

Buxton, W. (1997). Living in Augmented Reality: Ubiquitous Media and Reactive Environments. In K. Finn, A. Sellen & S. Wilber (Eds.). *Video Mediated Communication*. Hillsdale, N.J.: Erlbaum, 363-384. An earlier version of this chapter also appears in *Proceedings of Imagina '95*, 215-229.

Living in Augmented Reality: Ubiquitous Media and Reactive Environments Redux.¹

William A.S. Buxton
Computer Systems Research Institute, University of Toronto
&
Alias | Wavefront Inc., Toronto

Abstract

One thread of this chapter presents an approach to the design of media. It is based on the notion that media spaces can be thought of as the video counterpart of ubiquitous computing. The combination of the two is what we call Ubiquitous Media. We go on to discuss the synergies that result from approaching these two technologies from a unified perspective.

The second thread is of a practice and experience nature. We discuss Ubiquitous Media from the perspective of having actually "lived the life." By basing our arguments on experience gained as part of the Ontario Telepresence Project, we attempt to anchor our views on practical experience rather than abstract speculation.

Introduction

In 1991, Mark Weiser, of Xerox PARC, published an article that outlined a vision of the next generation of computation (Weiser, 1991). He referred to this model as Ubiquitous Computing, or UbiComp. UbiComp was based on the notion that it is inappropriate to channel all of one's computational activities through a single computer or workstation. Rather, Weiser argued that access to computational services should be delivered through a number of different devices, each of whose design and location was tailored to support a particular task or set of tasks. It is on this notion of delivering computational services throughout our work, play and living spaces, that the ubiquity in the name is based.

In addition to ubiquity, UbiComp assumes that the delivery of computation should be transparent. There is a seeming paradox that arises between the principle of ubiquity and that of transparency. The examples which follow will hopefully demonstrate how this seeming paradox can be resolved.

¹ This version of the paper (June 2020) incorporates revisions to the published version. These are mostly typographical or wording changes to improve clarity. Two photos were added which were only referred to in the published version but had appeared in another chapter which I also wrote for the same book. Adding them makes this version free-standing. The added example is that on Front-to-Back Video Conferencing.



Figure 1: Xerox PARCtab. (Photo: Xerox PARC)

Around the same time that Weiser and his colleagues were developing the ideas that were to emerge as UbiComp, others down the hall at Xerox PARC were developing video-based extensions to physical architecture, so-called Media Spaces (Bly, Harrison & Irwin, 1993). These were systems through which people in remote offices, buildings, and even cities, could work together as if they were in the same architectural space. While prototypes, these systems enabled one to work side by side at one's desk with someone in a remote location. You could call out of your door and ask, "Has anyone seen Sara?" without thinking about whether the answer would come from Portland, Oregon or Palo Alto, California. Nor did it matter at which of these two centres either you or Sara were at. The technology supported a sense of shared presence and communal social space which was independent of geographical location. The result can perhaps best be described as a social prosthesis that afforded support of the links that hold together a social network - links which are typically only maintainable in same-place activities.

Reading Weiser's paper gives no hint of the activities of the Media Space group, and vice versa. However, I increasingly began to see the two projects as two sides of the same coin. Consequently, in my work with the Ontario Telepresence Project (at the University of Toronto, partially supported by Xerox PARC), I began to consciously apply the tenets of UbiComp to the media space technology. Thus, just as UbiComp deems it inappropriate to channel all of your computational activity through a single workstation, so in Ubiquitous Video (UbiVid) did we deem it inappropriate to channel all of our communications through a single "video station" (viz., camera, video monitor, microphone, loudspeaker). And as in UbiComp, the location, scale and form of the technology was determined by its intended function. And while ubiquitous, our focus was to render access to the services of these communications technologies transparent.



Figure 2: Shared open office via Media Space (Photo: Xerox PARC)

UbiComp and UbiVid - let us call them collectively Ubiquitous Media - represent an approach to design that is in contrast to today's multimedia computers, in which functionality is inherently bundled into a single device, located at a single location, and operated by a single individual. Ubiquitous Media, on the other hand, is an architectural concept in that it is concerned with preserving, or building upon, conventional location-function-distance relationships.

Ubiquitous Media can also be understood in relation to Artificial Reality. Rather than turning inward into an artificial world, Ubiquitous Media, encourage us to look outward. It expands our perception and interaction in the physical world. (For example, in the attempt to find Sara, consider the augmentation of the social space to include the physical space of both Palo Alto and Portland. The augmentation was socially transparent. There was no "user interface" other than that used in conventional architecture: one just called blindly out the door.) In contrast to "virtual" or "artificial" reality, we consider our use of Ubiquitous Media as Augmented Reality (Wellner, Mackay, & Gold, 1993).

In what follows, we discuss our experience living in such an environment over the past seven years. From this experience emerge insights that we believe have important implications to the future

deployment of media - insights that we feel are doubly important in this period of technology convergence, especially since they are derived from actual experience, rather than theoretical speculation.

UbiComp: A Brief Overview

Introduction

As described by Weiser, UbiComp can be characterized by two main attributes:

- *Ubiquity*: Interactions are not channeled through a single workstation. Access to computation is "everywhere." For example, in one's office there would be 10's of computers, displays, etc. These would range from watch sized Tabs, through notebook sized Pads, to whiteboard sized Boards. All would be networked. Wireless networks would be widely available to support mobile and remote access.
- *Transparency*: This technology is non-intrusive and is as invisible and as integrated into the general ecology of the home or workplace as, for example, a desk, chair, or book.

These two attributes present an apparent paradox: how can something be everywhere yet be invisible? Resolving this paradox leads us to the essence of the underlying idea. It is not that one cannot see (hear or touch) the technology; rather, that its presence does not intrude into the environment of the workplace (either in terms of physical space or the activities being performed). Like the conventional technology of the workplace (architecture and furniture, for example), its use is clear, and its physical instantiation is tailored specifically for the space and the function for which it is intended. Central to UbiComp is a break from the "Henry Ford" model of computation which can be paraphrased as:

You can have it in any form you want as long as it has a mouse, keyboard and display.

Fitting the square peg of the breadth of real needs and applications into the round hole of conventional designs, such as the GUI, has no place in the UbiComp model.

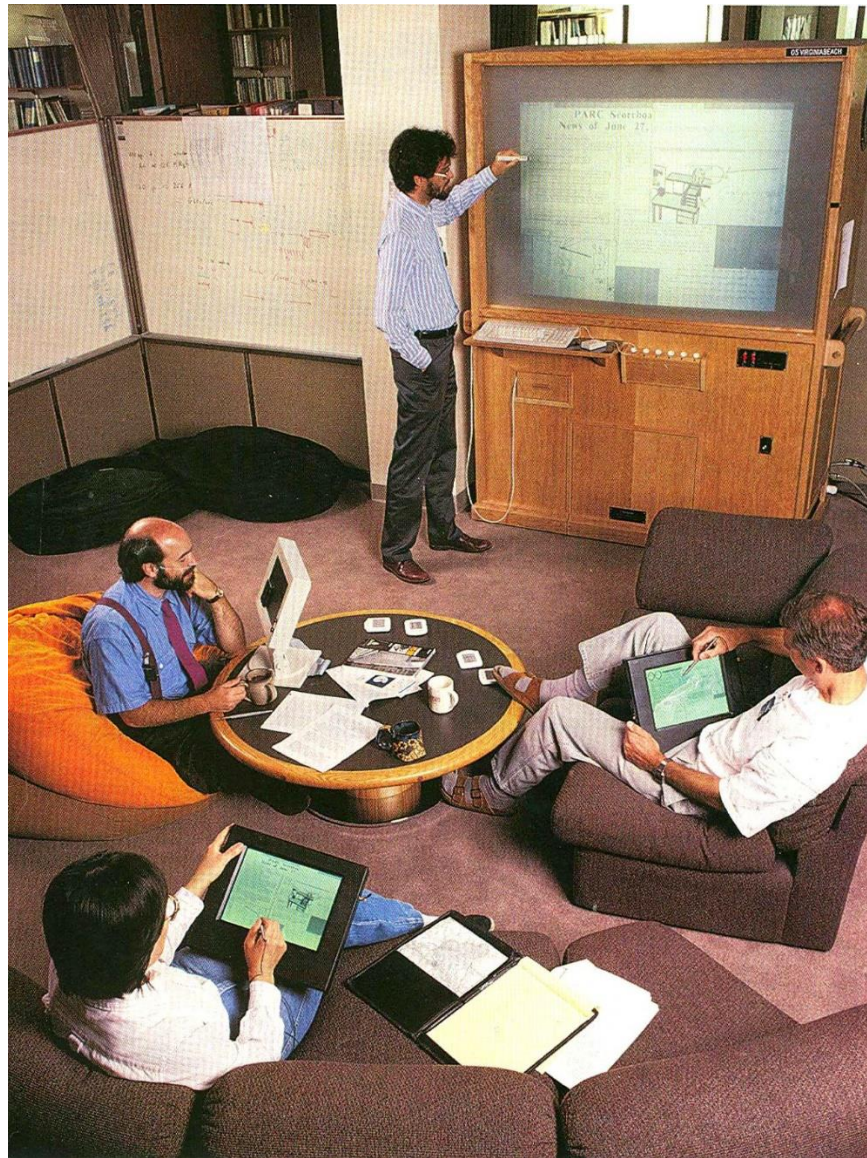


Figure 3: Xerox Liveboard and PARCpads (Photo: Xerox PARC)

Technology Warms Up

We can most easily place Weiser's model of computation in historical perspective by the use of an analogy with heating systems. In earliest times, architecture (at least in cold climates) was dominated by the need to contain heat. Special structures were built to contain an open fire without burning down. Likewise, in the early days, special structures were built to house computation. These were known as "computer centres."

As architecture progressed, buildings were constructed where fires were contained in fireplaces, thereby permitting heat in more than one room. Nevertheless, only special rooms had fire since having a fireplace required adjacency to a chimney. Similarly, the analogous generation of

computation was available in rooms outside of computer centres; however, these required access to special electrical cabling and air conditioning. Therefore, computation was still restricted to special "computer rooms."

The next generation of heating system is characterized by Franklin stoves and, later, radiators. Now we could have heat in every room. This required the "plumbing" to distribute the system, however. The intrusion of this "plumbing" into the living space was viewed as a small price to pay for distributed access to heat. Again, there is an analogous generation of computational technology (the generation in which we are now living). In it, we have access to computation in any room, as long as we are connected to the "plumbing" infrastructure. And like the heating system, this implies both an intrusion into the space and an "anchor" that limits mobility.

This leads us to the newest generation of heating system: climate control. Here, all aspects of the interior climate (heat, air conditioning, humidity, etc.) is controllable on a room-by-room basis. What provides this is invisible and is likely unknown (heat-pump, gas, oil, electricity?). All that we have in the space is a control that lets us tailor the climate to our individual preference. This is the heating equivalent of UbiComp: the service is ubiquitous, yet the delivery is invisible. UbiComp is the computational analogy to this mature phase of heating systems: in both, the technology is seamlessly integrated into the architecture of the workplace.

Within the UbiComp model, there is no computer on my desk because my desktop is my computer. As today, there is a large white board on my wall, but with UbiComp, it is active, and can be linked to yours, which may be 3000 km away. What I see is way less technology. What I get is way less intrusion (noise, heat, etc.) and way more functionality and convenience. And with my Pads and Tabs, and the wireless networks that they employ, I also get far more mobility without becoming a computational "orphan."

Media spaces And Ubiquitous Video

Introduction

UbiVid is the video complement to UbiComp in that it shares the twin properties of ubiquity and transparency. In "desktop videoconferencing," as it is generally practiced, what we typically see is a user at a desk talking to someone on a monitor that has a video camera placed on top. This is illustrated in Figure 2: Shared open office via Media Space (Photo: Xerox PARC). Generally, the video interactions are confined to this single camera-monitor pair.

In UbiVid, we break out of this, just as UbiComp breaks out of focusing all computer-mediated activity on a single desk-top computer. Instead, the assumption is that there are a range of video cameras and monitors in the workspace, and that all are available. By having video input and output available in different sizes and locations, we enable the most important concept underlying UbiVid: exploiting the relationship between (social) function and architectural space.

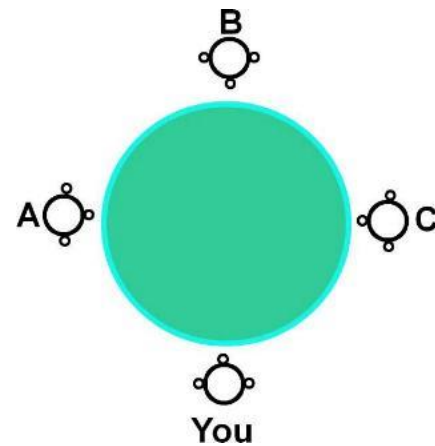


Figure 4: A 4-way round-the table conversation. By dedicating Hydra unit for each person, each occupies their own personal space at the virtual table. Their Hydra's camera constitutes their surrogate eyes, its speaker their surrogate mouth, and its microphone their surrogate ears. By preserving the "round-table" relationships illustrated schematically on the right, conversational acts found in face-to-face meetings, such as gaze awareness, head turning, etc. are preserved

One example can be seen in Figure 4. This illustrates our Hydra multiparty conferencing system (Sellen, Buxton & Arnott, 1992). Since each participant has their own personal space, many of the social mores of face-to-face meetings are preserved. For example, the design affords gaze awareness and eye-contact. Furthermore, one can lean over and whisper an aside to another participant, all the while maintaining the normal face-to-face social checks and balances. The Hydra units are just one of many examples to come, and they too will return in another context. Along with our examples and discussion, we will also articulate some of the underlying design principles, beginning with the following:

- Design Principle 1:** *Preserve function/location relations for both tele and local activities.*
- Design Principle 2:** *Treat electronic and physical "presences" or visitors the same.*
- Design Principle 3:** *Use same social protocols for electronic and physical social interactions.*
- Design Principle 4:** *The box into which we are designing our solutions is the room in which you work/play/learn, not a box that sits on your desk.*

Example: My Office

Let us work through an example to illustrate how these principles apply in a specific context, namely my office. A floorplan of my office is illustrated in Figure 5. It indicates 3 zones in my office, each of which has a distinct social function. (A) is my desk, where I work alone, or interact one-on-one with colleagues, students and visitors. (B) is the doorway, through which people may just pop their heads in to see if I am free, or to ask a quick question, or deliver a brief message. (C) is around a coffee table, around which, conversations tend to be informal, compared to those at my desk.

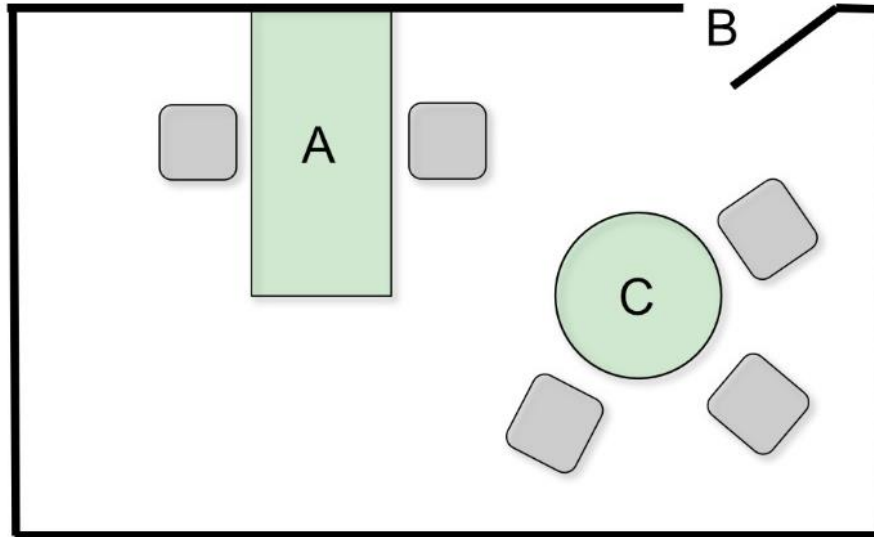


Figure 5: My office showing key locations: desk (A), door (B) and meeting table (C).

If it is important enough to set my office up specifically to employ space to support such diverse types of interactions, and I want to accommodate remote participation in any or all of those functions, then it seems reasonable that the same attention to the design and placement of the telepresence technology must receive the same level of attention as the placement of the furniture received.

Hence, for a remote person to work with me at my desk, they can appear right beside the monitor of my computer, as seen in Figure 6. It is important that they appear on a dedicated display, thereby leaving my computer screen free to dedicate to any shared documents which we may be working on. Likewise, the camera, speaker and microphone are positioned such that eye contact and gaze awareness are established, and their voice comes from the location which they occupy.



Figure 6: Remote face-to-face collaboration at the desktop. Despite the differences of our individual physical paces, each of us has close physical presence in the space of the other. Our shared work appears on a separate display, thereby establishing a clear difference between person and task space – and one which enables eye contact and gaze awareness (a theme which will come up again more than once, due to its importance).

If someone wants to glance into my office to see if I am available, they can do so from the door (location "B" in Figure 5), whether they come physically or electronically. A camera mounted above the door gives them approximately the same view that they would have if they were glancing through my physical door. This is illustrated in Figure 7. I can see who is "at the door" on a small monitor mounted by the camera, and - as in the physical world - I can hear their approach down the digital corridor by means of an auditory icon, or *earcon* – the sound of approaching footsteps.



Figure 7: Interactions at my office door: physically (a) and electronically (b).

Rather than superfluous, such capabilities afford the adherence to normal social conventions of approach and departure. It can also make a strong difference in the experience of using the technology. As an example, in 1993/4, Hiroshi Ishii visited us from NTT for a year. When he first came, this "door cam" was not deployed. After he had been with the project for a while, he explained to me that when he first came he was reluctant to use the system to contact me because he felt that it was rude to just "arrive" on my desktop. His reasons were partially due to not knowing me that well at the time, and partially out of "respect" for my position as director of the project. To him, the distance and

means of approach afforded by the "door-cam" was important to his comfort and effective use of the system. Our claim is that the need for such social sensitivities is not rare.

As mentioned, the third function-sensitive location in my office is around the coffee table, where informal meetings tend to take place. These may involve up to five or six people. Frequently these include a remote participant. To enable participation from an appropriate location in the room, a special "seat" is reserved for them around the table. This is shown in Figure 8.



Figure 8: An informal meeting with remote participation.

By appearing in a distinct and appropriate location, participants physically in my office are able to direct their gaze at the remote participant just as if they were physically present. Likewise, the remote participant has a sense of gaze awareness, that is, who is looking at whom, and when. The reason is that the remote participant has a physical presence in the room - a presence afforded by the location of the video surrogate through which they communicate.

In our discussion, we have mainly dealt with social function and distance in relation to fixed locations. These are issues, however, which normally have a strong dynamic component. People move. In so doing, functions change. In this regard, our system is still lacking. One can move from location to location within the room, but the transitions are awkward. This is an area that needs improvement. But before one can work on movement, one requires places to move to. Establishing such places within the room has been our focus to date.

Having lived in this environment in this form for almost three years, perhaps the most striking thing is a seeming paradox. By adding this extra equipment into the room, there actually appears to be less technology and far less intrusion of the technology in the social interactions that it mediates. Our argument is that this is due to the technology being in the appropriate locations for the tasks undertaken in the room. In a single desk-top solution, for example, one would be twisting the camera and monitor from the desk to the coffee table when switching between desk-top and group meetings. As well, due to the multiple cameras and monitors, we avoid the contention for resources that would otherwise result. For example, I can be in a desk-top conference on one monitor, monitor a video which I am copying on another, and still not prevent someone from appearing at my electronic door.

As we have pointed out in the examples above, through increased ubiquity, we have achieved increased transparency. This last point is achieved, however, only through the appropriate distribution of the technology - distribution whose foundations are the social conventions and mores of architectural location/distance/function relationships.

Example: Back-to-Front Videoconferencing

Another example of using spatially distributed video is the implementation of "back-to-front" videoconferencing at the University of Toronto. In conventional videoconferencing rooms, the camera and monitors are placed at the front of the room. This location works well if the remote participant is the presenter; however, if the remote party is intended to be part of the audience, then the location is inappropriate. In this case, the remote party should be located with the rest of the audience. If both social functions are to be supported, then so should be access to both locations. Figure 9 illustrates how we enable remote participants to "take their place at the table" with the other members of the audience.

The scenario shown in the figure illustrates the notion of transparency. Due to the maintenance of audio and video reciprocity, coupled with maintaining "personal space," the presenter uses the same social mechanisms in interacting with both local and remote attendees. Stated another way, even if the presenter has no experience with videoconferencing or technology, there is no new "user interface" to learn. If someone raises their hand, it is clear they want to ask a question. If someone looks confused, a point can be clarified. Rather than requiring the learning new skills, the design makes use of existing skills acquired from a lifetime of living in the everyday world.



Figure 9: A back-to-front videoconference: the remote attendee sits at the table.

Concept: *Video Surrogate: Don't think of the camera as a camera. Think of it as a surrogate eye. Likewise, don't think of the speaker as a speaker. Think of it as a surrogate mouth. Integrated into a single unit, a vehicle for supporting design Principles 1 & 2 is provided.*

Premise: *Physical distance and location of your video surrogate with respect to me carries the same social weight, function, and baggage as if you were physically in your surrogate's location. Furthermore, the assumption is that this is true regardless of your actual physical distance from me.*

Qualification: *This equivalence is dependent on appropriate design. It sets standards and criteria for design and evaluation.*

Example: Front-to-Back Videoconferencing

Looking in the other direction in the same conference room, to the front, we see a similar accommodation, one consistent with the principles underlying the design of the technology for my office.

In this case, the presenter has their own "place" at the front from which to deliver their material, be they there physically or remotely. Likewise, the material presented is viewed on an electronic display in order that it is viewable, independent of the view of the presenter, regardless of being remote or not.



Figure 10: On the left, paper materials are being presented using an overhead projector, while on the right, computer-based slides are being employed. Presentations are displayed electronically on a large display at the front of the conference room. They are therefore accessible as a separate channel from the view of the presenter, which remote attendees see from their “seat” at the back of the room. In each case, the presenter interacts with the materials directly, and these are visible both in the conference room and remotely,

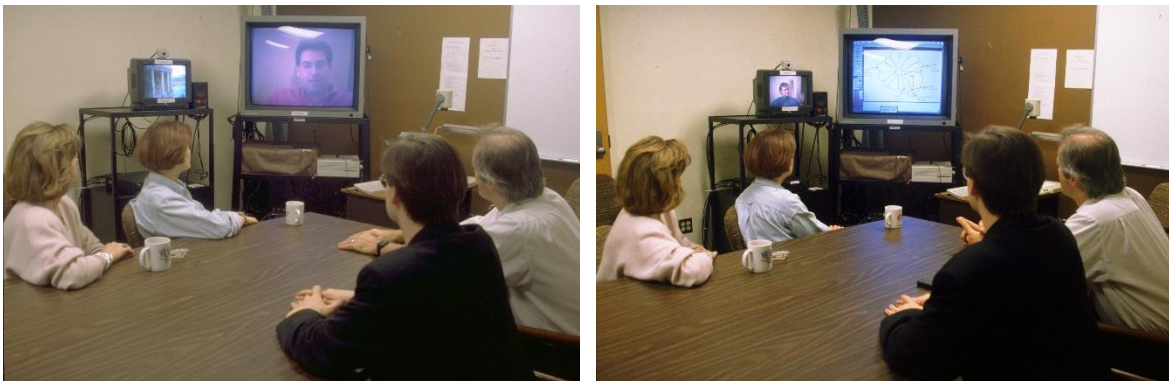


Figure 11 Both images show an example of the presentation being made remotely. In cases where no support materials are being used, the presenter appears on the large display. On the other hand, if slides or documents are included in the presentation, they appear on the large display, and the presenter appears on the smaller display to the left. As in same-place meetings, people and presentation materials, appear in their own independent space – for local and remote attendees, alike. More on this below,

While the layout and deployment of the technology in this room is different than that seen in my office, the underlying design is consistent. That is, the purpose and use of my office is different than a conference room. Each is configured accordingly, hence, through the design reflecting purpose, the consistency lies in appropriateness to place and function.

Ubiquitous Media: UbiComp + UbiVid

Introduction

To this point, we have discussed computation separately from the media spaces. Clearly, however, these two classes of technology coexist. They complement each other in a number of ways. First, there is a dependence relationship: it is only through the computational resources that the control and operation of media spaces can be deployed.

Second, there is a cumulative relationship. In collaborative work, the media space technology provides the shared space of the people, and the computers the shared space of electronic documents. Both types of shared space are required to establish a proper sense of shared presence, or telepresence.

When used together, a sense of awareness of the social periphery is afforded - as sense which would otherwise only be possible in a shared corridor or open concept office.

In the remainder of this section, we will give examples which illustrate each of these cases.

Making Contact

It is misleading to think about computation as separate from collaboration via media spaces. The notions of UbiComp and UbiVid go hand-in-hand. This is most easily seen in the example of how the computer is used to mediate interactions within a media space.

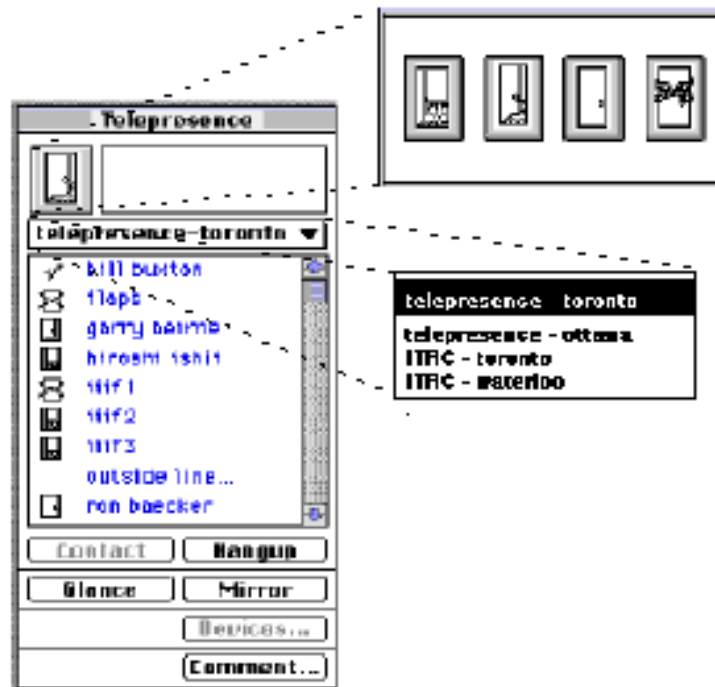


Figure 12: The Telepresence Client: Making connection and specifying accessibility.

Figure 12 shows the user's view of the main application used to mediate connections among people and resources. (The cross-disciplinary process which led to this design is documented in Harrison, Mantei, Beirne & Narine, 1994.) The left panel in the figure is what users normally see. It is primarily a scrolling list of names of the people and resources to which I can connect. Operationally, one selects the desired name, then selects the "Contact" button, shown in the lower portion of the panel. Notice that beside each name in the list is an icon of a door. The door icon can be in one of 4 states. Each indicates a different degree of accessibility for that name. If it is open, you are welcome to "pop in." If it is ajar, you can glance in and determine availability, but you must "knock" if you want to enter. If it is closed, you must knock and wait for a response before entering, and glancing is not possible. Finally, if the door is boarded shut, you can only leave a message.

I can set my door state by clicking on the door icon in the top left corner of the panel. This causes the menu shown on the upper right to pop up, from which I select one of the four icons. The icon beside my name is then accordingly updated for all users. Hence, a means is provided to control accessibility which is based upon everyday social protocols.

The application enables me to contact people at various sites, not just those who are local. In the example, residents at the Toronto Telepresence site are displayed in the name list. However, selecting the site name displayed above the name list causes a menu listing all sites to pop up. This is illustrated in the lower right panel. Selecting a site from this menu causes its residents' names to appear in the name list.

Shared Presence of Person & Task

In an earlier publication (Buxton, 1992), we argued for the need to support a sense of shared presence of both task and person. The main argument was that being able to move seamlessly from one to the other was important in undertaking several common interactions. This point has been argued by others, especially Ishii, Kobayashi, and Grudin (1992).

Figure 13 is one example of our efforts to support the seamless and natural redirection of gaze, and gaze awareness through the affordances of our design.

The next two examples illustrate how we can support this seamlessness, but in a manner appropriate to different application needs, by integrating UbiComp and UbiVid technologies with industrial design.

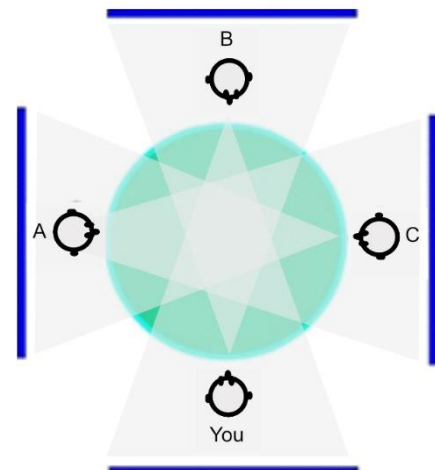
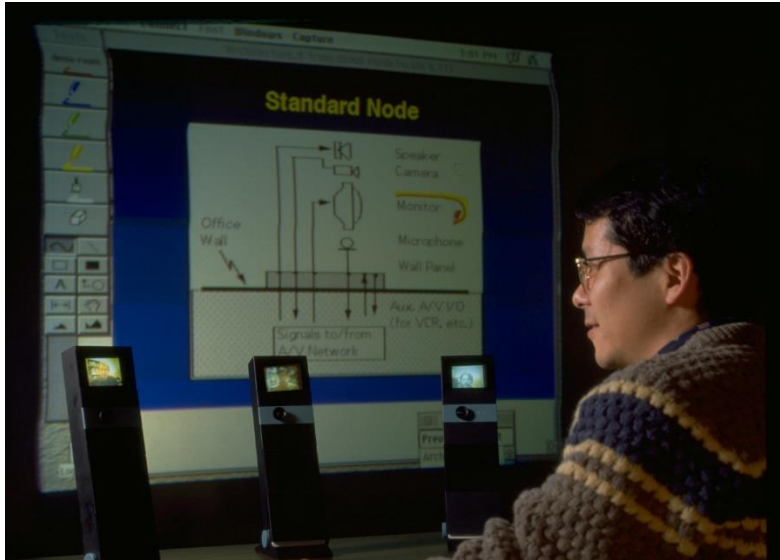


Figure 13: Seamless Integration of Person and Task Space. The photo on the left also shows a 4-way video conference using the Hydra units. However, this time, a large electronic "whiteboard" containing the information being discussed appears behind the units. As illustrated in blue in the schematic on the right, the same display can appear behind the units at each of the four sites, thereby giving each participant ideal sight lines to the "same" whiteboard (something that does not occur in same-place round-table meetings). Furthermore, gaze awareness now extends to whether one is looking up at the "whiteboard" or at a person, thereby seamlessly blending person and task space.

Example: The Active Desk

Figure 14 illustrates a collaborative design application which further illustrates distinct areas for each of the remote person(s) and the shared work. The design appears on what we call the Active Desk (developed jointly by the Ontario Telepresence Project and the Arnott Design Group of Toronto). This is a large electronic drafting table. It employs a 100x66 cm rear projection display. One interacts with the system with a stylus (4000x4000 points resolution), or a keyboard. In the illustration, a Hydra unit is mounted on each of the two upper corners of the desk (only one visible in the figure), thereby supporting three-way collaboration. The design being worked on is visible to all parties, and all parties can interact with it, such as pointing, annotating, or editing.

Little technology is visible, and the form of the system is specifically tailored to the task: graphic design. Industrial design has rendered the technology transparent to the task. In essence, there is no desktop computer and no desktop metaphor. The desktop is the computer.

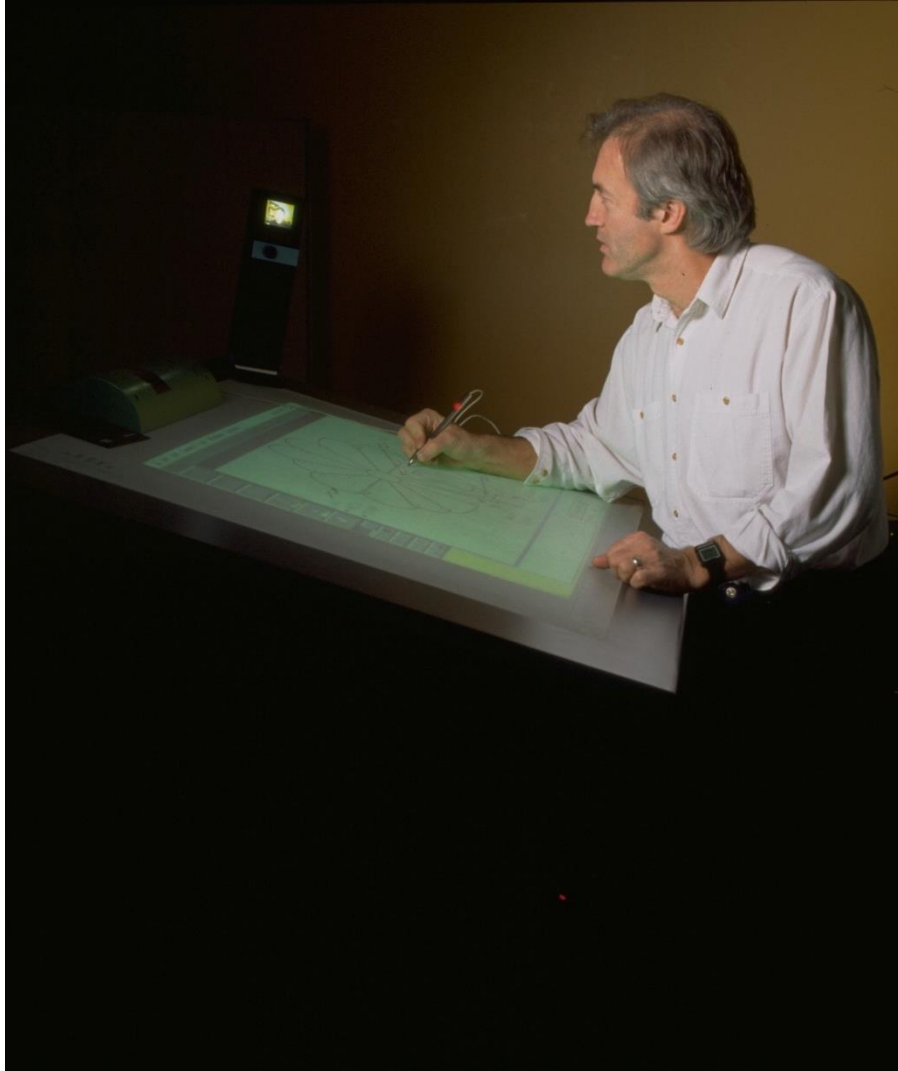


Figure 14: The Active Desk, equipped with a Hydra unit.

Example: Sitting Across the Desk

The next example, illustrated in Figure 15, is the UbiVid equivalent to sitting across the desk from someone. Here, through rear projection, the remote participant appears life-size. What we are trying to capture in this example are two people working together on a joint project, such as a drawing or budget. While not the case in the example, if there were shared documents, they would be on the surface of the table, using – for example - a technology similar to the Active Desk seen in Figure 14.



Figure 15: Face-to-face, life-size, across the desk. Size affects both sense of presence and supports balance between parties. Here, the display of the Hydra unity is replaced by the large one behind. Its camera, speaker and microphone are still being used.

First, notice that having one's counterpart displayed this way is not like seeing them on a regular video monitor. Because of the scale of the image, the borders of the screen are out of our main cone of vision. Hence, the space occupied by the remote person is defined by the periphery of their silhouette, not by the bezel of a monitor. Second, by being life size, there is a balance in the weight or power exercised by each participant. Third, the gaze of the remote participant can traverse into our own physical space. When the remote party looks down on their desk, our sense of gaze awareness gives us the sense that they are looking right onto our own desktop, and with their gaze, they can direct us to look at the same location. This power of gaze awareness is so strong that people have even argued that the eyes emitted "eye rays" that could be sensed (Russ, 1925). It is this same power of gaze that we have tried to exploit to achieve an ever more powerful sense of Telepresence.

Portholes and Awareness of the Social Periphery

Portholes was a system jointly developed by Xerox PARC and Rank Xerox EuroPARC (Dourish & Bly, 1992). Its purpose was to provide a sense of peripheral awareness and social proximity among a group of geographically distributed co-workers. The approach to doing so was to present a tiled view of video snapshots of all members of the distributed workgroup, with the images updated about once a minute.

With the cooperation of the original developers, we further developed the system. In our case, these snapshots are updated every five minutes. Unique to the Ontario Telepresence Project implementation is the superimposition of the door icons on the snapshots. A typical display of our version of *Portholes* is shown in Figure 16.

The snapshot and door state icon provides information as to both the presence and/or activities of group members, as well as their degree of accessibility. Furthermore, the snapshots provide a user interface to certain functions concerning individuals. For example, after selecting the snapshot of me on your screen, you can then click on the Info button on the top of the frame to get my phone number, name, address and email address. Or, double clicking on my image, or selecting the contact button asserts a high bandwidth connection to me (thereby providing an alternative means to make a connection to that illustrated in Figure 12).

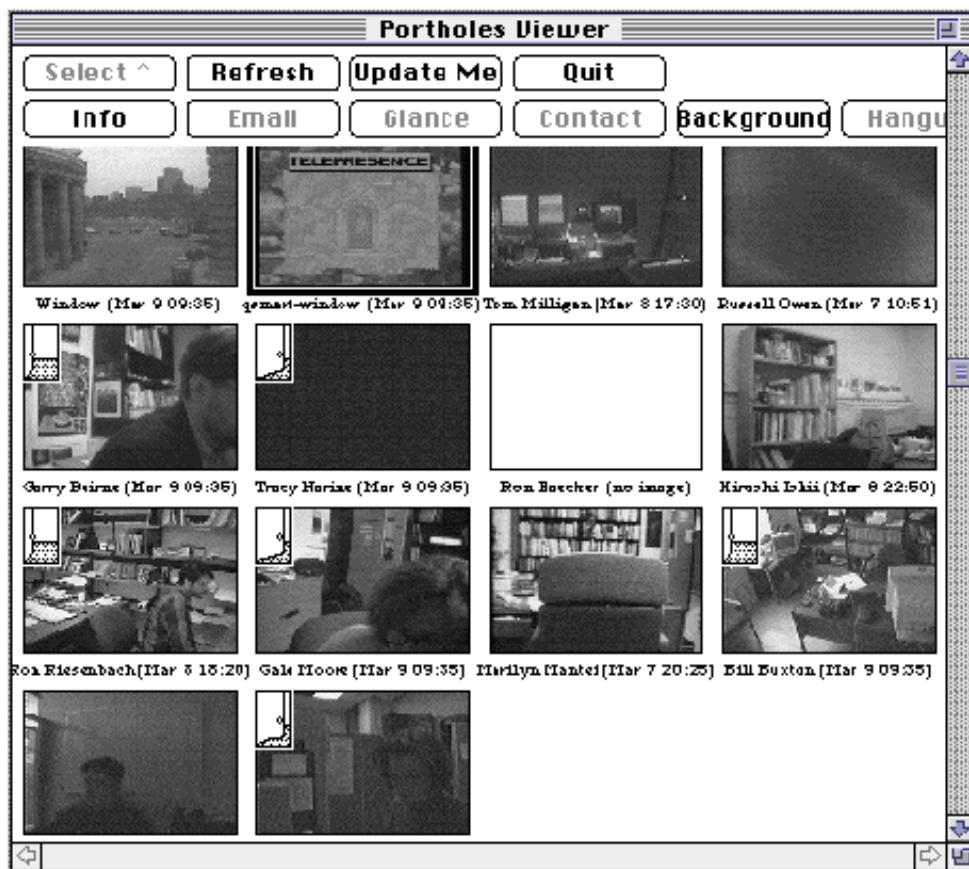


Figure 16: The Telepresence implementation of Portholes. Every 5 minutes, a snapshot of each member of the workgroup is distributed to all other members. In the Telepresence implementation, this is accompanied by an icon of that member's door icon, which indicates that person's degree of accessibility. The resulting tiled image of one's workgroup affords a strong sense of who is available when. It also can serve as a mechanism for making contact, finding phone numbers, and avoiding intruding on meetings. (Figure 5 of Buxton, Sellen & Sheasby, 1995).

Portholes takes advantage of the fact that each office has a video camera associated with it. This goes beyond the stereotyped notion of desktop video as simply a videophone. Rather, it supports a very important sense of awareness of the social periphery - an awareness that normally is only available in shared office or shared corridor situations. It introduces a very different notion of video on demand and delivers its potential with a transparent user interface.

Finally, discussions about Portholes always touch upon the issue of privacy. "How can you live in an environment where people can look in on you like that?" we are frequently asked. There are a couple of responses to this. First, Portholes is not an "open" application. It embodies a sense of reciprocity within a distinct social group. People cannot just randomly join a Portholes group. Members know who has access to the images. Secondly, even within the group, one can obtain a degree of privacy, since the distribution of your image can be controlled by your door state. Finally, remember that the images have no motion and no audio. What is provided is less than what would be available to someone looking through the window of your office door. This is especially true if the snapshot is taken from the "door camera", such as illustrated in Figure 7.

Active Sensing and The Reactive Environment

Introduction

In the examples thus far, the use of computational and video technologies has been complementary. However, the net effect has been cumulative. Our argument is that the benefits go well beyond this. In this section we will show that there is a synergy that occurs when these two classes of technologies are used together. The result can be far more than the sum of the parts. We will illustrate this in the examples in this section, of what we call proximal sensing, reactive environments, and context-sensitive interaction (Buxton, 1995).

Video, Portholes and "Call Parking"

We can leverage the video and computational technologies of Ubiquitous Media by recognizing that the same cameras that I use for video conferencing can give my computer "eyes." Furthermore, the same microphone through which I speak to my colleagues can also provide my computer with an "ear."

Design Principle 5: *Every device used for human-human interaction (cameras, microphones, etc.) are legitimate candidates for human-computer interaction (and often simultaneously).*

By mounting a video camera above the Active Desk, and feeding the video signal into an image processing system, one can use the techniques pioneered by Krueger (1983, 1991) to track the position

of the hands over the desk. This is illustrated in Figure 17, which shows a prototype system developed by Yuyan Liu, in our lab. In the example, the system tracks the position and orientation of the left hand as well as the angle between the thumb and forefinger. The resulting signal enables the user to "grasp" computer-generated objects displayed on the desk's surface.

Such use of video is relatively non-intrusive. One need not wear any special gloves or sensors. The system sees and understands hand gesture much in the same way as people: by watching the hands or body. Furthermore, the Digital Desk of Wellner (1991) has shown how the position objects, as well as the hands, and their relationship to the displayed data can be used as the basis for interaction.

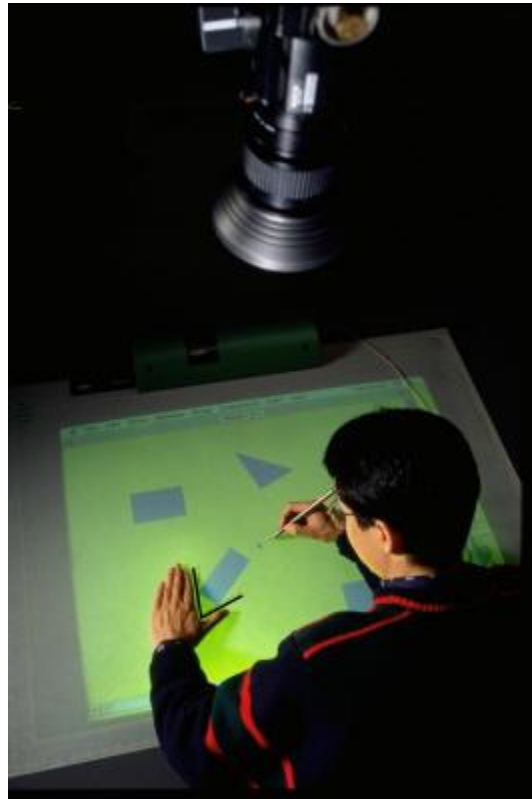


Figure 17: Using video to enable the computer to react to hand position and gesture.

Another simple, yet effective, use of video to support interaction can be demonstrated by an extension of the Portholes application. A prototype written by Luca Giachino, a visiting scientist from CEFRIEL in Milan, Italy, demonstrated this. The underlying observation is that two Portholes images in a row can constitute a motion detector.

By comparing two frames, if more than 40% of the pixels change, there has been motion. Hence, one can have a rather reliable indication whether there is someone there. By keeping 1 bit of state for each frame, one can - within the limits of the frame rate which images are captured - if someone is still there, still away, come in or gone out.

With this observation and the resultant code, the mechanism for a new type of "call parking" is provided. If I want to call you, I can look up at Portholes and see if you appear available. If so, I just double click on your image to assert a connection. If you are not there, one mouse click would instruct the system to let me know when you appeared to be available. The application would then monitor if you were in your office, as well as your availability as indicated by your door state. This, then, enabled me to get on with my work without the distraction of repeatedly checking on your availability. By enabling the selection of more than one person, the utility dramatically decreased the transaction cost of convening small, impromptu meetings and conference calls.

Doors Revisited: the "Door Mouse"

The cameras and microphones found in the office are not the only sensory devices that can be taken advantage of in the domain of Ubiquitous Media. Other alternatives include the full repertoire of motion and proximity sensors used in home automation and security. Let us revisit an earlier example, the specification of door state, as a case in point.

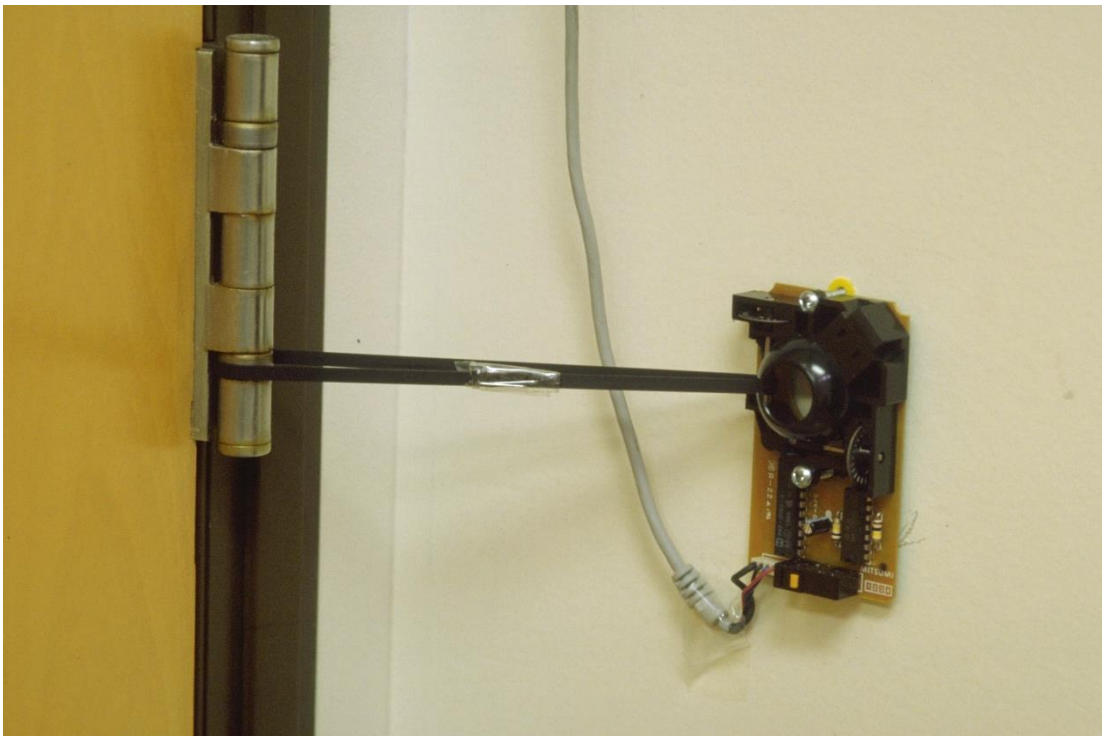


Figure 18: The "Door Mouse".

Specifying door state using the mechanism illustrated in Figure 12 preserves the protocols of the physical world by metaphor; however, while still useful, the link between the material and virtual doors is conceptual, not functional. Opening and closing one has no impact on the other. The two function in parallel universes.

Figure 18 illustrates how these two universes can come together. An ingenious prototype of how to do so was conceived and implemented by a student, Andrea Leganchuk. It involved mounting the mechanism from a standard mouse on the wall beside the door hinge. The mouse ball was removed, and one of the shaft encoders was then driven by an elastic loop which wrapped around it and the door hinge. The output of this second mouse (or “door mouse”) was fed to the computer, where it was hijacked by the software and used to control the door-state of the Telepresence and Portholes clients, seen previously in Figure 12 and Figure 16, respectively. Hence (naturally subject to the ability to override defaults), closing my physical door is sensed by the computer and prevents people from entering physically or electronically (by phone or by video). One known everyday action and protocol controls both the material and virtual worlds.

Observation: *A door is just as legitimate input device to a computer as are a mouse or a keyboard.*

Proximal Sensing and Context

What characterizes the previous examples is the increased ability of the computer to sense more than just the commands that are typed into it. Our experience suggests that computation is moving towards a future where our systems will respond to more and richer input.

One hint of this today is remote sensing, the gathering of data about the earth and environment by sensors in satellites. What we are describing is similar, except the sensors are much closer, hence the term proximal sensing. In this case, it is the ecology and context of the workspace which is being sensed.

When you walk up to your computer, does the screen saver stop and the working windows reveal themselves? Does it even know if you are there? How hard would it be to change this? Is it not ironic that, in this regard, a motion-sensing light switch is “smarter” than any of the switches in the computer, AI notwithstanding?

We see this transition as essential to being able to deliver the expanded range of functionality being promised as a result of technological convergence. Our perspective is that if considerable complexity is not off-loaded to the system, much (if not most) of the promised functionality will lie beyond the complexity barrier, or the users threshold of frustration. Our final example briefly introduces some of our ongoing work which is based on this premise.

Reactive Environment

The way in which proximal sensing and context-sensitive interaction can help reduce complexity while supporting new services is illustrated in our final example, an augmented meeting room. Much is promised in the way of meeting support by new technologies. Videoconferencing, electronic whiteboards, audio and video-based meeting capture and annotation and electronic presentations that support video and computer graphics are just some examples. The components for nearly all these services are now commercially available. And yet, our ability to deliver them in a way that augments a meeting, rather than intruding upon it, is limited, to say the least. Their being delivered to a techno-novice in a walk-up-and-use conference room is virtually unthinkable.

The reason is the amount of overhead associated with changing the state of the room to accommodate the changing demands and dynamics of a typical meeting. Take a simple example. Suppose that you are in a video conference and someone asks, "record the meeting." This turns out to be nontrivial, even if all the requisite gear is available. For the meeting to be recorded, the audio from both sites must be mixed and fed to the VCR. Furthermore, the video from each site must be combined into a single frame using a special piece of equipment, and the resulting signal also fed to the VCR. Somehow, all of this must happen. And recognize that the configuration described is very different than if just a local meeting was to be recorded, a video played back locally, or a video played back so that both a remote and local site can see it.

In each of these cases, let us assume that the user knows how to perform the primary task: to load the tape and hit record or play. That is not the problem. The complexity comes from the secondary task of reconfiguring the environment. However, if one takes advantage of proximal sensing, the system knows that you put a tape in, which key you hit (play or record), and knows if you are in a video conference or not, and if so, with how many people. Hence, all of the contextual knowledge is available for the system to respond in the appropriate way, simply as a response to your undertaking the simpler primary task: loading the tape and hitting the desired button.

Over the past year, we have been instrumenting our conference room (the one seen previously in Figure 9, Figure 10 & Figure 11), in such a way as to react in such a way. Furthermore, we have been doing so for a broad range of conference room applications, in order to gain a better understanding of the underlying issues (Cooperstock, Tanikoshi, Beirne, Narine & Buxton, 1995).

Summary and Conclusions

We have hit the complexity barrier. Using conventional design techniques, we cannot significantly expand the functionality of systems without passing users' threshold of frustration. Rather than adding complexity, technology should be reducing it, and enhancing our ability to function in the emerging world of the future.

The approach to design embodied in Ubiquitous Media represents a break from previous practice. It represents a shift to design that builds upon users' existing skills, rather than demanding the learning of new ones. It is a mature approach to design that breaks out of the "solution-in-a-box" super appliance mentality that dominates current practice. Like good architecture and interior design, it is comfortable, non-intrusive and functional.

To reap the benefits that this approach offers will require a rethinking of how we define, teach and practice our science. Following the path outlined above, the focus of our ongoing research is to apply our skills in technology and social science to both refine our understanding of design, and establish its validity in those terms that are the most important: human ones.

Acknowledgments

Virtually all of the work described in the above examples was designed and implemented by the members of the Ontario Telepresence Project, of which I had the pleasure to be Scientific Director. The excellence of the work comes partially from the excellence of the team itself, and partially from

strong support and collaboration with colleagues at Xerox PARC and Rank Xerox EuroPARC. To all of those who have helped make these such stimulating environments, I am very grateful.

The research discussed in this paper has been supported by the Ontario Government Centres of Excellence, Xerox PARC, Hewlett-Packard, Bell Canada, the Arnott Design Group, Object Technology International, Sun Microsystems, NTT, Bell Northern Research, Hitachi Ltd., Adcom Electronics, IBM Canada and the Natural Sciences and Engineering Research Council of Canada. This support is gratefully acknowledged.

References

- Bly, S., Harrison, S. & Irwin, S. (1993). Media Spaces: bringing people together in a video, audio and computing environment. *Communications of the ACM*, 36(1), 28-47.
- Buxton, W. (1992). Telepresence: Integrating shared task and person spaces. *Proceedings of Graphics Interface '92*, 123-129.
- Buxton, W. (1995). Integrating the periphery and context: A new model of telematics *Proceedings of Graphics Interface '95*, 239-246.
- Cooperstock, J., Tanikoshi, K., Beirne, G., Narine, T., Buxton, W. (1995). Evolution of a reactive environment. *Proceedings of CHI '95*, 170-177.
- Dourish, P. & Bly, S. (1992). Portholes: Supporting awareness in a distributed work group. *Proceedings of CHI '92*, 541- 547.
- Elrod, S., Hall, G., Costanza, R., Dixon, M. & Des Rivieres, J. (1993) Responsive office environments. *Communications of the ACM*, 36(7), 84-85.
- Fields, C.I. (1983). Virtual space teleconference system. United States Patent 4,400,724, August 23.
- Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K. & Buxton, W. (1992). Realizing a video environment: EuroPARC's RAVE system. *Proceedings of CHI '92*, 27-35.
- Harrison, B., Mantei, M., Beirne, G. & Narine, T. (1994). Communicating about communicating: Cross-disciplinary design of a Media Space interface. *Proceedings of CHI '94*, 124-130.
- Ishii, H., Kobayashi, M. & Grudin, J. (1992). Integration of inter-personal space and shared workspace: Clearboard design and experiments. *Proceedings of CSCW '92*, 33 - 42.
- Krueger, Myron, W. (1983). *Artificial Reality*. Reading: Addison-Wesley.
- Krueger, Myron, W. (1991). *Artificial Reality II*. Reading: Addison-Wesley.

- Mantei, M., Baecker, R., Sellen, A., Buxton, W., Milligan, T. & Welleman, B. (1991). Experiences in the use of a media space. *Proceedings of CHI '91*, 203-208.
- Russ, Charles (1925). An instrument which is set in motion by vision. *Discovery*, Series 1, Volume 6, 123-126.
- Sellen, A. (1992). Speech patterns in video mediated conferences. *Proceedings of CHI '92*, 49-59.
- Sellen, A., Buxton, W. & Arnott, J. (1992). Using spatial cues to improve videoconferencing. *Proceedings of CHI '92*, 651-652. Also videotape in CHI '92 Video Proceedings.
- Stults, R. (1986). Media Space. *Systems Concepts Lab Technical Report*. Palo Alto, CA: Xerox PARC.
- Vowles, H. (1992). Personal Communication regarding the TELEMEET Project, Feb. 1970, United Church Berkely Studio, Toronto.
- Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3), 94-104.
- Wellner, P. (1991). The DigitalDesk Calculator: Tactile manipulation on a desktop display. *Proceedings of the Fourth Annual Symposium on User Interface Software and Technology (UIST '91)*, 27-33.
- Wellner, P., Mackay, W. & Gold, R. (Eds.)(1993). Computer-augmented environments: Back to the real world. Special issue of the *Communications of the ACM*, 36(7).