

MAKING MEDIA SPACES USEFUL: VIDEO SUPPORT AND TELEPRESENCE

Dennis C. Neale & Mike K. McGee

Department of Industrial and Systems Engineering
Human-Computer Interaction Laboratory
Virginia Tech
Blacksburg, VA 24061-0118
dneale@vt.edu

Brian S. Amento & Patrick C. Brooks

Department of Computer Science
Virginia Tech
Blacksburg, VA 24061-0118

ABSTRACT

Media spaces support remotely based cooperative activities and in some cases promote telepresence. Often, however, video use surrounding these systems has not been completely adequate or even appropriate for the tasks typically investigated. Research has primarily focused on “talking heads” video—video images of participants conversing face to face. Presently, the findings suggest that use of video in this way has limited application for supporting communication and cooperation. In contrast video has rarely been considered as a tool for sharing physical objects and workspace, even though video use in this form has enormous potential for exploiting media space. Furthermore, little is known about optimally configuring camera views and their presentation given the various uses of video conferencing. This paper describes observations of participants using media spaces, focusing on system aspects of the desktop video conferencing. Multiple camera views were manipulated and their presentation for negotiation and physical object assembly tasks. In two studies video was assessed for its value in face-to-face and workspace configurations. The objective of this investigation was to evaluate the effectiveness of the various equipment configurations for the selected tasks, including an examination of telepresence. Examples of both successes and failures of

system efficacy are reported. Implications for future studies of media spaces, video conferencing systems, appropriate task contexts, and telepresence are discussed.

KEYWORDS

Computer-supported cooperative work, human-computer interaction, media spaces, desktop video conferencing, telepresence

INTRODUCTION

The availability of technology to support communication with video systems has existed since the early sixties, for example, the commercial development of the videophone (Rosen, 1996) and Picturephone (Wish, 1975). The introduction of this technology to support people working at a distance has had many failures (Egido, 1988), and realization of such systems has been substantially less than what has been predicted (Angiolillo, Blanchard, and Israelski, 1993). In spite of technological advances in recent years driving new uses for video conferencing systems, the crucial question of how video conferencing supports communication and genuine work has not been adequately addressed. Researchers are only beginning to form a knowledge base for understanding the impact of these systems on the physical, social, communication, and task environments.

Three principle conditions perpetuating the

problems discussed above have been (1) the lack in addressing appropriate task contexts that can benefit from video, (2) the lack of sufficiently configuring equipment to support different tasks associated with video, and (3) the lack of support for telepresence that affords users a “common ground” for collaborative working. This paper will address all three of these issues in the context of media spaces that support communicative-rich collaborative work.

New requirements emerging due to key changes in technical advances and social work practices are driving greater demands for solutions to the problems outlined above. Changes in work practices have arisen which place increasing demands on people to work in groups, often displacing them over time and place. This has also led to culturally and technically diverse work groups. Technical developments facilitating video-mediated communication approaches include computationally viable digital video due to advances in computer processing speeds and better compression algorithms, advances in networking approaches with higher bandwidth; and an array of new displays, cameras, and audio equipment. The convergence of these consumer electronics, communication technologies, and personal computers have resulted in feasible and useful video-mediated systems capable of supporting collaboratively dispersed work groups.

Establishing a shared telepresent work environment has been argued as being essential to effective video-mediated communication (Buxton, 1992). However, little is known about telepresence as a construct of computer-supported cooperative work (CSCW). Research needs

to be aimed at ascertaining an adequate understanding of what telepresence is, how to measure it, and whether it supports objective and subjective task performance necessary for effective video conferencing.

This paper will examine several different types of media spaces, focusing specifically on the video conferencing aspects of these systems. The goal is to understand the relationship between various technical configurations used to create these systems, their impact on the task contexts utilized, and how both the equipment and task components affect telepresence.

MEDIA SPACES

Media spaces, forged from recent developments in information technology that integrate video with groupware, are technologically based spaces (virtual spaces) created to support both social and technical practices of contemporary collaborative work. These spaces precipitate peripheral awareness of other workers and activities, chance encounters, the capability to locate colleagues, and group discussions (Bly, Harrison, and Irwin, 1993). While media spaces often include video, audio, and text for communicating, along with shared computer tools, video is a distinguishing feature, establishing context and providing a broader range of activities and behaviors. The desktop video conferencing aspects of these systems can include point-to-point or multipoint computer networked machines with video and audio capabilities. Groupware is multi-user software that allows users to collaborate when working with computer artifacts, which is one of many activities that can occur in media spaces. While video conferencing is realistically restricted to synchronous

communication, groupware can support synchronous and asynchronous collaboration.

It is useful to make a content/process distinction between the different media technology used to provide for interaction. The technology used to create media spaces can be viewed as supporting communication about work, or it can represent the work itself (see Olson, Card, Landauer, Olson, Malone, and Leggett, 1993). Typically, video has been investigated in the context of supporting communication surrounding other activities, including the use of groupware. How video supports interaction where it represents the work itself, as in the sharing of physical artifacts, has not received as much attention. Investigations into groupware, on the other hand, have focused mainly on the conditions when it represents the work itself, for example, in the case of a shared spreadsheet. This is somewhat a natural line of investigation stemming from the nature of the different technologies, but also a result of the bias in CSCW research and its focus on office work and the talking heads uses of video surrounding it. Considering the conditions under which technology supports collaborative work content and process will have important implications for the effectiveness of media spaces, such as how different media need to be presented to support the different contexts.

Face-to-Face Video

One taxonomy provided by McGrath (1993) characterizes common office worker tasks and has received considerable attention in CSCW. McGrath has categorized eight types of tasks, including planning, creative, intellectual, decision making, cognitive

conflict, mixed motive, competitive, and performance tasks. These tasks are categorized into four general processes: generate alternatives, choose alternatives, negotiate among alternatives, and execution. Research in CSCW, for the most part, has inappropriately restricted itself to these or similar office worker tasks when considering the appropriateness of video. This paradigm narrowly focuses on interactions that occur during face-to-face communication. Hence the issue has become one of whether video adds to computer-mediated communication through face-to-face video interactions. Some controversy exists in the literature as to the value of video under these conditions.

In face-to-face communication it has been observed that making eye contact, interpreting facial expressions, conveyance of gaze, and attending to body gestures all contribute important information (Argyle, 1975; Argyle, 1976; Harrison, 1974; Rutter and Stephenson, 1977; Mantei, Baecker, Sellen, Buxton, Milligan, and Wellman, 1991; Buxton, 1992; Heath and Luff, 1992). More recently, others (see Tang and Isaacs, 1993; Fussell and Benimoff, 1995) have further argued that video, as the result of the visual cues afforded, is critical in communication. Considering the types of information listed above that are provided by video, Isaacs and Tang (1994) have found that desktop video conferencing improves the ability to show understanding, forecast responses, give nonverbal information, enhance verbal descriptions, manage pauses, and express attitudes. Williams (1977) and Angiolillo et al. (1993) claim that video improves performance for interpersonal collaborative tasks, but not for information exchange tasks. For example, Williams found interpersonal tasks requiring negotiation and

persuasion to be assisted by video. Clearly video provides a great deal of information, but is this information necessary, crucial, or unimportant for CSCW?

The added effectiveness of video over simply audio channels for performing interpersonal collaborative tasks has been questioned (Cook and Lalljee, 1972). A series of early studies investigating the effectiveness of video-based communication repeatedly found the audio channel to be the most important in communication processes (Chapanis, 1971; Chapanis, 1975; Ochsman and Chapanis, 1974; Weeks and Chapanis, 1976; Pye and Williams, 1977). Nardi, Schwarz, Kuchinsky, and Leichner (1993) claim that studies evaluating the benefits of face-to-face video for collaborative work have failed to show any utility for its use.

Video-As-Data

The discussion of the role of video use in media spaces until this point exemplifies the first of three central problems outlined in the introduction: the lack in addressing appropriate task contexts that can benefit from video. This can have restrictive consequences for understanding all the parameters associated with media spaces in general and telepresence specifically. Certain communicative, social, and technical principles significant to media spaces can be misinterpreted by focusing narrowly on office tasks, for example, the findings outlined above concluding that the audio channel provides the most important communicative information. The effectiveness of video in conferencing systems should be addressed for tasks where the information afforded by video is useful rather than just merely a novelty. These conditions will arise under situations where

visual information is imperative for coordinated activities. One clear case where these conditions exist is the sharing of objects that are physically remote from one or more of the group members but are paramount to the coordinated activities.

Although the above discussion addressing the value of face-to-face video is an important line of investigation and one that continues need of resolution, *video-as-data* under many conditions is a more germane line of inquiry. Others are beginning to advocate this very position (Whittaker, 1995; Ramsay, Barabesi, and Preece, 1996). Video-as-data focuses on video images of shared workspace and objects that considerably extend the shared context provided by talking heads' video. Several researchers have shown the value of video-as-data in industrial and medical applications. For example, video has been used to monitor remote locations of industrial plant operations (Tani, Yamaashi, Tanikoshi, Futakawa, and Tanifuji, 1992). Whittaker (1995) and Nardi et al. (1993) have demonstrated the value of video for surgical teams when coordinating operating room activities.

Even though we have argued that video-as-data has enormous potential for supporting CSCW and is more pertinent than talking heads' video in many circumstances, the fact that media spaces have the potential for both, in addition to groupware tools that provide shared computer artifacts, stresses the need for considering all the types of information provided in these environments. Why? All three will be used in combination for most types of applications that fully utilize media space potential. As such, it is important that we address the second central

problem limiting the effectiveness of video use in media spaces: the lack of sufficiently configuring equipment to support different tasks, which include face-to-face and video-as-data.

Multiple Camera Viewpoints

Cameras never provide the same amount of information as being co-located with the persons or objects of interest. Many technological limitations of cameras reduce their capability to reproduce face-to-face visual performance or the sharing of physical objects and workspace: resolution, frame rate over networks, field of view, lack of stereoscopic images for depth information, scanning capabilities, etc. To account for some of these problems, cameras can be mounted to provide maximum contextual information (long shot), in face-to-face configurations, or in other context specific views. Although other views can be useful, most media spaces offer little more than face-to-face video.

Additionally, any single camera position provides unique information, but at the cost of limiting information from the other views possible. One solution for overcoming visual limitations due to camera use is to furnish multiple camera positions. Gaver, Sellen, Heath, and Luff (1993) developed a system called Multiple Target Video with four switchable cameras that allowed remote participants to extend the shared context by providing access to task-related physical objects and workspace. They found users preferred one view over another depending on the task, and face-to-face views were rarely preferred. They argue that face-to-face views of remote participants do not affect intellectual, decision-making tasks. Tasks that involve social cues such as

negotiating, however, will be affected by multiple views.

Another central limitation with camera use, even with multiple cameras, is they are stationary or only movable locally. There are a number of solutions that could be used to provide more appropriate and flexible access to remote locations through camera control. First, cameras that automatically track the remote participant(s) could be used to provide continual access to remote user's actions. While this solution may be most beneficial for face-to-face orientation or remote user actions, often the focus of both the local and remote user is oriented toward objects or workspace that is not contiguous with either participant. In this case remote-controlled pan, tilt, zoom, and focus cameras could be used for maintaining orientation to participants and workspace. Although this solves the problem of maintaining focus on several aspects of a remote space, it may be awkward to establish shared focus with this technique as opposed to simply moving the head and eyes when users are co-located. Also, it is often desirable to maintain a focus on what the collaborator is attending to in the remote location. This is naturally obtained when users share the same physical space by attending to body gestures, orientation, and visual attention, but may be difficult with the remote access techniques discussed above. Another option is to have what the remote user sees automatically feed to the other participant. One approach for providing this is to use a head camera mounted on the remote participant that consistently tracks what is being viewed.

Presentation of Multiple Viewpoints

All the solutions outlined above have advantages and disadvantages. Each one

provides visual context for different conditions, while limiting context from one of the other views possible. It is simply not feasible from a practical standpoint to provide the same visual access to remote space as is possible when being physically located in that space. When multiple views are available, design solutions must also be provided for presenting the visual information to users.

Several solutions are possible for presenting multiple inputs: A single monitor can be used for switching between views; a single monitor can display more than one signal at a time (multiple windows); multiple monitors can be used for displaying multiple inputs; or a combination of the methods above can be employed. It is likely that the utility of the different views and the necessity for moving between them will depend heavily on the type of task being performed and the collaboration surrounding it. For example, Gaver et al. (1993) found that, when four video inputs were presented on a single monitor, the duration spent in any single view and number of times switching between views depended strongly on the type of task being performed by collaborators. When view switching dominated use of the system, users had problems determining relations between views and had problems maintaining shared orientation surrounding the task. They suggested using multiple monitors for presenting different views.

Directional gaze cues in face-to-face conversation have been estimated to occur 60 percent of the time and mutual gaze 30 percent of the time (Argyle, 1975). Argyle and Cook (1976) and Kendon (1967) have identified several functions that gaze serves

in facilitating communication: regulating conversation flow, gaining feedback from the listener, expressing emotions, establishing the nature of the interpersonal relationship, and avoiding excessive information input. Sellen (1992) compared speech patterns between same room conversations and in two video systems, one with a single camera and monitor and the other with multiple cameras and monitors. Although no differences were found for speech patterns between the single and multiple monitor conditions, Sellen maintains that an in-depth analysis of the systems will reveal differences supporting the multiple monitor condition due to the ability to facilitate selective listening and gaze.

Sharing task space on a monitor(s) with video input and computer artifacts can be even more demanding. Smith, O'Shea, O'Malley, Scanlon, and Talor (1990) provided two monitors for sharing computer artifacts and visual images of the other user. A pattern emerged where eye contact occurred more at the beginning and end of tasks. Similar issues can arise when working with multiple task spaces. When using multiple monitors versus a single monitor with multiple windows for a complex task like those involved in flying an airplane, St. John, Harris, and Osga (1997) found that head movement to a second monitor was as fast as a single key press in a single monitor condition, and less disruptive of concurrent tasks. It is clear from the findings of this research that when multiple views of visual information are possible, view presentation is as significant as providing the information initially.

In even the simplest of media spaces, there are at least two signals that need to be

presented to the user: one camera view and one shared groupware application. The number of cameras or viewpoints, their orientation, and their presentation style to the user, along with shared computer artifacts, will substantially affect the usefulness and usability of media spaces. These characteristics will also impact the sense of telepresence experienced by the user. It is hypothesized that telepresence will also impact the usefulness and usability of media spaces. Therefore, it is important that we now turn to the last of three central problems discussed earlier associated with video use in media spaces: the lack of support for telepresence, which affords users a “common ground” for collaborative working.

TELEPRESENCE

One of the primary goals of media spaces is to create a sense of telepresence. For example, CAVECAT (Computer Audio Video Enhanced collaboration and Telepresence), a media space developed at the University of Toronto, was specifically designed to support shared personal presence (Ishii and Miyake, 1991; Kling, 1991). Many other media spaces have similar goals for creating telepresence (Gaver, Moran, MacLean, Lovstrand, Dourish, Carter, and Buxton, 1992; Mantei, et al., 1991; Buxton and Moran, 1990; Root, 1988). Despite this common goal of media spaces, the governing principles of CSCW telepresence have not been well defined, researched, or understood.

Computer-mediated and video-mediated communications technology typically have been compared to face-to-face meetings. Media richness theory (Daft and Lengel, 1986) and social presence theory (Short,

Williams, and Christie, 1976) consider the face-to-face paradigm the richest media at one continuum end, and written documents at the other. The degree to which video and other technology emulate face-to-face interactions will predict their effectiveness as a communication medium and likely their ability to produce a sense of telepresence, at least for some aspects of media spaces.

Telepresence has been defined as "the degree to which participants of a telemeeting get the impression of sharing space with interlocutors who are at a remote physical site." (Muhlbach, Bocker, and Prussog, 1995). Reflected in the definition above, the literature in CSCW provides only a "loose" interpretation of telepresence. The terminology used to define the concept has varied widely; this may reflect the lack of discrimination between the components of telepresence. For example, listed below are a few of the terms used to describe telepresence in the CSCW literature:

- Telepresence
- Presence
- Shared presence
- Social presence
- Shared personal presence
- Shared interpersonal presence
- Virtual presence
- Spatial presence
- Communicative presence
- Co-presence
- Virtual co-presence
- Shared space

Gaver et al. (1992) point out that virtual space created by telepresence technologies is both discontinuous and arbitrary: users are not surrounded by information, information bandwidth is limited, and users have

difficulty navigating in the space. As a result of this, media space telepresence experiences are determined by many factors, such as context of video use and the technology supporting its application discussed in previous sections. The type of telepresence discussed by authors is often defined by the technology of the system they study. In teleoperation environments where users remotely manipulated physical objects by video, and where the term telepresence originated (Minsky, 1980), the focus is on the experience of being at a remote physical environment, typically by only a single user.

Media spaces, on the other hand, create three distinct types of collaborative spaces, each with their own associated telepresence. Media spaces can also involve anywhere from two to several users. The first experience of presence these systems create is *shared workspace telepresence*. This space is characterized by groupware applications. Users experience the presence or telepresence of another user(s) based on actions performed individually or jointly across shared screens—a common workspace.

The second major focus of media spaces has been to create *interpersonal workspace telepresence* (Ishii, Kobayashi, and Grudin, 1992). This aspect of telepresence is created with video teleconferencing. By providing visual access to remote users, a sense of telepresence is achieved. However, rather than feeling as if the user is at a remote location, the feeling is more one of togetherness in the individual locations of each participant (any single user feels, to varying degrees, that the other user(s) is there with them at their location).

Lastly, a *remote workspace telepresence* can occur when video is used to visually share physical objects in one or both of the participant's remote locations. The latter two types of media space telepresence occur through the use of video, with remote telepresence being the closest to the use of the term in teleoperation. This type of telepresence is also most closely linked to the use of video-as-data.

Buxton (1992) makes a similar distinction between shared and interpersonal telepresence that he refers to as person and task telepresence. Person telepresence corresponds to interpersonal telepresence, and task telepresence corresponds to shared workspace telepresence. The technology used, its configuration, and the demands of a particular task environment can all affect how, when, and what type of telepresence the user will experience. Buxton maintains that the seamlessness of transitions made between different types of telepresence will affect usability, usefulness, and acceptance of media spaces. Telepresence is undoubtedly multidimensional in nature. Many factors of the video conferencing system, environment, and tasks interact with each other to produce varying degrees and types of telepresence.

Display size may also affect the telepresence of remote locations and participants. Scale has been an often overlooked factor in the consideration of telepresence. In fact, the scale of images are often reduced many times to account for small display sizes. This problem is exacerbated when multiple windows are used on a single display. Wall size displays can be used that present remote locations and participants in correct scale. Buxton (1992)

found that the sense of telepresence was so compelling when a large-sized video projection screen was used that participants referred to objects on their desk space as if it was shared space, when in fact it was not. Gestures can also be facilitated simply by using larger monitors, thus making movement more salient (Heath and Luff; 1992).

The field of CSCW encompasses many aspects of group work. Johanson (1988) has distinguished the different situations under which group work occurs along the dimensions of time and place (Figure 1). This model is useful for understanding how group work can occur

		<i>Time</i>	
		Same (Synchronous)	Different (Asynchronous)
<i>Place</i>	Same	Face-to-face Meetings	Project Rooms, Shift Work
	Different	Tele- and Video-conferencing	Email, Annotated Drafts

Figure 1. Characterization of group work (Adapted from Ellis, Gibbs, and Rein, 1991).

under different situations; however, the characterization is problematic for distinguishing among different systems because some systems now provide capabilities that span across quadrants (Baecker, 1993; Olson, et al., 1993). This is especially true with media spaces since they encompass both the combination of video and groupware that can span all four quadrants. This makes the investigation of telepresence in CSCW particularly difficult because a multitude of conditions exists for

which telepresence can occur, each having specific characteristics that are both independent as well as interdependent of other factors. As an initial investigation, this research specifically addresses telepresence under the classic same time, different place paradigm of CSCW.

THE MEDIA SPACE STUDIES

The studies reported here investigated the role of video in media spaces and its impact on telepresence. Selected to contrast each other, a negotiation task was studied as a traditional office environment task, and an object assembly task was examined as a highly visual task more reflective of video-as-data.

Sixty-four university students, most having little if any experience with media spaces, participated in both studies. The research was performed in the electronic conferencing suite of the Usability Methods Research Laboratory (UMRL) at Virginia Tech, which is comprised of three rooms: an experimenter room in the center with observational equipment, and two subject rooms on either side seen through one-way mirrors (Figure 2).

Both subject rooms were equipped with 200 MHz Pentium computers, 15 and 17 inch color monitors, microphones, and speakers (Figure 3). Video images were projected on the monitors using high quality, analog signals. All audio communication among the three rooms used high quality, full-duplex connections. Besides being able to observe the participants directly through the one-way mirrors, all the audio and video communications were captured in the observation room for later analyses (Figure 4). Microsoft NetMeeting was used to

share the drawing tool between the participants.



Figure 2. Electronic conferencing suite of the UMRL.



Figure 3. Typical equipment configuration for the subject rooms.



Figure 4. Observation and recording equipment in the UMRL.

The Action-Centered Task Video System
In an attempt to provide more flexible and appropriate visual access to remote environments through video-as-data, we designed and built an experimental system called Action-Centered Task Video (ACTV, pronounced “Active”). The system used a miniaturized camera (dimensions 1 by 3/4 by 1/2 inches) mounted to a pair of clear-lens eyeglasses. Figure 5 shows a close-up of the glasses with the miniaturized camera.



Figure 5. The miniaturized camera mounted to a pair of clear-lens eyeglasses.

The entire "headcam" setup was small and unobtrusive. Participants described it as being no more obtrusive than wearing eyeglasses or sunglasses. This camera view provides remote users with a gaze-directional view of what the person wearing the headcam is seeing.

The ACTV system provides what we have termed WISIWYG ("What I See Is What You Get") access. It is pronounced "wise-e-wig," in contrast to WYSIWYG (pronounced "wiz-e-wig") that stands for "What You See Is What You Get." It is also distinguished from WYSIWIS, which stands for "What You See Is What I See." WYSIWIS applies to interfaces where the shared context appears the same to all users. While WISIWYG does provide the same advantages of WYSIWIS in the sense of providing strong shared context, it can be differentiated along several dimensions. First, the human visual system and the camera/display system trying to reproduce it are significantly different for all the reasons identified as camera limitations discussed previously. The context also offers only unilateral control visually and physically to the context of interest. This is not the case in WYSIWIS, even when it is in a state that is referred to as "relaxed" WYSIWIS where users can modify their individual views independently to some varying degree. The user viewing the headcam image is to a large degree depending on the cooperative state of the individuals collaborating, reliant on the person wearing the headcam for access to the shared context. For this reason the system does introduce limitations resulting from disparate access to the shared context. Nevertheless, we believe that the ACTV system has tremendous potential for providing access to video-as-data for all

parties involved in the collaborative effort.

The Negotiation Study

A 2 x 2 factor design was constructed for the negotiation study with two levels of camera view and two levels of presentation method (Figure 6). Figure 7 shows the media space configuration for all the conditions. Camera allocation varied depending on the room. In the enhanced virtual space room three video

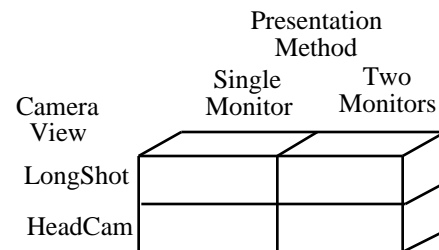


Figure 6. Experimental design for the Negotiation Study.

cameras were used: a bird's eye view camera (longshot), a face-to-face camera (facecam), and the miniaturized camera mounted on a pair of glasses worn by the user (headcam). The miniaturized camera produced only black and white images. All other cameras were full color. Only a single camera, a facecam, was provided in the limited virtual space room.

Half the participants in the study were provided with two monitors; the other half of participants received only a single monitor. Monitor allocation applied to both rooms for participant dyads. In the single monitor conditions, the 17" display was always used. When two monitors were used, the 15" monitor was always dedicated to the facecam view in both rooms, and the 17" display was used for all other views.

Camera view conditions, the longshot and

headcam, only applied to the participant in the limited virtual space room. Half of these participants received views from the

participants switched between the facecam and groupware view. In the dual monitor conditions, both views were simultaneously

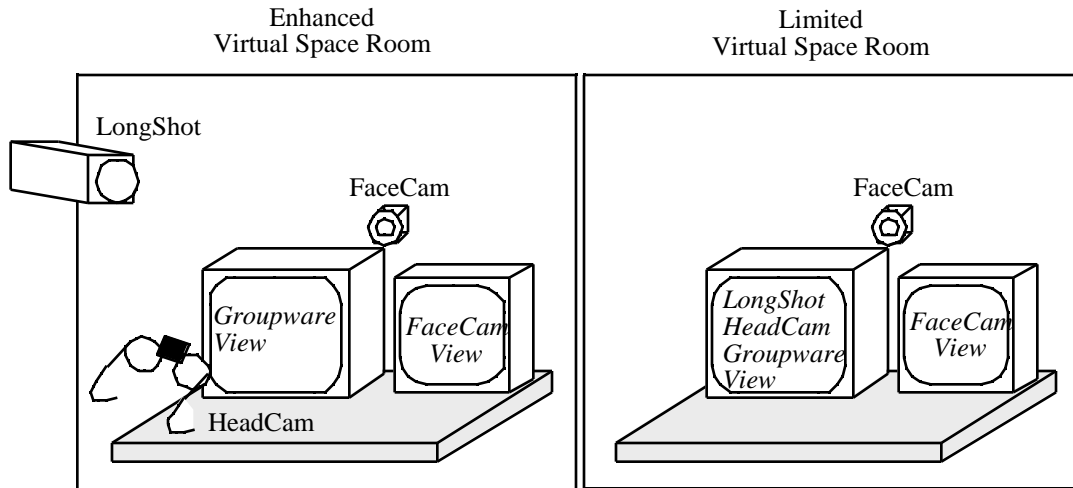


Figure 7. Media space configuration for the Negotiation Study.

longshot, and the other half received views from the headcam. However, participants in both rooms always received the face-to-face view. Therefore, participants in the limited virtual space room, viewing the enhanced virtual space room, always had three views: the groupware application, the face-to-face view, and a context view, either the longshot or headcam view depending on the condition. In the single monitor conditions, participants were required to switch between all three views using tabs at the bottom of the screen. In the two monitor conditions, the participants switched between the context video view and the groupware application on the 17" display, while the facecam view was presented to the 15" monitor.

The participants in the enhanced virtual space room, viewing the limited virtual space room, always received two views: the groupware application and the face-to-face view. In the single monitor conditions,

presented and no view switching was required.

Task

Users were required to negotiate over design objectives for an electronic meeting room. Participants were asked to cooperatively design the layout of an electronic meeting room using a computer-based shared drawing tool and the various camera views. A room layout as well as several objects that needed to be configured in the room were provided: a podium, four cameras, a large-screen TV, a projection screen, a chalk board, and a choice among three tables (Figure 8). Two of the table options were conference tables (one rectangular and one oval), and the third configuration was a set of eight individual workstations. Only one table configuration could be chosen, but all the other objects had to be used.

In negotiating the design of the electronic

meeting room, it was heavily stressed that participants keep five objectives in mind:

- 1) maintain maximum flexibility for conducting meetings
- 2) support different types of meetings
- 3) facilitate interaction
- 4) support access to and from remote participants
- 5) provide access to people with disabilities.

Participants were told that it was imperative

negotiating over layout configurations based on the design objectives.

View presentation methods. The multiple views available to the participants (video and shared drawing tool) created a situation where it was possible for one participant to refer to an item that the other participant was not attending to, especially when there were fewer displays than views available. Gaver et al. (1993) describe this as the lost in space phenomenon. For example, one

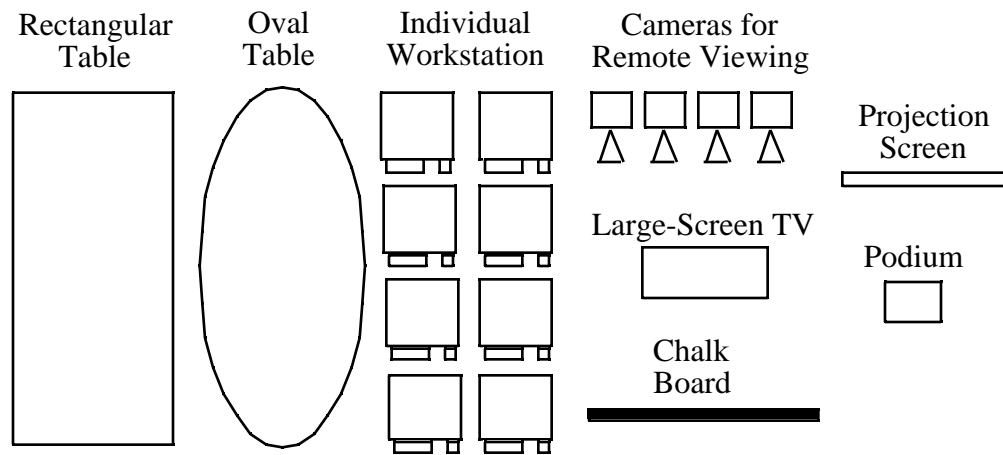


Figure 8. Objects for designing the electronic meeting room.

to meet all the design objectives, and to do this they had to work cooperatively with the other participant when negotiating an optimal design solution. Participants were given a maximum of twelve minutes to complete the task.

Results and Discussion

Overall, participants in both rooms found the systems easy to use even though most had never used media spaces or similar systems. Participants also were able to focus on the objectives for the task when designing the layout of the electronic meeting room, resulting in users continually

participant could be moving objects in the shared drawing tool to explain an idea, while the other was watching the face-to-face view. Many times such mismatches went completely unnoticed by both participants, which potentially led to miscommunication and misunderstanding.

Explicit and implicit language cues provided a means to avoid the mismatch between participants' views. Explicit questions and requests were made, such as "Can you see what I am doing?", and "Switch to the shared drawing tool, and I will show you." More commonly, participants avoided mismatches

by using implicit language cues. Phrases such as "this wall," "that camera," and "move this over here" allowed participants to know where they needed to direct their attention based on the other participant's comments. Many times view switches were not fast enough to completely see an action, and a participant would ask the other to repeat the behavior. The problems in maintaining a common frame of reference described above were primarily observed when there was only one monitor available. Although problems maintaining a shared reference in the two monitor conditions did occur, they were much less frequent. The two monitor conditions in essence expanded simultaneous access to the shared virtual space, eliminating many of the problems relating to shared reference.

Another significant factor limiting the single monitor conditions was the overhead involved in switching between views. Switching between multiple views manually on a single display required too much additional effort, especially considering the break downs in context and the need to follow the other participant's focus and activities that were described above. Because of this, users tended to "settle down" on the shared drawing tool once having initially explored the other views possible. Having two monitors extensively increased the use of other views, increasing the information bandwidth and therefore the resulting communication and cooperation. For example, users greatly increased their use of the face-to-face view dedicated to the second display in the two monitor conditions.

In the two monitor conditions with a simple glance, users could interact face to face, or

they could attend to the work. This also eliminated problems of reference because participants could easily determine the other participant's focus. If one user was turned away from the face-to-face view, the other participant knew they were focusing on the task. This created a situation very similar to face-to-face meetings: Both participants focused primarily on the work itself, but at anytime they could glance at each other to confirm the communication process through nonverbal cues. It was much easier and more natural for participants to simply glance at the face-to-face view, rather than switch between views manually.

However, participants in the limited virtual space room reported that the system was more difficult to learn and the task more difficult to complete than participants in the enhanced virtual space room. Presumably this was because they had three views possible at any given time but had only two monitors for viewing them. This made the system more difficult to use for these participants. In fact, because in the two monitor conditions the enhanced virtual space room participants had only two views, each dedicated to their own monitor which required no manual view switching, these participants reported having a greater awareness of events occurring in the other person's space. Presentation of views must be compatible with the number of views possible. Requiring users to manually toggle through views with on-screen tabs makes the system considerably more unnatural and difficult to use.

Problems also developed due to monitor allocation while coordinating activities with the shared pointer/selection tool controlled by the mouse. The mouse could be

controlled by only one participant at a time in the shared drawing tool. This forced participants to explicitly communicate over taking control of the mouse when they wanted to interact with computer artifacts. Some participants coordinated control of the mouse through language cues like "I am going to take control of the mouse and move this object." Other participants competed for control of the mouse by initiating actions opportunistically. With a single display it was difficult for participants to naturally share the tool. Turn taking became more cumbersome. With dual displays participants had more cues for when it was appropriate to take control. For example, the shared application could always be left visible, indicating when the other user was performing actions. The face-to-face view also could be used to indicate when control was desired. In the two monitor conditions participants could move naturally from face-to-face discussion to action taking.

Camera views. Camera views were utilized differently depending on the activities. It was apparent that in many cases the nature of the task dictated how the different views were used. Participants through trial and error gravitated to the shared drawing tool as the common space, not surprisingly considering that the coordinated activities were centered on the groupware application. This was almost entirely the case in conditions where only one monitor was available, and was true for both participants even though they could view the shared drawing tool or the face-to-face view. The participant in the limited virtual space room also had the longshot or headcam view available to them.

Some explicit language cues were used to

request switching to the face-to-face view, but these were rare. Implicit cues, aside from motion, also were rarely used to communicate the need to switch to the face-to-face view. Thus the face-to-face view became more of a channel to periodically check on the other participant, instead of a channel to meet and discuss issues, especially in the single monitor conditions. One reason that explicit meetings on the face-to-face view were rare may be because the participants focused on the shared workspace, which was the representation of the work itself rather than the social communication process about the work. However, as discussed above, in the two monitor conditions the face-to-face view was used much more extensively and naturally. Subtle visual cues about negotiated decisions may have been all that was needed from the facecam views.

The high quality of the audio link also may have reduced the need to use the face-to-face video links. With lower quality audio participants may have been compelled to use the face-to-face views to communicate more through expressions and gestures than spoken language. The high audio quality allowed the participants to detect voice inflections, background noises, and other cues about the co-participant's activity and agreement with negotiated decisions.

The longshot camera view afforded to the limited virtual space room participant offered little benefit in either the one or two monitor conditions, not surprisingly since the work objects themselves were computer-based representations. After initially checking this view, the participant all but abandoned it.

Another common communication breakdown occurred often when participants would point at the screen, and the gesture would go completely unnoticed by the other participant. This is a common phenomenon that occurs in media spaces. A significant lack of support for gesturing to common objects, either shared computer artifacts or video-as-data, occurs when users communicate and coordinate activities surrounding remotely based shared space created from information technology. In the case of the participant afforded the longshot, the view provided little benefit for accessing remote gestures. Little meaningful gesture information was communicated because the camera view placement (distance) prevented details on the screen from being discriminated. The longshot did not provide information specifically pertinent to the task.

The headcam in this case created a unique vantage point. Participants provided with the headcam view could see gestures being made to objects in the shared drawing tool by the person wearing the headcam. Figure 9 shows a user wearing the headcam



Figure 9. User gesturing toward the screen.
gesturing to a shared drawing tool object.

However, the ACTV system was not used in this capacity as much as hoped. Why was the headcam not utilized for this task? In the single monitor conditions, as for the face-to-face view, too much overhead was involved in switching between views. For the two monitor conditions, it is hypothesized that the face-to-face view afforded more information than the gestures offered by the headcam. Also, recall that three views were provided to the limited virtual space room participant, and even in the two monitor conditions, the participant would be required to now switch between two video views.

Had we provided three monitors, the story may have been quite different. And we suspect that this information would have been heavily utilized because participants in both rooms were observed gesturing to the screen. Participants in neither room, however, were observed using physical gesturing information from their co-participant. For the participant not afforded a headcam view, however, the information was simply not available. It is also worth noting that participants could use the telepointer to gesture and direct attention, and they did use this information considerably.

This was the only aspect of the negotiation study where video-as-data could be applied. But as we have seen from lack of its use for gesturing information, the visual information must be imperative for communication and coordinated activities. The equipment configuration must also appropriately support its use. Had the view from the headcam been easily accessible, distinctive from telepointer gesturing, and not in conflict with other information more relevant

because of access problems, support for video-as-data may have been more beneficial under conditions where the task was primarily centered around computer use.

It is, however, interesting to point out that, while participants reported the task being more difficult to complete with the headcam than the longshot, they found the headcam more useful than the longshot and reported it as being more beneficial when having multiple camera views. The headcam provided some video-as-data even when dealing with computer artifacts. And this video-as-data was more useful coming from a gaze directional viewpoint rather than from a fixed long shot of the user's physical space.

Telepresence. A strong pattern emerged where participants optimized information from the most advantageous views but without compromising telepresence. Above it was described how users had problems maintaining a common frame of reference across the different types of media space. These mismatches broke down the ability of participants to develop shared, interpersonal, and remote workspace telepresence. The reduced telepresence especially occurred in the single monitor conditions.

Ultimately, the shared drawing tool became the dominant telepresent space in the single monitor conditions. Because of the overhead involved in switching between views and the loss of shared context that resulted, the sense of presence was being compromised across the types of telepresence possible. Therefore, users opted for the most effective workspace that also provided a stable telepresence.

Buxton (1992) describes the seamlessness of these transitions between shared telepresent spaces as being one of the most important factors in system usability, usefulness, and acceptance. Ishil, et al. (1992) found that the ability of users to switch easily between eye contact and shared task context contributed to a feeling of telepresence. These conclusions were confirmed in this study. In the dual display conditions participants used the two views extensively because telepresence was not being compromised. Participants also used the face-to-face view considerably more in the two monitor conditions, leading us to conclude that telepresence was much greater in these conditions. Because the behavior of continually checking facial expressions is the norm for physically co-located collaborators, witnessing this behavior with the video interactions in the media space is compelling evidence of increased telepresence. For example, normally in video-mediated interactions asymmetries are introduced into interpersonal communication, transforming behaviors that would normally occur in the face-to-face communication process (Heath and Luff, 1992). Much of the asymmetry was eliminated with two monitors.

As described in the Camera View results section above, users in the enhanced virtual space room afforded with two monitors reported being more aware of events occurring in the other person's space. These users experienced a greater sense of telepresence than the users in the limited virtual space room because they were never forced to manually switch between views. Making virtual spaces as accessible as possible, mimicking co-located physical space access, increases telepresence. Providing more virtual space does not

necessarily increase telepresence and may in fact reduce telepresence if adequate access to the virtual space is not provided as well.

Participants also reported experiencing a greater sense of telepresence when using the face-to-face view in the headcam conditions than in the longshot conditions. Apparently, the headcam view increased a sense of sharing space for face-to-face interactions by providing more access to the remote space in other views. It is somewhat unclear why this would be the case, but because the headcam offered more useful access to the space than the longshot view, it may have increased the total sense of telepresence experienced in the face-to-face view. If this speculation turns out to be true, the experience across shared, interpersonal, and remote telepresence may

camera view and two levels of presentation method (Figure 10). Figure 11 shows the media space configuration for all the conditions. Camera allocation varied depending on the room. In the enhanced virtual space room three video cameras were used: a task space camera (overhead), a face-to-face camera (facecam), and the miniaturized camera mounted on a pair of glasses worn by the user (headcam). The overhead camera was created by placing the

Camera View	Presentation Method	
	17" Monitor	Wall-size Screen
OverHead		
HeadCam		

Figure 10. Physical Assembly Study.

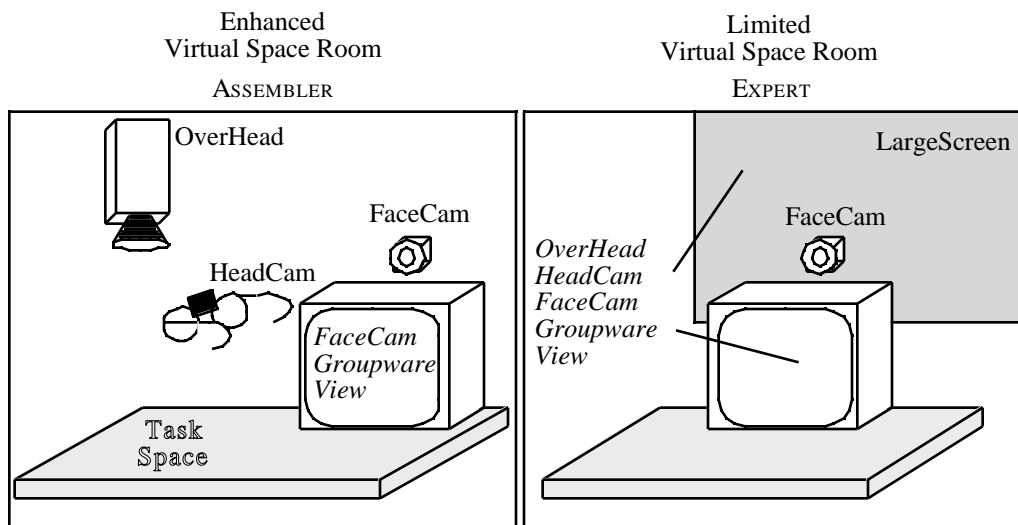


Figure 11. Media space configuration for the Physical Assembly Study.

be greater than the sum of each type of telepresence taken individually.

The Physical Assembly Study

A 2 x 2 factor design was constructed for the physical assembly study with two levels of

headcam in a box overhanging the physical task space (Figure 12). As in the negotiation study, only a facecam was provided in the limited virtual space room.

In this study both rooms were afforded only

one 17" monitor. However, for the presentation method conditions, half the participants in the limited virtual space room received the 17" monitor, the other half a high resolution liquid crystal display (1024 x768) that was projected onto a 4 by 5 foot wall-size screen (Figure 13).



Figure 12. Overhead camera configuration for the enhanced virtual space room.



Figure 13. Wall-size screen in the limited virtual space room.

As in the negotiation study, only the participants in the limited virtual space room received the camera view conditions. In this case half the participants received the overhead view, the other half the headcam view. Consequently, the limited virtual

space room participants were afforded three views: groupware application, face-to-face view, and either the overhead or headcam view depending on condition. The participants in the enhanced virtual space room once again received only the groupware and face-to-face views. Since only one display was used in all conditions, participants switched between views using tabs at the bottom of the screen.

Task

In this study participants cooperatively assembled a small LEGO car (Figure 14). The participant in the limited virtual space



Figure 14. The Technic 8207 LEGO car used in the assembly task.

room was designated the "expert" and given a set of step-by-step, pictorial instructions for assembling the car (instructions contained no textual descriptions). The participant (assembler) in the enhanced virtual space room was required to physically assemble the car based strictly on verbal and graphical (shared drawing tool) instructions from the expert. The expert was not allowed to show the instructions using video images. Participants were allowed to take as long as needed to finish the car assembly; however, they were

instructed to work as fast and accurate as possible.

The unassembled car contained dozens of small, intricate parts. Many of these parts served a mechanical purpose. For the car to work properly the parts had to be precisely and correctly put together. Figure 15 shows a set of the instructions for constructing the steering mechanism of the car, and Figure 16 shows a user assembling the steering mechanism while using the headcam.

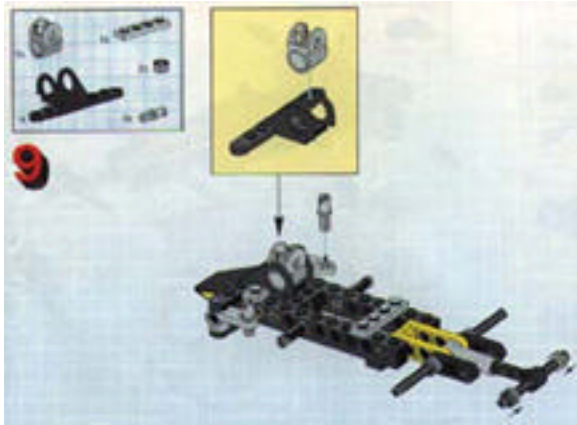


Figure 15. Example of the instructions for constructing the LEGO car.



Figure 16. A user assembling the car's steering mechanism while wearing the headcam.

It was a highly complex visual task requiring a great deal of coordination between the co-participants. Because the task was realistic in terms of being visually demanding, it had a high degree of face validity for being representative of video-as-data. The car assembly is very similar to real-world tasks in manufacturing, training, or surgery that require a great deal of visual access to shared space and objects where coordinated activities are imperative. These sorts of applications are ideal for the use of video-as-data, requiring remote access to visual information for collaborative work.

Results and Discussion

All of the subjects were able to complete the task within 25 minutes. Many, however, did struggle with assembling some of the part sequences correctly the first time. It often only became apparent on latter stages of constructing the car that previous steps were completed incorrectly. Participants would then return and correct the error. While the task was demanding, it was reasonably workable if participants communicated adequately about their coordinated activities.

Communication and coordination were significantly different for this task than in the negotiation study described above. Participants had to rely on each other continually for the coordinated activities to progress. Also, because of the task's highly visual nature, smooth coordination only occurred if participants communicated with each other visually as well as verbally. Shared access to the space visually was imperative. In this case video images became shared artifacts themselves, especially with use of the headcam. Participants shared, used, communicated about, and coordinated

the headcam view images.

The development of a “task-specific” language was crucial to successfully completing the assembly. Almost all participants created a terminology to describe car parts. The more elaborate the terminology, the more successful participants were. The use of terminology was particularly important for the expert who could see the objects and assembly in progress and had the instructions for describing what pieces looked like, where they went, and how they fit together. The most successful participants categorized the parts into different groups based on shape, color, or function. This was usually directed by the expert who had to describe the parts and their placement.

Explicit instructions were much more effective than vague commands. Participants frequently reported assembly pieces being “very hard to describe.” In one example that highlights participants who failed to develop an adequate task specific language, an expert continually repeated the phrase “put the piece in the other piece,” which thoroughly confused and frustrated the assembler. One participant stated that the most difficult portion of the task was developing a common terminology to discuss the assembly pieces.

View presentation methods. View presentation methods included the 17” monitor versus the large LCD projection screen. Very little in terms of behavior seemed to be impacted by the larger LCD screen. In both cases participants performed the task similarly. However, participants did report that they thought the system was easier to use and the task easier to complete

with the monitor than the LCD. This finding may reflect the apparent resolutions of the two monitors. While both monitors had a 1024x768 pixel resolution, the LCD is considerably less bright once projected onto a wall screen. This in effect reduces the apparent resolution due to brightness and contrast reductions. Also, because a LCD projection system must be projected at distance to get the large screen size ratio compared to the monitor condition, participants were seated at approximately 6 feet from the display. Having to use the keyboard and facecam directly in front of them while viewing the screen at a distance may have been awkward for participants.

Camera views. Performance in the assembly task varied depending on strategies employed by the participants. Approaches that involved drawing items in the shared drawing tool to explain instructions took much longer than those that focused on the video views of the car assembly alone. Performance was noticeably worse for people who appeared to be non-artistic or lacking skills with computer drawing applications. In most cases shared drawing tool use followed a sequence of communication failures. Once verbal descriptions failed, the expert would try to illustrate assembly parts and associated actions with the drawing application. When the drawing tool was effective, partial drawings were all that was needed to communicate concepts. Generally speaking participants continually reduced their use of the drawing tool, abandoning the drawing strategy altogether as the task progressed and the difficulty of drawing the pieces was realized. This stemmed partially from the tool’s ineffectiveness, and because as most users developed a task-specific language for

parts and assembly procedures, the task got considerably easier. A few users—mostly reflecting the experts' approach—painstakingly chose to continue to use the shared drawing tool regardless of its obvious drawbacks. This dramatically increased assembly performance time for these users. Finally, regardless of the camera view condition, there were no differences between how the participants viewed the usefulness of the shared drawing tool. In other words, the drawing tool did not differentially affect system usefulness based on whether users had the headcam or overhead camera.

The facecam was used significantly more by the assembler than by the expert. This could be expected since (1) the assembler's only video was a facecam view, and (2) the assembler was the recipient of the instructions for much of the communication process. Since the assembler was typically receiving information from the expert, he or she would continually glance to the face-to-face view. The expert, on the other hand, was more interested in a view of the assembly process. Experts were required to switch between views with screen tabs since only one monitor was available. Switching between views on a single monitor, as in the negotiation study, required too much additional attention and required a loss of context to the primary workspace. Experts inferred communication effectiveness from the actions taken by assemblers. However, if we had provided users with multiple monitors the pattern of camera view use may have been significantly different. Experts especially, but assemblers as well, may have used more face-to-face communication, changing not only communication patterns but likewise the type of coordinated activities.

Participants who developed advanced techniques for optimizing camera utilization completed the task more efficiently. For example, in one technique we observed participants raising either the pieces or the partially assembled car to the overhead camera or headcam for closer inspection. Even more effective was to use the higher quality (color and resolution) face-to-face camera instead of the overhead camera or headcam for examining parts. This effectively turned the face-to-face view into a shared task workspace rather than an interpersonal workspace for simply communicating. This proved especially effective when using the color of the pieces for instructional purposes. Only a few of the participants discovered this method, however. The lack of color images was frequently cited by participants as the main problem with the headcam/overhead camera views. These views were nearly universally praised by participants, however, for providing unique information necessary for effective task completion.

The overhead camera provided the best overall visual perspective of the task. This view provided the expert with information about all the pieces involved in the assembly, allowing the expert to scan the workspace and direct the assembler to the correct car parts. This was substantial for the assembly task because locating the correct parts was a significant component of the entire task. Once parts were located, many times the assembler was asked to bring parts closer to the lens for a more detailed view. Keep in mind that many of the parts were quite small, requiring intricate placement during the car construction. One of the more notable drawbacks of the overhead camera was that the expert lacked a

full understanding of the assemblers attentional focus. Aside from hand gestures visible in the overhead field of view, the expert was unaware of what the assembler was directing her or his attention toward.

The ability of the headcam to provide very direct and specific feedback to the expert on what the assembler was focusing on during the task made the headcam an overall superior context view. The headcam view was always directed to the exact point where users were concentrating their efforts. In some ways this provided more direct feedback than even being physically co-located. When sitting next to or near another person physically, one has to infer or make an estimate as to where another person's gaze is directed, and this estimate is usually determined from a single vantage point. In addition, this information must be continually checked visually by sharing attention between checking the gaze and visually focusing on where the gaze is directed. Although this does occur normally with relative ease, it is by no means as exact as "looking through the eyes of another." Seeing through the precise viewpoint of another user gives a first-person perspective that is a decisively powerful form of understanding shared visual space. It is worth noting here that providing foveal eye region information that could be mapped translucently onto the headcam view using eye tracking technology would make the information provided by the ACTV system even more useful. This technique would give the person receiving the headcam view precise information about the other participants visual focus.

Conversely to asking the assembler to move parts closer to the overhead camera, experts

sometimes requested that the assembler move their viewpoint away from the workspace with the headcam. Experts requested this action to see a wider perspective of the task space. One possible improvement to the ACTV system that could account for viewing distance problems would be to provide the user receiving the headcam view with zoom capabilities. This could to some degree emulate the visual system of the person wearing the headcam, and thus provide more equal access to the shared visual information.

The users receiving the headcam view reported the ACTV system as being more useful in performing tasks than those who wore the camera system. It appears that the users wearing the headcam failed to fully appreciate its usefulness. Participants also reported the face-to-face view to be more beneficial when using the overhead camera than the ACTV system. These findings reflect less reliance on the face-to-face view when the ACTV system was used. The ACTV system provided more specific information regarding coordinated actions and context, which reduced the need for the face-to-face view.

Telepresence

One of the main factors being considered for telepresence in this study was the impact of a large wall-size display compared to that of a standard monitor. The hypothesis behind scale issues associated with large displays is that by making images of remote participants and workspace true-life scale users will experience a greater sense of telepresence. Unfortunately, display size did not appear to affect telepresence in this study. However, we believe that we erred in the implementation of the large screen

display, not that the issue of scale was unimportant per se. Because we used a LCD projection system, the user had to sit at a reasonable distance from the display for it to be visible, and this turned out to be 6 to 8 feet. Aside from resolution problems discussed above that may have also contributed to the lack of an effect for this variable, the field of view of display images between the monitor and LCD did not change significantly. In other words, images were much larger on the LCD display, but because the distance from the observer to the display itself increased greatly for the LCD, the field of view of the images on the monitor and large screen were approximately the same. If we had used a back-lit display that could be viewed at close distances, scale may have been a significant factor for the experience of telepresence.

Participants using the ACTV system reported experiencing a greater sense of sharing the same space with their co-participants than participants who used the overhead camera. It appears that head coupling a viewpoint to a remote co-participant, a viewpoint linked to workspace and objects that are jointly associated with coordinated activities, gives the person receiving the view a richer experience of actually being located in that space. This is an intriguing finding since in neither case did the person receiving the remote workspace view have control over the viewpoint, not directly in any case. The increased experience of telepresence in the headcam conditions may, however, be explained from the ability to indirectly control the viewpoint. First, users receiving the headcam view could request the other participant to direct their gaze in a desired manner. Second, by giving explicit

instructions about actions to be taken with the assembly process, implicit instructions were being given for viewpoint control. The ability to control viewpoint in these ways created a sense of telepresence for the user.

It is worth noting that participants reported similar experiences of telepresence in both the limited and enhanced virtual space rooms, and these experiences were similar regardless of camera view condition. What is interesting is that the participants in the different rooms were at times experiencing the same type of telepresence, but for the majority of the task, they were experiencing different types of telepresence. The participants in the enhanced virtual space room could only experience shared and interpersonal workspace telepresence since they were merely afforded the groupware and face-to-face camera view. At the same time the limited virtual space room participants were provided the groupware, face-to-face view, and either the headcam or overhead camera view. The primary state of the users' workspace amounted to the assembler being in the physical space (referring to the assembler's attentional focus) where the car parts were being assembled, and the expert being located in the remote workspace via the headcam or overhead view.

Recognize that for the majority of the time this resulted in one user experiencing no telepresence (they were physically present) and the other user experiencing remote workspace telepresence. Furthermore, although the expert was mostly in the remote workspace, the assembler moved in and out of the interpersonal space while instructions were being given by the expert. Participants also met in the shared

workspace when one of the participants was trying to illustrate concepts with the drawing tool. These findings have important implications for the use of video-as-data and the telepresence associated with it. Remotely collaborating participants often require different virtual and physical space access needs when working with video-as-data. For coordinated activities to go smoothly, both parties rely heavily on the ability of the individual remotely accessing the physical objects and space to be successfully telepresent in that space.

CONCLUSIONS

In this report we have outlined three fundamental issues surrounding media space use as they relate to video-mediated communication and collaboration: (1) the appropriateness of video use needs to be considered in video-as-data contexts to fully appreciate the benefits of video applications; (2) how equipment is configured to support collaborative video use can significantly affect the usability and usefulness of video systems; (3) media space systems need to be designed to support telepresence.

The two studies described in this report demonstrated the usefulness of video under a variety of conditions. Contrary to current research findings, the results from the negotiation study illustrate the advantages of video even for applications where users are primarily performing joint activities with computer artifacts. Equally significant, video-as-data unequivocally demonstrated the value of video for coordinating activities that primarily deal with physical artifacts remote from one or more of the co-participants. These types of activities would virtually be impossible to perform without video conferencing capabilities.

Equipment configurations for the different media spaces made a significant impact on how the technologies were used and whether they were effective. Conclusions have been drawn regarding the ineffectiveness of video for computer-supported communication and collaboration without due consideration of how technology needs to be designed to support video use. For example, how useful video is and how effectively it is utilized depends heavily on supporting task-critical access to video images in relation to other workspaces available. Having multiple monitors to support multiple workspaces is crucial for the effective use of video in this regard. The advantages of face-to-face video can be disregarded and misinterpreted if its presentation is not considered in a context where its access matches similar affordances found in physically co-located face-to-face communication. In addition, the ACTV system provided access to remote workspaces in a manner not supported by more conventional approaches to video, establishing robust support for context-sensitive visual communication. Eye gaze viewpoint information was used to communicate actions and context in a form not supported by other video systems or actually being physically co-located. Video images generated from the ACTV system became artifacts that were used jointly and independently, communicated with and about, and coordinated for understanding and manipulating workspace shared by network-based collaborative working teams.

One of the central goals of this research was to validate a new model of telepresence for video-based computer-supported cooperative work. These initial findings substantiate the claim that users experience *shared, interpersonal, and remote*

workspace telepresence as they interact with different aspects of media spaces similar to the ones presented here. The extent to which users experience telepresence depends largely on how the technology is configured to support the different interaction styles with video, for example, providing multiple monitors to access different virtual spaces. How users make transitions between the different workspaces affects usefulness, usability, and telepresence of virtual workspaces created from video. One of the working hypotheses generated from these studies is that the three types of telepresence outlined in this paper taken together may be greater than the sum of the types of telepresence considered individually.

The media spaces were universally praised by participants for ease of use and enjoyment. One user exclaimed, “initially it was difficult, but towards the end, I wanted it in my office.” That statement summarizes participant expressions toward the technology in the two studies. Problems with the systems were encountered, but overall the media spaces designed and implemented in these studies provided an effective means for video-based CSCW.

ACKNOWLEDGMENTS

The authors would like to thank John Kelso for his assistance in setting up the media spaces. This research was sponsored in part by the National Science Foundation and the Southeastern University College Coalition for Engineering Education.

REFERENCES

Angiolillo, J. S., Blanchard, H. E., and Israelski, E. W. (May/June, 1993). Video telephony. *AT&T Technical Journal*, 7-20.
 Argyle, M. (1975). *Bodily communication*. New

York, NY: International Universities Press.
 Argyle, M, Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, U. K.: Cambridge University Press.
 Bly, S. A., Harrison, S. R., and Irwin, S. (1993). Media Spaces: Video, audio, and computing. *Communications of the ACM*, 36(1), 29-47.
 Buxton, W. (1992). Telepresence: Integrating shared task and person space. In *Proceedings of Graphics Interface '92*, pp. 123-129. San Francisco, CA: Morgan Kaufmann.
 Buxton, W., and Moran, T. (1990). EuroPARC's integrated interactive intermedia facility (iiif): Early experiences. In *Proceedings of the IFIP WG8.4 Conference on Multi-User Interfaces and Applications*. Herakleion, Crete.
 Chapanis, A. (1971). Prelude to 2001: Explorations in human communication. *American Psychologist*, 26, 949-961.
 Chapanis, A. (1975). Interactive human communication. *Scientific American*, 232(3), 36-42.
 Cook, M., and Lalljee, M. G. (1972). Verbal substitutes for visual signals in interaction. *Semiotics*, 3, 221-231.
 Daft, R. L. and Lengel, R. H. (1986). Organizational Information requirements, media richness and structural design. *Management Science*, 32(5), 554-571.
 Egidio, C. (1988). Video-conferencing as a technology to support group work: a review of its failures. In *Proceedings of the Conference on Computer-Supported Co-operative Work*, 13-24.
 Fussell, S. R., and Benimoff, N. I. (1995). Social and cognitive processes in interpersonal communication: implications for advanced telecommunications technologies. *Human Factors*, 27(2), 228-250.
 Gaver, W., Moran, T., MacLean, A., Lovstrand, L., Dourish, P., Carter, K., and Buxton, W. (1992). Realizing a video environment: Europarc's rave system. In *Proceedings of ACM CHI '92 Conference on Human Factors in Computing Systems*, pp. 27-35. New York: Association for Computing Machinery.
 Gaver, W., Sellen, A., Heath, C., and Luff, P. (1993). One is not enough: Multiple views in a media space. In *Proceedings of INTERCHI '93 Conference on Human Factors in Computing Systems*, pp. 335-341. New York: Association for Computing Machinery.
 Harrison, R. P. (1974). *Beyond words: An introduction to nonverbal communication*. Englewood Cliffs, NJ: Prentice-Hall.
 Heath, C. and Luff, P. (1992). Media space and

- communicative asymmetries: Preliminary observations of video-mediated interaction. *Human-Computer Interaction*, 7, 315-346.
- Ishii, H., Kobayashi, M., and Grudin, J. (1992). Integration of inter-personal space and shared workspace: ClearBoard design and experiments. In *CSCW '92: Computer Supported Cooperative Work*, pp. 33-42. New York: Association of Computing Machinery.
- Ishii, H., and Miyake, N. (1991). Toward an open shared workspace: Computer and video fusion approach of Team WorkStation. *Communications of the ACM*, 34(12), 37-50.
- Isaacs, E. A., and Tang, J. C. (1994). What video can and cannot do for collaboration: A case study. *Multimedia Systems*, 2, 63-73.
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 32, 1-25.
- Kling, R. (1991). Cooperation, coordination and control in computer-supported work. *Communications of the ACM*, 34(12), 83-88.
- Mantei, M. M., Baecker, R. M., Sellen, A., Buxton, W. A. S., Milligan, T., and Wellman, B. (1991). Experiences in the use of a media space. In *Proceedings of ACM CHI '91 Conference on Human Factors in Computing Systems*, pp. 203-208. New York: Association for Computing Machinery.
- McGrath, J. E. (1993). A typology of tasks. *Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*, pp. 165-168. San Francisco, CA: Morgan Kaufmann.
- Minsky, M. (1980). *Telepresence*. *Omni*, 6, 15-51.
- Muhlbach, L., Bocker, M., and Prussog, A. (1995). Telepresence in videocommunications: A study on stereoscopy and individual eye contact. *Human Factors*, 37(2), 290-305.
- Nardi, B. A., Schwarz, H., Kuchinsky, A., and Leichner, R. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. In *Proceedings of INTERCHI '93 Conference on Human Factors in Computing Systems*, pp. 327-334. New York: Association for Computing Machinery.
- Ochsman, R. B., and Chapanis, A. (1974). The effects of 10 communication modes on the behavior of teams during co-operative problem-solving. *International Journal of Man-Machine Studies*, 6, 579-619.
- Olson, J. S., Card, S. K., Landauer, T. K., Olson, G. M., Malone, T., and Leggett, J. (1993). Computer-supported co-operative work: research issues for the 90s. *Behaviour & Information Technology*, 12(2), 115-129.
- Pye, R., and Williams, E. (1977). Teleconferencing: Is video valuable or is audio adequate? *Telecommunications Policy*, 6, 230-241.
- Ramsay, J., Barabesi, A., and Preece, J. (1996). Informal communication is about sharing objects and media. *Interacting with Computers*, 8(3), 277-283.
- Root, R. W. (1988). Design of a multi-media vehicle for social browsing. In *Proceedings of CSCW '88*, 25-38. New York: The Association for Computing Machinery.
- Rosen, E. (1996). *Personal videoconferencing*. Greenwich, CT: Manning.
- Rutter, D. R., and Stephenson, G. M. (1977). The role of visual communication in synchronizing conversation. *European Journal of Social Psychology*, 7(1), 29-37.
- Short, J., Williams, E., and Christie, B. (1976). *The social psychology of telecommunications*. London: John Wiley & Sons.
- Smith, R., O'Shea, T., O'Malley, C., Scanlon, E., and Taylor, J. (1990). *Preliminary experiments with a distributed multi-media, problem solving environment*. Unpublished Manuscript. Cambridge: Rank Xerox EuroPARC.
- St. John, M., Harris, W. C., and Osga, G. (1997). Designing for multi-tasking environments: Multiple monitors vs. multiple windows. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, pp. 1313-1317. Santa Monica, CA: Human Factors and Ergonomics Society.
- Tang, J. C., and Isaacs, E. (1993). Why do users like video? Studies of multimedia-supported collaboration. *Computer Supported Cooperative Work (CSCW)*, 1, 163-193.
- Tani, M., Yamaashi, K., Tanikoshi, K., Futakawa, M., and Tanifuji, S. (1992). Object-oriented video: Interaction with real-world objects through live video. In *Proceedings of ACM CHI '92 Conference on Human Factors in Computing Systems*, pp. 593-598. New York: Association for Computing Machinery.
- Weeks, G. D., and Chapanis, A. (1976). Cooperative versus conflictive problem solving in three telecommunication modes. *Perceptual and Motor Skills*, 42, 879-917.
- Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, 84(5), 963-976.
- Whittaker, S. (1995). Rethinking video as a technology for interpersonal communications

theory and design implications. *International Journal of Human-Computer Studies*, 42, 501-529.

Wish, M. (1975). User and non-user conceptions of PICTUREPHONE service. In *Proceedings of the Human Factors Society 19th Annual Meeting*. Santa Monica, CA: The Human Factors Society.