

Telepresence: Integrating Shared Task and Person Spaces

William A. S. Buxton

Computer Systems research Institute
University of Toronto
Toronto, Ontario, Canada M5S 1A4

Abstract

From a technological and human perspective, shared space in remote collaboration has tended to focus on shared space of either the people or the task. The former would be characterized by traditional video/teleconferencing or videophones. The latter could be characterized by synchronous computer conferencing or *groupware*.

The focus of this presentation is the area where these two spaces meet and are integrated into what could be characterized as video-enhanced computer conferencing or computer-enhanced video conferencing.

From the behavioural perspective, the interest lies in how - in collaborative work - we make transitions between these two spaces. For example, in negotiating, the activity is mainly in the shared space of the participants themselves, where we are "reading" each other for information about trust and confidence. On the other hand, in preparing a budget using a shared electronic spreadsheet, for example, the visual channel is dominated by the task space.

How well systems affords natural transitions between these spaces will have a large impact on their usability, usefulness, and acceptance. Consequently, we investigate the design space and some of the issues affecting it.

Keywords: Human-computer interaction, CSCW, Videoconferencing, Groupware.

Introduction

Groups play an important role in our work-a-day life. Physical proximity facilitates interaction among group members. Even splitting groups across two floors of the same building can have a negative effect on group dynamics (Kraut & Egido, 1988), yet in many organizations groups are distributed across campuses, cities, countries, or even the globe. The health of these organizations is tightly coupled to the ability to maintain a sense of "group," despite such distances. Our interest lies in developing *telepresence* technologies appropriate for fostering such maintenance.

As we use the term, telepresence is the use of technology to establish a sense of shared *presence* or shared *space* among geographically separated members of a group. The topic is of particular interest now due to the ongoing convergence and affordability of the requisite computer,

telecommunications and audio/video technologies; however, if these technologies are going to be deployed in anything other than a tail-wagging-the-dog technology-driven manner, we must first develop a better understanding of what we mean by "shared space" or "shared presence" in the context of group interactions.

In what follows, we begin to investigate what is shared in various types of group interactions, and some of the technological implications of supporting such sharing. Our purpose is "consciousness raising" rather than the presentation of formal theories or models. Our case is made primarily through the use of examples. Our hope is to provide some foundation for making better design decisions and better exploiting the potential of existing and evolving resources.

Starting from the Known

The terms "meeting" or "group interaction" are almost devoid of information since they encompass such a broad range of activities. Each has its own set of properties and purposes. Only by understanding these properties can we hope to design the appropriate affordances into supporting technologies.

This is nothing new. Take architecture as an example. Because it is a mature discipline, we think of it as part of the general ecology of work, rather than as a technology. Yet a technology it is, and very much a technology to support group activities. Consider, then, the different types of group activities that are a part of our everyday work, and how the affordances of this technology have been designed to support them. We clearly understand differences of purpose, and choose the space (office, lounge, laboratory, board room, gym, lunch room, etc.) accordingly.

Because the technology is mature, we have a good sense of how to match the activity to the space. In order to be considered mature, the electronic meeting spaces of telepresence must meet the same dual criteria of supporting a comparably rich range of group activities and doing so in such a way that users have the same transparent sense of appropriateness of space-to-activity.

To speak of "videoconferencing" or "telepresence" is analogous to speaking about "buildings." While having some value, the grain of analysis is too coarse to foster an under-



standing of what goes on "inside." While we would never do so with rooms in a building, our current level of (im)maturity with electronic spaces has a tendency towards "one size fits all." This is something that we must break out of. The range of electronic meeting spaces, like the range of spaces in a well-designed building, must match the richness and range of meeting types. As a start to achieving this, we can move from the level of "buildings" to that of "rooms" and try to gain some insight into the nature of some of the different spaces that we want to share.

Person and Task Spaces

In what follows, we are going to consider presence in terms of two spaces: that of the person and that of the task. From even such a simple cut, several interesting insights emerge.

What we call shared *person space* in telepresence is the collective sense of copresence between/among group participants.¹ This includes things like their facial expressions, voice, gaze and body language.

By shared *task space* we mean a copresence in the domain of the task being undertaken. If we were doing a budget, for example, this might mean that each of us has the budget in front of us in the form of a shared spreadsheet. Despite the distance, each of us can act upon it to make changes, annotations, or just to indicate cells that are the subject of discussion.

Sometimes the person and the task spaces are the same. One example would be in negotiations or counseling. Here a major part of the task involves "reading" the other person, such as to evaluate confidence or trust². In other cases, such as our budget example, person and task spaces are more distinct. In what follows we shall see that different technologies lend themselves to differing degrees in supporting these two spaces. The point that we are leading to is that one of the most important attributes of a system is the *seamlessness* of their integration (Ishii & Miyake, 1991), and how well they match the needs of the activity to be supported.

Video Conferencing and Person Space: Some Examples

Traditional videoconferencing is a fairly good example of attempting to establish shared person space. While nobody would ever be fooled into thinking that the remote parties were actually in the same room, one can at least maintain an awareness of who is present and get a general reading of their body language, for example. The absence of checks like, "Are you still there Marilyn?" that are characteristic of telephone conferences is an example of what video contributes to maintaining a sense of personal presence.

¹This is in contrast to "personal space" which carries the connotation of privacy, not sharing. Thanks to Hiroshii Ishii for making this point and prompting me to change my terminology.

²This is sufficiently important that we might well refer to these as *trustification*, rather than *communication technologies*.

Fig. 1, illustrates one example of how video can be used to maintain a sense of personal presence in a four-way meeting.



Figure 1: A videoconference involving four participants.

The quality of the shared person space can be improved through design, however. Below, we give some examples that illustrate the breadth of the available design space. While many of these techniques are well known, few have found their way into mainstream videoconferencing. If establishing a strong sense of person space is important, then perhaps current practice needs to be reexamined.

For example, traditional videoconferencing is typically afflicted by an inability to establish eye contact among participants. This is because of the discrepancy of the position of the image of your eyes on my monitor and the position of your effective (surrogate) eyes, the camera, which is typically located on top of the monitor.

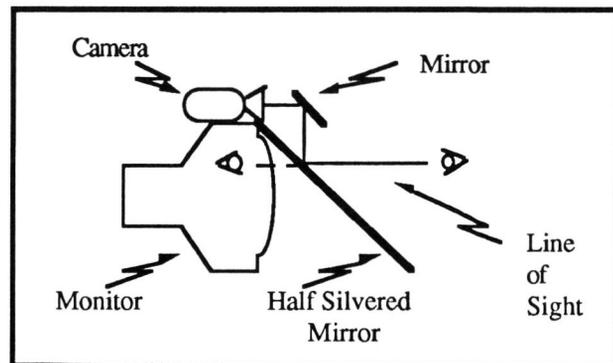


Figure 2: The Reciprocal Video Tunnel. Through the combination of a mirror and half silvered mirror, there appears to be direct eye-to-eye contact. The mirrors effectively place the camera right in the line of sight. A close approximation to reciprocal eye contact can be obtained if both parties are using such an arrangement (from Buxton & Moran, 1990).

By adopting teleprompter technology from the broadcast industry, this problem of eye contact can be largely over-



come. The technique is shown in Fig. 2, as it was implemented by William Newman at Rank Xerox EuroPARC. Two mirrors, one of which is half silvered, are used to reflect what is in front of the screen up to the camera, which is mounted on top of the monitor.

The use of such teleprompter-like technology to obtain eye contact is not new. It was patented in 1947 (Rosenthal, 1947), has been studied by Acker & Levitt (1987) and used by Newman (as mentioned above), and more recently in a novel form in the *Clearboard* system (Ishii & Kobayashi, 1992). While its use is not widespread in videoconferencing, users report greater comfort and naturalness in face-to-face meetings carried out using the technique.

Portrait painting provides the lead for another approach to augmenting the nature of personal presence using video. Video monitors have what is called a landscape *aspect ratio* (the ratio of the width to the height of a video monitor), because of their horizontal orientation. A very simple trick is to turn the camera and monitor at both ends of a conference onto their sides. The result, illustrated graphically in Fig. 3, is a *portrait style* aspect ratio.

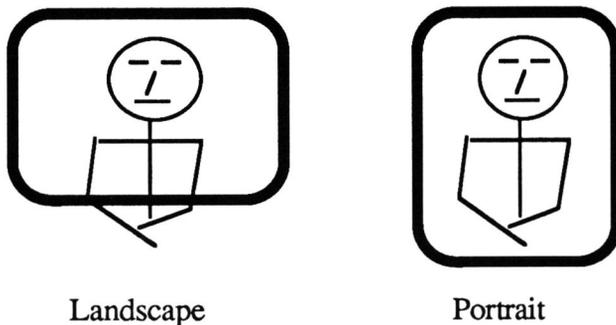


Figure 3: The effect of switching from Landscape to Portrait aspect ratios in person-to-person video conferences. Note that, all other things being equal, in the portrait orientation, the hands and desk-top are visible, thereby adding to the ability to use a richer vocabulary of body language in the dialogue.

When the image of a single person is to be transmitted, more of that person's body is visible without changing the size or resolution of the face. Consequently, in the example, the hands of the participant are visible in the portrait version, as would be the desk-top. The design affords access to a richer vocabulary of body language. As a prototype unit built by colleagues from the University of Ottawa has shown, this approach can be particularly effective where screen size is constrained, such as with small desk-top units, since a larger screen surface is available for a given width of package.

Next, let us consider the case of where we want to have a meeting involving the participation of more than two sites. At the University of Toronto, we have developed a system called *Hydra*, in which each remote participant is represented by a *video surrogate* (Sellen, Buxton & Arnott, 1992; Buxton & Sellen, 1991)³. The technique involves having a separate camera, monitor and speaker for each remote participant. As we have implemented it, these com-

³ After the fact, we have become aware that this approach was first developed by Fields (1983).

ponents are housed in very compact desk-top units, as shown in Fig. 4.



Figure 4: A user is seated in front of three Hydra units. In the photo, the Hydra units sit on the table in the positions that would otherwise be occupied by three remote participants. Each Hydra unit contains a video monitor, camera, and loudspeaker. A single microphone conveys audio to the remote participants (From Buxton & Sellen, 1991).

Using this arrangement, the notion of person space is preserved. Because of this it is potentially much easier to maintain awareness of who is visually attending to whom, and to take advantage of conversational acts such as head turning. The idea behind the design is to take advantage of existing skills used in the work-a-day world. For example, in comparing this technique to other approaches to supporting multiparty conferences (Sellen, 1992), the *Hydra* units were unique in their ability to support parallel conversations, which naturally occurred in the face-to-face base-line condition.

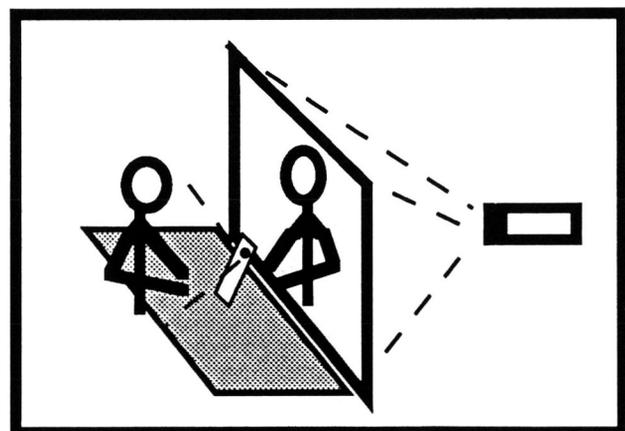


Figure 5: Using a projected image to obtain a life-sized cross-table presence. Participants are captured using a miniature camera on the desk-top, so as to minimize obstruction of the projected image. In our installation, we use one of the *Hydra* units (camera only), illustrated in Fig. 4.



Finally, the effect of scale has been little explored as a factor that influences a sense of presence. There is a strong possibility that if the video images are life size, that social relationships, such as power, may be more balanced and natural. We have observed this informally where, with head-and-shoulder shots, a projected image is presented at human scale.

Recently, we have been experimenting with projection techniques to achieve the effect of cross-table conversations. In this case, a video projection screen is placed directly against the desk, as illustrated in Fig. 5. The remote participant is then rear projected life-size. The result is powerful. The sense of presence is so strong that there is a compulsion to refer to things on the desk, despite the fact it is not really visible to the remote participant. This leads us to the topic of shared task space: what might be on the desk to discuss in the first place?

Shared Task Space

It takes very limited power of observation to note that we are sharing more than ourselves in face-to-face group interactions. I may be showing you my new sneakers, video or latest budget. Alternatively, we may both be scribbling madly on the whiteboard trying to brainstorm about the design of a new piece of software.

As there is a range of shared "accessories" and how they are used, so must there be a range of technologies in our repertoire to support similar sharing in telepresence. Like shared person space, the design space is rich and largely unexplored. The examples which follow touch the surface to give a feel for some of the issues and alternatives.

The (technically) simplest way to share some things that form part of the task space is to use the same channels as the person space. In videoconferencing, for example, we might just make sure that the subject of interest is visible to the camera. This is illustrated in the video frame shown in Fig. 6, where the participants are discussing the design of a PC board.



Figure 6: Using videoconferencing as a forum for discussing the design of a PC board. Both participants are shown one on either side of the frame. (from Shomi Corp., San Diego, CA.)

In many cases, this approach is effective and appropriate - but not always. Consider the difficulty if both participants didn't have the circuit board. Without the physical object, how would the person on the left in Fig. 6 point to problem areas, or indicate where changes should be made? While there is a *telawareness*, for the task at hand, there is clearly is not a *telepresence*.

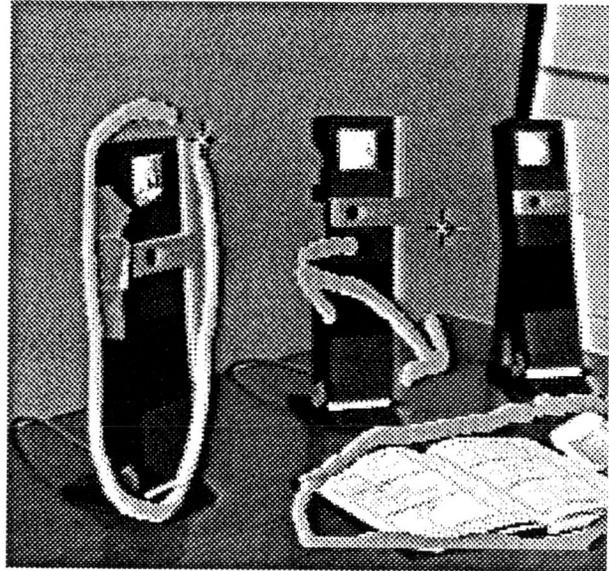


Figure 7: Distributed shared drawing on video to enhance communications in a videoconference. Here the marking have to do with the space occupied by the Hydra units (seen in Fig. 4) and other articles on the desk.

There are techniques that can be applied to this situation. One is a variation on a technique frequently used by television sportscasters: using a computer paint program to draw on, or *annotate* a video clip. The variation is to permit each participant in the conference to do so. This is illustrated in Fig. 7 which shows a frame from a conference where two participants are discussing the usage of the Hydra units (seen previously in Fig. 4)⁴.

This technique is extended even further by Millgram and Drascic (1990). They use two video cameras mounted side-by-side (like a pair of binoculars) to capture the object under discussion. By alternating between the frames from each camera, they transmit a stereo image of the view. This is overlaid with computer-generated stereo-pair graphics (such as pointers and markers) which permits participants to work in 3D.

At a certain point, or in certain cases, however, the video channel is inappropriate for supporting shared task space. If, for example, the task was to debug some code, then it may well be more appropriate to have the software in ques-

⁴ Note that the technique described differs from that found in many videoconferencing systems. In such systems, a still video image is transmitted, and one frequently cannot point at or mark-up the image. The technique described makes use of full-motion video, and may well (perhaps temporarily) use the same channel as the face-to-face communication.



tion available, rather than some video image. Here is a situation appropriate for complimenting video conferencing with shared synchronous computation.

Using dial-up telecommunications links, or computer networks, there are a number of ways that multiple users in remote sites can work together on a single computer application. A number of firms use such software, combined with teleconferencing, to provide remote product support.



Figure 8: *Liveboard* (Weiser, 1991): by combing large-screen interactive displays with advanced networks and distributed software, shared "whiteboards" can be provided to support brainstorming sessions and other collaborative work from remote sites.

Environments such as the X window system, coupled with large interactive displays, such as Xerox PARC's *Liveboard* (Weiser, 1991) are leading towards technologies to support distributed brainstorming sessions that preserve many of the properties of same-room sessions based around a whiteboard.

What we see from the examples is that we can use a range of techniques to support both shared and person spaces, and that being able to do so is important to supporting group activity across distances. What we haven't seen - to this point - is very much on how these two types of spaces work together, or relate.

Integrating Shared Task and Person Spaces: Two Examples

Shared ARK (Smith, O'Shea, O'Malley, Scanlon & Taylor, 1990) was one of the first studies to be undertaken at Rank Xerox's Cambridge EuroPARC (Buxton & Moran, 1990). It was an investigation of joint problem solving: subjects had to determine - through the use of a computer simulation - whether one stayed dryer by running or walking in the rain. Subjects were in separate rooms. They had a high fidelity voice link and a video link, implemented using the reciprocal video tunnel shown diagrammatically in Fig. 2.

The simulation was a distributed application presented to each user on a networked workstation, and took two people to operate. Within the task space, each user was "visible" by way of an identifiable cursor in the form of a hand. The

relationship of the workstation and video tunnel is shown in Fig. 9. Note that the position of the video tunnel is akin to having the remote participant sitting right beside you. Eye contact can be established by a simple turn of the head, and voice contact can be maintained throughout.



Figure 9: *Shared ARK* (Smith, O'Shea, O'Malley, Scanlon & Taylor, 1990): The shared task space is on the computer display on the left. The shared person space is via the video tunnel on the left which is an implementation by William Newman of the design shown in Fig. 2.

As with working on a paper on your desk with someone by your side, you couldn't look at your collaborator's face and the computer screen at the same time. So one aspect of interest was determining when subjects visually attended to the computer display, and when they established eye contact through the video tunnel. A pattern did emerge in which eye contact was established especially when they were initially negotiating how to proceed and at the end when checking results. When actually running the simulation - which was a visually vigilant task - the video tunnel was seldom used except for short glances.

Remember, however, that the video tunnel was not the only vehicle for establishing shared person space. While attending to the computer display, each user's surrogate "hand" provided a (limited) visual personal presence through its pointing and gesturing capability. This was supplemented by the voice channel (and in a later study, Gaver, Smith & O'Shea, 1991, nonspeech audio). When visual attention was directed at the computer screen, the speech and nonspeech audio established a shared space which was more effective than the highest fidelity video display.

While what we have described is an over simplification of the experiment, it is adequate to establish that subjects moved between task and person spaces as they moved through different components of the overall task. What we take from this is the observation that some (many or most?) complex tasks require a range of channels and modalities of communication in order to be effectively supported. The reason that *Shared ARK* was so effective was because the methods and overhead in switching contexts (such as from computer screen to eye contact) had the same *overhead* and *action* as is used in analogous work-a-day tasks. That is, they were built on existing everyday skills that subjects al-



ready possessed, resulting in a natural behaviour. This is evident to anyone watching the experimental tapes.

Videodraw (Tang & Minneman, 1990) and its successor *Videowhiteboard* (Tang & Minneman, 1991), are excellent examples of a smooth integration of shared personal presence in a distributed task space. The systems were concerned with providing tools to support design and brainstorming activities, such as one would encounter around a drawing pad or whiteboard, respectively.

Videowhiteboard's main power came from its sensitivity to the need to support both drawing and the body language and gestures that typically accompany design and brainstorming at a whiteboard. Consequently, the system cleverly enables participants to be visible one another on the drawing surface, much like in the face-to-face situation. This is illustrated in Fig. 10, which is a frame from a video of a work session with the system.

Summary and Conclusions

Through the use of examples, we have argued that effective telepresence depends on quality sharing of both person and task space. Through this, the interaction breaks out of being like watching TV, into a direct engagement of the participants. They meet each other, not the system.

The integration of these two types of space are important. The smoothness of transitions between them is critical. Without this, the natural flow of interaction is disrupted. If the flow is to be natural, then the overhead and styles of interaction used in everyday face-to-face meetings should set the standards and design basis for telepresence technologies.

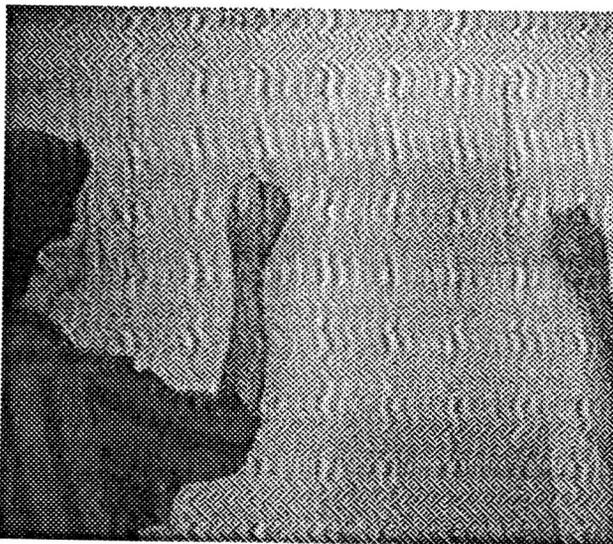


Figure 10: *Videowhiteboard* (Tang & Minneman, 1991): an excellent example of effectively blending shared person and task space. The remote participant appears as a shadow on the far side of the drawing surface. The approach supports a rich vocabulary of physical gesture, including the ability to anticipate intended actions.

What we hope the examples have illustrated is that, just as in traditional meeting spaces, one size doesn't fit all. There are a range of reasons that people meet and bonds that hold

groups together. Our technologies must reflect these reasons and bonds, and their richness. Current technologies do not excel in this regard. What we hope to have shown is that this need not be so.

The design space, as afforded by available and emerging technologies, is far richer than is evident by popular practice. Hopefully the examples help show the potential and provide some keys to how it can be untapped.

Acknowledgements

This paper reflects the results of countless discussions with colleagues at Rank Xerox EuroPARC, the University of Toronto, and at Xerox PARC. This contribution is gratefully acknowledged.

Our work in this area has been supported by the Ontario Information Technology Research Centre (ITRC), the Natural Sciences and Engineering Research Council of Canada (NSERC), Xerox Palo Alto Research Center (PARC), Rank Xerox EuroPARC, Cambridge, England, The Arnott Design Group, Toronto, Apple Computer's Human-Interface Group, Object Technology International, Ottawa, Digital Equipment Corp., Maynard, MA., and IBM Canada's Laboratory Centre for Advanced Studies, Toronto. This support is gratefully acknowledged.

References

- Acker, S. & Levitt, S. (1987). Designing videoconference facilities for improved eye contact. *Journal of Broadcasting & Electronic Media*, 31(2), 181-191.
- Buxton, W. & Moran, T. (1990). EuroPARC's Integrated Interactive Intermedia Facility (iiif): early experience, In S. Gibbs & A.A. Verrijn-Stuart (Eds.). *Multi-user interfaces and applications*, Proceedings of the IFIP WG 8.4 Conference on Multi-user Interfaces and Applications, Heraklion, Crete. Amsterdam: Elsevier Science Publishers B.V. (North-Holland), 11-34.
- Buxton, W. & Sellen, A. (1991). Interfaces for multiparty video conferences. University of Toronto. Submitted for publication.
- Fields, C.I. (1983). Virtual space teleconference system. *United States Patent 4,400,724*, August 23, 1983.
- Gaver, W., Smith, R. & O'Shea, T. (1991). Effective sounds in complex systems: the ARKola simulation. *Proceedings of the 1991 Conference on Human Factors in Computer Systems, CHI '91*, 85-90.
- Ishii, H. & Miyake, N. (1991). Toward an open shared workspace: computer and video fusion approach of TeamWorkStation. *Communications of the ACM*, 34(12), 37-50.
- Ishii, H. & Kobayashi, M. (1992). Clearboard: a seamless medium for shared drawing and conversation with eye contact. To appear in the *Proceedings of CHI '92*, May 1992.
- Kraut, R. & Egido, C. (1988). Patterns in contact and communication in scientific collaboration. *Proceedings of CSCW '88*, 1-12.
- Millgram, P. & Drascic, D. (1990). A virtual stereographic pointer for a real three dimensional video world. In D. Diaper et al. (Eds), *Human-Computer Interaction - INTERACT '90*. Amsterdam: Elsevier Science Publishers B.V. (North-Holland), 695-700.
- Rosenthal, A.H. (1947). Two-way television communication unit. *United States Patent 2,420,198*, May 6, 1947.



- Sellen, A. (1992). Speech patterns in video-mediated conversations. To appear in *The Proceedings of CHI '92*, May 1992.
- Sellen, A., Buxton, W. & Arnott, J. (1992). *Using spatial cues to improve desktop video conferencing*. 8 minute videotape. To appear in the *CHI '92 Video Proceedings*.
- Smith, R., O'Shea, T., O'Malley, C., Scanlon, E. & Taylor, J. (1990). Preliminary experiments with a distributed multi-media, problem solving environment. Unpublished manuscript. Cambridge: Rank Xerox EuroPARC.
- Tang, J. & Minneman, S. (1990). Videodraw: a video interface for collaborative drawing. *Proceedings of the 1990 Conference on Human Factors in Computer Systems, CHI '90*, 313-320.
- Tang, J. & Minneman, S. (1991). Videowhiteboard: video shadows to support remote collaboration. *Proceedings of the 1991 Conference on Human Factors in Computer Systems, CHI '91*, 315-322.
- Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3), 94-104.

