Multimedia Computing: Applications, Designs, and Human Factors

Scott M. Stevens
Software Engineering Institute'
Camegie Mellon University

ABSTRACT

Computer input and output has evolved from teletypewriters to text oriented CRTs to today's window oriented bit mapped displays. Beyond text and images, multimedia input and output must be concerned with constant rate continuous media such as digital video and audio. Today's multimedia applications are just beginning to explore the capabilities of multimedia computing. There is a growing understanding of the functionality required by multimedia application domains such as intelligent interfaces, collaboration environments, visualization systems, and virtual reality environments. Advanced applications dynamically operate on and modify multimedia objects, providing possibilities previously unavailable for human computer interaction. This chapter discusses some of these advanced multimedia paradigms through examples of today's nascent multimedia applications. With an understanding of multimedia's full capabilities, large, quantum improvements in the usefulness of applications are possible.

9.1 Introduction

Modes of communication, interfaces, between humans and computers have evolved from teletypewriters to text oriented CRTs to today's window oriented bit mapped displays. One of the next steps in this evolution, multimedia, affords the interface designer exciting new

t This work supported by the U.S. Department of Defense

User Interface Software, Edited by Bass and Dewan

channels of communication. However, the definition of multimedia is indistinct and rapidly evolving.

A graphical simulation of a falling body, such as a ball, associated with a text on Newton's Laws certainly seems to fit within the domain of multimedia. If each image of the simulation is a three dimensional perspective view of the falling ball, does the application continue to fit what we call multimedia or is it an animation [Sta93]? Is it multimedia, animation, or virtual reality if we create perspective views from two different locations and look at the same simulation through immersive displays [Esp93]? In a short chapter such as this it is impossible to survey every aspect of what may fit under the heading of multimedia. Instead, I will focus on one of the most talked about and probably least understood areas of multimedia, digital video and audio.

The distinction between digital and analog video is an important one. Over the years there have been many interesting, useful demonstrations employing analog video technology. Such projects span more than a decade from the Aspen project, a "multimedia" tour of Aspen, Colorado [Lip80], continuing to current efforts such as CIearBoard, using analog video in support of collaboration [IK92]. In each case, these high quality, high functionality analog video-computer systems have in common expensive, one-of-a-kind, complicated hardware. With the introduction of all digital video and audio systems application and interface designers have access to increased functionality at a fraction of the cost.

Having video and audio as another manipulable, digital data type afford the opportunity to create new interfaces and applications never before possible. Moreover, it becomes both easier and more important to design these applications so they conform to the needs of the user, rather than force the user to conform to the design of the system.

This chapter begins by describing the characteristics of digital video (it is a constant rate continuous time medium). It then illustrates the difficulties and the opportunities, both social and technical, of designing applications for this new medium through two prototypical application domains (multimedia electronic mail, and networked digital video conferencing). While there are great opportunities in multimedia, there is a danger in the rush to incorporate multimedia in today's computing environment. Some of the problems associated with low fidelity implementations of multimedia are discussed in Section 9.5. Finally, a broad hint at how multimedia applications may evolve is given in the concluding section.

9.2 Digital Video: Continuous Time Multimedia

Early interface designers certainly did not and perhaps could not anticipate the capabilities of today's computing environments. In a teletypewriter or CRT, synchronization and control issues were mainly ones of, "is the output device ready to receive the next character?" and "should each line send a line feed, carriage return, or both?" Today's typical multimedia applications ask synchronization questions such as "how should audio stream A and video stream B be synchronized with each other?" and "should video C, graphic image D, or both be presented on the display of text ET' The complexities and flexibility of today's multimedia systems go beyond placing a video window of arbitrary size on a computer screen. It is now possible to operate on video as a data type in order to conform the system to the needs of the user.

In an emerging, complex field such as multimedia, it is not surprising that most of today's applications have failed to take full advantage of the information bandwidth, much less the

capabilities of a multimedia, digital video and audio environment. Today's designs typically employ a VCR/VideoPhone view of multimedia. In this simplistic model, video and audio can be played, stopped, their windows positioned on the screen, and, possibly, other simple events such as the display of a graphic synchronized to a temporal point in the multimedia object. This is essentially the traditional analog interactive video paradigm developed almost two decades ago. Rather than interactive video, a much more appropriate term for this is "interrupted video."

Today's interrupted video paradigm views multimedia objects more as text with a temporal dimension [HSA89, YHMD88]. Researchers note the unique nature of motion video. However, differences between motion video and other media, such as text and still images, are attributed to the fact that time is a parameter of video and audio.

Every medium has a temporal nature. It takes time to read (process) a text document or a still image. However, in traditional media each user absorbs the information at his or her own rate. One may even assimilate visual information holistically, that is, to come to an understanding of complex information all at once. Even the creative process is subject to such "Ah ha!" experiences. Mozart said that he conceived of his compositions in their entirety, in one instant, not successively [Had45]. Yet, time is an intrinsic part of music that to most of us cannot be separated from pitch, melody, etc. Even though Mozart created and "saw" a whole composition at once, the temporal aspect of the music must have been present to him in that same instant.

Comparing the scrolling of text to viewing a motion video sequence illuminates the real difference between video and audio and other media. Let us assume a hypothetical application designer performs extensive human factor testing and determines that a mythical average user reads at exactly 400 words per minute. This designer then develops an electronic encyclopedia that continuously presents its text at 400 words per minute. Clearly, a user that reads at even 401 words per minute would soon be out of synch with the text. Even our designer's canonical reader will undoubtedly find text that, because of complexity, vocabulary, or interest, requires more time to read. It is unlikely that anyone would argue for fixed text presentation rates.

However, to convey almost any meaning at all video and audio must be played at a constant rate, the rate at which they were recorded. Granted, a user might accept video and audio played back at 1.5 times normal speed for a brief time. However, it is unlikely that users would accept long periods of such playback rates. In fact, studies described in Section 9.5, show that there is surprisingly significant sensitivity to altering playback fidelity. Even if users did accept accelerated playback, the information transfer rate would still be principally controlled by the system.

The real difference between video or audio and text or images is that video and audio have constant rate outputs that cannot be changed without significantly and negatively impacting the user's ability to extract information. Video and audio are a constant rate continuous time media. Their temporal nature is constant due to the requirements of the viewer/listener. Text is a variable rate continuous medium. Its temporal nature only comes to life in the hands of the users.

9.2.1 Searching Continuous Time Multimedia

Searching for information highlights one of the more significant differences between constant rate continuous time and variable rate continuous media. The human visual system is adept at quickly, holistically viewing an image or a page of text and finding a desired piece of

information while ignoring unwanted information (noise). This has been viewed as a general principle of selective omission of information [Res89] and is one of the factors that makes flipping through the pages of a book a relatively efficient process. Even when the location of a piece of information is known a priori from an index, the final search of a page is aided by this ability.

On the other hand, objects that have intrinsic constant temporal rates such as video and audio are difficult to search. There are about 150 spoken words per minute of "talking head" video. One hour of video contains 9,000 words, which is about 15 pages of text. The problem is acute if one is searching for a specific piece of a video lecture, or worse yet from audio only. Even if a comprehensible high playback rate of 3 to 4 times normal speed were possible, continuous play of audio and video is a totally unacceptable search mechanism. This can be seen by assuming the target information is on average half way through a one hour video file. In that case it would take 7.5 to 10 minutes to find! Certainly no user today would accept a system that took 10 minutes to find a word in 15 pages of text.

Detailed indexing can help. However, users often wish to peruse video much as they flip through the pages of a book. Unfortunately, today's mechanisms are inadequate. Analog videodisc scanning, jumping a set number of frames, may skip the target information completely. To be comprehensible, scanning every frame such as in a VCR's fast forward, often takes too much time. Accelerating the playback of motion video to, for instance, twenty times normal rate presents the information at an incomprehensible speed. And it would still take six minutes to scan through two hours of videotape!

Even if users could comprehend such accelerated motion, finding a specific piece of video would be difficult. A short two second scene would be presented in only one tenth of a second. With human and system reaction times easily adding to a second or more, significant overshoots will occur as the user tries to stop the video when the desired segment is found.

Playing audio during the scan will not help. Beyond 1.5 or 2 times normal speed audio becomes incomprehensible as the faster playback rates shift the frequencies to inaudible ranges [DMS92]. Digital signal processing techniques are available to reduce these frequency shifts_ At high playback rates, these techniques present sound bytes much like the analog videodisc scan. Listening to a Compact Disc audio scan is convincing proof that even without frequency distortion, rapid scanning of audio fails to be a practical search mechanism for large volumes of data.

Tools have been created to facilitate sound browsing where visual representations of the audio waveform are presented to the user to aid identification of locations of interest. However, this has been shown to be useful only for audio segments under three minutes [DMS92]. When searching for a specific piece of information in hours of audio or video other mechanisms will be required.

The Advanced Learning Technologies (ALT) project at CMU's Software Engineering Institute developed a multidimensional model of multimedia objects (text, images, digital video, and digital audio). With this model, variable granularity knowledge about the domain, content, image structure, and the appropriate use of the multimedia object is embedded with the object. In ALT an expert system acts as a director, behaving intelligently in the presentation of image, digital audio, and digital video data. Based on a history of current interactions (input and output) the system makes a judgment on what and how to display multimedia objects [CS92, Ste89].

Techniques using such associated abstract representations have been proposed as a mechanism to facilitate searches of large digital video and audio spaces [Ste92]. In this scheme,

embedding knowledge of the video information with the video objects allows for scans by various views, such as by content area or depth of information. When video objects are imbued with knowledge about their content, then partitioning them into many separate small files permits first pass searches to retrieve a small segment of one to two minutes of video. Continuous play of extended sequences is accomplished by seamlessly concatenating logically contiguous files.

Another aid to searching is found in compression schemes for digital video. Since significant image changes affect the compression and decompression algorithms, identification of visual information changes is relatively easy. Thus, scans of small digital video files can be performed by changes in visual information, such as by scene changes. This can be an improvement over jumping a set number of frames, since scene changes often reflect changes in organization of the video much like sections in a book. And in cases where this is not efficient, embedded knowledge about the content of scenes or even individual frames can substitute. In the end, appropriate application of these techniques will permit information-based scanning of multimedia material much like noticing chapter and section headings in a book while flipping pages.

The search and presentation of information stored as motion video and audio is the essence of today's interrupted video paradigm. Emerging multimedia applications move somewhat beyond this by the combination of presenting text, images, animations, and interrupted video. The structures of these paradigms are due to the historical fact that the first marriage of video and the computer was the combination of analog video with the computer's video output.

In these early systems, the analog video was presented on one screen while the computer output was on a second screen. Until the advent of digital video the only integration between video and computer output was through key color mixing. Here, computer graphics could be overlaid on the video image. But the analog video was still effectively separated from the computer and could not be affected by the computer, other than starting, stopping or covering it with graphics.

Even though today's digital video integrates the video and audio completely with other digital data, the interrupted video paradigm remains. This is in part due to current multimedia developers' lack of experience and limited number of more advanced models to emulate. It may also be due to the fact that video and audio are constant rate continuous time media. Yet just because video and audio have a constant rate does not mean they cannot be manipulated by the system or the user. The following sections investigate the limitations of simple multimedia paradigms and suggest the possibilities for human computer interaction afforded by state of the art multimedia designs.

9.3 Multimedia Electronic Mail

One of the simplest applications of multimedia is electronic mail. Most users appreciate the advantage of hearing a message and seeing the face of the speaker. Factors such as being more personal and eliminating the need to rely on punctuation to impart messages with affect are cited as benefits of multimedia email.

Multimedia email will impact business, professional, and personal communications. In each domain, multimedia email will afford similar advantages and potential disadvantages. Highlighting one application domain will serve to illustrate the complex social issues raised by combining multimedia with electronic mail.

It is hard to imagine an area in greater need of technological tools than education and training. The nation's schools and industry together spend between $400 and $600 billion per year on the business of education and training. Ninety-three percent of this expense is labor-intensive, approximately two times that of the average business, with no change in teacher productivity since the 1800s]Per90].

Advantages of multimedia email for education and training seem obvious. Clearly it is difficult for students to gain individual attention in today's classroom. An instructor has lecture time for a few questions at best. And those questions are usually posed by the most aggressive or outspoken students. The average student may wait until an office hour to ask a question, or more typically, ask another student, or never ask the question at all. In the classroom of the future students will be able to send their instructor a multimedia email question. The general hypothesis is that students will then have greatly increased access to their instructors.

But this hypothesis is based on the assumption that multimedia email is isomorphic to textual email with respect to how users will interact with it. Underlying multimedia email is constant rate continuous time media. Neither the audio nor the video will be able to be played back at a significantly greater speed than the speed they were recorded. Textual email is variable pace continuous media. This may produce some unexpected consequences.

For example, a hypothetical lecture has one hundred and twenty students. If each student were to ask only one multimedia email question per week with an average length of one minute, the instructor will spend a minimum of two hours per week just listening to the questions. Individual two minute answers from the instructor will add another four hours to the task. The six hours to listen and respond to students' questions assume no time was necessary to think out the answers or to interact with the video email interface! At a minimum, the instructor will need to edit responses. What will be the best form for this editor? No matter how easy a multimedia email system is to use, the time to listen to a question and compose a response will be more than the total time of the messages. Few faculty will appreciate trading one hour recitation and two office hours for six to ten hours of multimedia email.

Granted, numerous questions will be asked repeatedly. This will permit the instructor to prerecord his or her responses, helping reduce the time to respond to these questions. But we have seen that the time to listen to the multimedia email cannot be significantly reduced.

Moreover, it is not clear if this will be an improvement over today's system. Will a generic response miss nuances of a student's question? Might it then misdirect the student? Instructors will without a doubt create catalogs of answers to frequently asked questions (FAQ), just as there are FAQs for today's network news groups. All too many instructors today give lectures from notes that are used year after year. It will be difficult for the most conscientious instructors to keep their multimedia email responses up to date, much less those who continually reuse their lecture notes.

When multimedia FAQ answers are saved for years, how will students react to their instructor, appearing years younger, giving a canned answer in his out of style leisure suit? If a student asks more than one question in a single message and the instructor composes a response from two separate files how will the discontinuity affect the user? Section 9.5 in this chapter reports on questions of fidelity in multimedia. It will be seen that changes more subtle than these can have significant effects on users' understanding.

Technical solutions can help. For a classroom environment the students' multimedia email questions can be tied to a passage under study or a problem being solved in an electronic text. Knowledge about the content area that generated the question can be automatically attached to the students question in a machine readable form. Prospective prerecorded responses can

then be brought to a local information space for easier perusal and retrieval by the instructor. Creation dates can be checked by the system and messages that are too old, defined by some agreed to convention, can be flagged to encourage re-recording. Innovative tools are being developed to help manage this type of information flow in text form [LM9]]. Still, the potential is great for creating an environment that fosters mindless responses, using stale recordings.

These issues carry over to professional and business multimedia electronic mail as well. Managers high on the organizational tree may become inundated with time consuming messages and create "one size fits all" responses. What is the legal liability of a physician who sends a prerecorded response to a colleague that is not absolutely up to date? Certainly there are analogs to these issues in paper environments. But how much more prevalent they may be when multimedia computing aids the process is unknown.

Multimedia email as a solution to current communications problems raises many questions. It will not be a panacea for reducing the time it takes to communicate. The advantages of seeing and hearing the tone of an author will likely outweigh the potential problems. But the issues raised by such a seemingly straightforward application as multimedia email are suggestive of the power and complexity of multimedia applications for human computer interaction.

9.4 Networked Digital Video Multimedia Conferencing

When it becomes interactive and real-time, multimedia email evolves into digital video conferencing. Although teleconferencing and analog video conferencing have both been available for decades, new difficulties and opportunities arise when multimedia computing solutions are brought to bear on this old problem.

Today's analog teleconference can be multiple simultaneous telephone connections, more than one person using a speaker phone, or some combination. And a video conference may be effected between two or more remote sites each with the capability of receiving and transmitting both video and audio. (It is interesting to note that up to now common usage of the term "video conference" is often applied to the case of two parties making what might more appropriately be called a video phone call. As technology changes to permit every phone call to be a video call, the term video conference will no doubt be applied differently than it is today. AT&T and MCI's]992 introductions of video phones that permit real time compressed video to be transmitted over standard phone lines are the beginning of this change.)

Unlike a conference call wherein all audio signals are mixed, today's analog video conference uses a separate monitor for each video feed. Usually, when more than one person at one of the sites is participating in a video conference the camera and furniture are arranged so that all of the collocated participants are simultaneously framed in the video. Some video conference facilities have multiple cameras. When several people are involved in the conference from such a site, one camera is trained on each participant and then switched by an operator. A single video feed is then distributed to the other sites.

In successful collaborative meetings participants have specific roles [Off85]. These roles have titles such as facilitator, scribe, and reader. But beyond the participants and their roles, researchers have found the physical arrangement of the seats in the room to be important to meeting success [SS88].

In this respect analog video teleconferencing schemes have significant disadvantages. Probably the most obvious consequence of analog video schemes is what I call "The Brady Bunch Effect." In the opening sequence to the TV series the Brady Bunch each of the actors was placed

¡n postage stamp fashion in a portion of the screen as they are in today's video conferencing systems (See Figure 9.1).

Figure 9.1 Brady Bunch 2x2 video meeting screen

With the type of visual presentation illustrated in Figure 9.], problems with perceived social dominance or lack thereof, due to arbitrary spatial locations of images arise. Xerox PARC's Media Space and its successors such as CAVECAT (Computer Audio Video Enhanced Collaboration And Telepresence) exemplify an interrupted, analog video paradigm and illustrate this older paradigm's limitations with respect to visual organization of a meeting space:

> ...CAVECAT changed social status relationships due to the loss of the usual spatial and nonverbal cues which convey status information. In face-to-face meetings, the seating of the people in the room is usually indicative of a hierarchy with higher status people occupying more central positions or "head of the table" locations. The design of CAVECAT unintentionally introduced its own social status cues. In meetings of four individuals, CAVECAT arbitrarily positioned participants' images in a 2X2 grid. CAVECAT also configured the video images for a meeting based on who requested the meeting. This meant that if meetings were reconvened after a short break by a different person, the result was a different image configuration. This was highly disconcerting to the participants. It was as if everyone had left the room and returned to take new positions around the table [MBS +911.

It is unfortunate that this type of artifact has been needlessly carried to networked computer/video conferencing designs. An obvious and simple solution to the problem of random ¡mage placement is to allow screen placement of meeting participant's images to be deter-

mined by their role. When a meeting is continued at a later time, no matter by whom, the same arrangements for each user would be reconstructed. This is not to say that each viewer sees the same scene. In fact, that would be most unlike a face to face meeting where each person sees the room from own visual perspective.

This type of facility has been argued against, not on human factors grounds, but on erroneous technical grounds. The argument chain begins with the see yourself "mirror" window of systems such as Xerox's Media Space and CAVECAT. In the 2x2 meeting screen of Figure 9.1 one of the images is your own. This permits each user to insure that they have framed themselves in the camera properly and to point to other screen objects. Novice users are immediately struck by the fact that their image is in fact not a mirror image. The consequence of this is that visual cues in the media space's "mirror" are precisely backward from what the user expects. So that unlike an optical mirror, if you point to the left you see your image point to its left (your right). But what the user really needs to see is an image that is pointing in the same absolute direction as he or she is pointing (see Figure 9.2).

It has been argued that after a period of time users adapt to the situation. This is undoubtedly true. As part of a workshop on the development of intelligent digital video simulations, university faculty are put through similar experiences [SFC89]. Participants perform a number of writing tasks in a device that permits them only a mirrored view of their hand and paper. The purpose of this task is to give attendees personal experience in the accommodation of cognitive processes. During the one-half to one hour experience some people never accommodate. However, frequently someone will experience a gestalt shift and begin to write perfectly well in the mirror system. A frightening experience for some of these people is that after this, and sometimes for several hours, they cannot write normally without the mirror.

Arguing for the not mirroring the user's own view of their image in a media space is analogous to defending any design by saying users adapt. In other words, it is the Marie Antoinette `Let them eat cake' school of human factors design. In a digital video, multimedia environment one's own image can be mirrored locally and sent non-mirrored to external recipients. Thus, images, such as text, that were placed in front of the camera in such a system would not be mirrored for the other participants. Manipulations such as these are trivial in a digital video environment, but very difficult in an analog video environment.

The question arises whether this will cause a new problem: in such a system when a user points to another person in the media space, another quadrant on the screen, the rest of the participants would see the first user pointing in the wrong direction (see Figure 9.3). As can be seen in these figures, if placement of the images is arbitrary, the direction of pointing will be incorrect (see Figure 9.4). However, in a digital video multimedia paradigm there is no reason why the placement cannot be tailored for each user.

Moreover, the work of Stone and others suggests that the placement of these images should not be arbitrary but associated to some convention. Thus, in the simplest case the screens could be arranged as in Figure 9.5 where it is seen that each user sees a unique but basically correct representation of the meeting and the participants. In this design, not only would each user have a natural image of themselves (i.e. actually mirrored), but there would be consistent screens for reconvened meetings and effectively arranged "meeting chairs."

Since you do not see yourself in real meetings, it may in fact be preferable to have a small, inconspicuous image of the user presented locally. The principal use of the your own image is then for centering (framing) your self in the camera's field of view, not pointing to locations in the media space. In this case, if the user wishes to point to a location in the media space, rather than their own hand, a digital pointer can be made available.

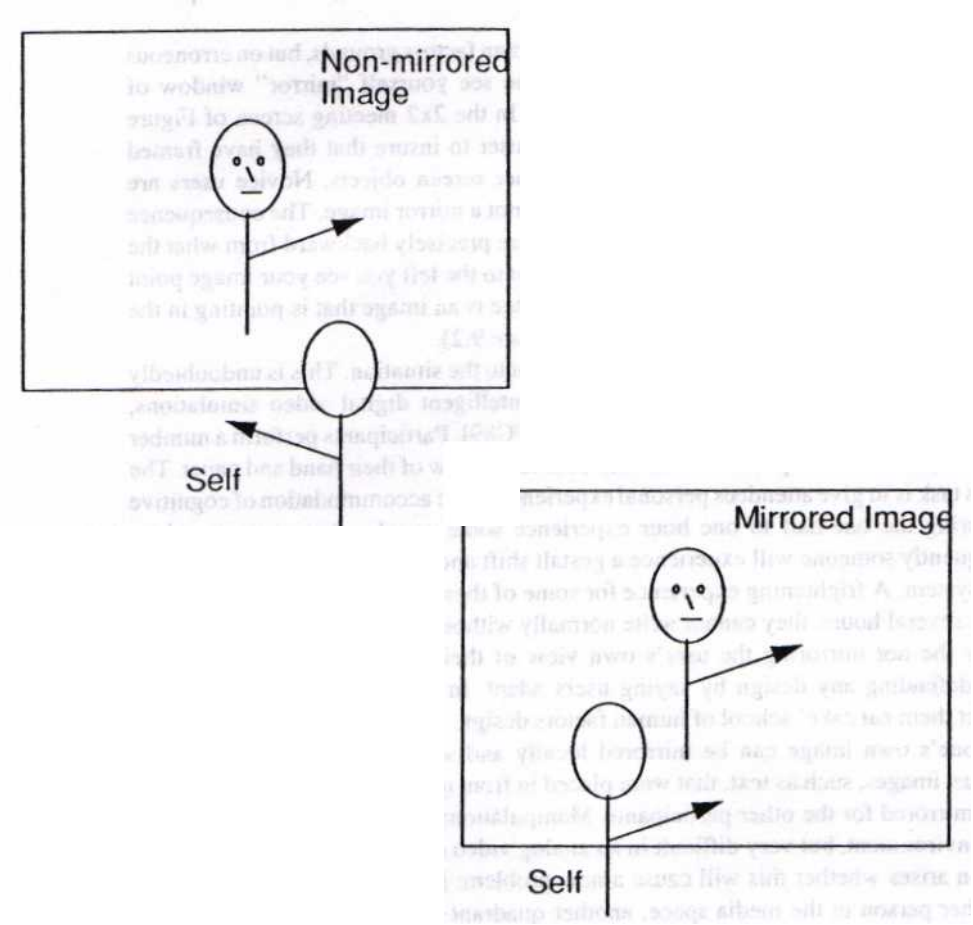

Figure 9.2 Effect of non-mirrored images on user gesticulation

Still more advanced representations of meeting spaces are possible. Figure 9.6 is an image from a simulation of a meeting [CS92, Ste89]. This image is actually made from five different images blended to form a single seamless scene (see Figure 9.7). While these images are prerecorded, there is no impediment to performing this type of synthesis live. With luminance keying users need not be seated in a "Chromakey blue" space. Thus, with few special architectural constraints placed on their offices, users could be placed in a virtual meeting space.

While digital video in multimedia computing permits natural, user centered designs, many questions remain. Can and should a system automatically frame the subject in the camera's field of view? What is the effect of camera angles on users' perception? How should subjects

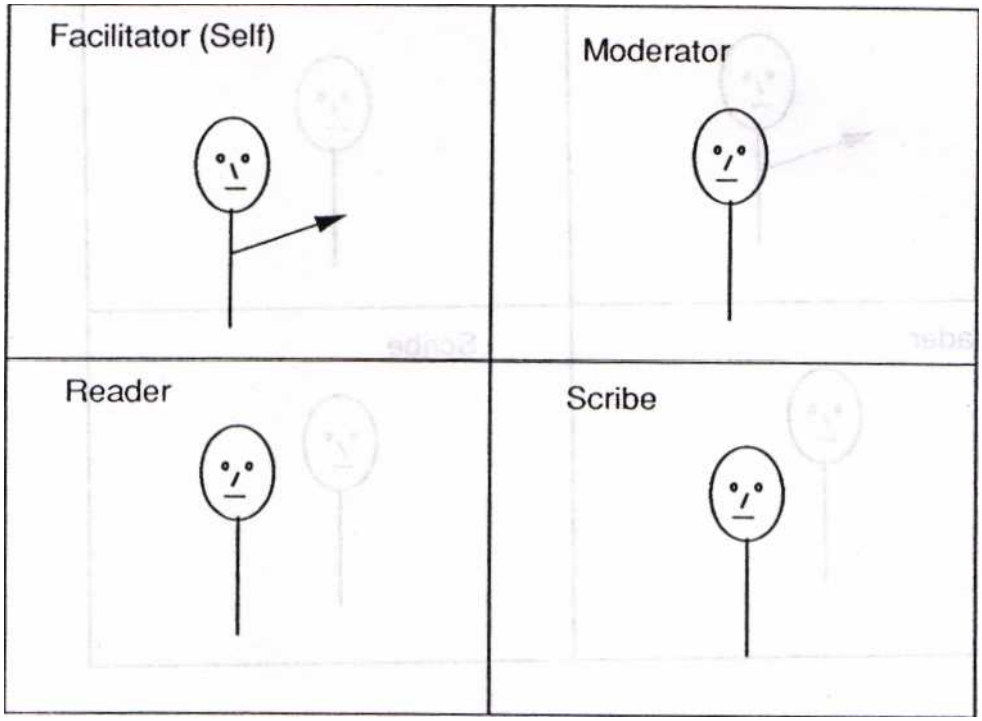

Figure 9.3 What user sees in a mirrored self image video conference

be properly lighted? Although the new questions generated by multimedia are complex, technologies suggesting solutions are available today.

At the consumer end of the scale $150 camcorder tracking devices are available that can keep subjects framed. At the high end of the scale NHK has developed SYNTHEVISION. Foreground and background images are coupled in SYNTHEVISION. Using data from the foreground image a background image is derived with appropriate perspective, taking into consideration panning, tilting, focusing, zooming, and dollying of a camera tracking the foreground image [NHK91]. This allows for users to be electronically placed in any arbitrary physical space with a totally convincing visual presentation.

The effect of camera angles and lighting on viewers' comprehension perception is well studied [Kra87, Kra88]. Mechanisms for capturing that expertise and automatically applying

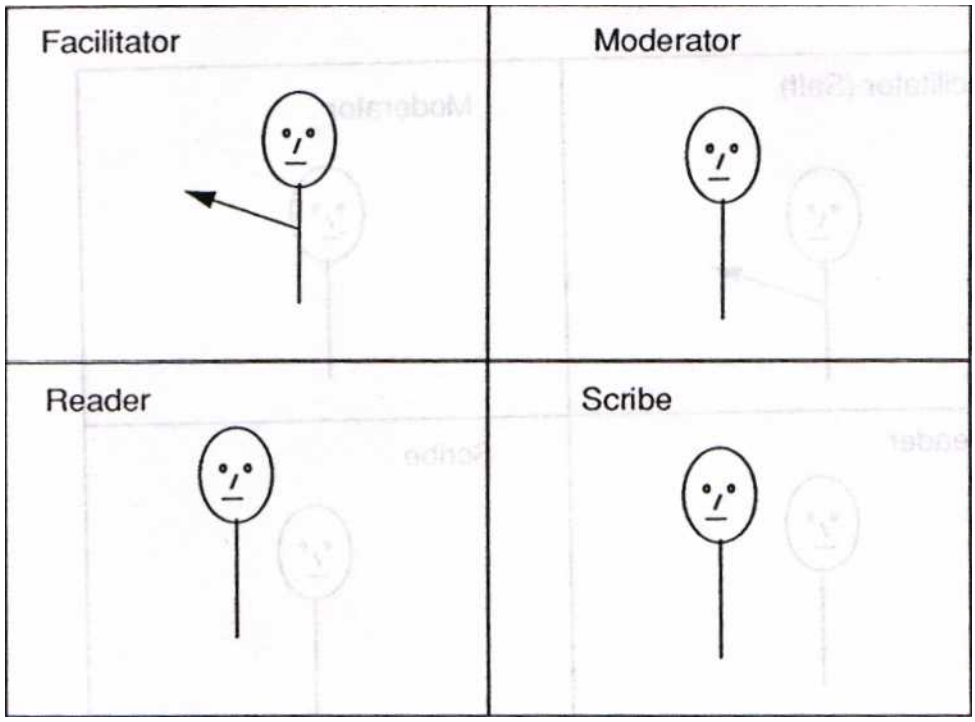

Figure 9.4 What others would see when the non-mirrored image is sent to others.

it in multimedia computing systems have been demonstrated [SFC89]. With capabilities such as these, multimedia meeting spaces can ultimately be as comfortable and productive as a well designed conference room.

9.5 Fidelity

So far this chapter has equated multimedia with digital video and implied that any video is better than no video. But is that really true? Are there differences between the closeness to reality (fidelity) of the design of a multimedia space? How much fidelity is enough? Does frame rate of video matter?

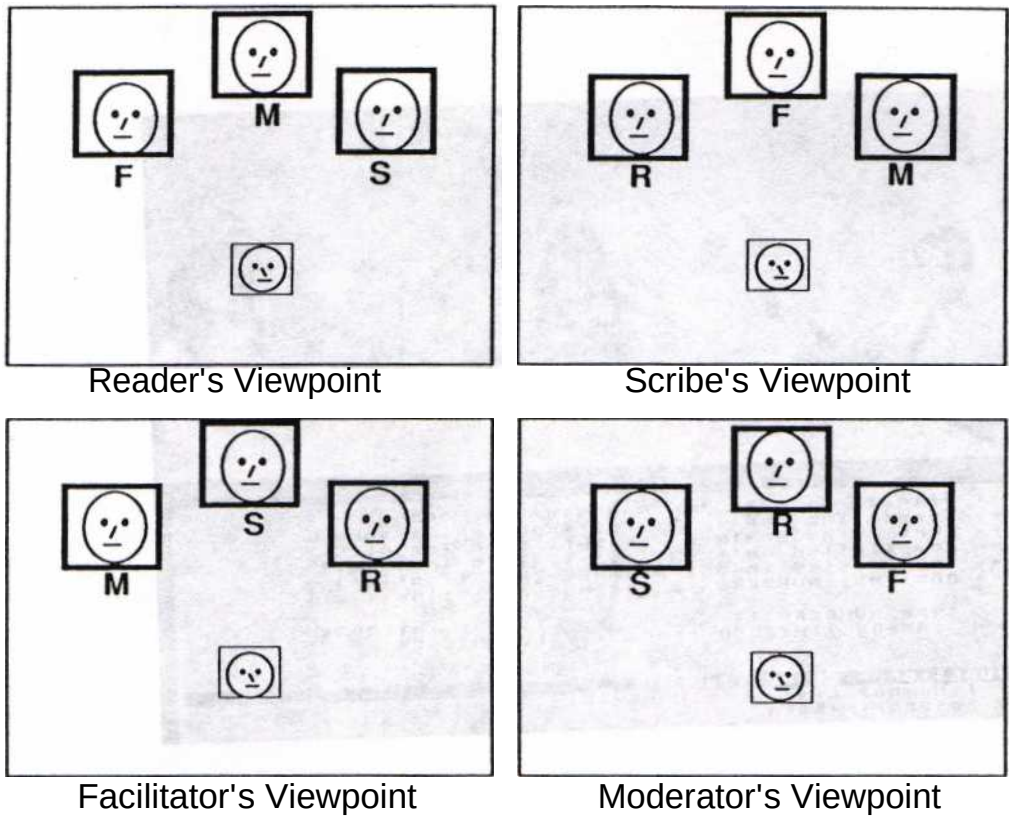

Figure 9.5 Variable placement of participants allows each user a unique scene

9.5.1 Design Fidelity

Michael Christel studied users of a virtual reality workspace for learning and experience software inspections [Chr91]]. Two methods of navigating through the world were evaluated, one a direct manipulation point and click floor plan of the space, and the other a surrogate travel interface where the user "walks" through the space and into the desired sub-world (see figure 9.8). Both groups of users liked, or disliked, the interfaces equally and used them in equivalent fashions to navigate the world. In itself this is surprising as one would expect the direct point and click interface to permit the users more freedom to move between sub-worlds. This is especially true as the surrogate travel interface took considerably more time to perform the equivalent task (moving the user from one sub-world to another).

More interesting than the fact that the users navigated between sub-worlds equivalently was

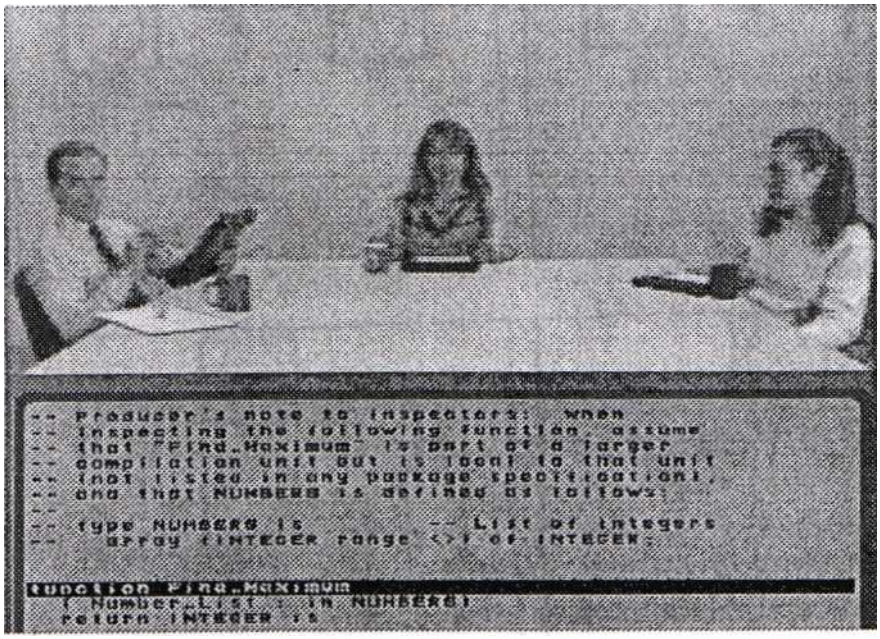

Figure 9.6 High visual fidelity simulation of a conference

their attitudes after the experience. Users with the surrogate travel interface came away from the experience with more positive opinions about the subject under investigation! While the surrogate travel interface was more cumbersome and slower, its users were more completely brought into the virtual world and developed better attitudes because of it.

Earlier evidence from a study of an interactive videodisc laboratory simulation also supported the view that more realistic interfaces had positive effects on users [Ste85]. That study showed that by creating a more visually abstract yet easier to use interface (analogous to the floor plan above), users tended to act as if the visually salient features of the interface were to be operated on rather than the functions those features represented.

Since the virtual conference room, multimedia meeting place described in Section 9.4 has yet to be implemented, the formal assertion that it is preferable to a "Brady Bunch" design

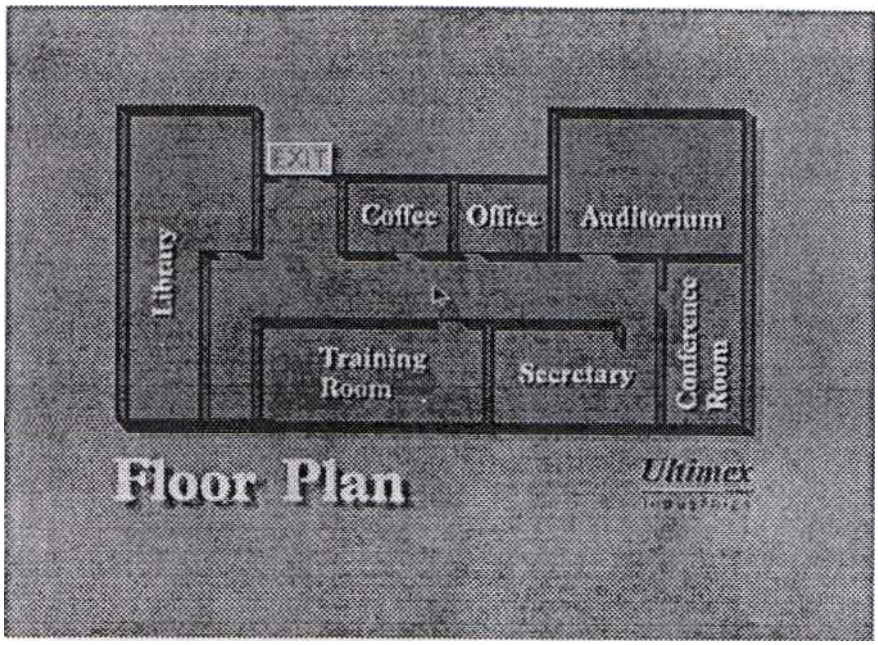

Figure 9.7 Component multimedia objects used to create the conference simulation

cannot yet be made. However, the work cited above suggests that in general a design will have positive effects if it more closely resembles a real-world environment. Such design fidelity more closely maps to users' day to day world and allows them to more naturally enter the computer's world.

9.5.2 Motion Fidelity

Besides design fidelity, multimedia applications must be concerned with spatial (resolution) and especially frequency (temporal or motion) fidelity. A simple view of resolution is that it is a function the number of lines per image and pixels per line. Assume a],000 by],000 pixel image completely fills a monitor. Now, if one cuts the number of lines and pixels per line each in half, there is one fourth of the information. However, if you display the new image on the same monitor, it looks like a smaller but equally clear image. Such operations in the frequency domain are not so benign. If we loose three out of four frames of thirty frame per second video there is no masking the effect. Until recently, how great an effect motion fidelity has was unknown.

Without special hardware many of today's multimedia operating systems and networks are incapable of thirty frame per second motion digital video. From an application's perspective these operating systems may make seemingly arbitrary presentation decisions, for example trading frame rate for resolution or synchronization. Is this lower motion fidelity, lower frame

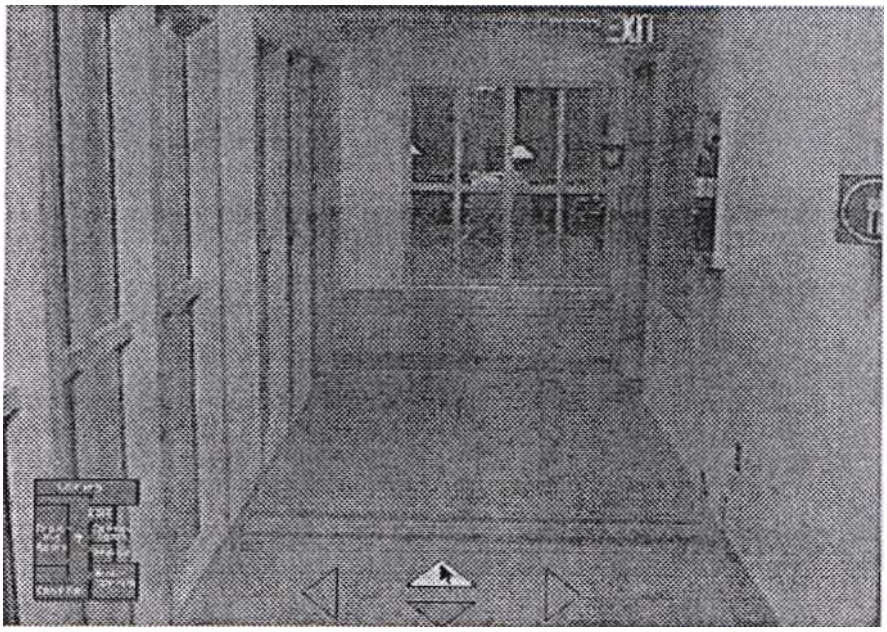

Figure 9.8 Point and click surrogate travel interfaces

rate video effective? Do differing levels of fidelity matter? Are four frames per second fast enough? Twenty-five frames? Thirty frames?

There are compelling reasons to believe that high frame rates are more than frills. In the same experiment cited above, Christel presented one group of users with thirty frame per second full motion video and audio in the various sub-worlds. Another group had the same experience except identical, but sequential still images, one every four seconds taken from the video, with audio. The users with the full motion video retained significantly more information even though the information tested on was contained in the audio.

In a second experiment, we compared five frame per second video to full motion video. The subjects were graduate students in a course on software engineering. The students were required to use the virtual workspace as part of the course requirements. With their consent, they participated in the experiment and were randomly assigned to one of two groups. One group's system presented full motion, thirty frame per second video. The second group had the same virtual workspace with identical video, except that the video was presented at five frames per second.

Seven out of ten users subjected to the five frame per second video refused to finish the experiment! Typical of the responses was one user's: "What did I do wrong in life that I had to go through this hell?" Such reactions should not be surprising. Photic stimulation in the one to eighteen cycle per second range has been shown to produce effects from migraines and

epileptic seizures to resonances of EEGs and induced visual imagery [FBK90, Gli87, RM90, WW49].

Even with text, a user's performance can be adversely affected when text is presented at less than fifteen characters per second [TAD85]. With video, users will not accept a "wait" of over one thirtieth of a second (or, according to PAL video standard advocates, one-twenty-fifth of a second) for the presentation of the next image.

The lesson seems to be clear. It is one thing to watch a 30 second QuickTime or Indeo movie running at five frames per second on a personal computer with no special hardware to improve multimedia performance. It is something quite different to sit for three hours viewing video presented at such a frame rate. In short, motion fidelity is critical. If a comfortable frame rate is not possible with the available hardware, users are better served with still images and audio. Beyond this, it is not just motion fidelity or design fidelity that makes a high quality multimedia application. Success comes from an understanding of users' needs, human factors, and system capabilities.

9.6 Outlook

There is a growing understanding of the appropriate design and functionality required by multimedia applications. Along with many other factors, developers must consider the complexity inherent in multimedia objects, create high fidelity designs, and factor in the temporal nature of multimedia.

In the future, multimedia applied to collaboration environments will permit multi-person workspaces that greatly enhance distributed groups' creativity and productivity. Beyond video conferences, multimedia computing will impact virtually every facet of human computer interaction. In each new domain, multimedia applications must be designed with a deep knowledge of both multimedia capabilities and human factors.

A seemingly simple multimedia object such as audio may be used or misused. In the analog world video is relatively difficult to mix. This has led to the Brady Bunch video interface even though in the digital world video is easily manipulated. Conversely, most digital audio components of multimedia systems make mixing digital audio difficult if there is only one DAC or DSP Requiring participants of a multimedia audio conference to "take the floor" before talking will certainly alter the meeting. There is no technical constraint that would require this. Digital audio data can be combined in an algorithmic fashion by the multimedia platform. Just because digital audio may be more difficult to mix is not sufficient justification to constrain the user. Developers must look to the requirements of the user.

Multimedia technology is becoming widely available and it can deliver more information, more effectively than any scheme developed to date. But more than just delivering information, effective design of multimedia systems will require a deep understanding of the roles of interaction with huge volumes of information, simulation, learning, and even virtual reality. Ultimately, multimedia computing will allow communication, access to information, and learning to become more active and compelling.

Both aiding other applications and as applications in their own right, multimedia visualization and presentation systems will intelligently present information based on users needs, human factors, and even cinematic principles [CS92, DSP9], Ste89]. Artificial intelligence applied to multimedia will permit intelligent interfaces, agents, guides, and anthropomorphic systems [Lau86].

This chapter has just begun to scratch the surface of multimedia computing's role in human computer interaction. Advanced multimedia applications require much more of developers and computing systems than do today's interrupted video. Rather than develop the multimedia equivalent of a teletypewriter I/O paradigm, there is today an opportunity to avoid the mistakes of the past. Through a creative, multi-disciplinary approach, a new multimedia paradigm will emerge that provides a quantum leap beyond today's GUIs.

References

[Chr91] M. G. Christel. *A Comparative Evaluation of Digital Video Interactive Interfaces in the Delivery of a Code Inspection Course.* PhD thesis, Georgia Institute of Technology, Atlanta GA, 1991.

[CS92] M. B. Christel and S. M. Stevens. Rule base and digital video technologies applied to training simulations. In *Software Engineering Institute Technical Review 92.* Software Engineering Institute, Pittsburgh, PA, 1992.

[DMS92] L. Degen, R. Mander, and G. Salomon. Working with audio: Integrating personal tape recorders and desktop computers. In *Proceeding of ACM CHI '92 Conference on Human Factors In Computing Systems,* 1992.

[DSP91] G. Davenport, T Smith, and N. Pincever. Cinematic primitives for multimedia. *IEEE Computer Graphics and Applications, July* 1991.

[Esp93] C. Esposito. Virtual reality: Perspecitves, applications, and architectures. In Bass and Dewan, editors, *User Interface Software,* Trends in Software Series. Wiley, 1993.

[FBK90] A. I. Fedotchev, A. T. Bondar, and V. F Konovalov. Stability of resonance eeg reactions to flickering light in humans. *International Journal of Psychophysiology,* 9, 1990.

[Gli87] J. Glicksohn. Photic driving and altered states of consciousness: An exploritory study. *Cognition and Personality, 6(2),* 1986-87.

[Had45] J. Hadamand. *The Psychology of Invention in the Mathematical Field.* Princeton University Press, Princeton, NJ, 1945.

[HSA89] M. E. Hodges, R. M. Sasnett, and M. S. Ackerman. A construction set for multimedia applications. *IEEE Software,* January 1989.

[IK92] H. Ishii and M. Kobayashi. Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proceeding of ACM CHI '92 Conference on Human Factors In Computing Systems,* 1992.

[Kra87] R. Kraft. The influence of camera angle on comprehension and retention of pictorial events. *Memory and Cognition,* 15(4), 1987.

[Kra88] R. Kraft. Mind and media: The psychological reality of cinematic principles. In D. Schultz and C.W. Moody, editors, *Images, Information and Interfaces: Directions for the 1990's.* Human Factors Society, New York, NY, 1988.

[Lau86] B. Laurel. Interface as mimesis. In D. Nomran and S. Draper, editors, *User Centered System Design,* chapter 4. Lawrence Erlbaum Assoc., 1986.

[Lip80] A. Lippman. Movie-maps: and application of the optical videodisc to computer graphics. *Computer Graphics,* 14(3), 1980.

[LM91] K-Y Lai and T.W. Malone. Object lens: Letting end-users create cooperative work application. In *Proceeding of ACM CHI '91 Conference on Human Factors In Computing Systems,* 1991.

[MBS+91) M. M. Mantei, R. M. Baecker, A. J. Sellen, W. A. S. Buxton, and T Milligan. Experiences in the use of a media space. In *Proceedings of ACM CHI'91 Conference on Human Factors In Computing Systems,* 1991.

[NHK91] NHK. *1991 Technology Open House.* NHK Engineering Services, Inc., Tokyo, Japan, 1991.

[Off85] Xerox Corporation's Quality Office. *Leadership Through Quality; Mining Group Gold; A Guide Providing Facilitation Techniques, Tips, Checklists and Guidesheets.* Multinational Customer and Service Education Reprographic Business Group, Xerox Corporation,

Rochester, NY, July 1985.

[Per90] L. J. Perelman. A new learning enterprise. *Business Week,* December 10 1990.

[Res89] H_ L. Resnikoff. *The Illusion of Reality*. Springer-Verlag, New York, NY, 1989.

[RM90] A. Richardson and F. McAndrew. The effects of photic stimulation and private self-consciousness on the complexity of visual imagination imagery. *British Journal of Psychology,* 81, 1990.

[SFC89] S.M. Stevens, R. G. Fuller, and M. G. Christel. *Workshop on Intelligent Tutoring Systems and **Digital Video.*** American Association of Physics Teachers and Software Engineering Institute, Pittsburgh, PA, 1989.

[SS88] D. Stone and A. Stone. The seat of power. *Carnegie Mellon Magazine,* Winter 1988.

[Sta93] J. Stasko. Animation in user interfaces: Principles and techniques. In Bass and Dewan, editors, *User Interface Software,* Trends in Software Series. Wiley, 1993.

[Ste85] S.M. Stevens. Interactive computer/videodisc lessons and their effect on students' understanding of science. In *National Association for Research in Science Teaching: 58th Annual NARST Conference,* Columbus, OH, 1985. ERIC.

[Ste89] S. M. Stevens. Intelligent interactive video simulation of a code inspection. *Communications of the ACM,* July 1989.

[Ste92] S. M. Stevens. Next generation network and operating system requirements for continuous time media. In R.G. Herrtwich, editor, *Network and Operating System Support for Digital Audio and Video.* Springer-Verlag, New York, NY, 1992.

[TAD85] J. W. Tombaugh, M.D. Arkin, and R. F Dillon. The effect of vdu text-presentation rate on reading comprehension and reading speed. In *Proceedings of ACM CHI '85 Conference on Human Factors In Computing Systems,* 1985.

[WW49] V.J. Walter and W -G. Walter. The central effects of rhythmic sensory stimulation. *Electroencephalography and Clinical Neurophysiology, 1,* 1949.

[YHMD88] N. Yankelovich, B. J. Haan, N. K. Meyrowitz, and S. M. Drucker. Intermedia: The concept and the construction of a seamless information environment. *IEEE Computer,* January 1988.