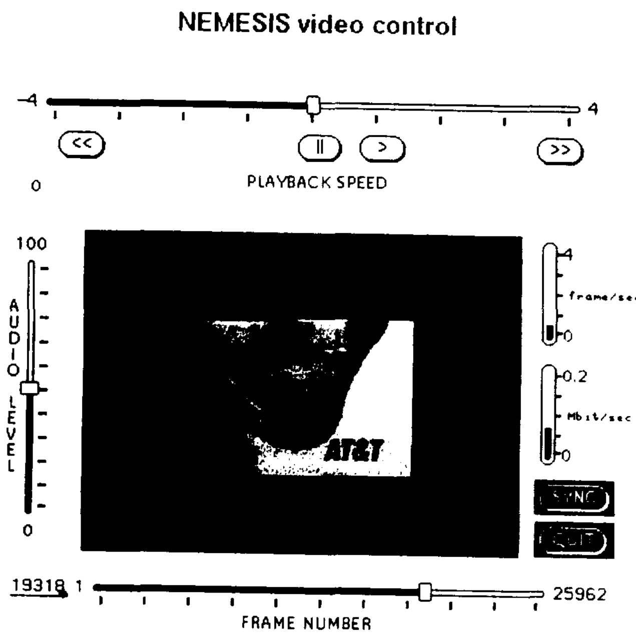
Richard Brandwein
Howard Katseff
Robert Markowitz
Robert Mortenson
Beth Robinson

AT&T Bell Laboratories
Holmdel, NJ 07733

Critical to the success of future multimedia services is the ability to provide fast access to stored information via communications networks. In the *Nemesis* project, we focus on application control protocols for delivering stored multimedia to a user. We are exploring adaptive rate control strategies for networks of variable bandwidths and predictive prefetch of information from remote storage servers as a strategy for coping with short-term network congestion. We also provide "better that being there" features such as variable playback rates with intelligible audio and synchronization and linkage of multiple media.



Our first prototype combines the Nemesis service with the NCSA Mosaic graphical information navigation tool to give access to corporate information in a variety of media, including text, image, sound and multimedia. The multimedia database contains talks and presentations given at AT&T, including accompanying documents and viewgraphs. Because talks are given at specific times and places, people often miss them either due to conflicts or because travel is too time-consuming and expensive. The Nemesis service provides for on-line archival storage and delayed, remote viewing of presentations. A key feature of Nemesis is the integration of other media with audio and video. For example, for auditorium talks Nemesis presents a separate window with the viewgraphs being used by the speaker.



We plan to make the data stored in Nemesis available on a wide variety of platforms ranging from workstations and personal computers lacking decompression hardware to multimedia workstations capable of displaying full-motion video streams. Additionally, the data can be accessed over networks with a range of capabilities such as ISDN, Ethernet, and ATM networks. The first prototype uses Sun SPARCstations for display, software decompression of JPEG-encoded video, and a nationwide corporate TCP/IP network.

ExperMedia/2: A Multimedia Expert System Shell for Domain Experts

Amir Hekmatpour
IBM Manufacturing Technology Center
Knowledge Engineering Services
9Y3/975–1, Zip B84, Essex Jct., VT. 05452
(802)769–9627, Fax: (802)769–8444, E–mail: rima@vnet.ibm.com

ABSTRACT

This video describes ExperMedia/2: an OS/2 multimedia expert system shell for development of intelligent multimedia/hypermedia computer–aided training and certification applications[1]. ExperMedia/2 (EM/2) takes full advantage of modern multimedia workstations. Hypertext, hypergraphics, high resolution images, audio, software motion video, animation, and expert system are technologies that in synergy are far more useful than each by itself. ExperMedia/2 greatly reduces the initial startup cost and allows domain experts (non–programmers) to develop the multimedia applications without much dependency on media production specialists. ExperMedia/2 applications reside entirely on hard disk or on removable, rewritable media (optical storage) and can be used as a stand–alone or in a network configuration.

EM/2's modular architecture gives the designer the flexibility to both add and modify applications as the need arises and the information becomes available. Additional modules for training, documentation, diagnosis and certification can easily be added to the system as the organization expands.

A major departure of EM/2 from previously reported computer–aided training and intelligent tutoring systems is the seamless integration of training, certification, and on–line documentation with the knowledge–based control and management[2]. In addition, EM/2 provides a visual programming environment for the knowledge–based diagnostic development, a WYSIWYG hypermedia editor, and a library of multimedia and hypermedia templates for developing interactive training and on–line documentation.

This video provides a brief overview of some of the capabilities of EM/2. The video includes the following sections:

1) Introduction
2) What is ExperMedia/2 ?
3) Interactive Multimedia Information Presentation
4) Interactive Hypermedia Training
5) Multimedia Knowledge–based Diagnostics
6) Summary

The following is a brief description of the contents of each section in the video.

1) INTRODUCTION

The introduction gives an overview of IBM semiconductor manufacturing plant in vermont, where advanced emerging technologies are used in various stages of manufacturing. EM/2 is used to develop state–of–the–art multimedia and hypermedia applications for operator training, tool diagnostics, process control and on–line documentation.

2) WHAT IS ExperMedia/2 ?

ExperMedia/2 is an OS/2 multimedia expert system shell. Its primary potential applications are tool maintenance and troubleshooting, process control and diagnostics, operator training and certification, as well as on–line documentation. EM/2 provides a seamless integration of training, certification, on–line documentation, and access management with knowledge–based diagnostics.

3) INTERACTIVE MULTIMEDIA INFORMATION PRESENTATION

Managing business information with speed and efficiency is essential to today's productive work environment. From distribution and storage to retrieval and use. EM/2 provides hypertext, hypergraphics, software video, and animation for creating efficient and interactive on–line documents. Using powerful hypertext and hypergraphics, a user can quickly access any portion of a document on a complex tool or manufacturing process with a simple "point–and–click". This section reviews several examples of such on–line documents developed in EM/2 and at use at IBM Vermont manufacturing plant. Hypergraphics is widely used in tool repositories where user can click on a specific component or a part of a tool and get its detailed description. Hypergraphics is also used to navigate inside complex tools and structures.

4) INTERACTIVE AND COOPERATING HYPERMEDIA TRAINING

Compared to traditional classroom training, training applications developed in EM/2 provide effective, consistent and less costly training. This section reviews several on–line training applications developed in EM/2. These applications encourage learning by discovery and take intimidation out of the learning/training process. The control and management modules monitor the progress of the trainee and schedule/recommend additional training modules based on the user area, assignments, and previous training records. Some of the advantages of such training are:

- *Shorter training process*
- *Improved training quality*
- *Consistent training*
- *Controlled training objectives*
- *Reduced trainee intimidation*
- *Active learning environment*
- *Learner–directed training*
- *Integrated certification*
- *Reduced instructor dependency*
- *Easily updated courseware*
- *Training customized to student*

5) MULTIMEDIA KNOWLEDGE–BASED DIAGNOSTICS

Accurately identifying the cause of equipment or process failure is critical to improving productivity and reducing tool down–time in manufacturing. This section reviews several examples of manufacturing tool diagnostic applications. The interactive diagnostic process guides the user to identify and fix equipment problems. Tool diagnostic modules can use still images, slide shows, animation, and motion video to improve the quality of diagnostic advice being delivered. In addition, the diagnostic modules can call upon the training and on–line modules at any time during diagnosis. This becomes necessary during 2ns and 3rd shift hours where the expert technicians and manufacturing engineers may not be readily available. The knowledge–based control and management modules ensure that the user is certified and capable of carrying the diagnostic tasks.

Process control is very similar to tool diagnosis described above. In process control, product progress is monitored. When any deviation from the desired specifications are observed, The operator is notified. Knowledge–based modules can then be consulted to identify the manufacturing process problem and possible corrections.

6) SUMMARY

EM/2 provides a visual programming environment for the knowledge–based diagnostics development, a WYSIWYG hypermedia editor, and a library of multimedia and hypermedia templates for developing interactive training and on–line documentation.

ACKNOWLEDGMENTS

The author wish to thank Jim Wilkinson–Ray of IBM communications for the video production and post–production and Curt Curtis for the narration.

REFERENCES

[1] A. Hekmatpour, E. Millham, L. Grant "ExperMedia/2: An OS/2 Multimedia Expert System Shell for Domain Experts," presented at the Conference on Educational Multimedia and Hypermedia Conference (ED–MEDIA 93), Orlando, June 1993.

[2] A. Hekmatpour, G. Brown, R. Brault, G. Bowen, L. Grant, E. Millham, "FTDD973: A Multimedia Knowledge Based System and Methodology for Operator Training and Diagnostics," Proc. of Conference on Intelligent Computer–Aided Training and Virtual Environment Technology, NASA/Johnson Space Center, ICAT/VET–93, May 1993.

Capturing and Playing Multimedia Events with STREAMS

Gil Cruz and Ralph Hill
Bellcore
445 South Street
Morristown, NJ 07960-6438

STREAMS is a prototype application designed and implemented at Bellcore to support the recording and playback of technical presentations, training sessions, and meetings. During playback STREAMS lets users make choices about how the recorded information is to be presented. To further aid users, STREAMS incorporates powerful searching techniques for locating information in audio and video streams.

STREAMS is based on digital storage of multiple streams. A stream is typically an audio track or a (silent) video track. Streams are stored independently, but share a common dimension — time. This common dimension is used to synchronize the streams during playback, browsing, and searching.

The driving paradigm of STREAMS is to empower the user to make intelligent choices. STREAMS lets the user (implicitly) allocate resources by controlling playback. For example, the user can turn off playback of a video stream, increase the quality of the audio, or decrease the frame size of another video stream. STREAMS also empowers the user by providing searching mechanisms designed to enable rapid browsing so users can quickly identify points of interest.

While our primary goal with STREAMS is to improve the capture and playback of events such as meetings and lectures, STREAMS can be viewed as a multimedia authoring tool. Multimedia authoring is currently an expensive task. It is only cost effective to author multimedia products that have a large market. It may be practical, however, to use STREAMS to prepare multimedia instructional material for narrow markets. An educator could capture a lecture and annotate it with prepared text and graphics. While this approach might not produce material that is as rich and compelling as a carefully authored multimedia presentation, it can rapidly generate large libraries of instructional multimedia presentations.

The video tape shows an auditorium at Bellcore that is one of the environments in which STREAMS events are captured. The auditorium has been outfitted with traditional

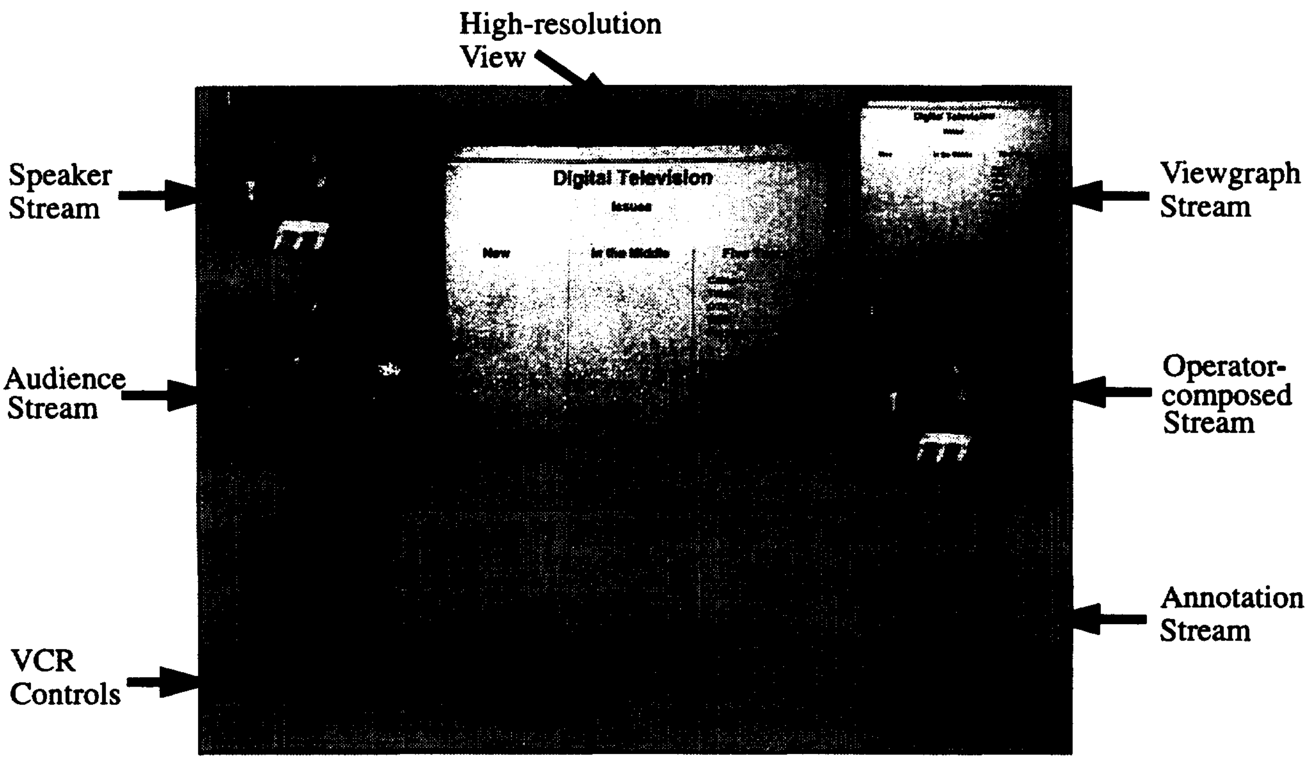


Figure 1. Sample interface for STREAMS prototype.

media capture equipment (i.e., video cameras and microphones), in order to capture multiple streams for each event. The video briefly describes the types of processing that are done to take the media from the original analog form to the digital form used by STREAMS. This includes media specific compression such as H.261 for motion video and source specific processing such as automatically removing redundant frames from video of the view graphs.

The video goes on to show how STREAMS can be used to play back a recorded event. STREAMS supports VCR-like traversal methods such as play, pause, and rewind, as well as random jumps. STREAMS goes beyond these traditional searching and accessing methods to provide media and source specific searching tools that allow users to rapidly locate sections of interest and then play back these sections at full fidelity. The searching tools are optimized to the type of media being searched (e.g., video, audio, text annotation) and to the source (e.g., video of speaker vs. video of view graphs).

Figure 1 shows the playback user interface for STREAMS as shown in the video tape. There are four visual preview streams. The user can select one of these streams for presentation at a higher resolution and frame rate. In addition, a stream showing audio and text comments inserted by other users who have previously viewed this record is displayed.

Media Streams: Representing Video for Retrieval and Repurposing

Marc Davis
Interval Research Corporation
1801-C Page Mill Road
Palo Alto, CA 94304
davis@interval.com

In order to enable the search and retrieval of video from large archives, we need a representation of video content. For the types of applications that will be developed in the near future (interactive television, personalized news, video on demand, etc.) these archives will remain a largely untapped resource, unless we are able to access their contents. Given the current state of the art in machine vision and image processing, we cannot now, and probably will not be able to for a long time, have machines "watch" and understand the content of digital video archives for us. Unlike text, for which we have developed sophisticated parsing technologies, and which is accessible to processing in various structured forms (ASCII, RTF, PostScript), video is still largely opaque. We are currently able to automatically analyze scene breaks, pauses in the audio, and camera pans and zooms, yet this information alone does not enable the creation of a sufficiently detailed representation of video content to support content-based retrieval and repurposing. In the near term, it is computer-supported human annotation that will enable video to become a rich, structured data type.

Over the past three years, members of the MIT Media Laboratory's Machine Understanding Group in the Learning and Common Sense Section (Marc Davis with the assistance of Brian Williams and Golan Levin under the direction of Prof. Kenneth Haase) have been building a prototype for the annotation and retrieval of video data. This system is called **Media Streams**. Media Streams is written in Macintosh Common Lisp and FRAMER, a persistent framework for media annotation and description that supports cross-platform knowledge representation and database functionality. Media Streams runs on an Apple Macintosh Quadra 950 with three high resolution, accelerated 24-bit color displays and uses Apple's QuickTime digital video format. With Media Streams, users create stream-based, temporally-indexed, iconic annotations of video content which enable content-based retrieval of annotated video sequences.

The system has three main interface components: the Director's Workshop (Figure 1); Icon Palettes (Figure 2); and Media Time Lines (Figure 3). The process of annotating video in Media Streams using these components involves a few simple steps:

In the **Director's Workshop**, the user creates iconic descriptors by cascading down hierarchies of icons in order to select or compound iconic primitives

As the user creates iconic descriptors, they accumulate on one or more **Icon Palettes.** This process effectively groups related iconic descriptors. The user builds up Icon Palettes for various types of default scenes in which iconic descriptors are likely to co-occur; for example, an Icon Palette for "treaty signings" would contain icons for certain dignitaries, a treaty, journalists, the action of writing, a stateroom, etc.

By dragging iconic descriptors from Icon Palettes and dropping them onto a **Media Time Line,** the user annotates the temporal

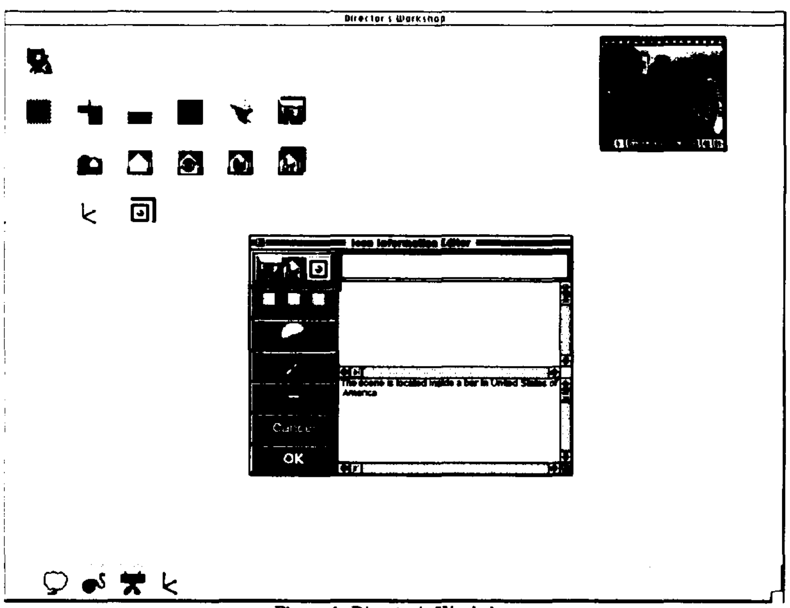


Figure 1: Director's Workshop

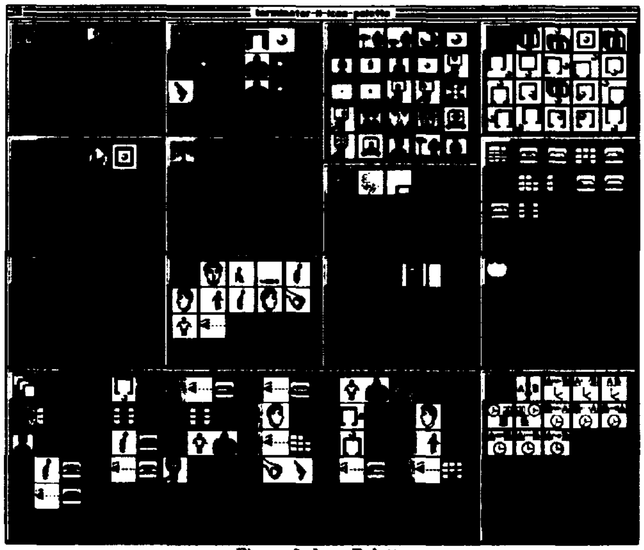


Figure 2: Icon Palette

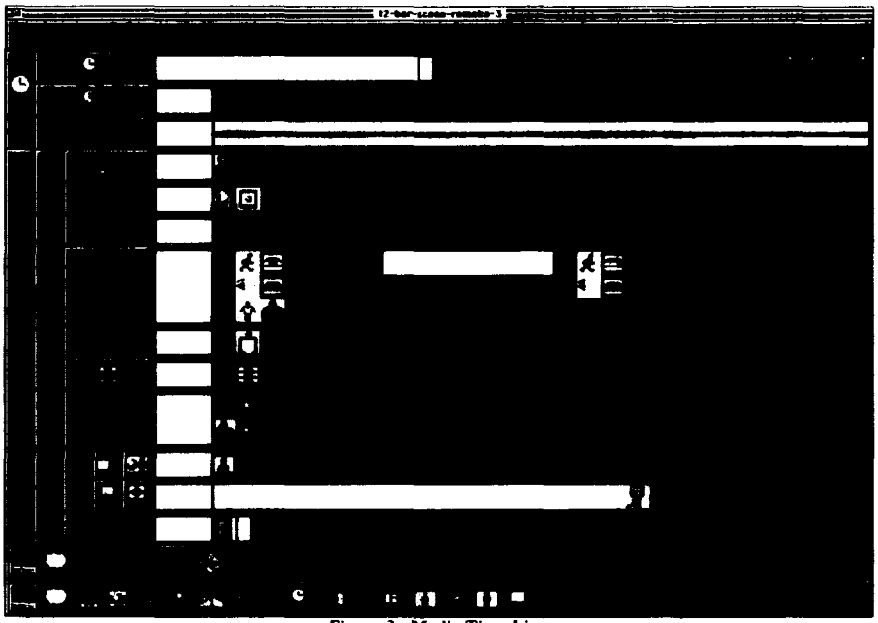


Figure 3 : Media Time Line

media represented in the Media Time Line. Once dropped onto a Media Time Line, an iconic description extends from its insertion point in the video stream to either a scene break or the end of the video stream. The user then ends the iconic description at the point in the video stream at which it no longer applies.

In addition to dropping individual icons onto the Media Time Line, the user can construct compound icon sentences by dropping certain "glommable" icons onto the Media Time Line, which, when completed, are then added to the relevant Icon Palette and may themselves be used as primitives. By annotating various aspects of the video stream (time, space, characters, character actions, camera motions, etc.), the user constructs a multi-layered, temporally indexed representation of video content.

The Media Time Line is the core browser and viewer of Media Streams. It enables users to visualize video at multiple timescales simultaneously, to read and write multi-layered iconic annotations, and provides one consistent interface for annotation, browsing, query, and editing of video and audio data.

With Media Streams, users can create shareable representations of media content which enable the construction of large archives of reusable temporal media. Without tools like Media Streams, a thousand hours of video content will be less useful than one. With tools like Media Streams, we move closer to the day when all the digital video in the world can become an accessible and reusable resource for human and computational imagination.

REFERENCES

Davis, Marc. "Media Streams: An Iconic Visual Language for Video Annotation." Telektronikk 4.93 (1993): 59-71.

Davis, Marc. "Knowledge Representation for Video." Forthcoming in: Proceedings of the 1994 National Conference on Artificial Intelligence in Seattle, Washington, 1994.

INFORMEDIA DIGITAL VIDEO LIBRARY

Michael Christel, Scott Stevens, and Howard Wactlar

Carnegie Mellon University
Pittsburgh, PA 15213
mac@sei.cmu.edu

Vast digital libraries of information will soon be available on the nation's Information Superhighway as a result of emerging technologies for multimedia computing. These libraries will profoundly impact the conduct of business, professional, and personal activity. However, it is not enough to simply store and play back video (as in currently envisioned commercial video-on-demand services); to be most effective, new technology is needed for searching through these vast data collections and retrieving the most relevant selections.

The Informedia Project is developing these new technologies for data storage, search, and retrieval, and in collaboration with QED Communications is embedding them in a video library system for use in education, training, and entertainment. The Informedia Project leverages efforts from many Carnegie Mellon University computing research activities, including:

- Sphinx-II speech recognition
- Image Understanding Systems Laboratory
- Center for Machine Translation (information retrieval)
- Software Engineering Institute (information modeling)

The Informedia Project is developing intelligent, automatic mechanisms that provide full-content search and retrieval from digital video, audio, and text libraries. The project integrates speech, image, and language understanding for the creation and exploration of such libraries. The initial library will be built using WQED's video assets.

LIBRARY CREATION

The Informedia system uses Sphinx-II to transcribe narratives and dialogues automatically. Sphinx-II is a large vocabulary, speaker-independent, continuous speech recognizer developed at Carnegie Mellon. With recent advances in acoustic and language modeling, it has achieved a 95% success rate on standardized tests for a 5000-word, general dictation task. By relaxing time constraints and allowing transcripts to be generated off-line, Sphinx-II will be adapted to handle the video library domain's larger vocabulary and diverse audio sources without severely degrading recognition rates.

In addition to annotating the video library with text transcripts, the videos will be segmented into smaller subsets for faster access and retrieval of relevant information. Some of this segmentation is possible via the time-based transcript generated from the audio information. The work at CMU's Image Understanding Systems Laboratory focuses on segmenting video clips via visual content. Rather than manually reviewing a file frame-by-frame around an index entry point, machine vision methods that interpret image sequences can be used to automatically locate beginning and end points for a scene or conversation. This segmentation process can be improved through the use of contextual information supplied by the transcript and language understanding. Figure 1 gives an overview of the InforMedia system.

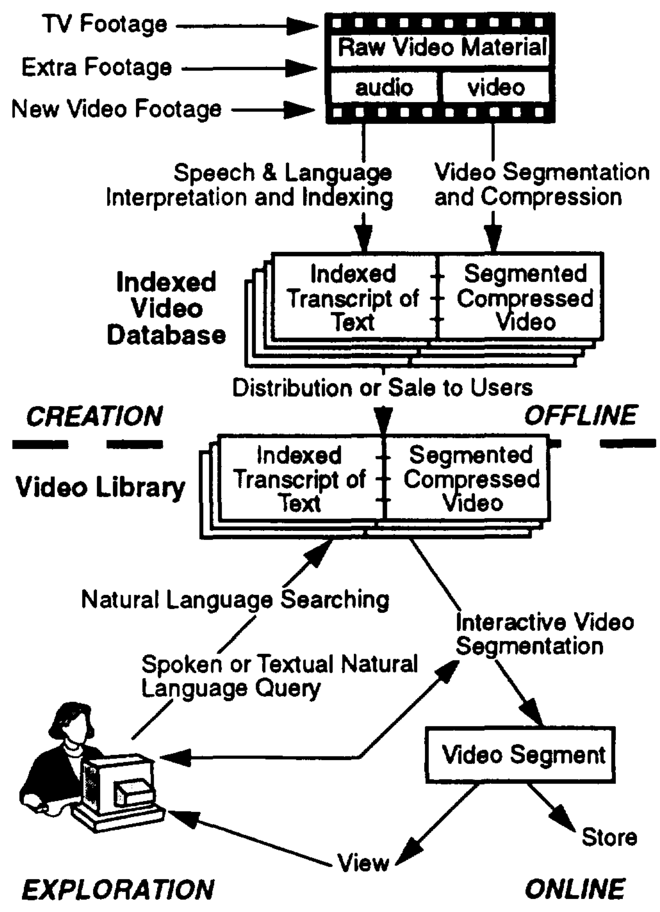


Figure 1. Informedia™ System Overview

LIBRARY EXPLORATION

Finding desired items in a large information base poses a major challenge. The Informedia Project goes beyond simply searching the transcript text and will, in addition, apply natural-language understanding for knowledge-based search and retrieval. One strategy employs computational linguistic techniques from the Center for Machine Translation for indexing, browsing, and retrieving based on identification of noun phrases in a written document. Other techniques from the Center include statistical weighting, term selection heuristics, and natural-language processing. More complex than individual words, these linguistic units provide a conceptually richer domain for subsequent processing.

The Informedia system is extending this technology for spoken language and applying it to correct and index the automatically-transcribed soundtracks. Other tasks will include identification of topics and subtopics in transcript collections, and a rich natural language retrieval interface. A second thrust is developing robust techniques for matching transcribed words and phrases that sound alike when spoken. This integrated approach will significantly increase the Informedia system's ability to locate a particular video segment quickly, despite transcription errors, inadequate keywords, and ambiguous sounds.

Along with improving query capabilities, the Informedia Project is researching better ways to present information from a given video library. Interface issues include helping the user identify desired video when multiple objects are returned, adjusting the length of video objects returned, and letting the user quickly skim video objects to locate sections of interest.

Cinematic knowledge can enhance the composition and reuse of materials from the video library. For example, the library may contain hours of interview footage with experts in a certain topic area. Rather than simply presenting a series of disassociated video windows in response to user queries, this interview footage could be leveraged to produce an interface in which the user becomes the interviewer. The natural language techniques mentioned above are used to parse the user's questions, and scenes from the interview footage are composed dynamically to present relevant answers. Such an interface is designed to engage the user into more fully exploring and interacting with the video library in search of information as an active interviewer.

LIBRARY DEMONSTRATION

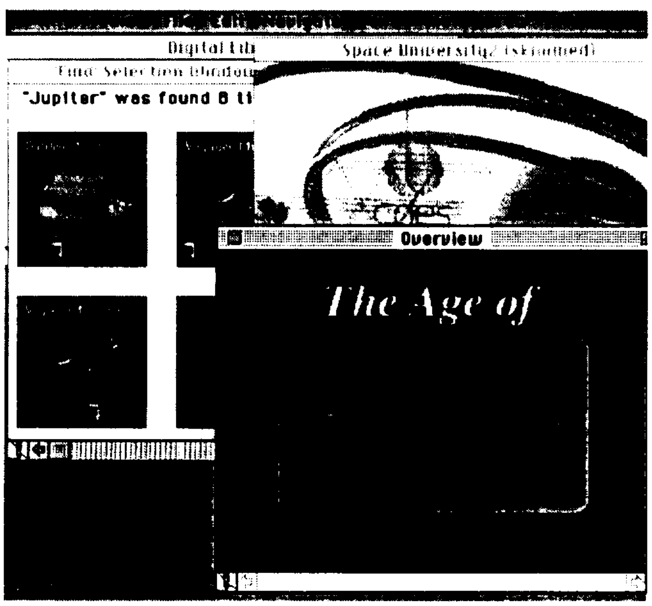


Figure 2. Screen Dump of Informedia™ Demonstration

The Informedia Project currently has a demonstration based on a small database (one gigabyte) of text, graphics, video, and audio material drawn from WQED's "Space Age" series. A sample display of this demonstration appears as Figure 2 for the reader's reference. The demonstration was carefully scripted to illustrate the following points (listed in temporal order, as they occur in the demonstration):

- parsing the user's input according to an appropriate grammar for that domain allows for more natural, less cumbersome queries
- natural language understanding of both a user's query and the video library transcripts enables the efficient retrieval of relevant information
- the location *within* a video object is identified relevant to a user query via the text transcript
- the video "paragraph", or size of the video object, is determinable based on language understanding of the transcript and image understanding of the video contents
- a larger video object can be "skimmed" in an order of magnitude less time, while coherently presenting all of the important information of the original object
- video clips can be reused in different ways, e.g., to create an interactive simulated interview, as shown in Figure 3

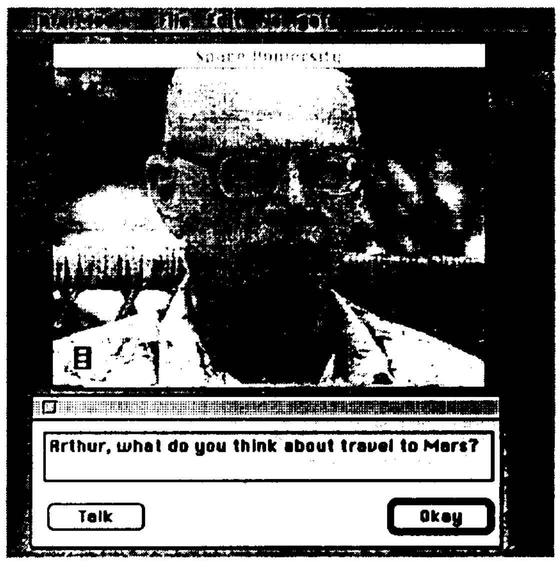


Figure 3. Simulated Interview with Arthur C. Clarke

The demonstration does not show Sphinx-II in action performing automatic transcription nor does it document the process of video segmentation or natural language parsing. It shows the benefits of such automatic indexing and segmentation, illustrating the accurate search and selective retrieval of audio and video materials appropriate to users' needs and desires. It shows how users can preview as well as scan video at variable rates of speed and presentation, akin to skimming written material. Finally, it demonstrates the concept of combining speech, language, and image understanding technologies to create entertaining educational experiences.

The MICE Project
Multimedia Integrated Conferencing for Europe (MICE)

The MICE consortium
Contact: Claus-Dieter Schulz
Computer Center University of Stuttgart
Allmandring 30, D-70550 Stuttgart, Germany

1. ABOUT THE MICE PROJECT

In 1992, the CEC agreed to finance a one-year piloting project called Multimedia Integrated Conferencing for Europe (MICE). The aim of the project was to pilot technology to improve inter working between European researchers. The technology base was to be a heterogeneous (open) hardware platform, and use existing software tools and the emerging 2 Mb European network infrastructure as far as possible. Another goal was to demonstrate interworking between European research sites and the US. Integration means that MICE technology allows multimedia conferencing - i.e. the use of audio, video and shared workspace tools - between conference rooms and workstation-based facilities, hardware and software codecs, packet-switched networks and ISDN, and uses both uni- and multicast technology. A detailed rationale for selecting these technologies, and the technical effort required to integrate them, is given in [Kirstein],[Handley].

The MICE partners set up an International Research Seminar Series between some of the partners to give organizers, technical support personnel and users the opportunity of evaluating the technology in regular service [Bilting]. Since the overall aim of the project was to further remote cooperation between researchers, a series of International Research Seminars constituted a "real-world task" for evaluation. Two of the MICE partners, University College London (UCL) and the Swedish Institute of Computer Science/ Royal Institute of Technology (SICS/KTH) in Stockholm organized a joint series of 11 weekly seminars on Multimedia, Communications and Networks, Distributed Systems and CSCW, starting in October 1993. The seminars were attended by researchers and students in the conference rooms in Darmstadt (Germany), London (UK), Stockholm (Sweden), and Oslo (Norway). Other MICE partners in Germany and France followed the seminars from their conference rooms or workstations, as did up to 25 individual researchers from Europe, the US and Australia.

1. THE VIDEO

1.1 Overview

The video is structured in three main parts. It shows in a first part our videoconference system as presented at the Interop '93 in Paris. As a second part it shows an example of a Multimedia Conference meeting on remote language teaching held between MICE partners. The third part is a seminar given by Steve Deering (Xerox PARC) during his stay at KTH (Stockholm, Sweden), which was part of a series of seminars broadcasted by the MICE project over the Internet.

1.1 MICE project demonstration at Interop'93, Paris

This part starts with some scenes from the conference site in Paris which lead to the stand of the German Telecom where our demonstration took place. The following parts of this section then show the view of the conference system as shown on the workstation an presented to the visitors.
On the screen you see 4 video streams presented in a quad-image. The lower right is the local view at Interop. Top right is coming from an office at UCL London. The upper left is send from INRIA (France). All of these are send to the Interop over the Internet. The lower left is a video input from an ISDN videophone from the BT (British Telecom) Helpdesk. Additionally visible on the screen are the other tools, like audio tool and the shared whiteboard. Video, audio and shared whiteboard are broadcasted during the conference to partners all over the world over the Internet Mbone (Multicast Backbone)[Casner] Infrastructure.

1.2 Example of a weekly MICE meeting

This section shows part of a MICE project meeting on network infrastructure planning for a remote language teaching project. This is an example of how multiparty conferencing is used within the MICE project. In additions to meetings on special topics like the one shown here, the MICE project does it's weekly project meetings using the same technology.
The partners participating this meeting were Mark Handley (UCL, London, UK), Ronny Nilsen (Univ. of Oslo, Norway) and Christian Wettergren (KTH, Stockholm, Sweden).
This parts shows several views from the different sites during the meeting. The partners use VAT (audio), IVS (video) and WB (shared whiteboard) during their discussions.
A map of Europe was preloaded in the whiteboard. The partners used this tool to add (draw) additional parts (like network links) during their discussion. This meeting was multicasted between the MICE partners over the Internet (Mbone).

1.3 Example of a weekly MICE seminar

This section shows part of a seminar given at the 'Informal Workshop on Multicast Conferencing' at KTH Stockholm, Sweden, by Steve Deering (Xerox PARC) on "State of the Art and Research Issues" as part of the MICE International Research Seminar Series. What you see is the view of this seminar at a remote workstation at the Computer Centre University of Stuttgart (RUS), Germany, showing how this seminar was received at this site. The shared whiteboard is used to present the slides during the talk. In addition the speaker uses this tool to explain and emphasize parts of his talk. This seminar was broadcasted worldwide over the Internet (Mbone).

2. Multimedia Conferencing tools used by the MICE project

Audio: VAT (Visual audio tool) written by Van Jacobson & Steve McCanne (Lawrence Berkeley Laboratory (LBL),USA)
Video: IVS (INRIA Videoconferencing Software) developed at INRA, France
Shared whiteboard: WB (WhiteBoard) Written by Van Jacobson, Steve McCanne (Lawrence Berkeley Laboratory (LBL),USA)

3. REFERENCES

[Bilting] U. Bilting et. al. "Remote Seminars through Multinedia Conferencing: Expereiences from the MICE project", Proceedings of the JENC/INET94
[Casner] S. Casner, Frequently Asked Questions (FAQ) on the Multicast Backbone (MBONE)., available by anonymous ftp from venera.isi.edu in the mbone/faq.txt, May 6th 93.
[Handley] M. J. Handley, P. T. Kirstein & M. A. Sasse, "Multimedia Integrated Conferencing for European Researchers (MICE): piloting activities and the Conference Management and Multiplexing Centre", Computer Networks and ISDN Systems, 26, 275-290, 1993.
[Kirstein] P. T. Kirstein, M. J. Handley, M. A. Sasse Piloting of Multimedia Integrated Communications for European Researchers (MICE) Proc. INET.

4. ACKNOWLEDGEMENTS

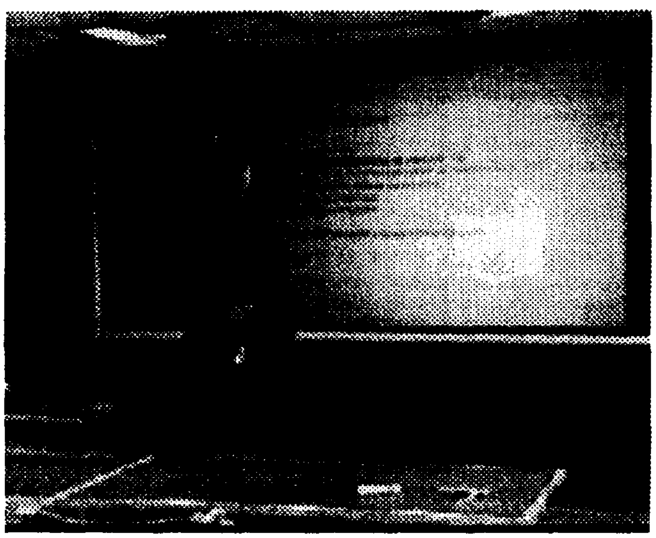
The author would like to thank all MICE partners who helped a lot during the making of this video in several ways, especialy the members of the MICE team who helped producing the MICE meeting session Mark Handley (UCL), Ronny Nilson (UiO), Christian Wettergren (KTK), Hans Eriksson (SICS). and last but not least our project manager Angela Sasse who encouraged us to make this video.
Jutta Sauer and Uwe Zimmat from the RUS videolab for their help during the postprocessing of the video and the additional timeslots for finishing the work in time.
Special thanks are due to Steve Deering of Xerox PARC who let us use a part of his seminar talk and to Van Jacobson of Lawrence Berkeley Labs, the creator VAT and WB, for making those tools available to MICE and quickly dealing with problems whenever we encountered them.

Combining Realtime Multimedia Conferencing with Hypertext Archives in Distance Education

Per Einar Dybvik, Håkon W Lie
Norwegian Telecom Research

Video conferencing systems are currently being used to distribute lectures over geographical distances. Through 2-way video and audio links, students can actively participate in classes without being physically co-located with the lecturer.

While this use of video conferencing systems may reduce travelling and make more courses available to students, the video link only conveys a limited part of the communication that naturally takes place during a course. Handouts, copies of tranparencies and high-quality images are examples of data that are not easily transferable over a video link.



Norwegian Telecom has established a high-capacity digital network for experimental applications. The bandwidth of the Supernet (33Mbit/s) allows for establishing real time video conferences over a packet-switched network using Internet protocols. This duplicates the functionality of a traditional video conferencing system while allowing us to integrate common Internet applications into the lectures.

A course at the University of Oslo has been using videoconferencing systems to make the lectures available to a larger group of students. Two electronic classrooms, one located at the main campus and the other at a satellite location has been linked over the Supernet allowing for video and audio transmission. Each classroom contains cameras and monitors, in addition, the rooms are equipped with large electronic white boards where the lecturer can present and interact with information.

Also, students connected to Internet are able to participate in the course. Through freely available, but lower quality videoconferencing software, students in Trondheim, Tromsø and Stockholm can follow lectures. Unfortunately, this has been a one-way link not allowing student interaction.

All students, wether in the electronic classrooms or at their machines, can follow the whiteboard presentation in a window. The shared whiteboard is based on the World Wide Web system, which has established itself as the hypertext system of choice on the net. An increasing number of information servers around the world allow users to access information through client programs.

The hypertext markup language, which is the native data format of the web, allows for simple, but effective, presentation of textual information. Also, client programs may support other data types such as images and sound. The whiteboard, which is a modified version of the Mosaic client software, is rendered at each participating site. As the lecturer moves forward in the presentation, a multicast protocol instructs each client to fetch the new transparency from a common server. Compared to sending a video image of the transparency, the presentation quality is significantly higher, while utilizing the power of distributed processing and high-resolution computer screens.

The shared whiteboard window has proved to be valuable for all participating students -- probably more so than the image of the lecturer or the participants. Still, audio seems to be most important.

By, putting the presentations into the web, transparencies are available to the students for review after the lectures, thereby replacing paper copies.

However, making the lecture transparencies electronically available has raised some questions with regard to the structuring of information. The sequential nature of a lecture and the corresponding transparencies does not exploit the capabilities of hyper-linked structures -- which users expect to find on web-servers.

The World Wide Web represents an information architecture that can be a basis for distributed groupware applications. By combining a hypertext system with real-time multimedia communication, we are seeing the contours of a rich, distributed groupware environment where distance education will thrive.

ACKNOWLEDGEMENTS

The classrooms and the technology are developed in a collaborative project between Norwegian Telecom Research (NTR), University of Oslo (USIT) and Center for Technology at Kjeller (UNIK). We would like to thank Bjørn Hestnes and Hjalmar Martinsen from NTR, Geir Pedersen and Ronny Nilsen from USIT and Rune Fløisbonn and Morten Sørdal from UNIK for helping out and letting us use the classrooms. We would also thank Tim Berners-Lee and his team at CERN for giving us World Wide Web.

REFERENCES:

D. Solvoll, et al: *Information Exchange in MultiTorg*, Telektronikk 4/93, Norwegian Telecom 1993

I. Hovig, H.W. Lie, Teleteaching in a Graduate Seminar: *Practical Experiences and a Look Ahead*, Proceedings of the IFIP TC3 International Conference: Teleteaching 93, Trondheim, Norway, 1993

T.J. Berners-Lee, R. Cailliau and J.F. Groff, *The World-Wide Web*, Computer Networks and ISDN Systems 25 (1992) 454-459. North-Holland. Presented at Joint European Networking Conference , Innsbruck, Austria, 1992.

C. Egido, *Videoconferencing as a Technology to support Group Work: A Review of its failure,* Proceedings of CSCW'88, pp. 13-24, Portland, Oregon, 1988

K. Bringsrud, G. Pedersen, *The MUNIN Project: Distributed Electronic Classrooms with Large Electronic Whiteboards,* Proceedings of the IFIP TC3 International Conference: Teleteaching 93, Trondheim, Norway, 1993

C. A. Ellis, S. J. Gibbs, G. L. Rein, *Groupware — some issues and experiences,* Communications of the ACM 32, 39-57, 1991

B. Olsen, *Third Generation Distance Education*, Proceedings of INET'92, pp. 359-367, Kobe, Japan, 1992

An Argo Telecollaboration Session

Hania Gajewska, Jay Kistler, Mark S. Manasse, and David D. Redell

Systems Research Center
Digital Equipment Corporation
130 Lytton Avenue
Palo Alto, CA 94301

ABSTRACT

The goal of the Argo system[1] is to allow medium-sized groups of users to collaborate remotely from their desktops in a way that approaches as closely as possible the effectiveness of face-to-face meetings. In support of this goal, Argo combines high quality multi-party digital video and full-duplex audio with telepointers, shared applications, and whiteboards in a uniform and familiar environment. The shared applications can be unmodified X programs shared via a proxy server, unmodified groupware applications, and applications written using our toolkit. Workers can contact each other as easily as making a phone call, and can easily bring into a conference any material they are working on. They do so by interacting with an object-oriented, client/server conference control system. The same conference control system is used to support teleporting, i.e. moving the desktop environment from one workstation's display to another (for example, from office to home).

This video tape provides a user's point of view for a single session with the Argo system. This session shows a variety of applications in use, displaying all of the different sharing mechanisms.

The tape starts from Hania's point of view, as she places a call to Rob. They share a Trestle-based user interface editor, and then add Mark to the conversation. Mark brings an X-based editor into the conference, and discusses the content of the editor buffer with Rob and Hania, using his telepointer to indicate points of interest. They get interrupted by a call from Dave, and Hania puts Rob and Mark on hold to talk to Dave. When she is done with Dave, Hania returns to the conference with Rob and Mark, and then they all join the conference that holds their weekly project meeting. In this conference, they use a shared whiteboard to discuss the meeting agenda. They also use the whiteboard to view and annotate a PostScript file.

The tape was shot off the screen of a DEC Alpha/AXP-based workstation; the participants are on Alphas or on DEC-manufactured MIPS R3000-based workstations.

The tape was shot and produced by Ken Beckman of the Systems Research Center, and features our 1994 summer intern, Rob DeLine, from Carnegie-Mellon University.

REFERENCES

1. H. Gajewska, J. Kistler, M. Manasse, and D. Redell. Argo: A System for Distributed Collaboration. In Proceedings of Multimedia '94, San Francisco, CA. October 1994.

ADAPTIVE, BEST-EFFORT DELIVERY OF LIVE AUDIO AND VIDEO ACROSS PACKET-SWITCHED NETWORKS

Video Abstract

Kevin Jeffay *Donald L. Stone*
University of North Carolina at Chapel Hill
Department of Computer Science
Chapel Hill, NC 27599-3175
{jeffay,stone}@cs.unc.edu

INTRODUCTION

This videotape is a demonstration of a transport protocol developed by the authors for the transmission of live audio and video streams. The goal of this work has been to understand the complexity of supporting applications such as desktop video conferencing when the network does not support real-time communication. We believe this problem is important because such networks will likely exist for the foreseeable future, hence the problems addressed by this work are fundamental in delivering continuous media in real-time across the "last mile" to the desktop.

Our protocol is a "best effort" protocol that attempts to ameliorate the effect of three basic phenomena: jitter, congestion, and packet loss, to provide low latency, synchronized audio and video communications [3]. This goal is realized through four transport and display mechanisms, and a real-time implementation of these mechanisms that integrates operating system services (*e.g.*, scheduling and resource allocation, and device management) with network communication services (*e.g.*, transport protocols), and with application code (*e.g.*, display routines). The four mechanisms are: a facility for varying synchronization between audio and video to achieve continuous audio in the face of jitter, a network congestion monitoring mechanism that is used to control media latency, a queueing mechanism at the sender that is used to maximize throughput with out unnecessarily increasing latency, and a forward error correction mechanism for transmitting audio frames multiple times to ameliorate the effects of packet loss in the network.

A key difficulty in evaluating our work has been the lack of metrics for comparing two given media transmission and playout scenarios. For example, performance measures such as end-to-end latency, frame transmission and playout rates, gap-rates, intermedia synchronization differential, *etc.*, are relatively easy to compute, but difficult to relate. For example, if scheme A results in lower end-to-end latency than scheme B, but B provides a lower gap-rate than A, which has performed better?

We do not provide any answers to this dilemma. Instead, we simply demonstrate, through the use of our protocol, the qualitative effects of varying and trading off performance parameters such as lip synchronization and gap-rate.

This videotape attempts to (1) demonstrate the quality of the audio/video streams delivered via our protocol on congested networks, and (2) give viewers a qualitative feel for the effects of varying various so-called quality-of-service parameters such as number of discontinuities (*e.g.*, gap rate), end-to-end latency, lip sync, and throughput.

DESCRIPTION OF VIDEOTAPE

Three demonstrations of transmitting digital audio and video across interconnected local-area networks are presented. The first illustrates the latency inherent in our video conferencing system. End-to-end latency is one of the most important performance parameters for a videoconferencing system as latency can severely impair and impede interaction between conference participants [2, 8]. At present there is some agreement that an end-to-end latency of no more than 250 ms. is acceptable [1]. In the best case, our system is capable of delivering synchronized audio and video streams with an end-to-end latency of approximately 170 ms. In the first demonstration we illustrate the effect of this latency by comparing our system with an analog conferencing system (with no latency). We show a split screen with analog video in one half and digital video in the other half (Figure 1). The digital video is shown after having been acquired by a workstation, compressed, transmitted over an idle network, received by a second workstation, decompressed, and dis-

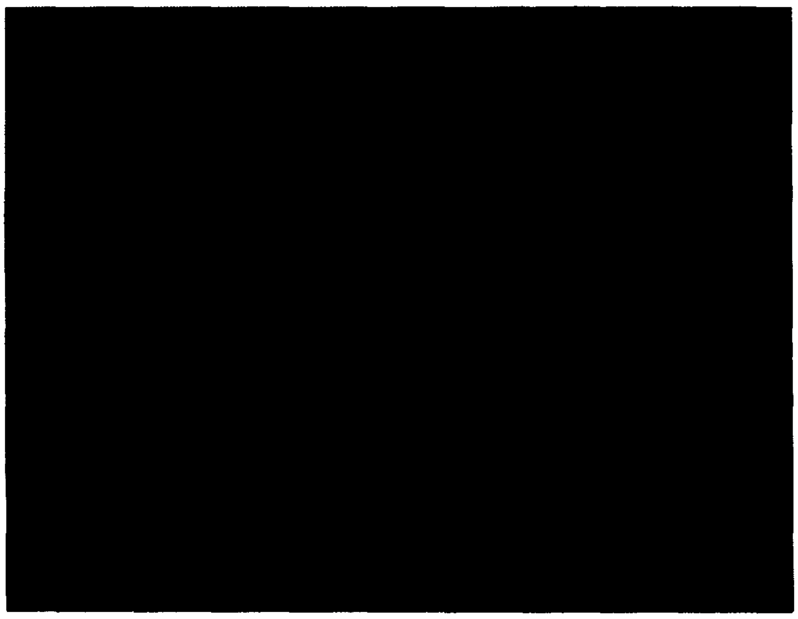


Figure 1: Demonstration of latency differential between analog (upper left) and digital (lower right) systems.

played. It takes approximately 170 ms. for a video frame to propagate from the camera to the display [3].

The second demonstration illustrates the effect of varying the synchronization between the audio and video streams. As described in [3], a useful technique for ameliorating the effect of network congestion is to purposely play audio and video out of exact synchronization; specifically, to play audio frames ahead (in time) of their corresponding video frames. Although this technique has proved effective in improving quantitative measures of video conference performance, playing audio "ahead" of video is unnatural. In nature the speed of sound is several orders of magnitude slower than the speed of light and hence whenever we view noise-emitting scenes from a distance, we perceive the visual information before the corresponding sonic information. Humans are therefore more tolerant of audio "behind" video. Our system assumes (somewhat arbitrarily although motivated by [7]) that users will tolerate a synchronization differential of at least 100 ms.

The second demonstrations varies the degree to which audio is played ahead of video while a person is speaking and while a person claps (Figure 2). This illustrates that while humans are, in general, relatively intolerant of audio ahead of video, our ability to perceive this to be the case depends on such (arbitrary) factors as the resolution of the image and composition of the scene. For example, it is much easier to discern the difference in synchronization in the clapping experiment than in the speaking experiment.

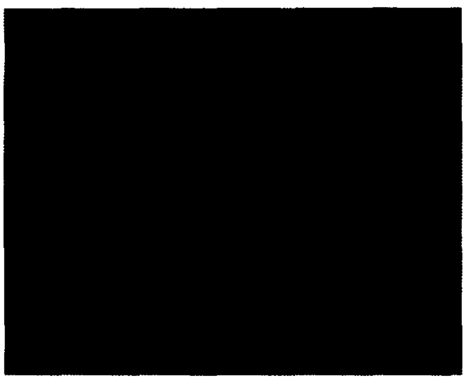


Figure 2: Demonstration of the effect of varying audio/video synchronization by clapping.

Finally, we demonstrate the effect of transmitting audio and video via our protocol and compare the protocol's performance to UDP. We present a set of controlled experiments wherein media is transmitted over a small internetwork while varying degrees of traffic are introduced into the network. In the first case UDP is used for transport. The video steam is jerky (because of loss) and audio has numerous gaps (because of jitter and loss). Next, our protocol is used for transport. In this case video is marginally better but audio is perfect (although in the case of video, the jerkiness is now due to fact that frames were never transmitted because the protocol is trying to avoid wasting network resources). A quantitative comparison of a similar set of experiments can be found in [3].

TECHNICAL DETAILS

The workstations used in this videotape are IBM PS/2 (20 Mhz x386 processor) personal computers using IBM/Intel ActionMedia I audio/video adapters. We use an experimental real-time operating system kernel and video conferencing application we have developed. The kernel is described in [4, 6]; the application in [4]. The adaptations used in the protocol for managing media streams are described in [3]. A more detailed description and analysis of the delay jitter management scheme used in this work is presented in [5].

The conferencing system generates 60 audio frames and 30 video frames per second. An average video frame is approximately 8000 bytes; an audio frame is approximately 250 bytes. This yields an aggregate data stream of approximately 2 Mb/s.

The network used in these experiments is a building-sized internetwork consisting of several 10 Mb Ethernets and 16 Mb token rings interconnect by bridges and routers. It supports approximately 400 UNIX workstations and Macintosh personal computers. The workstations share a common file system using a mix of NFS and AFS. The application mix running on these workstations should be typical of most academic computer science departments.

ACKNOWLEDGEMENTS

Terry Talley and Ta-Ming Chen helped construct the network and traffic generators used in the demonstrations in this video. David Harrison and Elliot Poger recorded and mastered the first versions of this video. Peggy Wetzel edited and produced the final version.

This work supported in parts by the National Science Foundation (grant numbers CCR-9110938 and ICI-9015443), and the IBM and Intel Corporations.

REFERENCES

[1] Ferrari, D., 1990. *Client Requirements for Real-Time Communication Services*, IEEE Communications, (November), pp. 65-72.

[2] Isaacs, E., Tang, J.C., *What Video Can and Can't Do for Collaboration: A Case Study*, Proc. ACM Multimedia 1993, pp. 199-205.

[3] Jeffay, K., Stone, D.L., and Smith, F.D., *Transport and Display Mechanisms for Multimedia Conferencing Across Packet-Switched Networks*, Computer Networks and ISDN Systems, Vol. 26, No. 10 (July 1994), pp. 1281-1304.

[4] Jeffay, K., Stone, D.L., and Smith, F.D., *Kernel Support for Live Digital Audio and Video*, Computer Communications, Vol. 16, No. 6 (July 1992), pp. 388-395.

[5] Stone, D.L., Jeffay, K., *An Empirical Study of Delay Jitter Management Policies*, ACM Multimedia Systems, *to appear*.

[6] Jeffay, K., Stone, D.L., Poirier, D., *YARTOS: Kernel support for efficient, predictable real-time systems*, Proc. Joint Eighth IEEE Workshop on Real-Time Operating Systems and Software and IFAC/IFIP Workshop on Real-Time Programming, Atlanta, GA, Real-Time Systems Newsletter, Vol. 7, No. 4, Fall 1991, pp. 8-13.

[7] Steinmetz, R., Meyer, T., 1992. *Multimedia Synchronization Techniques: Experiences Based on Different System Structures*, IEEE Multimedia Workshop, Monterey, CA, April, 1992.

[8] Wolf, C., *Video Conferencing: Delay and Transmission Considerations*, in Teleconferencing and Electronic Communications: Applications, Technologies, and Human Factors, L. Parker and C. Olgren (Eds.), 1982.

The Tenet Real-time Protocol Suite: A Demonstration

Edward Knightly, Fred Templin, Anindo Banerjea,
Bruce Mah, Hui Zhang and Domenico Ferrari

The Tenet Group
University of California at Berkeley
and International Computer Science Institute

A recent trend in computer networks is their use for the transmission of various forms of multimedia data, such as video, audio, and still images. When these data streams are transmitted for immediate consumption at the destination, the frames in these streams must be delivered regularly for reception quality to be high. For example, frames that arrive too late to be played in the proper order are to be considered lost [2]. Packet-switching networks occasionally suffer from traffic congestion and cannot guarantee such regular delivery. The Tenet *real-time protocol suite*, shown in Figure 1 along with the Internet protocol suite, transforms a congestion-prone network or internetwork into one able to offer the required performance guarantees to the real-time portion of its traffic. In the Tenet suite's connection-oriented, resource-reservation based paradigm, clients can specify what their distributed multimedia applications want, and obtain from the network explicit and binding guarantees that their needs will be met. Since the Tenet real-time protocols have been designed to coexist with the Internet protocols in the same network, they will transform a conventional network into an integrated-services one, which can transport several different kinds of traffic while satisfying the requirements of each kind. Further details about the Tenet protocols may be found in [3].

This videotape documents a live demonstration presented by the Tenet Group at UC Berkeley's 1994 Industrial Liaison Program. Since many of the anomalies common for live television broadcasts may appear in the videotape, we ask viewers to recognize that the immediacy of the live demonstration precluded stringent audio/video editing. The demonstration involves the transmission of live and pre-recorded video streams from San Diego to Berkeley over the Sequoia 2000 network (S2Knet) testbed [4] using the application `vic`, a video conferencing tool developed by Steve

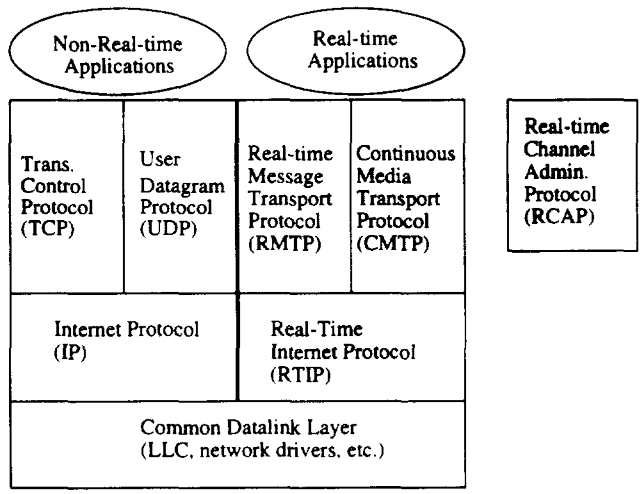


Figure 1: Tenet Real-Time Protocol Suite

McCanne of the Tenet Group. The testbed, shown in Figure 2, concurrently runs both the Tenet real-time protocol suite as well as the Internet protocol suite. Under various network loads, the demonstration qualitatively compares the service received by the application when using the Tenet protocols with the service received by the application when using the Internet protocols.

As shown in Figure 3, The same video stream is JPEG-compressed in hardware at a workstation in San Diego, and transmitted with both the UDP/IP Internet protocols and the RMTP/RTIP Tenet protocols under different network loading conditions. At Berkeley, the two streams are decompressed in hardware and displayed. During the experiment, the S2Knet is also transferring files from Los Angeles to Berkeley; this traffic interferes with the video traffic in the Santa Barbara and Berkeley routers and FDDI rings as well as on the Santa Barbara - Berkeley link. When the file transfer load is light, the qualities of the two video streams at the receiving end are equally good. At medium loads, the quality of the stream transmitted over UDP/IP becomes visibly worse than the quality of the RMTP/RTIP stream. With heavy network loads, the UDP/IP stream deteriorates profoundly, while the quality of the stream transported by the Tenet protocols still remains unaffected.

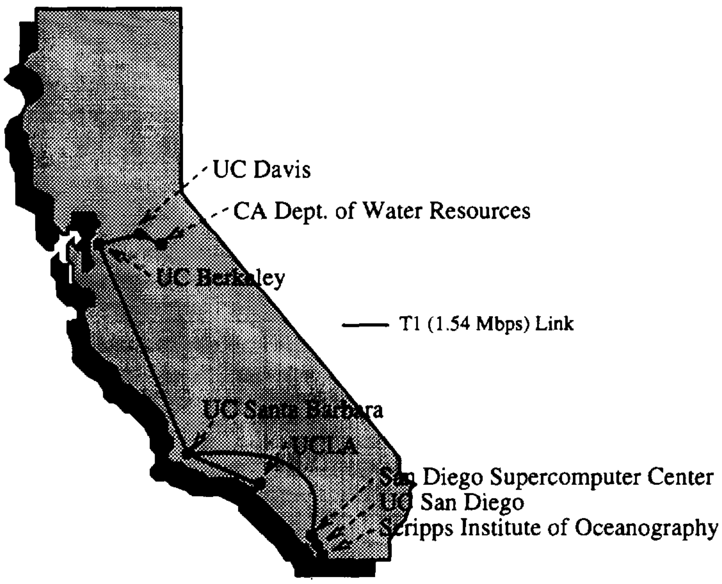


Figure 2: S2Knet Topology

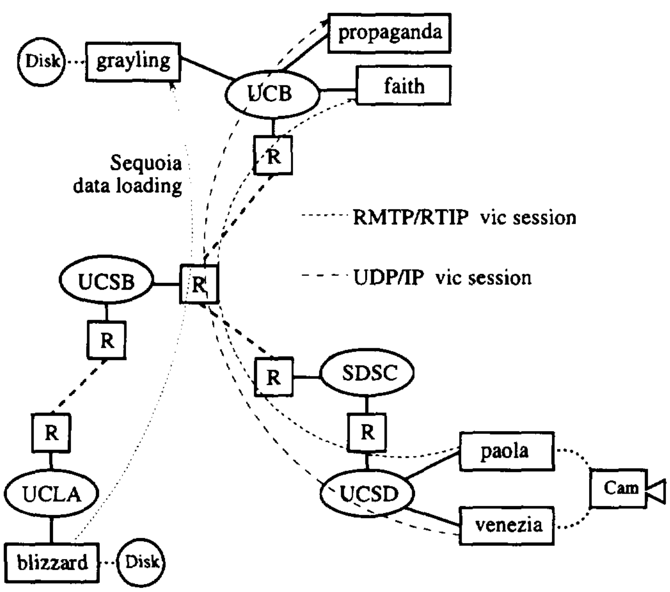


Figure 3: Demonstration Scenario

One unexpected result of the demonstration is the severe dissatisfaction expressed by the demo's observers regarding the audio drop-outs. These drop-outs are easily perceivable and nearly always provoked complaints and questions. The reason for the severe audio outages is that the audio tool we used (**vat**) uses UDP/IP rather than the Tenet protocols. This further emphasizes the need to use guaranteed performance protocols to protect audio quality from network load fluctuations. Further details of the demonstration and further experimental results may be found in [1].

The Tenet Group was established in 1989 and is jointly sponsored by the Computer Science Division (EECS Department) of the University of California at Berkeley and by the International Computer Science Institute, also located in Berkeley. Its research mission in the areas of real-time communication and multimedia networking applications is carried out in the context of three high-speed networking testbeds: BLANCA, one of the five national gigabit testbeds sponsored by the Corporation for National Research Initiatives, with support from the National Science Foundation, and the Advance Research Projects Agency of the Department of Defense, the Department of Energy, and AT&T Bell Laboratories; S2Knet, the network of Sequoia 2000, the flagship project of Digital Equipment Corporation; and the Bay Area Gigabit Network provided by Pacific Bell and supported by the CalREN Foundation.

References

[1] A. Banerjea, E. Knightly, F. Templin, and H. Zhang. Experiments with the Tenet real-time protocol suite on the Sequoia 2000 wide area network. In *Proceedings of the 2nd ACM International Conference on Multimedia*, San Francisco, CA, October 1994.

[2] D. Ferrari. Client requirements for real-time communication services. *IEEE Communications Magazine*, 28(11):65–72, November 1990.

[3] D. Ferrari, A. Banerjea, and H. Zhang. Network support for multimedia: a discussion of the Tenet approach. Technical Report TR-92-072, International Computer Science Institute, Berkeley, California, October 1992. Also to appear in *Computer networks and ISDN systems.*

[4] D. Ferrari, J. Pasquale, and G. Polyzos. Network issues for Sequoia 2000. In *Proceedings of COMPCOM 92*, pages 401–406, San Francisco, CA, February 1992.

DISTRIBUTED MUSIC TRIO - MEDIA SYNCHRONIZATION OVER WIDE AREA NETWORKS

Julio Escobar
Bolt Beranek and Newman Inc.

Multimedia synchronization is a subject of major interest within multimedia communications. This video documents a demonstration of the ability to display (play out) a musical trio at the conference site by remotely synchronizing the audio and video from musicians dispersed across the country. It explains the problem and highlights the current capabilities of multimedia synchronization.

In the demonstration portrayed in the video, the performances of two live players and one prerecorded part (retransmitted from a repeater site) are synchronized across the continental USA, thus delivering a synchronized performance of the music trio at the demo site in spite of delay variations through the internetworks that carry the separate audio and video flows. From the audio portion of the video, the viewer will be able to experience the difference between synchronized and unsynchronized music performances.

The video also shows the operation of the graphical user interface implemented to facilitate synchronization set-up. The GUI includes displays for monitoring of the measured and managed delays in the system. The displays show, in real-time, the way in which the synchronization protocol operates.

Operating System Support for Multimedia Applications

Clifford W. Mercer

School of Computer Science, Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
cwm@cs.cmu.edu

Abstract

This paper briefly describes a resource reservation operating system abstraction that supports real-time multimedia applications alongside non-real-time applications. This approach facilitates predictable behavior in applications while still maintaining the flexibility required by multimedia applications in a dynamic system. The accompanying video tape shows how to reserve processor capacity for multimedia applications and how to control processor usage even when real-time and non-real-time applications compete for access to system servers.

1 Reservation for predictability

We address the problems of providing predictable real-time behavior based on high-level quality of service specifications by structuring the system in two parts:

1. a quality of service layer which uses information about user preferences and the relative importance of various types of applications to make resource allocation decisions, and

2. a resource reservation mechanism which takes resource allocation specifications in the form of capacity requirements and schedules contention for the resource accordingly.

Our research focuses primarily on the resource reservation mechanism, specifically processor capacity reservation. The processor capacity reservation system has three components:

1. an admission control policy to determine whether new reservation requests can be accepted,

2. a scheduling policy to arbitrate contention for the resources, and

3. a reservation enforcement mechanism that measures resource usage to ensure that reserved programs don't use more than their reserved capacity.

Reservations can be requested from the system and terminated dynamically, and reservations can even be adjusted on the fly to fit the changing needs of their applications. Even with the flexibility of being able to change reservations, the system assures predictability – because between reservation requests, the application mix and its real-time requirements are static and known to the system.

An initial reservation system has been implemented in Real-Time Mach [3] (an extension of Mach 3.0 [1]) running on a Gateway 2000 66MHz 486-based machine with 16MB of main memory. The demonstrations shown in the video used this machine configuration.

The resource capacity reservation mechanism associated with the processor resource is described in more detail elsewhere [2].

2 The video

In the video, we show two clips of video applications running. The first shows three instantiations of a QuickTime video player displaying a short video clip from main memory. Figure 1 shows these three players as they appear on the screen in the video. These players are scheduled under a time-sharing policy, and their behavior is, as expected, unpredictable.

The second clip shows that the system can reserve processor time for multimedia applications, and then enforce that reservation in the face of competition for the processor. Also, this clip shows that reservations are flexible and can be adjusted dynamically. The three players, arranged on the screen as in the first clip (Figure 1), are running under the reservation system. One of the video players has a reservation of 90% of the processor capacity, so it has ample time to display its frames. The other two players are not reserved; they run under a "background" time-sharing scheduling policy. So their performance is again unpredictable.

During the second demo clip, we show the flexibility of the reservation system by changing the reservations of

This work was supported in part by a National Science Foundation Graduate Fellowship and by the U.S. Naval Ocean Systems Center under contract number N00014-91-J-4061. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of NSF, NOSC, or the U.S. Government.

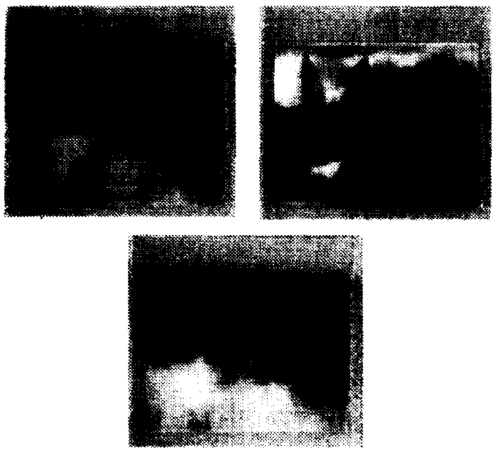


Figure 1: Video Players (from video clip)

the video players on the fly. We remove the 90% reservation from the reserved video players, and give the capacity to one of the previously unreserved players. Now the newly reserved player has the resources to produce a constant frame rate output while the previously reserved player suffers unpredictability in unreserved time-sharing mode. A quality of service manager which keeps track of all the reservations in the system facilitates the coordination of reservation changes.

The scenario depicted in this demonstration is similar to the situation in a video conference with several participants, where the user or a conference manager application would like to be able to change the quality of service parameters of various multimedia data streams based on information like which participant currently has the floor. Our reservation mechanism enables that kind of resource management and control.

3 System configuration

The demonstration shows how the reservation system handles the allocation of processor capacity, but we should note that the end-to-end predictability shown in the demo requires coordination among several system components. Figure 2 shows the structure of the software components in the demo. The players keep their short video clips in memory to avoid interaction with a file server or with a disk.

The important thing to note about this demo is that there is contention in the X server between the frames of the reserved player and the frames of the other players. To achieve end-to-end predictability, the X server must work within the framework of the reservation system; it must service incoming work according to the priority of the client, and it must charge its computation time to the client for whom the service is performed.

The inter-process communication mechanism that clients use to talk to the server must also understand reservations and deliver messages in the order prescribed by the reservation system, otherwise the reservation system can be circumvented, resulting in unpredictable performance.

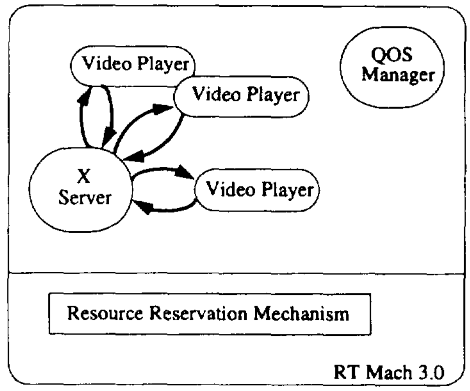


Figure 2: Demo System Software Configuration

4 Conclusion

The video shows that are processor reservation system facilitates predictable behavior in real-time multimedia applications, even in the face of competition for system resources from non-real-time programs.

Acknowledgements

The author would like to express his appreciation to Jim Kocher, Chen Lee, Ragunathan Rajkumar, and Jim Zelenka for their help in producing the accompanying video, and to the researchers who have worked on RT-Mach and its associated programming environment: Takuro Kitayama, Tatsuo Nakajima, Ragunathan Rajkumar, Stefan Savage, Hideyuki Tokuda, and Jim Zelenka.

References

[1] D. L. Black et al. Microkernel Operating System Architecture and Mach. In *Proceedings of the USENIX Workshop on Micro-kernels and Other Kernel Architectures*, pages 11–30, April 1992.

[2] C. W. Mercer, S. Savage, and H. Tokuda. Processor Capacity Reserves: Operating System Support for Multimedia Applications. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMCS)*, May 1994.

[3] H. Tokuda, T. Nakajima, and P. Rao. Real-Time Mach: Toward a Predictable Real-Time System. In *Proceedings of USENIX Mach Workshop*, pages 73–82, October 1990.

THE POETRY FILM WORKSHOP PRESENTS:
DR. EDMOND SKELLINGS
BY DAVID KRZYSIK
A PROJECT OF
THE NATIONAL POETRY ASSOCIATION

On this tape we are presenting one of Dr. Edmond Skellings' multimedia video poems: "To His Machine". To quote from Dr. Skellings writings: "'To His Machine' says goodbye to the typewriter, hello to the computer. Welcome to computer poetry."

The Poetry Film Workshop has produced a film/video festival for the past 18 years. The 19th Annual Festival is coming up in November at the Center for the Arts at the Yerba Buena Gardens in San Francisco, California. We are looking forward at another year of innovation and change, accepting CD-Rom for the first time and producing a television series from our vast catalog.

With the advent of super highway media technology, we can envision perhaps our next festival with phone hookups for competitions of images over the same poems with different treatments, music, sound effects, etc.

Even our television series will be embedded with pieces created on current CD-Rom technology so that it will have the capacity of interactive broadcast, when that technology becomes available.

These pieces by Dr. Skellings have been created in computer environments, yet true to the genre of the poetry film. The festival, in fact, the National Poetry Festival as well, have always presented truly multi-media and mixed-media artists and poets.

This technology also shows expansion into help for adult and youth educational tools. In 1930, only 4.3 percent of the population was illiterate. Now, six decades later, between a fifth and a quarter of the population is illiterate and in the last 50 years, the written vocabulary of the average 6 to 14 year old child in the U.S. has shrunk from 25,000 to 10,000 words, a three-fifths loss in the ability to make sense of the world through language. The CD-Rom technology, with its capabilities of embedded text holds the promise to educate, while it entertains in a popular fashion, with a literary based genre such as poetry film.

Dr. Skellings is the Poet Laurete of Florida.

PhenomenArts, Inc.
75 Kendall Road
Lexington, MA 02173 USA
Christopher Janney, Artistic Director
(617)862-6413/FAX: 862-6114

"PHILLY SOUND PARK" Using Animation as an Architectural Design/Presentation Tool

Philly Sound Park is a proposed design for a "talking park" in downtown Philadelphia. The video presentation uses computer models, animation's, stills, videotape and 4D sound surround to describe and document the project and the process of its development. The principal artists are Christopher Janney and Martha Schwartz. Associates involved in developing the multi-media techniques include Geoffrey Pingree and Stephen Langstaff.

The park itself, a site in Philadelphia, is divided into sixteen 15' x 15' squares. Within each square there is a lectern and bench made out of traditional Philadelphia brick and a Filbert tree. The bricks of the lectern are built on spacers, making them seem to 'float' and allowing sound and light to emanate. Inside each podium there will be a speaker, lights and two sensors. As people pass by, they will trigger the sensor, activating sound and light at different times of the day.

The sound score itself will be composed of many sounds comprising a "sonic portrait" of Philadelphia. Categories include the sound of the Liberty Bell, readings of historical documents, radio and television advertising slogans and famous groups and music from Philadelphia. Different sounds will be activated at different times through out the day and week, depending on what sensors will be triggered. The sounds of Philly Sound Park can be changed to accommodate new and as yet "unborn sounds," making a place for the "future history" of Philadelphia.

The videotape was created for presentation at a closed competition and as an example of how architects and artists can use multi-media techniques, both as a design and presentation tool.

Working with a number of computer companies for the last two years, PhenomenArts, Inc. has created a desktop computer animation system for under $25,000. This hardware for this system includes a Macintosh Quadra 950 with 56 megs of RAM, a PLI disk array and the Radius Video Vision Studio card. The software includes Adobe's Photoshop 2.5, ArchiCAD release 12, Strata-Vision Studio Pro and Adobe Premier 3.0. The music is mixed in MOTU Digital Performer and played into Premier for the final mix-to-pix. With the Radius card we print directly to tape from our computer.

Specific production methods included creating models and animation's in the computer for "walk-through's" of the site as well as 4D views of objects on the site. We also connected a video camera directly to the computer through the Radius board and recorded shots of "real" models as well as some early sketching sessions into the computer. As well, we audio-taped a number of conversations about the genesis of the project and loaded that into the computer. Then, using Premier, we cut the picture and sound together.

" COME, HUMAN, SPIN IN MY WEB!"

BY BEVERLY AND HANS REISER
MUSIC BY BILL FLEMING

"Come, Human, Spin In My Web! " is an interactive installation using sound, video, and computer graphics. It is a metaphorical reality that starts with an A.I. persona sending an invitation to humans. If the humans take up the invitation, they are then given a choice between visiting the Feral Forest (the world of Chaos), or the Cerebral Cathedral (a world of lawful mental constructs).

Modern culture thinks of "Mother Nature" as a goddess of balance and simplicity, but Chaos theory hints that She is often turbulent and complex. In the Feral Forest, a visitor can explore these issues. Naturalists are encouraging our society toward a more stable, renewable resource orientation - a self-sustaining bubble. But, do bubbles really survive in nature?

On the other hand, what if we choose to push the envelope of human evolution rapidly through the use of our technologies, will we cease to be what we now recognize as human?

Instead of the Feral Forest, the visitor might choose the man-made, lawful constructs of the Cerebral Cathedral. Before long, however, the visitor is buried under the weight of either worn-out paradigms or useless paradoxes. Does the visitor have the will to throw out the old paradigms and paradoxes and start anew? Or wll they attempt to cling to them finding comfort and reassurance in cultural rituals ?