JCMC 4 (3) March 1999

The Human Factors of MBone Videoconferences: Recommendations for Improving Sessions and Software

Andrew S. Patrick
Communications Research Centre, Ottawa

Table of Contents

Abstract

The "MBone" is the portion of the Internet that has implemented a new computer network technology called "multicasting". Multicasting supports efficient distribution of network traffic to multiple users simultaneously. Videoconferencing is the most common MBone application in use today and this paper reviews the human factors issues related to MBone videoconferencing. Three parameters of human factors concerns are defined: the task of the session (meeting, education, entertainment), the media used during the session (video, audio, shared workspace), and the communication modes involved (interactive vs. non-interactive and formal vs. informal). Videoconference sessions and software can be placed in this parameter space and this can provide valuable information about the technical and human requirements. The human factors research literature relevant to each of these parameters is reviewed and the current MBone tools are analyzed. Specific recommendations are made for MBone session organizers and software developers. These recommendations are not all specific to multicasting and will be of interest to people developing or using any videoconference system.

Introduction

Imagine a new Internet service where users can see and hear each other over the network in real time. Imagine that this service lets users talk, debate, share information, be entertained, and learn. These users can be located anywhere in the world and their picture and voice will be transmitted instantly. The users can come together in pairs or small groups for intimate conversations, or they can form very large groups, numbering in the thousands, to witness an important event. Dream that this new service is free, once the users are connected to the Internet, using software technology that is readily available for all modern computer systems. Finally, imagine that the service does not place astronomical loads on the Internet as more and more users learn to enjoy it.

This new Internet service is not a dream -- it is available today on some networks around the world and it is called "The MBone" (Kumar; 1996; Macedonia & Brutzman, 1994). It is based on two recent developments in computer and network technology. The first is an advance in multimedia codecs that allows audio and video to be encoded and compressed, transferred over the Internet, and uncompressed and decoded on standard desktop computers. The second is a new Internet technology for transferring multimedia data efficiently.

The "MBone" is defined as the portion of the Internet that has implemented these developments, and it is growing every day (Kumar, 1996). It is likely that this new service will have a very large impact on the Internet and how it is used, perhaps as large as the impact of the World Wide Web (see Mark Weiser's forward to Kumar, 1996). Thus, the MBone may be the next "killer application" for the Internet.

There are a number of factors that are contributing to the rate that the MBone is being deployed. First, implementing the MBone is a difficult technical task that requires advanced knowledge of Internet protocols and routers. Second, support for the MBone protocols must be provided in the network itself and cannot be installed by end users. Third, the MBone technologies have developed gradually and have been relatively unstable and unpredictable in the past, making many potential users hesitant. Fourth, the end user applications developed for the MBone are still immature and difficult to use especially in comparison to more highly developed multimedia applications. Fifth, the Internet is often congested and this leads to unacceptable audio and video quality. Sixth, there is little compelling content available on the MBone and too few people to support personal communications and too little high quality production material to support a mass audience. Seventh, the content that is available on the MBone is usually badly produced, often created by pointing a camera at the front of a classroom, making the service pale in comparison to television or films.

The goal of this report is to examine some of the reasons why the MBone is developing slowly and make recommendations that may accelerate its adoption. The MBone is very much like broadcasting and videoconferencing combined and yet many of the lessons learned in these areas have been ignored. There is much to be learned from examining previous research on videoconferencing systems in particular that can be applied to improving the MBone.

Gale (1992) suggests that "videoconferencing brings with it a history of failed promises" (pg. 517). Videoconferencing was very popular for a time and many corporations built special conference rooms and installed thousands of dollars worth of equipment. Promises were made about better communications and reduced travel budgets. Today, videoconference rooms sit empty and remote conferences have not replaced travel. The MBone is also very promising but it could suffer the same fate as videoconference rooms. By examining how videoconferencing systems are used and not used, and how the technology affects the people who use it, we can learn valuable lessons that may help to make the MBone more useful and valuable.

Multicasting and the MBone

Multicasting is a new computer network technology emerging on the Internet that allows the efficient distribution of network traffic to multiple users simultaneously. Multicasting can be understood by contrasting it with two other network technologies, broadcasting and unicasting (see Figure 1). In broadcasting, traffic sent from a source is distributed to all segments of the network, even portions where there are no receivers for the message. Broadcasting is efficient because only one copy of the data is sent, but it is inefficient because the data is sent to portions of the network where there may be no receivers.

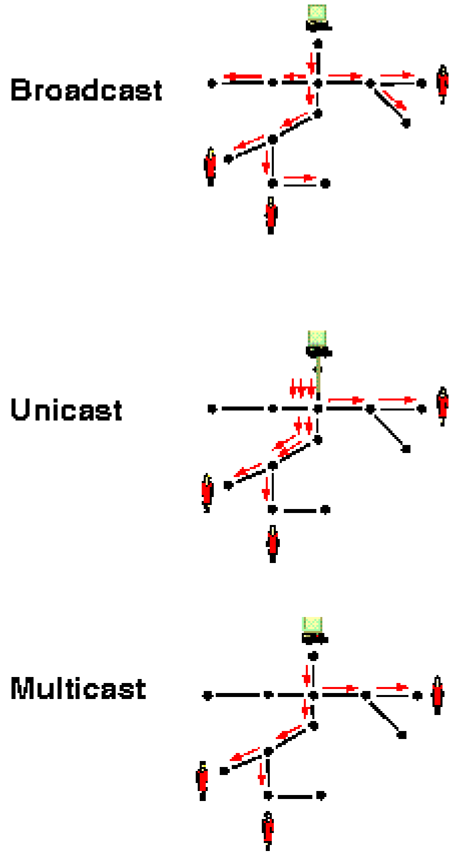

Figure 1: Comparison of broadcast, unicast, and multicast technologies. The computer icon represents a sender, and the person icons represent recipients.

Unicasting, on the other hand, sends data only to the portions of the network where there are users interested in receiving it, making it more efficient than broadcasting in that respect. With unicasting, however, a new copy of the data must be sent for each individual receiver and this makes the technology inefficient.

Multicasting combines the best aspects of broadcast and unicast. Only one copy of the data is sent and it is only sent to the portions of the network where there are users interested in receiving it. Multicasting is implemented by routing processes that keep track of the data streams in use and adjust the data routing rules accordingly.

A simple example was used in Figure 1 where there is one sender and multiple receivers. With all of these distribution methods, however, it is possible for all the users to act as both senders and receivers. This situation is shown for the multicasting technology in Figure 2. Here all the users are acting as peers, each sending and receiving network traffic. This situation, which is common during interactive videoconferences, is made much more efficient by the multicasting technology.

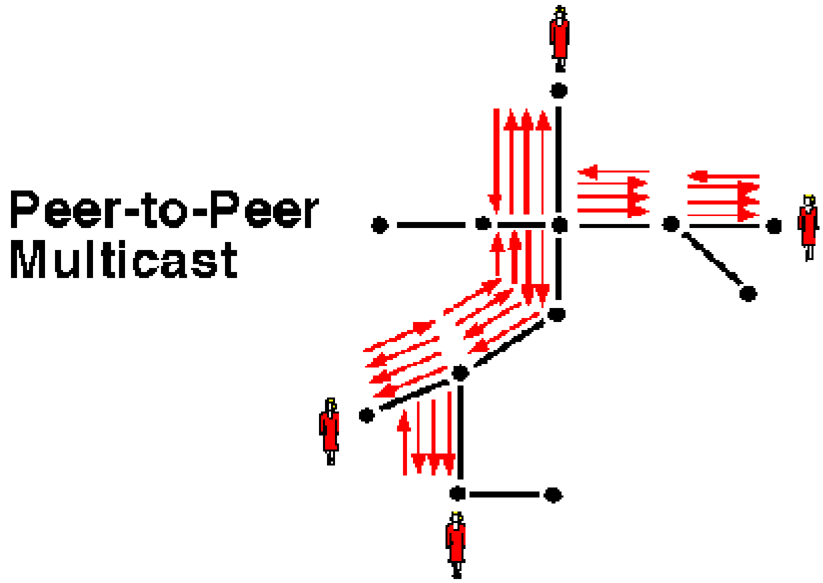

Figure 2: Multicasting with numerous sender/receivers.
The person icons represent sender/receivers.

Enabling multicasting does take time and technical expertise so the MBone primarily consists of large institutions with good network administration resources. Moreover, since the most popular multicast application, videoconferencing, does require significant amounts of bandwidth (500 Kb/s is considered a starting point for an institutional link), multicasting is most popular with institutions that have a large Internet access capacity. Even with these restrictions, however, the MBone has grown to include thousands of networks around the globe and is currently available to perhaps 50,000 Internet users. Some groups are working to encourage the deployment of multicasting more widely, including industry consortia and an IETF "MBone Deployment" working group. Multicasting will also be a standard component of IPv6, the next generation Internet protocol currently under development. Thus, multicasting is an efficient and useful technology that will be deployed on the Internet and it will likely be an important component when building network applications for the future.

MBone Videoconferencing (MVC)

Videoconferencing is the most common MBone application in use today. Typical videoconferencing applications digitize images and sounds at one site, encode them into a standard format, and transmit the resulting data over a network connection. At the receiving site the data is collected, decoded, and presented to the user. The multimedia data produced during videoconferencing is very bandwidth demanding. Audio encoding schemes typically require 71 Kb/s, while low quality video (2-5 frames per second) requires 128 Kb/s. Multicasting provides a distinct advantage for videoconferencing because of its ability to distribute data efficiently to multiple users without redundancy on the network connections.

MBone Videoconferencing (MVC) is currently being used in a variety of communication situations. One of the main uses is to transmit scientific and technical conferences to users around the world. For example, the Internet Engineering Task Force (IETF) is the forum where new Internet technologies are developed and standardized. Remotely attending IETF meetings, which take place 3 times a year around the world, was one of the initial motivations behind the MBone and the MVC applications. The MBone technologies are a product of the IETF and transmission of IETF conferences continues to be a significant use of the MBone. MVC is also used for similar sessions, including lectures and seminars from university campuses, concerts and other forms of entertainment, and current events (such as coverage of the NASA shuttle missions).

Each of these examples, however, involves a one-way form of communication -- there is one transmitter (the conference location or event organizer) and many receivers passively view the material. MVC can also be used in two-way situations where multiple users are both transmitters and receivers. Collaborative work meetings are the best example, where workers discuss topics and work on shared documents. Collaborative work sessions are not as common as one-way sessions on the MBone but they do occur. (It is possible that two-way MVC sessions are more common than first appears, as most collaborative sessions may not be advertised publicly.) For this paper, videoconferencing is defined broadly to include both one-way and two-way sessions, and sessions that may, or may not, involve a video channel and other communication channels.

The MERCI Project

MERCI
(for Multimedia European Research Conferencing Integration) was a research and development project coordinated by University College London (UCL) and supported by the European Commission Fourth Framework Telematics Application Programme. The goal of MERCI was to further the development of multimedia conferencing tools. The project consisted of a number of work packages, one of which was concerned with "usability and assessment". This package was to ensure that user needs were fully understood and the MVC tools were usable. This paper was prepared as part of the MERCI project to review the human factors literature with a focus on issues related to MBone videoconferencing. The review is not intended to be exhaustive but instead covers recent papers that are relevant to the MBone technology.

MBone Videoconferencing in Perspective

There is a long history of videoconferencing, perhaps beginning in the 1950's with AT&T's Picturephone. This device combined the normal audio functions of a telephone with a slow video transmission from each end of the link. The Picturephone was first demonstrated widely at the 1964 World's Fair and the interest was very high, with the public lining up for hours to see the new technology (Gale 1992). Later, Picturephones were installed in Pittsburgh and AT&T had hopes of deploying more than one million phones nationwide by 1980. The Picturephone was not successful, however. The equipment was too bulky and the controls were awkward to use. The pictures were small and the image quality was poor. Finally, there was some suggestion that people did not want to be seen when answering the telephone, especially in their homes.

In the 1980's and early 1990's there was great interest in room-based videoconference systems. Cameras and video monitors were installed in special meeting rooms linked together by telecommunication lines, often ISDN or dedicated leased lines. Special equipment was installed to control the devices and compress the multimedia data for transmission. A lot of money was spent on room-based videoconference systems, including money for custom rooms and furniture to accommodate the technology (e.g., special lighting and trapezoid-shaped tables). There were high hopes that the videoconference systems would allow people to communicate over vast distances, work better, and reduce travel budgets.

The reality is that room-based videoconference systems have not been very successful (Tang & Isaacs, 1992). The equipment is expensive, not standardized, and difficult to use. The systems require specialized connections and leased lines that are expensive and difficult to administer. On the human side, the conferences have to be booked in advance and this requires formal meeting planning and limits spontaneous use. Finally, room-based systems often have a limited ability to share the objects of work, data and applications, and this makes effective communication difficult.

The latest trend is desktop videoconferencing (which includes MBone videoconferencing), where the audio and video interaction is done in the users' home or office using multimedia desktop computers. Such systems eliminate some of the problems of room-based systems, such as scheduling problems and the need for special rooms. In addition, the desktop videoconferencing system can be integrated into the familiar computing environment and support sharing data and desktop applications. Since these systems are in the office, users can do other office tasks while they are conferencing. This can be an advantage or disadvantage depending on the situation.

Desktop videoconferencing systems use one of two communications schemes: circuit-based or packet-based (Willebeek-LeMair & Shae, 1997). Circuit systems use a dedicated connection for the videoconference, usually ISDN and the H.320 codec standards. In contrast, packet systems usually use general data connections for the videoconference, often Ethernet or Token Ring LANs. Circuit systems have the advantage of dedicated bandwidth for the videoconference, which means that other data traffic will not interfere. Circuit systems require additional equipment, however, when more than two parties want to join a conference. Multipoint Control Units (MCU) must be used to connect multiple parties into a single conference. The MCU controls how the audio and video is distributed to all the conference participants. The multimedia routing performed by the MCU can be simple or complex. For example, MCUs may simply distribute the video according to which party is currently sending audio so all users only see the speaker. More sophisticated MCUs can give the users some control over what they see during a conference, or even multiplex all the views into a combined video stream.

Packet systems have an advantage because MCUs are not required -- all the participants receive all the videoconference data. This does mean, however, that the end-user application must take on the role of the MCU and make decisions about who is seen and heard during a videoconference (e.g., decoding all the video or following the speaker). Packet systems usually do not have dedicated bandwidth for the videoconference so other network traffic can cause interruptions. Also, because there is no specialized videoconference channel and the multimedia bit rates can vary, synchronization of the various multimedia streams can be more difficult in packet systems. MBone videoconferencing is an example of a new packet-based system that uses an efficient protocol (multicasting) for distributing data to multiple receivers.

The Human Factors of Videoconferencing

In parallel to the history of videoconferencing, there has been a history of concerns about the human factors of mediated communication systems. Much of the development and use of these systems has been driven by technology advances with little consideration for actual human communication. It has become clear that successful systems will have to take into account what is known about human communication. Developers need to understand the nature of the communication that takes place in a particular location or context. Failure to do so may lead to technology that is not used or not trusted. For example, Harper et al. (1991) found that a new automated air traffic control system was under-used and distrusted because the designers failed to consider the ways controllers communicate in their jobs.

It is also important to examine what communication functions are enabled by various technologies, and what may be hindered. Tang et al. (1994) found that technologies that took too long to setup and were unreliable tended to hinder communication and users avoided them.

Consideration must also be given to how organizations adapt and change with the introduction of technology.

All of these concerns apply to MBone videoconferencing. After reviewing the MVC technology, we will consider the forms of communication that are facilitated and hindered by this technology. We will also consider the human context set by the tasks being undertaken when the technology is used.

MVC Software Tools

A variety of MBone videoconferencing tools has been developed. Most of the applications were developed by IETF committee members and are freely available, although some commercial applications are starting to appear. The most common platform for building MVC applications has been Unix and most of the interfaces have been built using Tcl/Tk. This combination gives the applications a modular form (separate programs for each function) and a similar "look and feel". MVC applications are now being ported to other platforms, most notably Windows95/NT, but they often retain the Unix tradition and visual appearance.

Video

The most common video application is "VIC" (for VIdeo Conferencing), developed by the Network Research Group at the Lawrence Berkeley National Laboratory (LBL). As can be seen in the right side of Figure 3, VIC displays a thumbnail-sized picture of each transmitting source along with identification and bandwidth information. VIC users can select one or more of these sources for a larger display, as is shown on the left side of Figure 3. Users can control their video transmissions by selecting options from a "menu" panel (not shown).

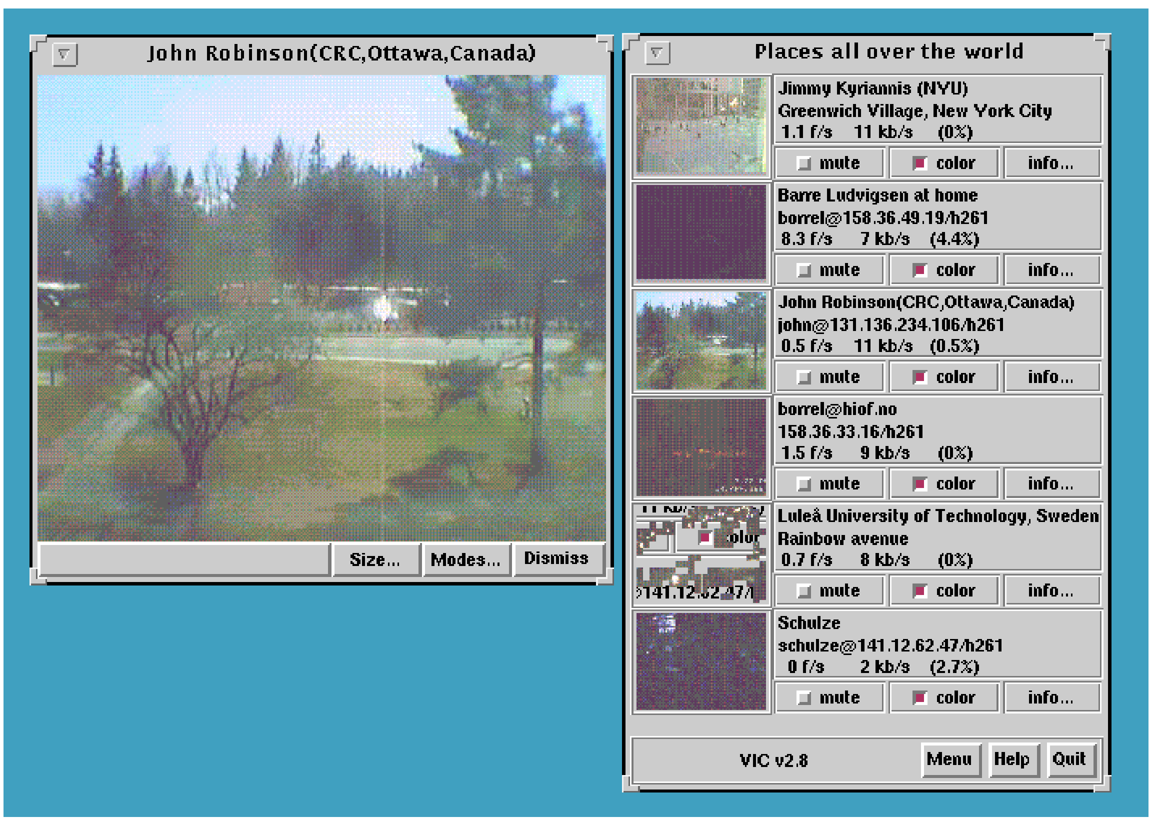

Figure 3: The VIC interface.

VIC supports a number of video encoding schemes, including JPEG, H.261, and CELL B. VIC also supports a variety of frame rates and bandwidth settings that are under the control of the sender. When the H.261 encoding is used with a typical bandwidth setting of 128 Kb/s the result is a video frame rate of 4-6 frames per second for most video material.

Audio

There are two audio applications in common use today. "VAT" (for Visual Audio Tool) was also developed at LBL. The VAT interface is shown in Figure 4. A list of users participating in the session is displayed on the left side of the tool while microphone and speaker controls are displayed on the right. VAT users can obtain more information about the other users by selecting a name in the list. Again, users control various audio options from a "menu" panel (not shown).

Figure 4: The VAT interface.

The VAT tool supports a number of audio encoding schemes that differ in their sampling rate and packet frame size: pcm (78 Kb/s), pcm2 (71 Kb/s), pcm4 (68 Kb/s), dvi (46 Kb/s), dvi2 (39 Kb/s), dvi4 (36 Kb/s), gsm (17 Kb/s), and lpc4 (9 Kb/s). VAT is most often used in pcm2 mode, which results in audio quality that is equivalent to a telephone call. When audio packets are lost or delayed on the network, however, the audio quality deteriorates rapidly.

"RAT" (for Robust-Audio Tool) was developed at University College London (UCL) and it uses an interface that is very similar to VAT (see Figure 5). Again, the list of participants is shown on the left and the audio controls are on the right. RAT also supports the pcm, dvi, gsm, and lpc audio coding schemes with different packet frame sizes. The main distinction between RAT and VAT is that RAT supports a redundant audio-encoding scheme that maintains audio quality over congested Internet connections. Redundant coding means that each audio packet is sent twice so that any missed fragments can be replaced before the audio is played to the user.

Figure 5: The RAT interface.

Another MVC audio tool is Free Phone, developed by INRIA in France. Free Phone is similar to RAT in that is supports redundant audio transmission. Free Phone also supports high sampling rates like those used in CD (44 kHz) and DAT (48 kHz) recordings.

Shared Workspace

Whiteboard

There is a shared whiteboard application for MVC called "WB" (for White Board), also developed at LBL. As is shown in Figure 6, WB users can draw on the large canvas on the left side of the tool and all the participants will see the results on their screens. WB users can select from a variety of drawing tools, shown on the right side of the tool, or simple type onto the shared space. Text files and graphical images can also be imported into WB, although this can be difficult to accomplish.

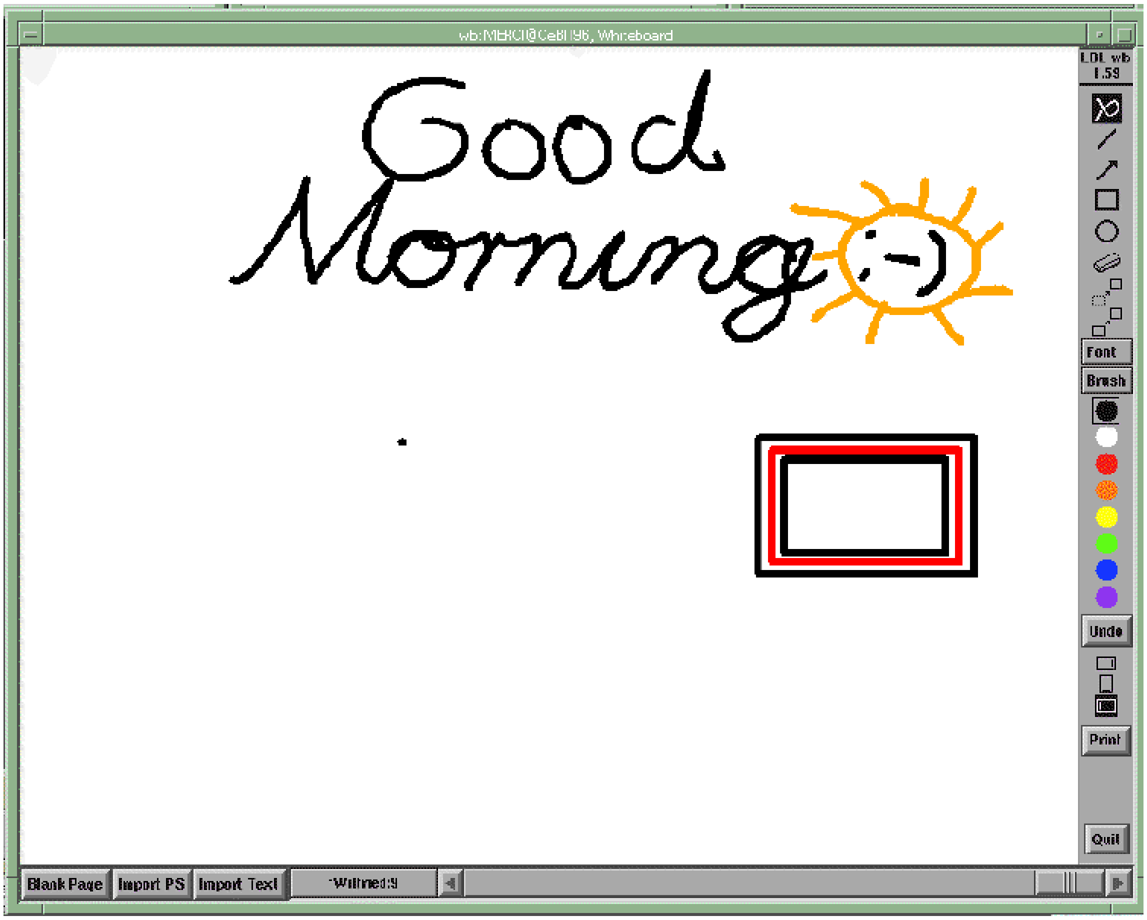

Figure 6: The WB interface.

Text Editor

A similar shared work tool is "NTE" (for Network Text Editor; sometimes called "nt"), developed at UCL. NTE is specialized for sharing text documents and it provides tools for editing and moving blocks of text that are not available in WB. Figure 7 shows that users work on text on the left side of the tool and information about the session participants and controls are on the right side.

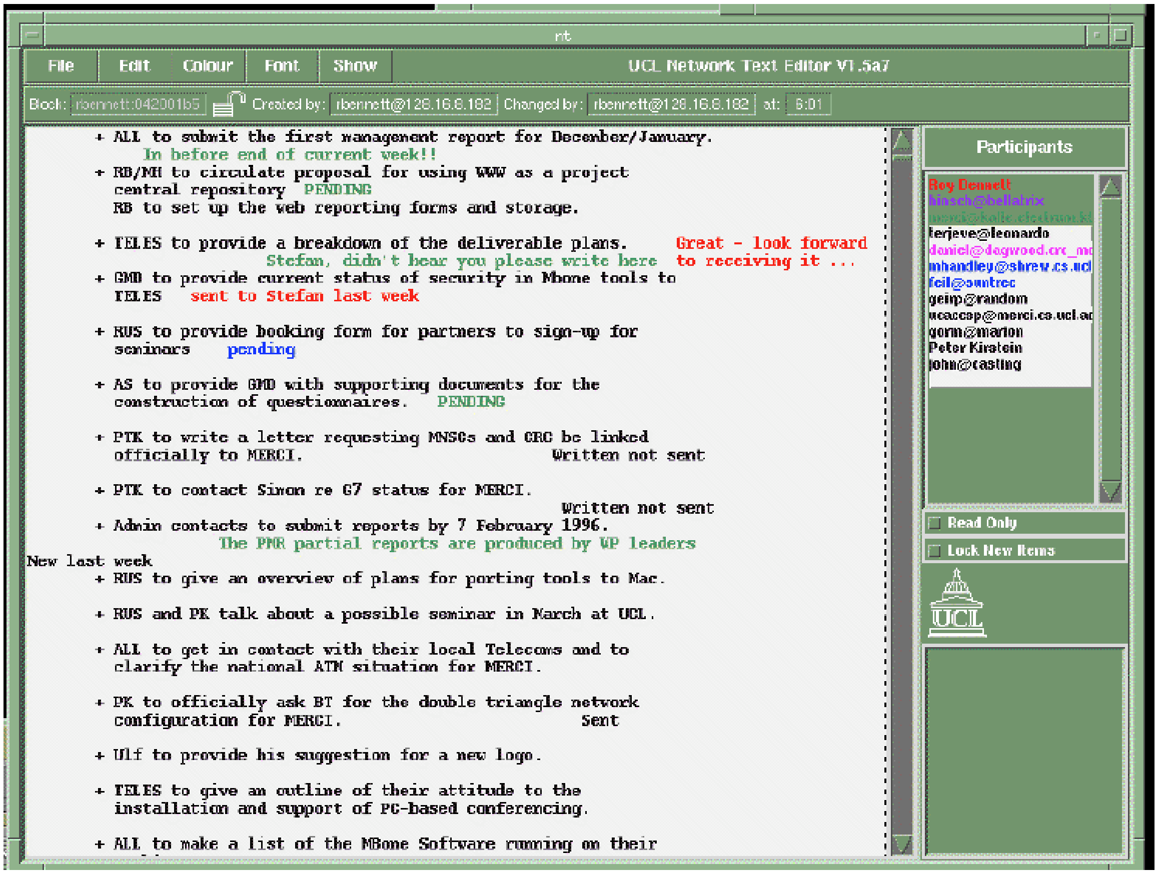

Figure 7: The NTE interface.

A Typical MVC Session

Each of these tools may be used during a typical MVC session. Users may communicate via audio, view each other using video, and work on a shared document at the same time. A screen layout for a typical session is shown in Figure 8. Here the VAT audio tool is located in the upper right corner of the screen. One of the participants' names is highlighted in VAT to indicate who is currently speaking. The NTE shared editor is on the lower right of the screen and various users have been working on a shared document using different colours of text. The VIC video tool is shown on the lower left and each of the thumbnail views of the participants have been enlarged and placed on the upper left of the screen.

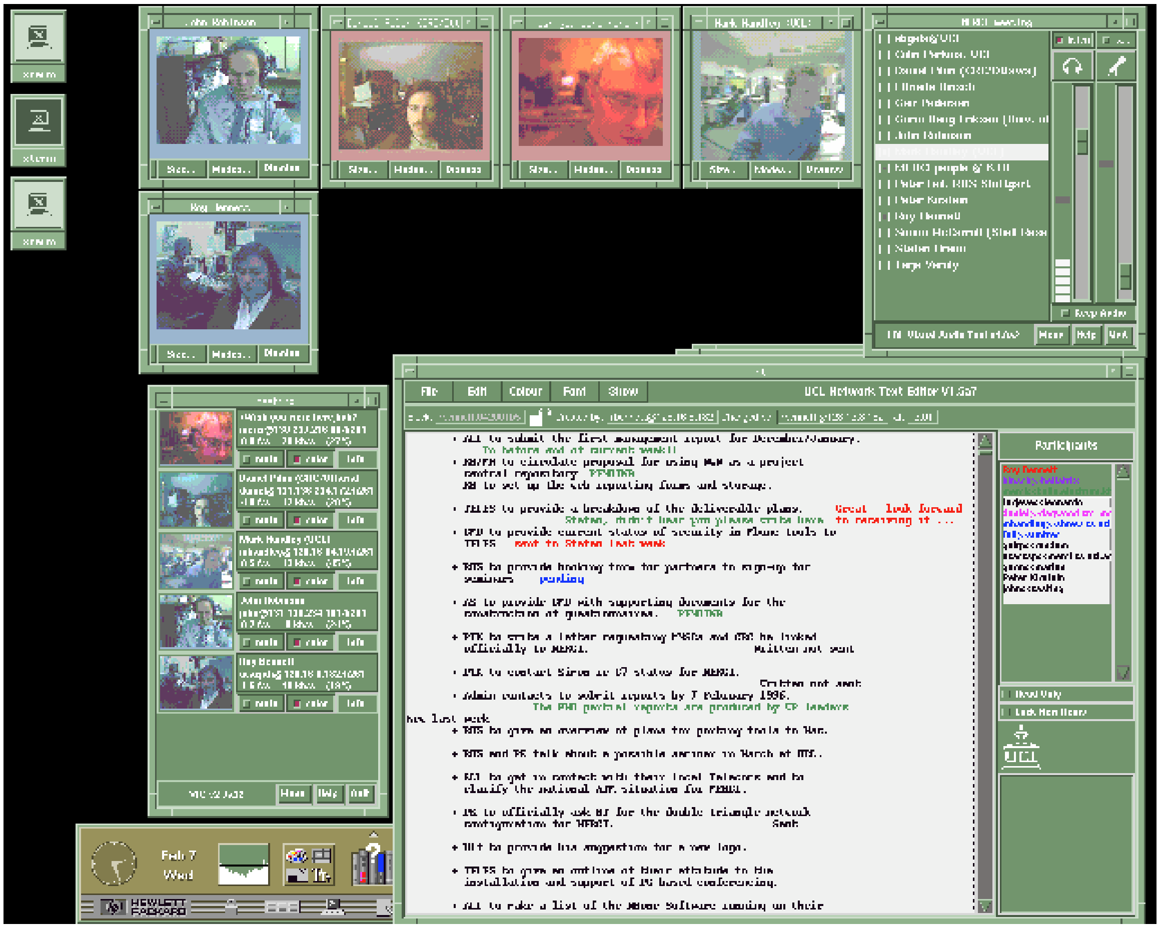

Figure 8: Screen capture of a typical MVC session.

The Session Directory

MVC sessions can be announced using "[SDR]", a session directory tool developed at UCL. SDR has a role similar to a "TV Guide" where users can learn about scheduled MVC sessions and create sessions of their own. The main SDR panel is shown in Figure 9. Here each MVC session is listed in alphabetical order and users can get more details about a session by clicking on the name. The detailed information panel (see Figure 10) includes a description of the session, its timing, and the types of media being used. SDR users can also view a calendar of MVC sessions, as is shown in the Figure 11. Each date on the calendar can be examined to determine when sessions have been scheduled.

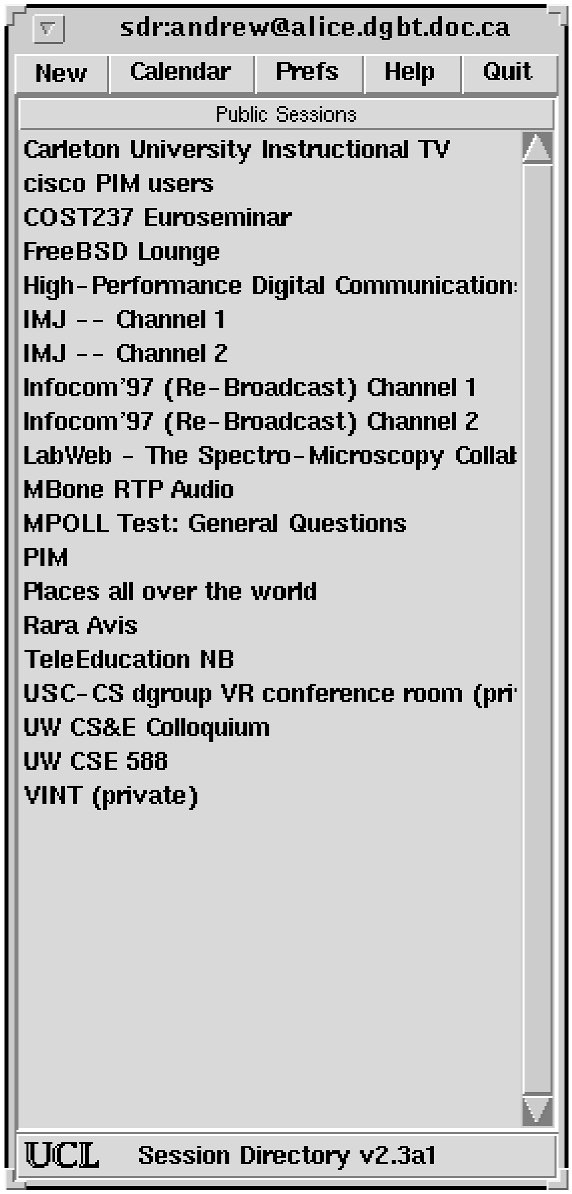

Figure 9: The main session list from the SDR interface.

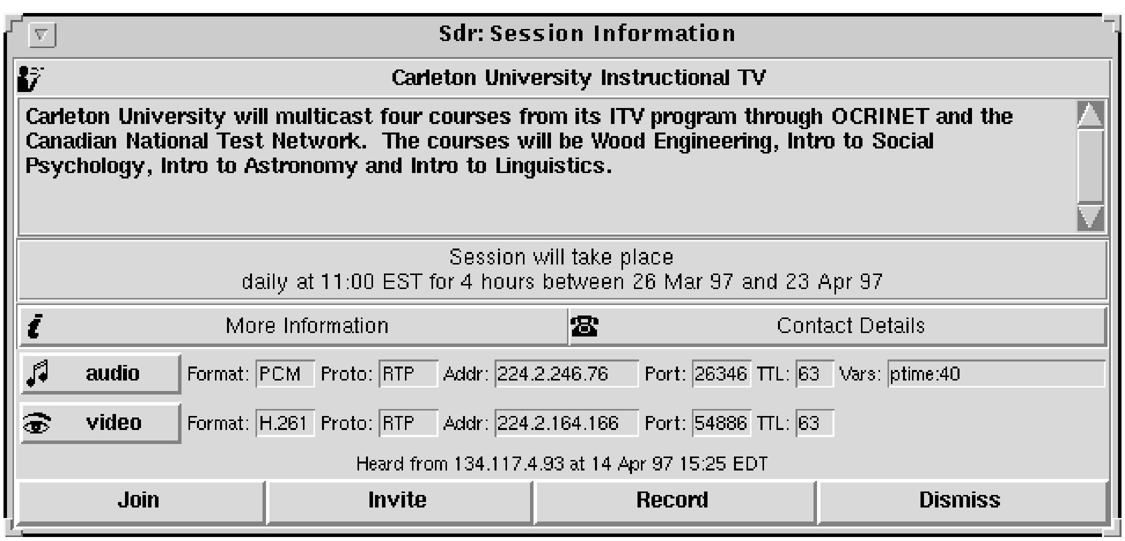

Figure 10: The session information panel of the SDR interface.

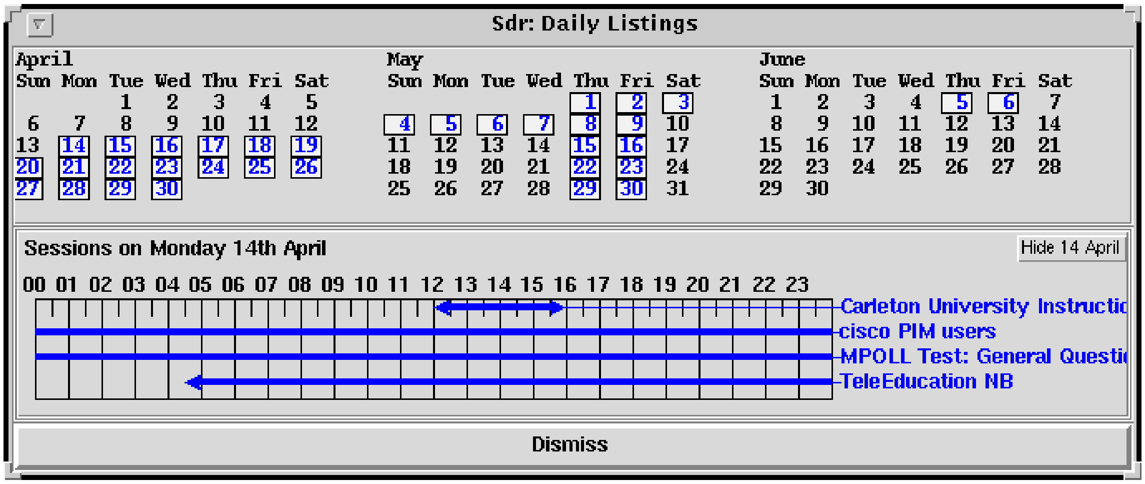

Figure 11: The calendar panel of the SDR interface.

Other Tools

Other software tools have been developed for MBone videoconferencing. For example, the author has created MPoll, a tool for collecting opinions and ratings during MVC sessions. MPoll can be used to present multiple-choice and short answer questions to users and the results are collected and displayed in real-time. MPoll has been used to collect quality and satisfaction ratings during some MVC sessions. It was also used during MERCI administration meetings to support decision-making. For example, MPoll was used to collect votes on approving the minutes of the previous meeting and to support management decisions.

mMosiac
is multicast-enabled WWW browser that permits users to share HTML pages. The URL for WWW pages of interest are multicast to all members of the session and this ensures that everyone is viewing the same page at the same time. This feature makes mMosaic a suitable application for slide displays during a remote presentation (and an alternative to WB).

Another software tool is the Image Multicaster (IMM), a tool for reliably distributing images to a group of multicast participants. This tool uses a repair scheme to ensure that data files are successfully multicast to all receivers even when there are data losses due to network congestion. IMM has been used on the MBone to distribute satellite weather photos and it is being expanded to support distribution of arbitrary data files. This would make IMM useful for multicasting news and software updates where many users want to receive the same data at the same time. It could also be used during MVC sessions to distribute software tools needed by the participants or for sharing binary document files (e.g., word processor files or spreadsheets).

Commercial Products

There have been some efforts to develop commercial MVC software but no current products support interactive videoconferencing (i.e., two-way communications). Precept Software Inc. has developed a product called IP/TV that uses multicasting to distribute audio and video to desktop computers (Windows 95 or NT). This software is based on a one-way model where a server transmits the programs and clients receive the transmission. There are no interactive features in the IP/TV system.

ICAST Corporation
has developed a similar multicast product called "The Company Channel" that also has separate modules for transmission and viewing. Again, there is no support for interactive use. Finally, Real Networks
has begun to support multicasting in their popular Real Audio and Real Video products. These products are also based on a one-way model and no interaction is possible.

Taxonomy for Understanding Videoconferencing

Videoconferencing is a complex service and any study of videoconferencing must examine the issues in a number of ways. Figure 12 shows a schematic representation of one way to organize a study of videoconferencing. Here three parameters are defined: the task of the session, the media used during the session, and the communication modes involved. These three parameters define a human-factors space for videoconferencing where the location in the space represents various demands and requirements on the technology and the users. Videoconference sessions can be placed in this parameter space and this can provide valuable information about the technical and human requirements, and the similarities between sessions. For example, the bi-weekly MERCI management meetings were characterized as a meeting task, an interactive and formal communications mode, and involving the video, audio, and shared workspace media. On the other hand, the constantly running "Radio Free VAT" session on the MBone is characterized as an entertainment task, a formal, non-interactive mode, and the audio medium. These two sessions have different human-factors characteristics and examining them according to these parameters is valuable for understanding those differences.

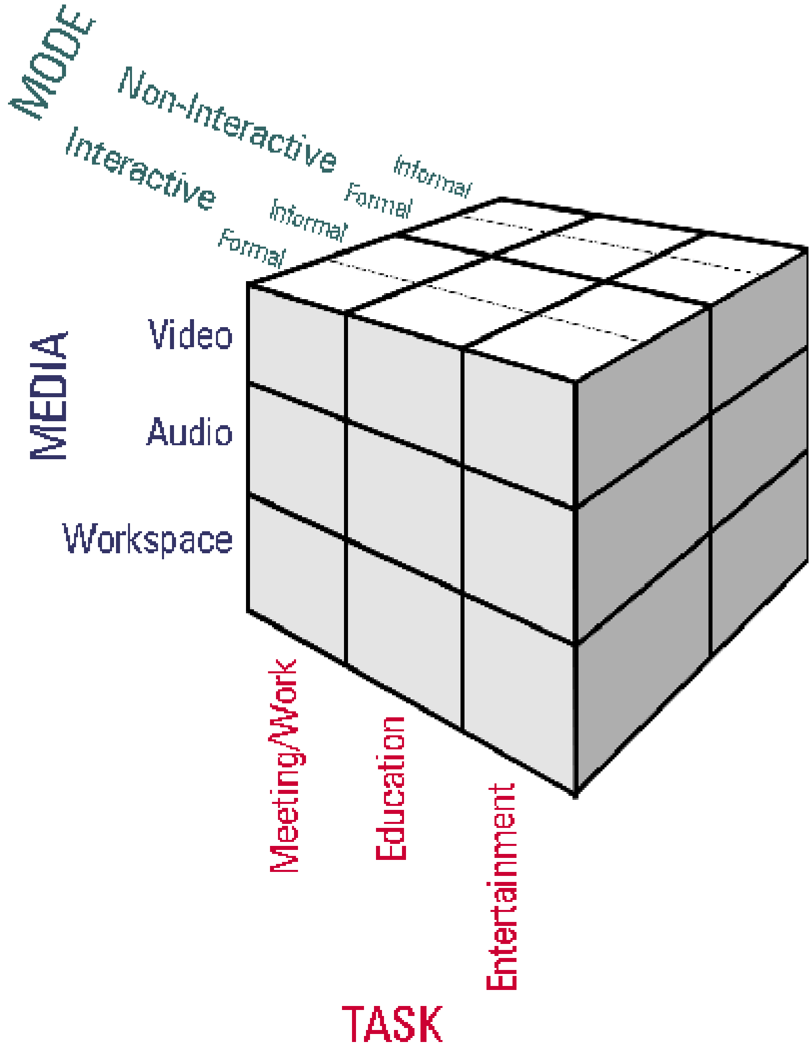

Figure 12: Schematic representation of videoconference taxonomies.

The media parameter makes a distinction between the video, audio, and shared workspace components. Although these media types can influence each other during a videoconference, research has often been conducted on one media type at a time. For example, there has been a great deal of research on the role of video in remote communication by comparing conferences with and without a video component (e.g., Ochsman & Chapanis, 1974; Gale, 1990). Other research has focused on the audio component, examining such issues as quality and synchronization (e.g., Watson & Sasse, 1996; Tang & Isaacs, 1992).

The mode parameter focuses on the styles of communication that are used in videoconferencing, including the sub-parameters of interactive versus non-interactive and formal versus informal. Research has shown that these different modes of communication are important for understanding the use and success of videoconferencing. For example, Rettinger (1995) suggests that "lecture" and "collaborative" are two distinct modes of remote interaction with different demands on a videoconference system. In addition, Wilbur and Ing (1996) suggest that the degree of formality and the amount of interaction are crucial for understanding how and when videoconferencing will be successful.

The third parameter is the task undertaken when using the videoconferencing system (e.g., meetings or collaborative work, distance education, and entertainment). Differences between the tasks can have a large impact on how videoconferencing systems are used. For example, meetings are usually more task-oriented and interactive than entertainment sessions and thus have different requirements for response times and collaboration tools.

In the next sections, the research on the human-factors of videoconferencing is reviewed according to these three parameters: media, mode, and task. Some research addresses only one of these parameters while other studies are important for two or more of the parameters.

Research Pertaining to MVC Media

Video

The role of video in videoconferencing is enigmatic. On the one hand, video is essential for videoconferencing by definition -- it is what makes the service unique from traditional telephone calls. Video can give a sense of presence that is not available with audio alone and seeing a speaker allows for nonverbal forms of communication, including expressions and gestures, that can be very important for effective interaction. Television, which is distinguished from radio by its use of video, is perhaps the most successful media service of all time.

On the other hand, several video services are notable in their failures. The Picturephone, which added video to traditional telephone calls, was a complete failure. Videoconference rooms, which devote a lot of equipment and bandwidth to the video medium, have also had limited success. Moreover, research on the role of video in remote communications has often found no significant effects if video is added or removed from a conference. For example, Ochsman and Chapanis (1974) examined performance on group problem solving tasks using various remote communication technologies and found that adding a video channel had no significant effect on communication times or communication behaviour. Gale (1990) examined collaboration on experimental tasks and found that there was no difference when a video channel was added in the quality of the output or the time taken to complete the task. How do we reconcile these product failures and research results with the apparent value that video adds to remote communication?

The Presence or Absence of Video

Recent research on the role of video in videoconferencing has revealed some positive effects of the video channel, but these effects are often subtle and indirect. For example, Tang and Isaacs (1992) evaluated the role of video over the long term by examining groups that were already working together (making this a naturalistic approach). Tang and Isaacs used a quasi-experimental design that involved a pre-test phase, where users worked without a remote conferencing tool, followed by a test phase where users used a videoconferencing tool that included a video link. In the third phase the video component was removed from the conference tool. The subjects in this study strongly preferred the phase where a video channel was included. The researchers observed that the video channel was useful for interpreting long pauses (e.g., a remote user is interrupted to answer the phone). They also found that body gestures were used to facilitate smooth interactions when the video channel was present (e.g., the speaker could see reactions to what was being said, including facial expressions and shrugs). This study also showed that gaze awareness was important because users could use the video to determine when a remote participant was paying attention to them. Tang and Isaacs reconcile their results with previous findings by suggesting that previous research had looked at the *product* of remote sessions and found no effect of the video while this study looked at the *process* of the session and did find a difference.

This interpretation was supported by research by Smith et al. (1989) who found that adding a video channel when doing problem solving tasks resulted in more discussions about the task rather than discussions about the mechanics of using the remote communication tools. More recently, O'Malley et al. (1996) reported no difference in task performance with and without a video channel but did observe that remote participants who did not have a video component experienced difficulties in mutual understanding because the users were not sure they were being understood. This resulted in increased 'overhead' communications that were not needed in video mediated and face-to-face interactions.

We can conclude that video is important for videoconferencing but the main effects are in the nature of the interactions and the perceptions of the users, not in the performance of tasks. There are some situations, however, where visual communication can be crucial to performance of the tasks. This occurs when the video channel is used to convey some key task-related information, such as slide material during a lecture. In these situations the video media is very important for the success of the session. This is often a weak area for MBone videoconferences. Often a slide display that is crucial for understanding a remote lecture is transmitted over the video channel by pointing a camera at a projection screen. This leads to low-quality images if the room lighting, camera alignment, and material on the slides is not perfectly coordinated. Moreover, a single camera is often used to capture both the speaker and the slides, moving back and forth between these two subjects.

Care must be taken when capturing crucial video information. Some measures that would improve MVC sessions include: using conferencing tools that are more appropriate for the slide materials, such as the whiteboard tool (WB) and shared WWW pages; using a separate, dedicated video channel for the slides that has proper lighting, camera alignment, and codec settings for the material; and using a video switcher to alternate between two or more cameras that are capturing the slide material and the speaker. These simple changes would improve many MVC sessions that are seen on the Internet today. They do require a sincere effort to present the video material in the best manner for MVC and avoiding adding videoconferencing to an event as an afterthought.

Video Quality

Other research has examined the importance of video quality during videoconferences. Most videoconferencing systems use codecs to heavily compress the video images. The result is small image sizes (often 320 X 240 pixels compared to the television standard of 640 X 480 pixels) and slow screen updates (often less than 10 frames per second [fps] compared to the television standard of 30 fps). Human factors researchers have been interested in determining if such low video quality affects remote communication.

Looking at frame rate, Isaacs et al. (1995) reported that video at 4 fps can be "distracting" and cause a loss of audio-video synchronization. Watson & Sasse (1996) also found that the perception of audio-video synchronization was impaired if the video frame rate was less than 5 fps. In addition, Kies et al. (1996) found that a frame rate of 1 fps (not unusual in low-bandwidth MVC situations) was very distracting. On the other hand, Tang and Isaacs (1992) examined video at 5 fps and found that users rated it acceptable.

Kies et al. (1996) performed a detailed study of video quality in videoconferencing and found that various levels of video quality did not affect understanding in a distance education situation, but it did affect users' satisfaction with the system. These authors recommend a video rate of 6 fps or more and a video resolution of at least 320 X 240. They also suggested that measuring performance is not enough and studies of videoconference systems must also measure satisfaction. Kies et al. also found large individual differences in the tolerance of low quality video that may be related to the users' prior experience, expectations, and other factors.

Bruce (1996) has examined why an acceptable level of video quality is important for effective videoconferences. One reason is that facial cues, especially lip movements, are important aids to speech perception. People with normal hearing read lips to some extent and are able to tolerate more noise in the audio signal if a clear view of the face is available. Bruce goes on to emphasis that a video channel can only be beneficial for speech perception if there is a good view of the face, which is not always true during videoconferences because of bad camera placements. Moreover, any benefit in auditory perception resulting from a view of the face vanishes if the two media are not synchronized and Bruce suggests a maximum audio delay of 80 msecs is tolerable. Also, a reasonable video frame rate is needed to convey enough dynamic facial information, and Bruce's suggests that at least 17 fps is required.

In other research Bruce (1996) has examined the role of facial information in identifying speakers and communicating emotions. In a series of studies Bruce demonstrates that the video quality can be reduced a great deal (e.g., using only one bit per pixel to create monochrome images) while still maintaining good identification and emotion expression. This suggests that low quality video, such as that used in MVC systems, can be effective for identification and emotion expression.

Improving MVC Video

So, video is important for MVC sessions and some minimal level of video is needed to have effective remote communications. Unfortunately, the typical MVC sessions that use 128 Kb/s and H.261 encoding for video do not provide the minimal video quality that is required. MVC sessions should be configured to transmit at least 6 frames per second with an image size of at least 320 X 240 pixels. This may not be possible today with the current limitations on Internet bandwidth and codec technologies, but it should be a goal for the future. Special attention should also be given when it is known that the video information is crucial for performing a particular task, such as the presentation of slides during a remote lecture. In addition, higher frame rates and better image quality will be needed when the content of the session is technical or novel for the users. Moreover, situations where the audio channel is difficult to understand, such as language learning, will be improved by increasing the video quality and ensuring synchronization with the audio channel. This suggests that distance education may be a situation where video quality is especially important, and this will be discussed below.

More attention must also be given to the production techniques that are used when capturing the video component of a conference. Care must be taken when configuring the camera angle and lighting, for example, and this is often not done. Producers of MVC content must also ensure that the video portion is interesting and satisfying for the users. In one study Inoue et al. (1995) compared the video techniques used during videoconferences with those found in comparable television programs and found the videoconferences to be static and boring. They examined the video shot selection used during television productions and found that most shots lasted less than 5 seconds, while this was much longer during videoconferences (perhaps indefinite if only one static camera is used). Based on their analysis of television production techniques, Inoue et al. developed an automated camera control system for videoconference rooms that attempts to ensure that shots are not too static, and include an appropriate mixture of speaker and background shots. They have not yet determined if this system actually results in videoconference sessions that are more satisfying to the user, but this promising approach should be pursued.

Audio

In contrast to video, the role of the audio channel in videoconferencing is well established.

Audio is crucial for most forms of communication because the audio channel often caries the bulk of the information, especially during interpersonal communications. Research on remote audio is important because videoconferences can experience a variety of problems in the audio channel, including low quality, audio loss, echo, and delay. For example, Jameson et al. (1996) found that audio was the most difficult medium in remote surgery lectures using videoconferencing. They found that maintaining audio quality and echo cancellation was very difficult even with high-speed ATM links and specially prepared conference locations. The MVC technology is prone to many audio problems and this can decrease the effectiveness of a videoconference session dramatically.

Audio Quality

One problem area is poor audio quality that occurs when the audio signal is captured in an ineffective way. This often occurs during MVC sessions because of bad microphone placements, the use of inexpensive microphones, and bad room acoustics. Audio quality is also effected by the codec that is used to digitize the sounds. The common MVC audio tools support a variety of coding standards but the most common scheme (PCM) can only provide a quality that is equivalent to a telephone call and this makes it only suitable for spoken voice. Moreover, given today's limited bandwidth it is often desirable to use a lower bandwidth codec and this adds to the quality problems.

Audio echo occurs when the audio that is received at a remote site is echoed back to the sender and the other participants through the receiver's microphone. Echo occurs when loud speakers and microphones are located too close together and are active at the same time (full-duplex mode). In general, users prefer full-duplex audio because it is more natural than a half-duplex mode where users have to "push to talk". For example, O'Conaill et al. (1993) found that full duplex audio was more conducive to normal conversations by enabling interruptions and back channels to a greater degree than half-duplex audio. (However, even with high-quality full-duplex audio, O'Conaill et al. found that these aspects of spoken communication were less frequent than seen during face-to-face meetings.) MVC tools support full-duplex audio but they are normally used in half-duplex mode because of echo problems. Echo problems can be avoided by wearing headphones but this is often inconvenient for the users (Tang & Isaacs, 1992). The other alternative is echo cancellation technology but this is expensive, difficult to operate, and sometimes ineffective.

Audio loss occurs when the some of the digitized audio signal is lost during transmission. On the Internet, audio packets can be lost for a variety of reasons but the most common are network congestion, where router buffers overflow and reject packets, and routing instability, where network paths become unavailable for periods of time. Packet loss is a large problem for today's MBone and this often leads to unusable MVC sessions. Until audio packets can be received reliably it is likely that MVC will only have limited success.

Audio delay occurs when there is significant time needed for the audio to travel from the source to the destination. Sources that contribute to the delay are encoding time, transmission time, and decoding time. Significant audio delay can be distracting for the users and delay makes turn-taking awkward, decreasing the effectiveness of remote interactions. For example, O'Malley et al. (1996) found that performance in a map task was reduced when video and audio was delayed for .5 seconds, as is common in video phones. The performance decrease was due to the users' interrupting each other more often in the delayed audio conditions. Audio delay also becomes a problem if it causes the audio to be out of synchronization with the other media channels, especially the video. In one study, Tang & Isaacs (1992) found that an audio delay of 0.57 seconds hindered effective communications. Subjects in this study preferred to eliminate any audio delay at the expenses of keeping the audio in synchronization with the video. A prototype system built by Tang & Isaacs that had a delay of .32-.44 seconds was found to be more acceptable to the users.

Improving MVC Audio

Watson and Sasse (1996) have conducted some important work on the audio problems experienced on the MBone. These researchers experimented with a redundancy-coding scheme that sends a second copy of audio packets, often at a reduced bandwidth, so the audio signal can be recovered in case of a packet loss. Adding the audio redundancy could greatly improve the intelligibility of the audio and subjective quality ratings. Watson and Sasse (1996) also noted that their users preferred full-duplex audio and audio-video synchronization could be important, especially for the language learning task that they studied. The audio redundancy concept has been built into some of the audio tools (RAT and Free Phone) but it has not gained widespread acceptance in the MBone community. The reasons for this slow acceptance are not clear, especially given the crucial nature of the audio component and the significant packet loss problems experienced on the Internet.

There are a number of parameters that can be set in the MVC audio tools that can affect how well the audio is received. For example, there is an audio playout buffer to hold incoming packets until they are played. The size of this buffer determines the tolerance to late packets since late arrivals can be re-ordered in the buffer before they are decoded and played to the user. A trade-off occurs with increased buffer size because larger buffers lead to greater audio delays while the packets are held to wait for any late arrivals. If long enough, these delays can cause the audio to be out of synchronization with the other channels and make interactivity difficult and awkward. However, given how serious the audio quality problems are on the Internet, users may prefer to have significant audio delays in some situations if it means a greater chance of receiving all the audio packets. Research should be done on the trade-off between audio quality versus delay and synchronization with other media channels. VAT users can manually set the size of the playout buffer (from the default of 6 seconds) and VAT and RAT have a "lecture mode" that allow for longer audio delays. MVC users need to pay attention to these audio modes since lecture mode is not the default for VAT (although it is for RAT) even though it may be the most appropriate setting for many situations (Rettinger, 1995).

MVC users should also pay attention to the priority of local versus network audio. The RAT and VAT audio tools have settings for "net mutes mike" or "mike mutes net" and careful use of these settings can avoid echoes and enhance interactivity (if that is what is required). Perhaps some of these audio tool settings should be handled by the Session Description Protocol used in the SDR tool. The creators of a conference session know what role audio will play during the conference and could suggest appropriate parameters for the audio tools, including lecture mode for non-interactive sessions and long playout buffers when synchronization and interactivity is not important.

Another parameter of the audio tools that receives little attention is silence suppression. The audio tools stop sending packets when they detect no incoming audio signal to encode, which prevents sending empty packets. However, problems in the algorithm used to detect silence or special situations, such as a quiet speaker or passage of music, can lead to lost audio. Moreover, any delays in detecting the audio signal and restarting the transmission can lead to audio clipping, and if the algorithm decides too quickly that the audio signal is gone the signal may be chopped. As a result of these problems, disabling silence suppression may be appropriate for certain kinds of sessions. This could be combined with the "net mutes mike" setting at the receiving end to avoid unwanted audio interruptions in broadcast situations (Rettinger, 1995).

Shared Workspace and Other Tools

A shared workspace can be very useful during a videoconference. A popular tool is a shared whiteboard that can be used to display sketches or notes that all users are able to manipulate. Shared workspaces are especially useful for giving directions or complex explanations. Tang and Isaacs (1992) found that a shared drawing space was the most desired feature in a videoconference room environment. Rettinger (1995) also reported that MVC users found the whiteboard tool (WB) to be "essential" during remote learning sessions.

Rettinger (1995) found, however, that the WB tool was very difficult to use during distance learning sessions. Using the WB requires advanced planning because the charts and graphs must be converted into PostScript format and pre-loaded. WB also has limitations on the size and nature of the PostScript files it can handle so it is necessary to pre-process the files with a utility program. In addition, a separate "wbimport" program is needed to load the slides in sequence and ensure that all the recipients are viewing the correct slide at the correct time. For these reasons, the video channel is often used for slide materials, and this is usually not effective (see Video Section earlier).

Another popular shared workspace tool is a text editor. Olson et al. (1995) observed people using a shared text editor during a design task and found that in face-to-face work sessions the groups that used the editor produced better designs than groups that used normal meeting room tools (e.g., whiteboard, pencil, paper). Interestingly, people who used the shared editor were less satisfied with their work than those who used the traditional tools. A shared text editor that is available for MVC session is NTE. NTE has proven to be very effective for certain kinds of remote meetings. There does not appear to be any research available, however, on how it has been used.

Another application is polling where participants are asked questions and the results are collected and displayed. Isaacs et al. (1994, 1995) included a polling function in a prototype remote-lecturing tool called Forum. Here the speaker could ask multiple-choice questions of the audience and they could give their responses. These responses were displayed in a shared histogram that was updated whenever a new response was given. Isaacs et al. reported that the polling tool was valuable for the speaker to maintain a sense of contact with the audience during the remote presentation. It was also useful as a tool for determining the interests and background of the audience. Research on remote meetings in our laboratory found that decision support tools are needed. Having a tool to present an issue, list the alternative solutions, and collect the opinions is a very valuable for conducting effective remote meetings. Based on these findings, the author has developed an MVC polling application called MPoll that has been adopted by some MBone users.

Remote conferencing sessions sometimes require a floor control tool to coordinate turn taking among the participants. Isaacs et al. (1994, 1995) include a "speakers list" function in their Forum system so users who wanted to comment or ask questions could be placed in a queue. A floor control feature was also included in SSD, the Simplified Session Directory, that allows a conference chair to control who joins a conference and mute participants if necessary to ensure proper interaction. There has also been some work done on a floor control application for MVC sessions called "Questionboard" (Malpani & Rowe, 1997). Such control applications will likely be very important for larger remote meetings but no research has been done on their use and effectiveness.

In summary, shared workspace tools can be valuable when multiple users want to work on a drawing or document at the same time. The WB tool is a very popular shared workspace tool on the MBone. The WB tool can be used to present slide material but its limitations makes this difficult. More work is needed on making a whiteboard tool that is appropriate for slide material. A shared WWW tool such as mMosaic may be more suitable for slide presentations and this approach deserves more work and attention. In the mean time, originators of MVC broadcasts should take the time necessary to prepare the material for inclusion in the WB tool and install the "wbimport" application that makes effective slide presentations possible. Finally, it is likely that conference control will be a major issue when MVC gains popularity and more research is needed in this area.

Research Pertaining to MVC Communication Modes

Many examinations of videoconferencing note that there are different modes of communication that are related to the nature of interpersonal interactions. Wilbur and Ing (1996) suggest that the "degree of formality" and the "amount of interaction" are important dimensions for understanding the communication modes in videoconferencing. During interactive videoconference sessions the participants contribute relatively equally, while most participants are passive recipients during non-interactive sessions. Formal sessions often involving agendas and specific roles for the participants while informal sessions do not have these constraints. In addition, Rettinger (1995) suggests that "lecture" and "collaborative" are two distinct modes of remote interaction. Understanding these different modes of communication is necessary for understanding the demands on a videoconferencing system and how it will be used.

Interactive and Non-Interactive

In many ways, the term "non-interactive videoconference" is a paradox since one of the main features of a videoconference service is multi-way interaction in real time over a remote distance. If a session was completely non-interactive then a broadcasting technology may be more appropriate and often cheaper (especially with large audiences). Nevertheless, either by design or by usage, many videoconference sessions turn out to be non-interactive. In these sessions one participant or location does all of the transmitting and the other participants are passive recipients. This is especially true for remote lectures or conference sessions, which are very common on the MBone.

Non-interactive sessions are appropriate for some videoconferencing tasks (see below), but these sessions have different requirements than interactive sessions. For example, for interactive sessions the maximum end-to-end delay that can be tolerated is about 0.3 seconds (Willebeek-LeMair & Shae, 1997). Longer delays can be tolerated for non-interactive sessions because there will be no turn taking by multiple contributors. Thus, it is not appropriate to use the same MVC software with the same parameter settings for interactive and non-interactive sessions. Originators of MVC sessions should consider the nature of the session they are planning and control or suggest software parameters that are appropriate. Precept Software and ICAST have taken this to the extreme by building software that can only be used for non-interactive sessions.

Even when interaction is desired and expected it can be very difficult to achieve. For example, Jameson et al. (1996) described problems in making an interactive session during remote teaching of surgery. They found that students were reluctant to ask questions of the lecturer and this decreased the satisfaction for both the students and the lecturer. One technique for improving interactivity was to have a student facilitator at each remote site who was familiar with the equipment encourage and assist the students to be interactive. Another suggestion was to have a warm-up session where students could become familiar with the equipment. It is not clear if these techniques would be successful for MVC sessions but they may be worth exploring.

Another way to encourage interactivity is to explicitly involve the audience. Lecturers can do this by asking questions of their remote participants. Isaacs et al. (1994, 1995) found that certain features in their Forum software, such as a speakers list, written questions, and audience polling, did encourage interaction during remote lectures but the interaction was still different than face-to-face presentations. Isaacs et al. also found that the speakers in their study missed the feedback they get from a live audience. In particular, the speakers complained about the lack of non-verbal feedback like head nods, applause, and laughter that show understanding and appreciation.

Formal and Informal

A related characteristic of videoconference sessions is the degree of formality. Formal sessions often involve a fixed agenda and specific roles for the participants, while informal sessions are more free form and personal. Informal communications are often personal, oral, brief, and spontaneous. Gale (1992) suggested that, due to turn-taking problems, videoconference sessions tend to be more formal than face-to-face sessions. This is reflected in meetings that are more often agenda-based, tightly organized, and shorter than face-to-face meetings. These may be beneficial features for some situations but studies of work patterns show that informal communications are often the key to success. Groups that are close in physical proximity usually have more informal communications and more successful collaborations than groups not in close proximity. Gale's research leads to two conclusions: (1) videoconference systems are only appropriate for formal situations and will therefore have limited value, or (2) videoconference systems must be improved to support informal communications. Other research has supported both of these conclusions.

Formal communication tasks were found to be more suitable for videoconferencing in a study by Kies et al. (1995). They found that in videoconferencing situations where there are clear roles, such as remote lectures and formal meetings (a thesis defense), there were fewer problems in speaker turn-taking than situations without these formal aspects (a remote laboratory tour). Analysis at the linguistic level has also found more formality in videoconference interactions. For example, O'Conaill et al. (1993) found fewer back channels (such as saying "uh huh" while another person is speaking), less turn-taking, longer speaker turns, and more formal hand-overs when comparing videoconference and face-to-face conversations. It is not clear if the videoconference system required the formal communication style or if the users adopted the style for other reasons.

As a result of findings like these, much effort has been devoted to improving videoconference systems so they can better support informal communications. Wilbur and Ing (1996) suggest that new technologies often evolve from relatively formal, non-interactive systems into something that is less formal and more interactive. For example, the telephone was first envisioned as a system for relaying distant concerts. When telephones were eventually installed in homes they were left for the servants to answer and used rather formally. It was only later that informal telephone calling came into common practice. Wilbur and Ing argue that electronic mail has evolved in a similar way -- from a limited, formal means of communication to a wide-spread, ubiquitous service with a range of informal uses. The suggestion is that videoconferencing needs to experience its own evolution and anything that will help it along will be beneficial.

Various projects have examined supporting informal communications in videoconferencing systems. One approach has been to build a shared media space that closely resembles a shared physical space. For example, the CAVECAT system developed at the University of Toronto (Mantei et al., 1991) linked offices with permanent audio and video channels. Users of the system could then communicate as they would with physically present colleagues. Another approach has been to develop a virtual environment that resembles a well-known physical environment and then allow the users to interact in this virtual space. For example, the CRUISER system (Root, 1988) uses a virtual hallway that users can travel to find colleagues and interact using multimedia links. It is important to note that a physical space metaphor may be too limiting because people often interact in multiple social environments at the same time and remote communications tools must accommodate this (Kaplan et al., 1997). A third approach is to encourage chance meetings between users based on shared activities. In these systems a monitoring program watches for users who are doing the same activity (e.g., editing the same file) and presents an opportunity for real-time interaction.

A number of social issues arise when informal communication tools are used. Heath & Luff (1992) looked at the issue of privacy when videoconferencing systems are used for informal communications and found that users need to be able to understand and control when the remote link is active. This has led to devices such as chimes that sound when the link is activated so people know they are visible. Buxton (1995) took this further by wiring office doors with switches that also controlled the accessibility via the shared media space. In this way, users' remote accessibility was correlated with their physical accessibility. It is clear from this research that the social implications of informal communications will be an important factor for determining the success of any videoconferencing system.

There has been some work done on supporting informal communications on the MBone. For example, the SDR tool includes a "Quick Call" function that allows informal sessions to be created and users invited to these sessions. A "Session Invitation Protocol" has been proposed as a standard method for such informal communications but it has not been widely adopted. A related tool is CONFMAN, developed at the University of Hanover in Germany. This tool allows MVC users to initiate informal conferences and includes an address book for maintaining a list of frequent contacts. CONFMAN also allows users to indicate if they are available for a conference. A similar tool is TELEPORT, developed at Loughborough University of Technology in the UK. TELEPORT allows conference users to learn if a remote participant is busy or available for a conference. Unfortunately, TELEPORT and CONFMAN were limited duration projects and the software is not widely used. Support for informal communications should be a high priority since it is an important mode of communication that will likely lead to quicker acceptance of the MVC applications.

Research Pertaining to MVC Tasks

The third dimension that is important for understanding the human factors of videoconferencing is the task undertaken when using the system. Observation of MVC sessions over a 1 year period suggests that there are three main types of tasks being attempted: (1) meetings and collaborative work, (2) distance education, and (3) entertainment. These different tasks present different demands on the videoconference system. For example, meetings often require more interactivity and faster response times than distance education sessions. In addition, different sub-tasks within these main categories also present different demands on the videoconference system. For example, Gowan and Downs (1994) found that a corporate videoconference system was only effective for certain types of business meetings (e.g., tactical but not operational). In addition, Kies et al. (1995) found large differences in the use and value of videoconferencing depending on the type of education task being undertaken (lecture, thesis defense, or lab tour). Thus, understanding the task is important for understanding the use and value of a videoconferencing system.

The SDR tool that is used for advertising MVC sessions does have a parameter for describing the task of a session. Currently, the software allows a session to be described as a "meeting", "broadcast", or "test". However, this task information is not used for anything other than controlling some icons when the session details are displayed (e.g., Figure 10 shows the broadcast icon displayed in the upper left of the SDR panel). This task information may be useful for setting tool parameters that are most appropriate for the type of session and this possibility should be explored further.

Meetings and Collaborative Work

Meetings and collaborative work sessions are usually oriented around a specific goal or agenda. These sessions tend to be relatively small and interactive. They can be formal or informal and often include one of the shared work tools (whiteboard or text editor). These sessions are usually focused on getting something accomplished so the most important measure of success is usually progress toward a goal.

There has been surprisingly little research on the value of videoconferencing for meetings and collaborative work tasks. In one study Olson et al. (1995) examined performance at a design task where users collaborated in one of three conditions: face-to-face, audio conference, and audio plus video conference (all groups used a shared editor). Analogue audio and video links were used in this study so the media quality was very high. The important result was no difference in the quality of the designs produced in the face-to-face and audio/video conference situation. This suggests that videoconference systems can be very effective for doing collaborative work. However, Olson et al. did find differences in the nature of the interactions between the two groups. Groups using the videoconference system spent more time managing the meeting and clarifying what was meant than the face-to-face groups. Thus, videoconference systems may be less efficient than face-to-face communications while still being suitable for getting the task done.

Distance Education

Videoconferencing would seem to be an ideal technology for distance education tasks. Educational institutions are interested in serving students in a variety of ways and they are often motivated to expand their geographic appeal. Videoconferencing on the Internet is particularly attractive because educational institutions often have good connections to the Internet and a willingness to experiment with new technologies. Also, using the Internet may be less expensive than traditional broadcasts over cable TV or satellite. In addition, the interactive nature of videoconference systems is attractive for pedagogical reasons. However, actual use of videoconferencing systems for distance education has led to limited success.

Kies et al. (1996) conducted some experiments on using videoconferencing systems for distance learning tasks. They found that varying the level of video quality had no effect on quiz performance, which suggests that a low quality system like the MBone might be suitable. Kies et al. did find, however, that students had difficulty maintaining attention when the video quality was low and their satisfaction decreased. This research also found a high degree of variability among the students in their acceptance of the videoconference technology that was related to their enthusiasm, expectations, and prior experience. Thus, it is likely that learning may suffer for some students if the material is more demanding. In a second study Kies et al. examined use of the MVC tools in a distance learning task between two universities. They again found problems in maintaining attention in the remote classroom and low levels of satisfaction.

Another problem area was asking questions from the remote location.

In a similar study Rettinger (1995) examined the effectiveness of the MVC tools for distance learning. She also examined remote teaching between two universities and found significant problems using the MBone videoconference technology. Students found the video quality to be marginal and long audio delays made interactivity difficult. As a result, there was minimal interaction between the sites and little feedback from the remote audience. There were also turn taking problems because it was difficult for the remote audience to indicate that they wanted to speak. This suggests that a floor control tool might be useful for this task.

On a more positive note, Watson & Sasse (1996) found that language learning was at least as effective using MVC technology than conventional small group classes. This research demonstrated that high-quality audio was essential and that it needed to be synchronized to the video channel, but with these factors in place the students were able to learn effectively.

Isaacs et al. (1995) did a direct comparison of distance learning between their "Forum" software and live presentations. These researchers used surveys, videotapes, and computer logs to compare how the speakers and audiences acted during live and remote presentations. They found that the remote presentations had larger audiences but the remote audience paid less attention to the talk (65%) than the live audience (84%) did. Significant problems were noted in the distance learning situation because of low video quality (4 fps) and no audio/video synchronization. Speakers also complained about the lack of feedback from the remote audience. In addition, there was more audience interaction in the live situations, with more questions and follow-up discussion. Overall, however, users preferred the videoconference presentations to the live presentations.

One important finding from Isaacs et al. (1995) was that the Forum system was well suited for some types of presentations and ill suited for others. The best presentation for videoconferencing was from the Human Resources department of a company because users could attend to the talk as much or a little as they wished. An executive's vision presentation was also well received as a videoconference because it was relatively formal and there was no expectation of significant interactions between the audience and the speaker. The worst type of presentation for videoconferencing was a small-group talk because the videoconference system limited interaction and social contact among the audience. Thus, the type of task is important for understanding the value of videoconferencing.

Entertainment

There has been little research on using videoconference systems for entertainment even though it is one of the common session types found on the MBone. A variety of radio and TV stations can be found transmitting on the MBone. These broadcasters are also heavy users of unicast streaming technologies. Given the ability of broadcasters to create compelling content and the inefficient nature of unicast schemes on the Internet, perhaps more attention should be given to supporting entertainment services on the MBone.

Interactive systems can also be entertaining, as is evident in the popularity of Multi-User Dungeon (MUD) games and "chat" groups on the Internet. Some attempts have been made at building multicast entertainment applications, such as a multicast chat service called "MultiTalk" and an interactive video game called "MiMaze". However, research on the success and limitations of this networking technology for these applications is lacking.

One area of research that is relevant to entertainment is video production techniques. A study by

Inoue et al. (1995) showed that typical videoconferences produce video images that are static and boring when compared to television productions. This suggests that in order for videoconferences to be successful forms of entertainment more work is going to be needed on the video production techniques.

Summary of Recommendations

This review of the human factors research of videoconferencing systems has led to a number of recommendations that are summarized in this section. These recommendations can be divided into those appropriate for MVC session organizers and those appropriate for software developers. These recommendations are not all specific to multicasting and many may be of interest to people developing or using other videoconference system.

Session Organizers

Video

Careful consideration must be given to the video component of MVC sessions. Session organizers should consider the role that video will play during the session and adjust the session parameters accordingly. For example, if the visual information is crucial for understanding and enjoying the session then the video channel must be of good quality. Good quality video means at least 6 frames per second of material that is well produced (i.e., proper lighting and camera work). Achieving such quality may require over-riding VIC's default setting of a maximum bandwidth of 128 Kb/s for global sessions (TTL >= 127), which can be done with the -B command line option (e.g., -B 200). Selecting a non-standard video bandwidth must be done with caution since high bandwidth multicast traffic can cause congestion on local and global Internet links, but at times it may be appropriate.

If the visual information is secondary then the video bandwidth is better used elsewhere (e.g., sending redundant audio). In these cases, a bandwidth setting that is less than the 128 Kb/s may be appropriate. This is especially true when there is limited bandwidth and a trade-off between video and audio is necessary. Since audio is almost always crucial to a MVC session, it should take priority over video when trade-off decisions are made.

It may be appropriate to conduct some research on users' reactions to such audio/video tradeoffs. For example, given a fixed total bandwidth and today's congested networks, do users prefer a configuration of 128 Kb/s for video and 71 Kb/s for audio (the typical MVC session parameters) or an opposite configuration of 71 Kb/s for video and 128 Kb/s for audio (using redundant PCM encoding)? Will users appreciate the high-quality audio and consider it to be a good trade-off for the reduced video quality?

Video production techniques are also very important. This includes such factors as lighting, camera placement, and camera control. For interactive sessions it is important to direct the camera at the participant's face. This allows participants to use gaze information to learn who is attending to the session and who is distracted or doing other things. It can also provide important non-verbal communication information (e.g., gestures, facial expressions, and lip movements) that can aid understanding. Using a selection of carefully composed video shots is also important, especially for non-interactive sessions. All too often the video material in a conference is static and boring and this detracts from the potential of the video channel. It may be necessary to have a camera operator or multiple cameras routed through a video switcher to achieve a high quality MVC session.

Special attention must also be paid to any visual presentation materials (slides or diagrams) that are used during a conference. Too often MVC session organizers attempt to use a video channel for this material, with disastrous results. The most common technique is to point a camera at a projection screen being used in a lecture hall and this usually results in completely unusable material for the remote participants. If video must be used for presentation material then a good quality image must be insured. This can be done using a copy stand and dedicated camera to provide a clear, close-up image of the slides. For live presentations this does require having a second copy of the speaker's materials available, but it is well worth the effort. The video images of the presentation material should either be transmitted as a separate video stream, or alternated with other views using a video-switching device. Precept has built such an alternating scheme into their IP/TV software.

Audio

High quality audio is essential for a successful MVC session. Audio often carries the bulk of the information during remote sessions and audio problems can make a session annoying or useless. Many audio problems begin long before any computers and compression schemes come into play. The analogue audio signal must be of good quality before it is subjected to compression and transmission. This means using high-quality microphones and audio equipment and spending the time necessary for proper setup and testing. Microphone placement is very important and many MVC sessions suffer from not having microphones devoted to everyone who is speaking (e.g., the speakers, the local audience, and the moderator). In addition, microphones that are not directly in front of the speakers (or attached to them) often result in bad room acoustics and unwanted echo.

Once a good analogue audio source has been created, the next consideration is bandwidth. Bandwidth must be devoted to the audio channel in accordance to the importance of audio for the session. For example, music programs and small group interactive meetings may wish to devote large amounts of bandwidth to the audio channel to ensure that everything is heard. On the other hand, programs that are mainly visual in nature may wish to devote smaller amounts of bandwidth to the audio component. Often a trade-off is necessary between the bandwidth devoted to audio and video. The session organizers must decide the relative importance of these two media and use the tools accordingly.

Audio bandwidth is determined in the MVC tools by selecting a compression scheme and possibly a redundancy scheme. VAT and RAT offer a variety of compression schemes ranging from LPC to PCM and these schemes result in relatively low and high quality audio respectively. The Free Phone tool also supports higher quality audio up to CD and DAT sampling rates. In addition, RAT and Free Phone support a redundant coding scheme where audio packets can be reconstructed if necessary, and empirical research has shown that this can greatly improve the perception of the audio material. Given the high levels of packet loss currently encountered on the Internet and the availability of RAT and Free Phone, audio redundancy schemes should be used in MVC sessions whenever possible.

The audio buffers should also be configured depending on the nature of the session (this is easily done in Free Phone). Larger playout buffers and tolerance for longer delays may be appropriate for non-interactive sessions, and the tools should be used accordingly. Again, session organizers should consider the type of session they are creating and recommend or control the software settings as required.

Another audio consideration is the priority given to local and remote audio. For interactive sessions full-duplex audio is preferred because there is some likelihood that multiple speakers will speak at the same time (especially given the turn-taking problems that are usually encountered in videoconferencing sessions). It is usually necessary to use headphones or echo cancellation devices to avoid feedback when using full-duplex audio and these should be made available. If headphones are not used then half-duplex audio is necessary to avoid feedback. The RAT and VAT tools allow the user to give priority to the local microphone or the remote audio whenever there is contention. In "mike mutes net" mode the priority is given to local audio and in "net mutes mike" mode priority is given to remote audio. The choice among these audio modes should be made according to the amount of interactivity anticipated during the session and the role taken by the user. In interactive sessions the "mike mutes net" setting may be most appropriate to allow effective turn taking. For non-interactive sessions the active user should also use the "mike mutes net" setting to avoid interruptions, while the passive audience should use the "net mutes mike" setting to avoid any accidental interruptions of the speaker.

Another important audio setting is silence suppression. When turned on, the audio software will stop transmitting when incoming audio is not detected. This can be useful for avoiding wasted bandwidth that would be used to transmit silence, but it is not appropriate for certain types of material, such as quiet music. Silence suppression should be turned off when the audio material is expected to be constant, and/or when the audio levels may be low. Turning off silence suppression is often appropriate in videoconferences where there are multiple people at a remote site who all have to be heard, or when the background audio is important for the session (e.g., a telepresence application designed to give a feeling of being there).

Workspace

The WB tool is capable of handling presentation material and it allows users to import PostScript slides. In addition, the 'wbimport' utility will pre-load the slides and ensure that all the users are viewing the correct slide at the correct time. Also, during the session the WB tool can be used to annotate the presentation material and for simple interactive communications. Unfortunately, converting slides to the PostScript format can be a laborious task and WB has a default file size limit (32,768 bytes). In addition, WB is hard to install (especially the PostScript support) and only available for UNIX computers. The result is that WB is not used often enough for visual materials.

Some of these limitations in WB can be overcome with effort. For example, the file size limitation imposed when importing PostScript files can be overridden with a little-known run time option "-P". Thus, running "wb -P 100000" will allow a user to import PostScript files up to 100,000 bytes. WB users must do this with caution because the tool can be slow when transmitting large files and cause network congestion. However, given the serious problems encountered when using video for presentation materials, using this WB option must be seriously considered. Some network delay problems can be avoided if the slides are made available before the MVC session and users are encouraged to page-through the material and pre-load the images before the session begins.

The task of preparing material for presentation via WB can also be improved with some advanced work. For example, the author has developed a procedure for preparing WB material from PowerPoint presentation files. The first step is to prepare a PostScript printer file using an appropriate printer driver and the "archive format" of PostScript compatibility. The resulting printer file should then be split into separate files for each page, compressed, and indexed for the 'wbimport' program. A Perl script has been developed to do this automatically and it can be accessed at http://debra.dgbt.doc.ca/mbone/pstowb.txt. This script uses the 'pssplit' and 'lzps' utility programs that come with wbimport.

A new tool is being developed by Julian Highfield in the UK to eliminate some of these problems with WB. WBD is compatible with WB but does not enforce the strict PostScript file size limits found in WB. WBD is currently in the Alpha stage of development and not suitable for global conferences yet, but it may be a good alternative in the future.

Another alternative to WB is NTE, the network text editor. This tool would be suitable if the presentation material can be converted to a textual format, which may be the case for speaking notes and point-form charts.

Another approach for handling presentation materials is to use HTML files and a shared WWW tool. HTML is a commonly used layout format and WWW tools allow the inclusion of any arbitrary graphics (not just PostScript). There have been a few shared WWW tools developed for the MBone, including Shared Mosaic and mMosaic. However, the most promising application is WebCanal, a tool that is capable of multicasting the content of WWW pages to session participants (not just the URL). The MBone community has not adopted WebCanal yet but an application like this may be the best alternative for displaying presentation material during MVC sessions and it should be investigated further.

Modes

It is clear that the same software parameters are not appropriate for interactive and non-interactive sessions. MVC session organizers should consider the mode of their session and arrange for the software tools to be launched, or configured by the users, with the appropriate settings. For example, for non-interactive sessions half-duplex audio with large buffers and "receive only" tool settings are appropriate.

When an interactive session is planned, the organizers should be aware that explicit measures might be needed to encourage the interaction. Organizers should consider such techniques as warm-up periods, tool familiarization, and direct prompting to ensure that interaction does take place. Some software tools are also becoming available for supporting interaction. In particular, tools that support floor control, a speakers' list, question asking, and audience polling should be considered when planning interactive sessions.

Tasks

Session organizers should also consider the task being undertaken during the MVC session. Some tasks seem to be more appropriate for videoconferences, including tasks that are relatively formal and structured, have clear roles and responsibilities, are easily managed, and involve relatively well understood material and concepts. If a task has opposite characteristics, session organizers may have to take extra steps to ensure a successful session, or even consider another type of communication.

Software Developers

There are a number of recommendations for software developers resulting from this research. First, there is major distinction between interactive and non-interactive use of videoconference software. Having a non-interactive mode for a software tool will reduce expectations of interactivity when it is not appropriate and reduce conduct errors, such as transmitting when receive-only is expected. The appropriate software settings for interactive and non-interactive use should be grouped for easy configuration and made accessible as a command-line option and user-adjustable feature.

Conference organizing tools, such as SDR, should be enhanced to use information about the modes and tasks involved in MVC sessions. SDR has started to include some of these features but more information should be collected about the type of session and more tool parameters should be controlled at start-up. In addition, because informal communication is very important, more support is needed for rapid informal session creation. Tools will also be needed so users can control their accessibility for new sessions and their visibility during established sessions.

Video software should be developed that provides at least 6 fps of video over reasonable bandwidths. Synchronization with an audio channel is also important and should be included in any new software.

There is a clear need for more work on a reliable and easy method for presenting slide-type materials. Such software might be based on the WWW protocols, and efforts like WebCanal seem to be promising. Conference support tools are also needed for handling such tasks as turn taking, question queues, and audience polls. Again, some work has been done in these areas but often the tools have not been completed or publicized widely.

Network congestion and packet loss are serious problems. Quality of Service (QoS) schemes will help the problem by allowing bandwidth to be dedicated to a session, but users will have to support the protocols and be willing to pay for the extra service. In the short term other solutions must be found. The RAT and Free Phone tools provide redundant audio encoding that can greatly improve the audio communication. This is a good example of what needs to be done.

Another approach is to dynamically adjust the MVC tools depending on the congestion that is encountered. Busse et al. (1996) explains how this can be done using the RTCP protocol that is used in most of the MVC tools. They demonstrate an algorithm that adjusts the video transmission bandwidth according to reception loss over both Internet and ATM networks. They also discuss how multiple recipient reports, such as in a broadcast application, would have to be handled (e.g., average the loss reports or tune for worst case reception). This scheme is starting to be explored in some MVC tools, most notably the Free Phone and Rendez-Vous tools being developed at INRIA, and this is very promising work that should be pursued.

In addition, users need feedback about how well they are being received so that conference overhead messages (i.e., "can you hear me") are avoided (Vetter, 1995). Such feedback should include measures of loss and delay. Feedback information is being distributed in the RTCP protocol that is used by some MVC tools, and this should be exploited to build more informative tools.

Finally, use of the MVC software and protocols has primarily been restricted to academic events and demonstrations. Software developers should consider entertainment uses of these new technologies, since this may lead to more rapid adoption of the protocols and standards.

References

Bruce, V. (1996). The role of the face in communication: Implications for videophone design. *Interacting with Computers, 8*, 166-176.

Busse, I., Deffner, B., & Schulzrinne, H. (1996). Dynamic QoS control of multimedia applications based on RTP. *Computer Communications, 19,* 49-58.

Buxton, B. (1995). Scientific director's report: Living in augmented reality. *Ontario Telepresence Project Final Report.* http://www.dgp.toronto.edu/tp/tphp.html

Gale, S. (1990). Human aspects of interactive multimedia communication. *Interacting with Computers, 2,* 175-189.

Gale, S. (1992). Desktop video conferencing: Technical advances and evaluation issues. *Computer Communications, 15*, 517-526.

Gowan, J.A., & Downs, J.M. (1994). Video conferencing human-machine interface: A field study. *Information & Management*, *27*, 341-356.

Harper, R., Hughes, J., & Shapiro, D. (1991). Harmonious working and CSCW: Computer technology and air traffic control. In J. Bowers and S. Benford (eds.), *Studies in Computer Supported Cooperative Work: Theory, Practice and Design*. Amsterdam: North Holland, 225-234.

Heath, C., & Luff, P. (1992). Media space and communicative asymmetries: Preliminary observations of video-mediated interaction. *Human-Computer Interaction, 7*, 315-346.

Inoue, T., Okada, K., & Matsushita, Y. (1995). Learning from TV programs: Application of TV presentation to a videoconferencing system. *Proceedings of UIST'95, ACM Conference on User Interface and Software Technology*, Pittsburgh PA, Nov. 14-17, 147-154. http://www.acm.org/pubs/articles/proceedings/uist/215585/p147-inoue/p147-inoue.pdf

Isaacs, E.A., Morris, T., & Rodriguez, T.K. (1994). A forum for supporting interactive presentations to distributed audiences. *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '94)*, Oct. 22-26, Chapel Hill, NC, 23-34. http://www.acm.org/pubs/articles/proceedings/cscw/192844/p23-tang/p23-tang.pdf

Isaacs, E.A., Morris, T., Rodriguez, T.K., & Tang, J.C. (1995). A comparison of face-to-face and distributed presentations. *Proceedings of CHI '95, ACM Conference on Human Factors in Computing Systems*, May 7-11, Denver, CO, 354-361.

Jameson, D.G., Hobsley, M., O'Hanlon, P., & Buckton, S. (1996). Real-time interactivity on the SuperJANET network. *Interacting with Computers, 8*, 285-296.

Kaplan, S.M., Fitzpatrick, G., Mansfield, T., & Tolone, W.J. (1997). MUDdling through. *Proceedings of the Thirtieth Annual Hawaii International Conference on System Sciences (HICSS '97), Vol II*, IEEE Computer Society Press, 539-548.

Kies, J.K., Kelso, J., & Williges, R.C. (1995). The use of scenarios to evaluate the effects of group configuration and task on video-teleconferencing communication effectiveness. Paper presented at the *Third Annual Mid-Atlantic Human Factors Conference*, Blacksburg, VA, March 26-28. http://hci.ise.vt.edu/~hcil/MidAtlantic_Paper.html

Kies, J.K., Williges, R.C., & Rosson, M.B. (1996). *Controlled laboratory experimentation and field study evaluation of video conferencing for distance learning applications.* HCIL Hypermedia Technical Report HCIL-96-02. Human-Computer Interaction Laboratory, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg VA. http://hci.ise.vt.edu/~hcil/htr/HCIL-96-02/HCIL-96-02.html

Kumar, V. (1996). *MBone: Interactive multimedia on the Internet*. Indianapolis, IN: New Riders.

Macedonia, M.R., & Brutzman, D.P. (1994). MBone provides audio and video across the Internet. *IEEE Computer,* 30-36.

Malpani, R., & Rowe, L. (1997). Floor control for large-scale MBone seminars. *Proceedings of the Fifth Annual ACM International Multimedia Conference*, Seattle, WA, November 8-14. http://www.acm.org/sigmm/MM97/papers/malpani/qsbmm97.html

Mantei, M.M., Baecker, R.M., Sellen, A.J., Buxton, W.A.S., Milligan, T., & Wellman, B. (1991). Experiences in the use of a media space. *Proceedings of CHI'91, ACM Conference on Human Factors in Software*, 203-208.

Ochsman, R.B., & Chapanis, A. (1974). The effects of 10 communication modes on the behaviour of teams during co-operative problem-solving. *International Journal of Man-Machine Studies, 6*, 579-619.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction, 8*, 389-428.

Olson, J.S., Olson, G.M., & Meader, D.K. (1995). What mix of video and audio is useful for small groups doing remote real-time design work? *Proceedings of CHI'95, ACM Conference on Human Factors in Software,* 362-268.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers, 8*, 177-192.

Rettinger, L.A. (1995). *Desktop videoconferencing: Technology and use for remote seminar delivery*. Unpublished master's thesis, North Carolina State University, Raleigh, NC. http://www2.ncsu.edu/eos/service/ece/project/succeed_info/larettin/thesis/abs.html

Root, R.W. (1988). Design of a multi-media vehicle for social browsing. *Proceedings of the Second Conference on Computer-Supported Cooperative Work,* Portland, Oregon, 25-38.

Smith, R.B., O'Shea, T., O'Malley, C., Scanlon, E., & Taylor, J. (1989). Preliminary experiments with a distributed, multi-media, problem solving environment. *Proceedings of the First European Conference on Computer Supported Cooperative Work: EC-CSCW '89,* London, UK, September, 19-34.

Tang, J.C. & Isaacs, E.A. (1992). Why do users like video? Studies of multimedia-supported collaboration. *Computer Supported Cooperative Work (CSCW), 1*, 163-193.

Tang, J.C., Isaacs, E.A., Rua, M. (1994). Supporting distributed groups with a montage of lightweight interactions. *Proceedings of CSCW 94* (Chapel Hill, North Carolina, USA), 23-34.

Vetter, R.J. (1995). Videoconferencing on the Internet. *Computer,* (January), 77-79.

Watson, A., & Sasse, M.A. (1996). Evaluating audio and video quality in low-cost multimedia conferencing systems. *Interacting with Computers, 8*, 255-275.

Wilbur, S., & Ing, S. (1996). Real-time video for informal workgroup communication: A survey of recent advances. *Computer Networks and ISDN Systems, 28,* 491-497.

Willebeek-LeMair, M.H., & Shae, Z. (1997). Distributed video conferencing systems. *Computer Communications, 20*, 157-168.

About Author

Dr. Andrew Patrick was born in Ontario in 1958. He holds a Ph.D. in Experimental Psychology from the University of Western Ontario. Dr. Patrick is currently a Behavioral Scientist at Nortel Networks in Ottawa. He conducts research on the psychological and social aspects of advanced communications systems, and current areas of interest include users' responses to different levels of network service quality.
Address: Communications Research Centre, 3701 Carling Ave., Ottawa ON CANADA K2H 8S2