Experiences in the Use of a Media Space

Marilyn M. Mantei, Ronald M. Baecker, Abigail J. Sellen, William A.S.Buxton, and Thomas Milligan

Department of Computer Science

Barry Wellman

Department of Sociology
University of Toronto
Toronto, Ontario M5S 1A4

ABSTRACT

A media space is a system that uses integrated video, audio, and computers to allow individuals and groups to work together despite being distributed spatially and temporally. Our media space, CAVECAT (Computer Audio Video Enhanced Collaboration And Telepresence), enables a small number of individuals or groups located in separate offices to engage in collaborative work without leaving their offices. This paper presents and summarizes our experiences during initial use of CAVECAT, including unsolved technological obstacles we have encountered, and the psychological and social impact of the technology. Where possible we discuss relevant findings from the psychological literature, and implications for design of the next-generation media space.

KEYWORDS

Computer-supported cooperative work, groupware, media spaces, desktop videoconferencing.

INTRODUCTION

Although Engelbart and English (1968) provided the first demonstration of a media space, the current wave of activity began with the Xerox PARC Portland Experiments (Goodman & Abel, 1986; Abel, 1990) and continued with recent developments including those at Xerox PARC (Stults, 1986, 1988; Bly & Minneman, 1990; Tang & Minneman, 1990), Bolt, Beranek and Newman (Thomas, Forsdick, Crowley, Schaaf, Tomlinson & Travers, 1988), Olivetti (Lantz, 1988), Bellcore (Root, 1988), and Rank Xerox EuroPARC (Buxton & Moran, 1990).

Despite marked differences in technology and approach, these experiments suggest common themes:

- Media spaces define new methods of communication, with novel and unforseen uses and potentialities. Communication through a media space is more than an approximation of face-to-face communication — it has a richness and complexity all its own.
- The effective realization of media spaces requires one to solve serious architectural and implementation problems in distributed computing.
- Group working environments contain an enormously rich collection of communication protocols. The subset of communication metaphors built into existing media spaces only begin to reflect the possibilities.
- Media spaces raise serious ethical issues such as those of surveillance and privacy.

We have constructed a media space that enables a small number of individuals and groups located in separate offices to meet and collaborate without leaving their offices. This paper presents initial observations based on several months use of the system. Our goal is to contribute to the emerging dialogue on the potential, appropriate design, impact, and implications of media spaces. After a brief introduction to our system, we present our observations organized in terms of unexpected affordances, technological obstacles, and social and psychological impact. Each of our observations is discussed in terms of applicable underlying theories and suggested design recommendations.

THE CAVECAT SYSTEM

The CAVECAT (Computer Audio Video Enhanced Collaboration And Telepresence) system consists of a number of enhanced workstations connected by a digital+audio+video network. Each workstation consists of a personal computer, a TV monitor, a TV camera, a pair of speakers, and a microphone. A 4 x 1 video board allows the display of composite images of up to 4 sites (Figure 1). In some locations, video boards can place a lower resolution video image directly on the workstation's screen so that a separate monitor is not necessary.

The heart of the system is the switching network (Figure 2), patterned after the IIIF Server developed at Rank Xerox EuroPARC (Buxton & Moran , 1990; Milligan, 1989). Audio and video transmission is analog, but is switched digitally by the IIIF Server software residing on a workstation. Personal workstations in each office send messages via Ethernet to the IIIF server requesting connections. The IIIF Server also examines privacy settings for each office to determine if requested access by another office is permissible.

Figure 1. Video image of a CAVECAT meeting.

A server agent resides on each personal workstation. The user interface to this agent permits each office occupant to select a variety of communication metaphors: task oriented, (e.g., calling a meeting); spatially oriented, (e.g., walking into someone's office); or object oriented, (e.g., turning off the microphone in your office) (Louie, Mantei & Buxton, 1990).

We are developing shared software to support the computer communication aspect of the media space. These packages include a shared drawing tool and a shared text editor. Until this software is in place, we are using commercial software such as Timbuktu (Farallon, 1989), and ShrEdit, an experimental shared editor (Olson,Olson, Mack & Wellner, 1990).

UNEXPECTED AFFORDANCES

In order to understand the impact of the media space on its users, we applied it to ourselves by setting up CAVECAT nodes linking two faculty offices, the system programmer's office, and a graduate student work area. For the communication interface, we used a spatial metaphor consisting of a layout of the offices involved. We digitized video images of the CAVECAT users and placed these miniaturized images inside their owners respective onscreen offices Moving one or more of these images from one virtual office into another establishes a visual and acoustic link with the office or offices of choice.

Meetings of groups of groups

We had intended our setup to work primarily as a communicating device for one person located in each office. Our camera setups and camera angles were not designed for video conference meetings. However, in reality there was a natural demand for such a facility and it was used in this way. Individual members of the group used CAVECAT to introduce their visitors to others without going through the effort of physically walking the visitor over to the other individual's office for a more time-consuming interruption.

Mirror function

Although the system was designed for displaying other meeting participants, unexpected benefits came from displaying oneself. We used this "mirror" facility to make sure we were properly framed in the camera. The mirror function was included automatically in the split-screen display of 4-way conversations (Figure 1).

Monitoring function

Another surprising use was for the purpose of being virtually in one's own office. Instead of using the media links to place oneself virtually in another's office, we could also use the links as windows into our own offices when we were not there. We could monitor who was looking for us and when the phone rang. We could also use the system for security.

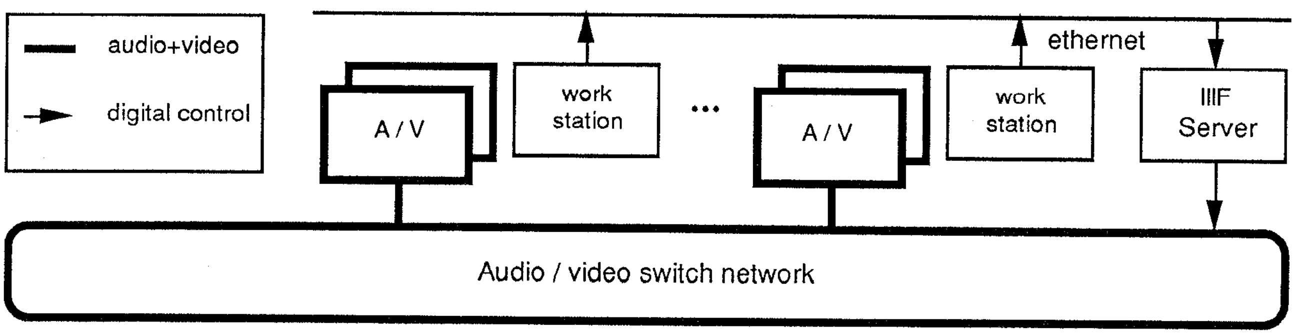

Figure 2 . Schematic diagram of the layout of the CAVECAT network.

TECHNOLOGICAL OBSTACLES

In our experience, a number of technological obstacles presented usability problems:

System response time delays

We initially built a prototype software package to run our virtual office connections. The time to establish a connection or to enter or leave an ongoing meeting was nearly two seconds. This delay quickly became intolerableAlthough system response times improved with a new version of the software, we must note that the two second wait time corresponds well with network switching and satellite delays for very long distance communications. This poses serious problems for the design of virtual offices spanning long distances.

Audio levels and noise

Ambient noise in the speakers' offices presented a major problem for sound quality. Different furniture arrangements, different numbers of people in the office, whether the office door was open or not, and where the office owner chose to sit in relation to speakers and microphones all had the potential to further degrade the quality of the sound. As a result, sound levels had to be continuously adjusted. When CAVECAT participants could not hear another participant, they tended to raise their voices, disturbing the audio levels more.

Obviously, it is inappropriate for us to tamper with the flexibility of individuals to decorate and move around their offices, nor do we have an intelligent device to automatically adjust levels. We are modifying the system by providing each participant with the ability to control their own audio, but we need to determine ways to make such adjustments easy and to guarantee that individual adjustments do not cause deterioration of the overall sound quality through feedback.

Sound localization

Participants of CAVECAT commented on how the sound in the shared communication seemed to come from "out of the air" rather than from the direction of the person speaking. When we had multiple participants communicating, the inability for participants to localize the sound sometimes made it difficult to determine who was speaking. Often, it was also difficult to know if one's phone was ringing, because of confusions with rings in other offices transmitted over the network.

Lighting and camera angles

The automatic light adjustments in our camera were intelligent but not intelligent enough. For example, some cameras were pointing at whiteboards located behind the occupants. The camera automatically adjusted for these white backgrounds, leaving the individuals in the foreground bathed in shadow. Some cameras were perched on bookshelves, while others sat on the side of their personal workstation. The location of the cameras, the lighting and color of the room's background, and the distance the individual chose to sit from the camera all affected the size and quality of the image transferred to the other offices.

Bad camera angles could distort impressions of speakers, which was particularly serious when one was not very familiar with a participant and when one was negotiating. Cameras with automatic focus continually zoomed in and out on the people moving about their offices, tending to make viewers in other offices slightly motion sick. It is clear that we need to consider carefully the placement of both camera and human, and to provide appropriate controls for presenting desirable video images.

PSYCHOLOGICAL AND SOCIAL IMPACT

Meetings between and within offices

Meetings of groups of groups of people were difficult to manage because people within an office were more "present" with each other than they were in the virtual office across the media. The physical closeness of people in the same office made them much more aware of their physical neighbors than of their video neighbors. This fact, combined with the poor acoustic quality across the network, encouraged people to address those in the same room rather than those in the other offices.

Two types of conversations often took place simultaneously. One conversation was public where people spoke to the camera. Private conversations were also being held among individuals in each office. Coordinating these two kinds of conversations and establishing the dominance of the public discussion when appropriate presented a challenge.

Another problem with such large meetings was that the displayed size of many individuals was so reduced that fine points of the interaction were often not visible. Facial expressions and nonverbal gestures were not as salient; interactions seemed less "real" than the ones taking place in the same room. When we switched from a meeting of multiple offices (a 2 x 2 video configuration of all participants) to a two-way communication (a single screen presentation of other participant), conversations again took place between offices rather than within offices.

Gaze and eye contact

Because participants were engaged in looking at the video image of their counterpart, they did not look directly into the TV camera. We did not use teleprompters or half-silvered mirrors to facilitate looking at the screen and the camera simultaneously. Thus eye contact was not established.

Gaze and mutual gaze are an important part of normal face-to-face communication. It is estimated that 61 percent of conversation involves gaze and 31 percent involves mutual gaze (Argyle, Ingham, Alkena & McCallin, 1973). Gaze serves at least five functions (Argyle et al., 1973; Exline, 1971): to regulate the flow of conversation; to provide feedback on how the communication is being perceived by the listener; to communicate emotions; to communicate the nature of the interpersonal relationship; and to reflect status relationships.

Rank Xerox EuroPARC (Buxton & Moran, 1990) used *video tunnels* — boxes containing cameras which pick up the video image of an individual via a half silvered mirror in front of their TV monitor. This solution makes it necessary for people to sit directly in front of big black boxes rather than in normal communicating positions within their office. EuroPARC has removed these video tunnels, but alternate solutions to achieving eye contact have not been devised. Hewlett-Packard embeds a miniature camera in the top of the workstation and uses on-screen video, but even this angle does not permit complete eye contact.

The best solution we have been able to achieve is produced by placing a camera with a wide-angle lens in front of and above the person and just above the monitor. The camera should not be very close to the person; zooming is used to make the person appear closer.

Status of meeting participants

Another interesting observation was that CAVECAT changed social status relationships due to the loss of the usual spatial and nonverbal cues which convey status information.

In face-to-face meetings, the seating of people in a room is usually indicative of a hierarchy with higher status people occupying more central positions or "head of the table" locations.

The design of CAVECAT unintentionally introduced its own social status cues. In meetings of four individuals, CAVECAT arbitrarily positioned participants' images in a 2 x 2 grid. CAVECAT also configured the video images for a meeting based on who requested the meeting. This meant that if meetings were reconvened after a short break by a different person, the result was a different image configuration. This was highly disconcerting to the participants. It was as if everyone had left the room and returned to take new positions around the table.

Meeting coordination

Our observed problems with loss of traditional status cues and generation of new cues speaks to the more general issue of control in discussions. When important cues are missing or degraded, there is a greater need for a moderator to control turn-taking and group decision processes. For example, people wanting to take control in conversations will often lean in to indicate their desire to speak. This cue is difficult to detect on video.

Our observation is that a moderator's success may depend on having "media presence" — a factor which does not necessarily come into play in face-to-face meetings.

Image size and personal impact

A participant's effectiveness within a conversation and the way each participant was perceived by others seemed to be, in part, determined by video image size. Participants with large images appeared to have more impact in the discussion. Participants with small images seemed distant and less effective in the conversation.

The size of the video image was determined by four factors: the screen size of the monitor, the distance of the viewer from the TV monitor, the distance of the person from the camera, and the zoom setting of the camera. Participants often had different sized images because these variables were rarely adjusted.

Video image and social distance

Image size and angle also interacted with people's perception of their social relationship to other participants. Inappropriate image size sometimes gave the sense of people being too personal or too impersonal in the conversation.

These observations are consistent with the social psychology literature which finds that interpersonal physical distance is predictive of relationships between people (Argyle & Dean, 1965). People who are only casually acquainted tend to maintain a distance of about 4 to 12 feet between them while interacting. Distances from 1 1/2 to 4 feet tend to be maintained for friends, while distances of less than 1 1/2 feet are reserved for intimate relationships. It is well established that people quickly become uncomfortable if the distance between them is perceived to be inappropriate for the relationship. Too close, and people feel their space is being violated. Too far, and people are also uncomfortable.

In the media space, what is relevant is the "perceived" interpersonal distance, a virtual distance rather than a physical one. Observation suggests that video images may be viewed as less personal and intrusive in general. In one hot summer's day usage, the participants talked freely with each other over the media space, but one individual immediately donned a lab coat to cover her shorts and tank top when meeting face-to-face with the same individual. One the other hand, occasionally a meeting participant reached for a book from a shelf or stood up, creating views several inches from the participant's neck or stomach, making an onlooker uncomfortable.

What is also unusual about a media space is that the interpersonal distance may be simultaneously different for any member of the group communicating. This is not the case for physical distance where distances between people are, in a sense, negotiated and shared. In CAVECAT, a participant's personal space can be invaded without the invader being aware of this.

Privacy and surveillance

When we first put the system in place, any node on the network could immediately connect with any other node via video and audio. The system was kept running semi-continuously because of the need to troubleshoot startup problems. This lack of privacy led to very strong protection behaviors on the part of two participants — one who was negotiating the secret sale of a company, and another who was negotiating problems in a personal relationship. The first individual unplugged or shut off all CAVECAT connections while the second worked shorter hours.

It became clear very early that "knowing" when you were connected to another office and being able to inhibit the connection were critical and necessary features. The media space, as it was, did not provide enough feedback to indicate that others were suddenly present in your office. In addition, although available, the privacy setting features in the IIIF Server were too complicated for easy use. One good approach to the provision of adequate feedback is through the use of non-speech audio cues (Buxton & Moran, 1990; Gaver & Smith, 1990).

IMPLICATIONS FOR FUTURE DESIGN

Our experiences begin to illustrate how technology can significantly alter the nature of human communication patterns. One important conclusion is that many of the cues implicit in face-to-face communication situations need to be taken into account and provided for in the design of the interface.

There are many communication variables that we had not considered in our original design. It is easy to take for granted aspects implicit in face-to-face communication such as the physical presence of someone in an office implying a desire to communicate, or nonverbal gestures of individuals in a meeting.

Another implication of our observations is that it is important to provide easy-to-use features that place some of the system variables under user control. For example, because there are many aspects of the visual image which affect the way participants perceive each other and interact, it is important that users are able to adjust for viewing and being viewed.

We have a number of specific plans based on our experiences to date:

- We are developing metaphors for communication and privacy protection that follow accepted communication practice. These metaphors consist of interface selections that allow the user to: (1) wait to see someone who is busy talking to someone else; (2) drop by to ask a quick question; (3) shut one's door partially or wholly; or (4) whisper something to a co-worker at a meeting. We are working with variables such as video image size, blurriness of the video image, duration of the video/audio connection and verbal and non-speech audio cues to create these communication protocols.

- We are building an underlying visual language for manipulating the parameters of the system so that its users can build their own protocols for adjusting the media space parameters.

- We are putting in new basic functionalities such as individual control of audio and comparative viewing of video images. We are also trying out automatic audio switching so that the person speaking in a meeting becomes the single image presented to all participants. This avoids our image size problems but may create new problems associated with not being able to view everyone in the meeting.

Despite our current problems, our media space has proved to be a successful tool for collaborative communication. We find that it is used extensively for communicating about software development. The system not only allows an approximation to face-to-face communication, but also confers many new advantages upon its users. We can have virtual open offices with the bad effects of continuous noise and disturbance removed and the good effects of proximity enhanced. Meantime we are continuing the process of iterative design in order to minimize the problems and capitalize on the advantages discussed in this paper.

ACKNOWLEDGEMENTS

For research support, the authors are indebted to the Natural Sciences and Engineering Research Council of Canada, the Information Technology Research Centre of Ontario, Apple Computer, Digital Equipment Corporation, IBM Canada and particularly to Rank Xerox EuroPARC, which contributed the code for the IIIF Server. In addition, we are grateful to the University of Michigan, which loaned us the object code for their shared editor. We also wish to thank the many students who have worked long hours on CAVECAT: Beverly Harrison, Jeffrey Lee, Gifford Louie, Iva Lu, Kelly Mawby, Tracy Narine, Ilona Posner, Michael Sheasby, and Ian Small.

REFERENCES

Abel, M.J,. Experiences in an exploratory distributed organization. In Galegher, Kraut & Egido (Eds), *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, Lawrence Erlbaum Associates, 489-510.

Argyle, M. and Dean, J. (1965) Eye contact, distance, and affiliation. *Sociometry*, 28, 289-304.

Argyle, M., Ingham, R., Alkena, F. and McCallin, M. (1973). The different functions of gaze. *Semiotica*, 7, 10-32.

Bly, S.A. and Minneman, S.L. (1990). Commune: a shared drawing surface. In *Proceedings of the Conference of Office Information Systems*, Cambridge, MA, April 1990. 184-192.

Engelbart, D. and English, W.K. (1968). A research center for augmenting human intellect. In Greif, I. (Ed.) . *Computer-Supported Cooperative Work: A Book of Readings*, Morgan Kaufmann Publishers, San Mateo, Calif., 81-105.

Exline, R.V. (1971). Visual interaction: The glances of power and preference. In J. K. Cole (Ed.) *Nebraska Symposium on Motivation* Vol. 19, 163-206, University of Nebraska Press.

Farallon Computing, Inc. (1989). *Timbuktu.* 2201 Dwight Way, Berkeley, CA, 94704 USA.

Gaver, W.W. and Smith, R.B. (1990). Auditory icons in large-scale collaborative environments. In Diaper, D., Gilmore, D. Cockton, G. and Shackel, B. (Eds), *Proceedings of Human-Compuer Interaction, INTERACT'90*, Cambridge, England, August 27-31, 1990, 735-740.

Goodman, G. and Abel, M. (1986). Collaboration research in SCL. In *Proceedings of the First Conference on Computer Supported Cooperative Work*, Austin, TX, December 86.

Lantz, K.A. (1988). An experiment in integrated multimedia conferencing. In Greif, I. (Ed.) . *Computer-Supported Cooperative Work: A Book of Readings,* Morgan Kaufmann Publishers, San Mateo, Calif., 533-552.

Louie, G., Mantei, M. and Buxton, W.A.S.(1990) Making contact in a multi-media environment. HCI Consortium on CSCW, Ann Arbor, MI., February 1991.

Milligan, T. (1989). IIIF: The Integrated Interactive Intermedia Facility design report - Revision 3 January 1989. Rank Xerox EuroPARC working paper.

Buxton, W.A.S. and Moran, T (1990). EuroPARC's Integrated Interactive Intermedia facility (IIIF): Early Experiences. *Proceedings of the IFIP WG8.4 Conference on Multi-user Interfaces and Applications,* Heraklion, Crete, September 1990. 24pp.

Olson, J.R., Olson, G.M., Mack, L.A. and Wellner, P. (1990). Concurrent editing: the group's interface. In Diaper, D., Gilmore, D. Cockton, G. and Shackel, B. (Eds), *Proceedings of Human-Compuer Interaction, INTERACT'90*, Cambridge, England, August 27-31, 1990, 835-840.

Root, R.W. (1988). Design of a multi-media vehicle for social browsing. In *Proceedings of the Second Conference on Computer-Supported Cooperative Work*, Portland, OR, September 1989, 25-38.

Stults, R. (1986). Media space. Xerox PARC technical report. 20 pp.

Stults, R. (1988). Experimental uses of video to support design activities. Xerox PARC technical report SSL-89-19.

Tang, J.C. and Minneman, S.L. (1990). VideoDraw: a video interface for collaborative drawing. *Proceedings of CHI '90,* 313-320.

Thomas, R.H., Forsdick, H.C., Crowley, T.R., Schaaf, R.W., Tomlinson, R.S., and Travers, V.M. (1988). Diamond: a multimedia message system built on a distributed architecture. In Greif, I. (Ed.) *Computer-Supported Cooperative Work: A Book of Readings,* Morgan Kaufmann Publishers, San Mateo, Calif., 509-532.