Buxton, W. (2006). Space-Function Integration and Ubiquitous Media. In M. Shamiyeh (Ed.). *Towards and Interactive and Integrative Design Process.* Linz, Austria: DOM Publications, 248-271.

# Space-Function Integration and Ubiquitous Media

William Buxton
Principal
Buxton Design
Toronto, Canada
bill@billbuxton.com
www.billbuxton.com

*Thoughts exchanged by one and another are not the same in one room as in another.*
Louis I. Kahn

# Abstract

As technology becomes ever more pervasive in our lives, one of the fundamental questions confronting us is how to resolve the increasing complexity that too often accompanies it – complexity which threatens to prevent our reaping the potential benefits that it offers.

In addressing this question, much of the literature has focused on improving the design and usability of the interface to the technologies themselves. In this paper we investigate another approach, one in which some of the complexity in using the devices is eliminated by exploiting some of the key properties of architectural and social space.

Our work is based on the observation that there is meaning in space and in distance. We can take advantage of this meaning in the design of the products that we produce. By doing so, we can reduce their complexity by making such knowledge known to them implicitly, rather than forcing the user to specify it to them explicitly.

The work reported was undertaken between 1989 and 1994 as part of the Ontario Telepresence Project at the University of Toronto, and was rooted in the Ubiquitous Computing and Mediaspace projects at Xerox PARC, and the IIIF/Rave project at Rank Xerox EuroPARC.

While more than a decade old, this work is not well known. We report it here in the belief that much of it is as relevant now as it was at the time it was done. The hope is that our experience may provide some insights that may be helpful to designers and practitioners today.

# Introduction

When you walk into a lecture hall at a university, even one that you have never been in before, and where you know nobody, you still know who is the professor and who are the students. If you see a photo of a dinner party, with everyone sitting around the dining room table, you know who are the hosts are and who are the guests. Walking in the park, you can tell if two people are in love, even if you see them only from a distance.

In each of these examples, we know what we know because of our literacy in the meaning of space. In the lecture hall, the professor is at the front, and the students in the chairs. We gain our understanding from the position of the people relative to the architectural space. With the dinner party, we can infer who are the hosts because they typically sit at the head of the table. In this case, it is position relative to a fixed object in the architectural space that provides the cues for interpreting the social relationship amongst the party. And finally, with the lovers in the park, it is their physical proximity relative to each other – regardless of if they are in the park, on a bus or on a boat – which leads to our conclusion about their emotional closeness.

What all of these examples illustrate is that from a lifetime of living in the everyday world, we have all built up a phenomenal depth of knowledge about the conventions of space and its meaning. This is knowledge that we exploit every day, in almost everything that we do, in order to make sense of, and function in, the world.

Furthermore, the space/function/distance relationships on which this knowledge is based can be exploited in the design of technologies that we place in the world. My favourite example is the exit door at the supermarket. When you walk out of the shop with your arms full of groceries, the door opens automatically, without any extra effort on your part. The door in the supermarket "knows" that it is at the exit of a supermarket, rather than the entrance to a bank vault, so it knows that it is OK to open automatically. It also "knows" when you are approaching and when it should open.

Now I am not suggesting that the door has any inherent intelligence. What intelligence there is in the system is that instilled by the designer, based on an understanding of the demands of the context, and the ability to exploit simple technologies to address those demands. A simple sensor, switch and motor make this door have more "intelligence" about you and your location than is exhibited by your typical PC.

All of these examples are intended to emphasize that there is inherent meaning in space, and furthermore, through judicious design, this meaning can be exploited in the design of the technologies that we introduce into our world. The reason that we might care about this is that the range of technologies that *are* being introduced into our society seems to be growing exponentially, both in terms of quantity and type. While almost every one of them would probably argue that they owe their existence to providing some benefit to their intended users, the reality is far less clear. The composite complexity resulting from this proliferation of technologies threatens to outweigh their potential benefits.

If these benefits are going to be reaped, then there is a need to adjust the balance of this equation. Our argument is that a key way to reduce the complexity, and even augment the potential usefulness of such devices, is to follow the example set by the supermarket door: to design for and exploit what we can know about the semantics of the space/function/distance/scale relationships of the social and physical context where they will be used.

We believe that if and when this is effectively done, the technology will recede into the background and become transparent, with the effect that you will be able to focus on your primary task of carrying your proverbial groceries to the car, rather than being distracted by operational complexities, such as opening the door.

The examples that we discuss all exploit two technology trends with which we were involved. The first was *Ubiquitous Computing* or *UbiComp*, and the other was *Mediaspaces,* which could otherwise be described as *Ubiquitous Video,* or *UbiVid.* Working in concert, as something that might well be called *Ubiquitous Media*, (since it is not just about computation or video), we implemented a fairly comprehensive system and lived in it for a number of years (1997).

Most of the examples have been implemented and used in practice. Take it for granted, however, that our approach was to do smart things with stupid technologies. Our purpose was not to make engineering breakthroughs, but rather to gain some experience living with these technologies *before* they were commercially viable. Our hope was that the human insights gained might help inform future design practice and development. Our mantra, while doing this work was as follows:

> *The only way to engineer the future tomorrow is to have lived in it yesterday.*

# Background

In the 1980s I was involved in two projects at Xerox PARC. One was the *Ubiquitous Computing* (UbiComp) project led by Mark Weiser, which was to have a major impact on our thinking about the future of computation (Weiser, 1991). The other was the *Mediaspace Project*, initiated by Bob Stults, Steve Harrison and Sara Bly (Stults, 1986; Bly, Harrison & Irwin, 1993).

The former had to do with digital computers, and as manifest at PARC at the time, primarily pen-based computing on three scales: palm-sized "tabs", slate-sized tablets, and whiteboard sized panels. All were networked using (then) uncommon wireless technologies (infrared and packet radio), and had high levels of interoperability.

On the other hand, the Mediaspace work had to do with augmenting computer networks with audio/video technologies that let designers, in particular, better collaborate at a distance. The idea was to use the technology to establish a persistent sense of presence amongst a community that was geographically distributed. The technologies used were decidedly "old school" in that conventional analogue video gear (albeit controlled by a novel computer interface) formed the foundation of the system.

Apart from existing at PARC, these two projects were very far apart, both physically and intellectually. Yet, in my mind, the two were actually the same project – two sides of the same coin.

What was clear was that UbiComp was going to extend beyond various forms of pen computing, and that the analogue technologies that we were using in the Mediaspace were simply pragmatic stop-gap solutions that let us envision and experience then what was inevitably going to be available using digital technologies in the not-too-distant future.

Between 1987 and 1989 I had my first chance to take an initial step in integrating some of the concepts that were emerging from PARC when I had the opportunity to design the media infrastructure for the new EuroPARC facility in Cambridge, what became known as the "IIIF" or "RAVE" system (Buxton & Moran, 1990; Gaver *et al.,* 1992). Then, from 1989 – 1994, I got a chance to go through another iteration when I set up the Ontario Telepresence Project in Toronto.

It is work undertaken as part of this project that forms the basis for this paper. However, it is important for me to provide the above historical context since it is very hard to separate what we did in Toronto from what was being done at PARC. This is not only because of my personal involvement in all of these projects, in many ways they *were* the same projects, since while I was scientific director of the Telepresence Project in Toronto, I was still working half time at

PARC as part of both the UbiComp and Mediaspace projects. Furthermore, the software which provided the foundation for the Telepresence project was that developed at EuroPARC. So this work was not just evolving from the PARC work, it was in many ways, part of it, both contributing to it, and benefiting from it.

# Ubiquity and Transparency

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

- *Ubiquity:* Interactions are not channeled through a single device. 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 work place 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. The point 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.

Our supermarket door example is as good of an illustration of this as any. What we have tried to do is extend this thinking beyond computation, and into mediaspaces. Take, for example, desktop videoconferencing. What we typically see is a user at a desk talking to someone who appears on a monitor that has a video camera placed on top of it, as illustrated in Figure 1. The video interactions are generally confined to this single camera-monitor pair.

**Figure 1: A Typical Desktop Video Conferencing Configuration.** *Conferencing is typically channeled through a video camera on top of a monitor on the user's desktop.*

Just as UbiComp breaks out of focusing all computer-mediated activity on a single desk-top computer, our approach tries to break out of all of our video transactions being restricted to this one configuration.  Consequently, the assumption is that there are a range of video cameras and monitors in the workspace, and that all are available and appropriate for certain types of use.  By having video input and output available in different sizes and locations, we able to exploit the relationship between (social) function and space.

In what follows, we explore the significance of this relationship.  We start by articulating some of the underlying design principles, and then proceed to work through a number of examples.

**Design Principle 1:** Respect conventional function/location relationships for tele 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.

## The Social and Spatial Anatomy of My Office

We can motivate these principles by working through some examples.

Figure 2 is a schematic of my old office at the University of Toronto. A number of specific locations in the office are labelled:

A. My chair behind the desk.
B. The chair across from my desk.
C. Beside my chair.
D. Chairs around the coffee table.
E. The doorway.

Even within this relatively simple space, very different social interactions or protocols are associated with each of these locations. Consider a meeting with a student, for example.

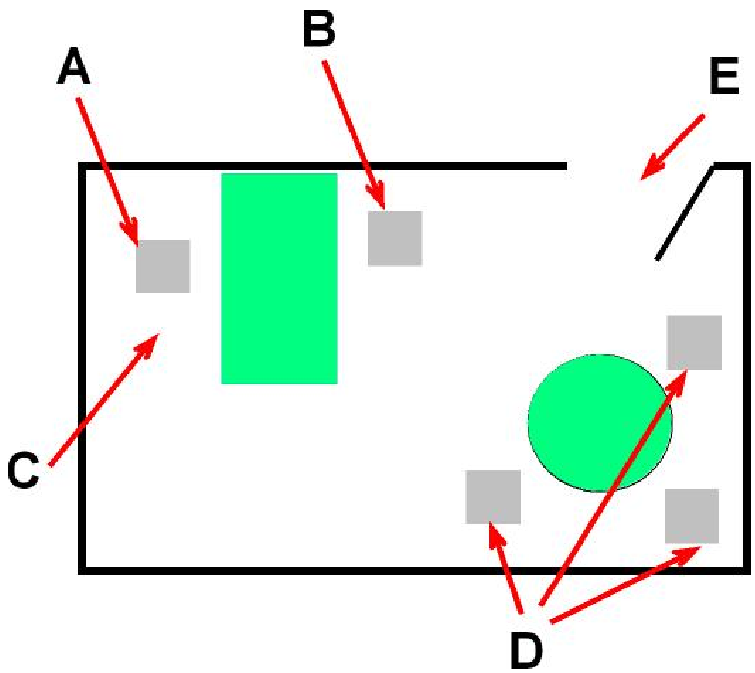


**Figure 2: Schematic of my Office.**
*A number of distinct locations in the office are indicated, including the chair behind my desk (A), the chair across from my desk (B), standing space behind my desk (C), and chairs around the coffee table (D). Different social functions are associated with each location. The deployment of any technology in the office to support collaboration or social activities should reflect and respect these differences.*

First, I might sit in my chair (A) and have them sit across the desk from me. In this case, I am Professor Buxton, and they are not. I might use this position if I were telling a student that they had failed, or if I was formally congratulating them on a great job.

Second, if I was working closely with the student on something, they might come behind my desk, to position “C”, while I sat in my chair. However, it would be very unusual for a stranger or someone with whom I was not working closely to go behind my desk.

Third, if I was having a casual meeting, or just chatting, we may sit around the coffee table. This would occur if the meeting was informal, and it would indicate that the relationship was more collegial than subordinate. It would be a meeting with “Bill” rather than “Professor Buxton.”

Fourth, I may be behind my desk working, and the student pop their head in the door to ask something. If I do not ask them in, the student would know that I was busy, and that the meeting was to be brief.

Finally, I would move certain types of meetings out of my office since it is simply an inappropriate space. One such scenario would be if I was conducting or attending a seminar, which is better situated in a conference room (which has its own set of conventions associated with the various locations within the room).

Our premise is that any technology introduced must reflect and respect these space-function-distance relationships. Consequently, activities have to take place at the appropriate location in the architectural space. The working assumption is that interaction with multiple technologies spatially deployed in various appropriate locations is much less intrusive that doing something on a single general-purpose technology that is deployed in a single, and therefore frequently wrong, location. While consistent with most of our experience in the physical world, this is rather different than mainstream practice in deploying technologies in the home or workplace.

We will now illustrate designing environments that follow some of these principles by working through some examples to support technologically mediated meetings.

## Example*:* Around the Desk *vs.* the Coffee Table

We can begin with the simple case of meeting with someone in your office via a video link. As we discussed earlier, most offices equipped to do this are set up much like the one illustrated in Figure 1. But such a configuration violates design Principles 1 & 2, since all video transactions occur on a single monitor in a fixed position. This means that location/function relationships cannot be exploited, as we shall see below. It also causes contention when there are overlapping demands for services (such as when someone wants to conference while I am watching a video).

That does not mean that it never works. Outside of the issues of approach, discussed below, this configuration can be adequate for handling meetings where I am at my desk and the social function of the remote person is similar to that associated with locations “B” or “C” in the floor-plan shown in Figure 2.

But what about an informal group meeting around the coffee table? In the standard desktop video configuration, how would the remote person assume their place at location “C”?

Our approach was to place a video “surrogate” at each location so that the remote participant would appear at the same location in the room as they would appear if they were there physically. Hence, there is a video system at my desk (Figure 1) for "reading" video documents and doing "up-close" work with a remote colleague. Then, there is also a system at the coffee table (Figure 3) where a visitor could “sit” and participate in around-the-table conversations.

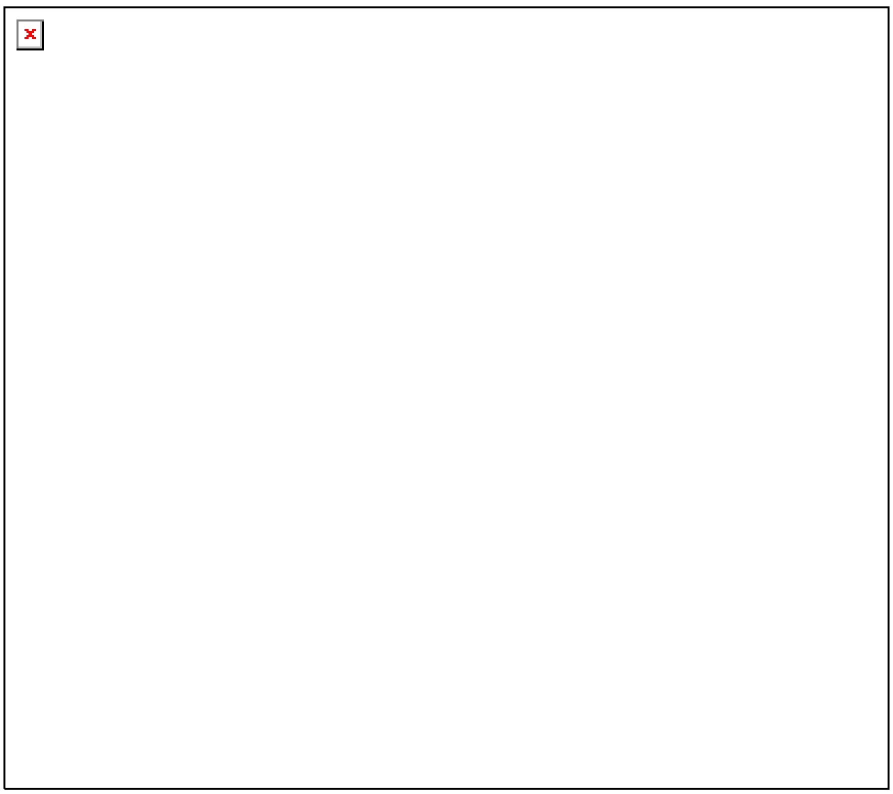

**Figure 3: Remote Participation in an Informal Group.**
*Here a group, including a remote participant (detail in inset), are sitting around the coffee table in my office having a casual meeting. (In position “D” relative to the schematic in Figure 1).*

In the example, function and space relationships are preserved. The "electronic" visitor sits where a physical visitor would. Likewise, the virtual office mate sits where a physical one would. If the equipment is properly placed, the visitor may well see the office mate, who could see the visitor, etc. Because of this distributed use of space, contention for resources is reduced and social conventions can be preserved.

## 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 contrast to traditional videoconferencing rooms, the camera and monitors are placed at the back of the room, as illustrated in Figure 4[1]. The intent here is to enable remote participants to "take their place at the table."

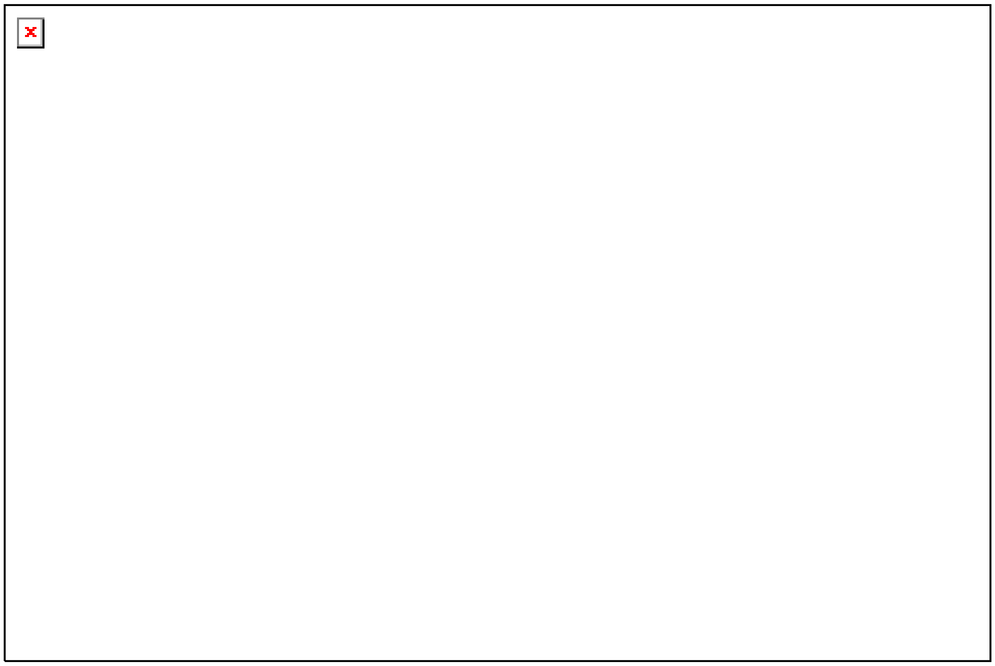

**Figure 4: Back-to-Front Videoconferencing**.

*Remote attendees to a meeting take their place at the table by means of video monitors mounted on the back wall. They see through the adjacent camera, hear via a microphone, and speak through their monitor's loudspeaker. The presenter uses the same conventional skills in interacting with those attending physically and those attending electronically. No new skills are required.*

The scenario shown in the figure illustrates the notion of transparency. A presentation is being made to five local and three remote participants. 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

[1] In fact, the room also supports traditional "front-to-back conferencing, which just pushes the issue of ubiquity even further.

question. If someone looks confused, a point can be clarified. Rather than requiring learning new skills, the design makes use of existing skills acquired from a lifetime of living in the everyday world.

***Example*: Hydra: supporting a 4-way round-table meeting**

In this example, we introduce a technique to support a four-way meeting, where each of the participants is in a different location. It was designed to capture many of the spatial cues of gaze, head turning, gaze awareness (Ishii, Kobayashi & Grudin, 1992) and turn taking that are found in face-to-face meetings. Consistent with the design principles outlined above, we do this by preserving the spatial relationships "around the table"[2]. This is illustrated in Figure 5.

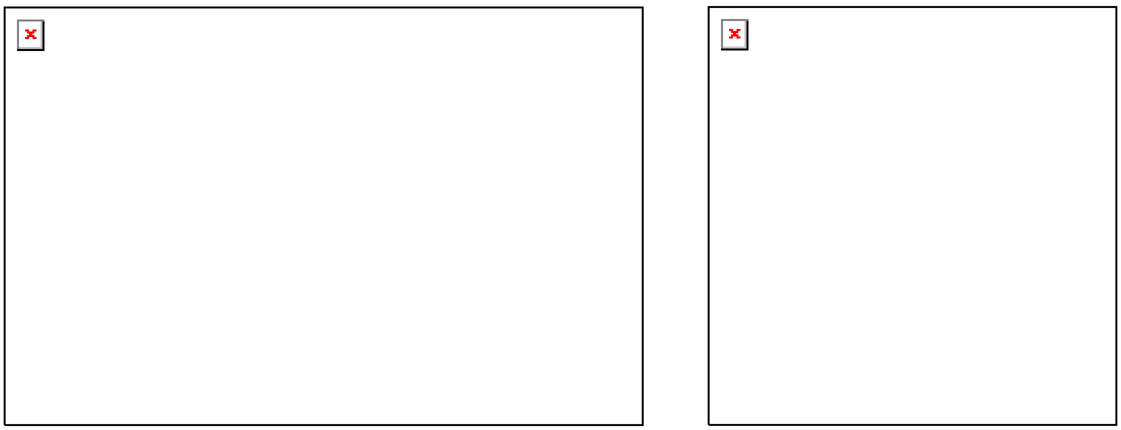

**Figure 5: Using Video "Surrogates"**.

*Figure 5(a) shows a 4-way video conference where each of the three remote participants attends via a video "surrogate." By preserving the "round-table" relationships illustrated in (b), conversational acts found in face-to-face meetings, such as gaze awareness, head turning, etc. are preserved.*

As seen in the left-hand figure, each of the three remote participants are represented by a small video surrogate. These are the small Hydra units seen on the desk (Sellen, Buxton & Arnott, 1992; Buxton, Sellen & Sheasby, 1997). Each unit provides you with a unique view of one of the remote participants, and provides each remote participant a unique view of you. The spatial relationship of the participants is illustrated by the "round-table" in the right-hand figure. Hence, relative to you, person A, B and C appear on the Hydra units to your left, front and right, respectively. Likewise, person A sees you to their right, and sees person B to their left.

---

[2] This idea of using video surrogates in this way for multiparty meetings turns out not to be new. After implementing it ourselves, we found that it had been proposed by Fields (1983).

Collectively, the units shown in the figure mean that the user has three monitors, cameras and speakers on their desk. Yet, the combined footprint is less than that of a conventional telephone. These Hydra units represent a good example of transparency through ubiquity. This is because each provides a distinct location for the source of each remote participant's voice. As a result, due to the resulting "cocktail party effect", the basis for supporting parallel conversations is provided. This showed up in a formal study that compared various technologies for supporting multiparty meetings (Sellen, 1992). The Hydra units were the only technology tested that exhibited the parallel conversations seen in face-to-face meetings.

The units lend themselves to incorporating proximity sensors that would enable aside comments to be made in the same way as face-to-face meetings: by leaning towards the person to whom the aside is being directed. Because of the gaze awareness that the units provide, the regular checks and balances of face-to-face meetings would be preserved, since all participants would be aware that the aside was being made, between whom, and for how long.

None of these every-day speech acts are supported by conventional designs, yet in this instantiation, they come without requiring any substantially new skills. There is no "user interface." One interacts with the video surrogates using essentially the same social skills or conventions that one would use in the face-to-face situation.

> ***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.*

Finally, we can augment the basic Hydra units by placing a large format display behind them. As shown in Figure 6, this is used to function like a large electronic "whiteboard" which enables the user to easily direct their glance among the other three participants and the work being discussed. Furthermore, if all four participants have their environments configured the same way, and the same information is displayed on each of the large displays, then each has optimal sight lines to the "whiteboard." Here is a case where the combination of electronic and physical space (Buxton, 1992) provides something that is an improvement on the traditional physical world where, if the physical whiteboard were across from you, it would be behind person "B" sitting opposite you. Furthermore, note that the awareness that each participant has of who is looking at who (so-called "gaze awareness") extends to the "whiteboard".

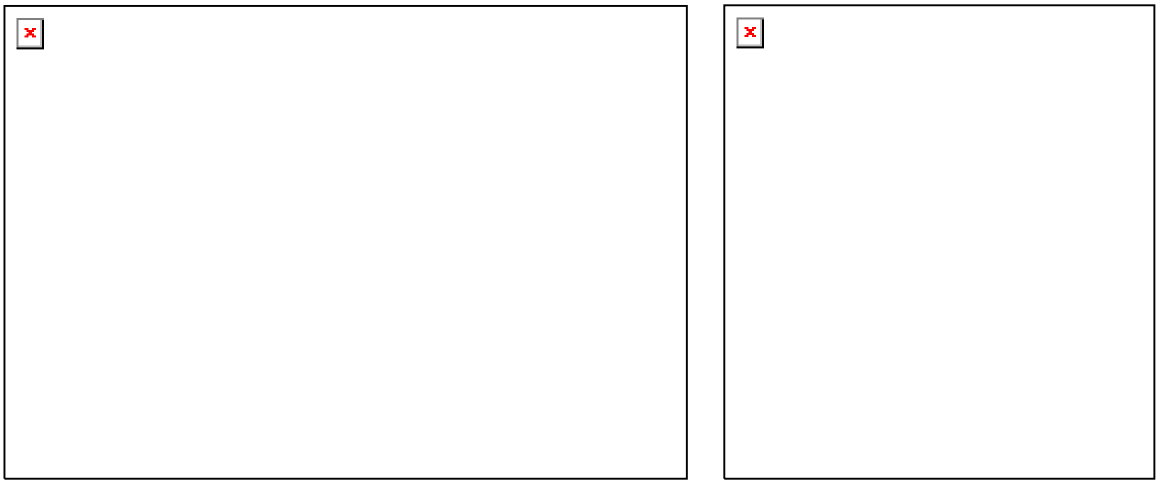

**Figure 6: Seamless Integration of Person and Task Space.**
*Figure 6(a) also shows a 4-way video conference using the Hydra units. However, this time, a large electronic "whiteboard" containing the information being discussed appears behind the units. As illustrated in blue in 6(b), the same display 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 not only to who is looking at who, but also to whether one is looking up at the "whiteboard" or at a person, thereby seamlessly blending person and task space.*

Our next example pushes even harder at the notion of using video surrogates to capture important relationships between physical space and social function.

## Example*:* Fly-on-the-wall View from the Door

The physical world occupies real space. Not only is there physical location and distance in this space, there is also the social space. Social graces are governed by our position and movement in the physical space. This is seen in things such as how we approach one another, or take leave. However moving through physical space involves a continuum, whereas making a connection via a video link does not. Therefore, with conventional desktop video techniques, such as illustrated in Figure 1, you are either there or not there, and when you are there, you are right in my face, you get there abruptly, and thereby violate normal social behaviour.

**Figure 7: Maintaining Social Distance**.

*In establishing contact, one appears by the door and has a from-the-door view via the camera, regardless of whether one approaches from the physical corridor (left image) or the electronic corridor (right image). People approaching electronically do so via a monitor and speaker mounted above the door (inset on right image). The social graces of approach are preserved, and the same social conventions are used for both physical and electronic visitors.*

Figure 7, above, illustrates our approach to addressing this problem. When you come to my office, you come via the door. If you come physically, then all is normal. If you come electronically, you also appear by the door, but on a small video monitor mounted above it. I hear your approach *via* an emitted "earcon" from a speaker by the door monitor *before* you appear or can see me. When you do see me (which is at the same time I can see you), you do so from a wide-angle low-resolution camera that is integrated into the same surrogate that incorporates the monitor and speaker. Thus, the glance that you first get is essentially the same as what you would get through the door. If I am concentrating on something or someone else, I may not see you or pay attention to you, just as would be the case if you were walking by in the hall (even though I may well hear that someone is there or has passed by). Appropriate distance is maintained. If you knock or announce yourself, I may invite you in, in which case you come in to the "visitor's" electronic chair, i.e., the visitor's monitor seen in Figure 3.

## ***Example:*** **Door State and Accessibility**

The previous example showed the preservation of distance for both electronic and physical visitors by preserving the social distance to the door.  We can extend this further.  The underlying assumption in what follows is that if we are going to build the means for wide access to our offices and our homes, then these must be balanced by complimentary means that enable us to exercise control over that access. This is an example of how that can be done in a manner that leverages everyday social conventions.

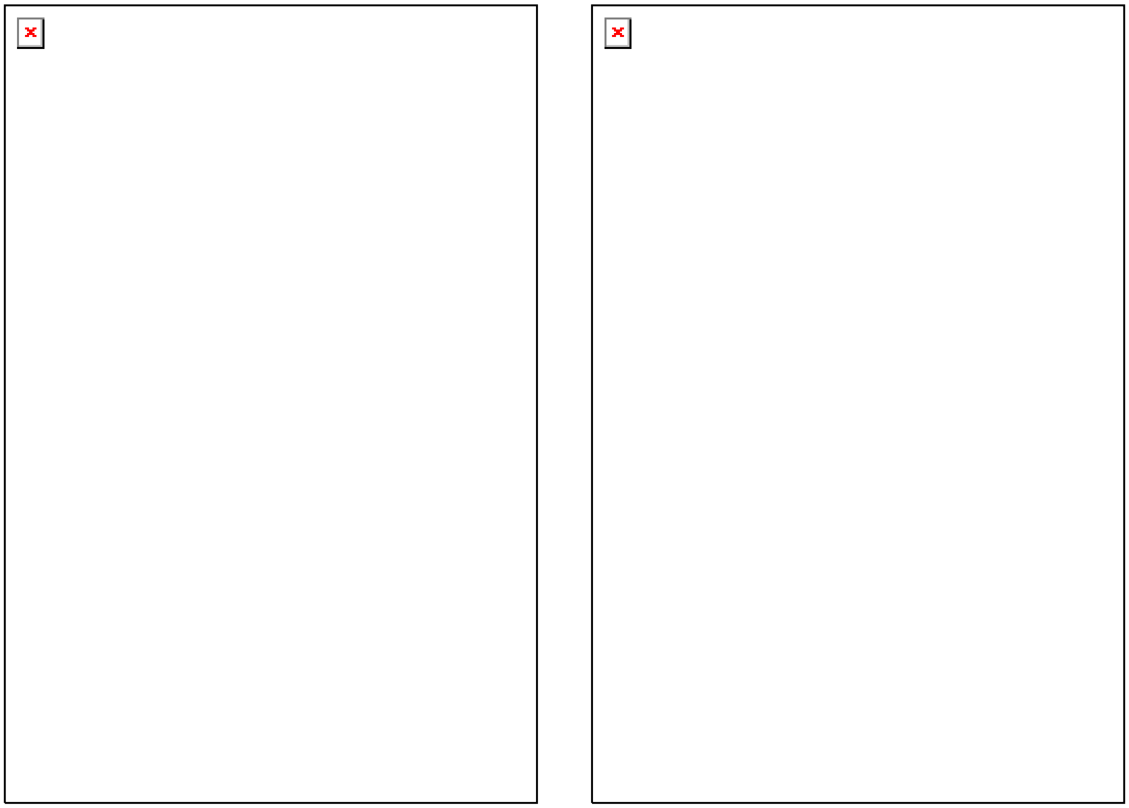

**Figure 8: Using "Door State" to Specify Accessibility.**
*The figures illustrates a technique for users in a media space to control their own accessibility following the same approach used in physical space, namely via the state of their door.  In this case, it is an icon of a door that can appear in any one of four states:  open, semi-open, closed, and boarded shut.  Each state indicates a different level of accessibility. Potential callers can determine a person's accessibility by the door state indicated beside each their name, as seen in the left image.  Each person can set their own door state, and therefore indicate their availability, using a simple menu, as shown in the right image.*

In our approach, the same basic mechanism that is used to control access for those approaching using the physical corridor – namely the door – is used as the basis to regulate access for those approaching along the electronic corridor. With physical visitors, for example, if my door is open, you are welcome to "pop in."  If it is ajar, you can peep in and determine if I am busy, but you will probably knock first if you want to enter. If it is closed, you will almost certainly knock and wait for a response before entering. If there is a "Do Not Disturb" sign on the door, you will not knock, but you might leave a message. According to Principle 3, so should it be for electronic visitations, regardless if one is approaching by phone or by video link.

Figure 8 represents an interface, first suggested by Abi Sellen, that we use to transfer these protocols to the electronic domain. With this interface, one sets one's own accessibility by selecting one of the four door states shown. One can even leave a "note" on the virtual door in order to pass on a message to visitors.

While preserving the protocols of the physical world by metaphor, this design, however, still fails to comply fully with Principle 3. The reason is that while the protocols are parallel, they are not “one”.

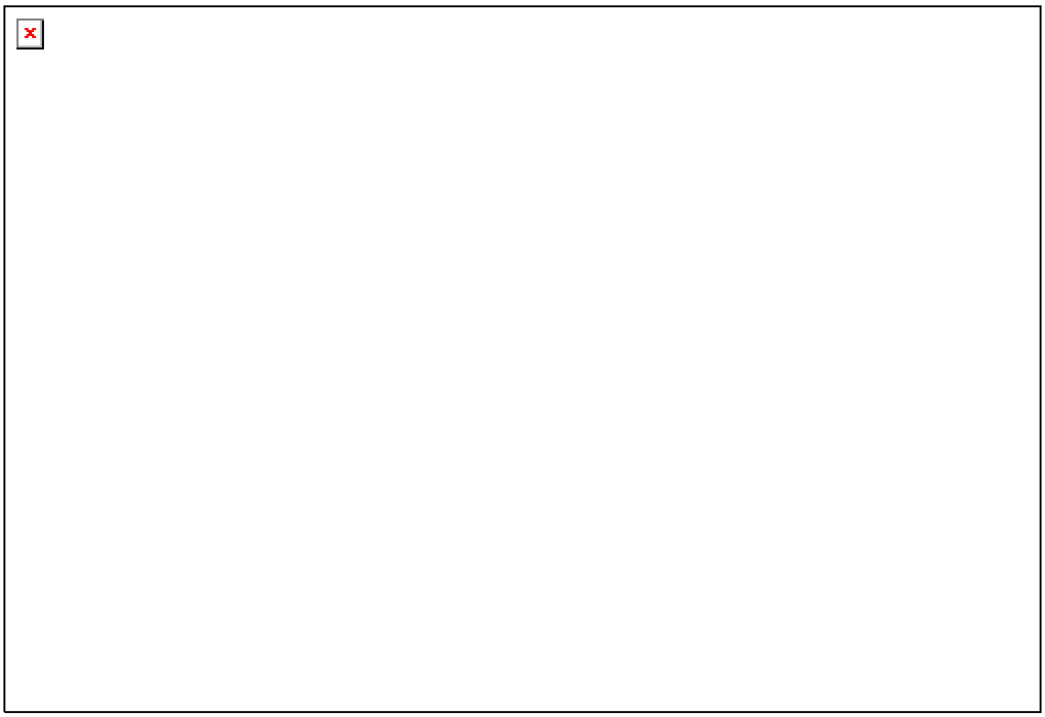

**Figure 9: The “Door Mouse.”**

*This is a MacIntosh mouse that has had the cover removed and been mounted on the wall by the door. A belt mechanism connects the door hinge to one of the mouse's shaft encoders. This enables the mouse to sense how open the door is and report this information to the computer so that it can set the door state (see Figure 8).*

This would be achieved if the physical door itself controlled the state of my accessibility for both electronic and physical visitors, alike. Hence (naturally subject to the ability to override defaults), closing my physical door could be sensed by the computer and prevent people from entering physically or electronically (by phone or by video).  A method of how this was implemented by one of our team, Andrea Leganchuck, is illustrated in Figure 9. Hence, we become consistent with Principle 3: one protocol controls all[3].

Much of the above is based on the notion that the physical location of participants has an important influence on social interactions in face-to-face meetings.  What we are driving at from a design perspective is that these same cues can be used or exploited in telepresence.  When we talk about distance between participants, therefore, it is important to distinguish between their physical distance from me, and the distance between their video surrogate and me.  The latter, rather than the former, is what determines social distance.

> ***Premise:*** *Physical distance and location of your video surrogate with respect to me carries the same social weight/function/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.*

# From Appliances to Architecture

Scale, as well as location is important in terms of its ability to affect the quality of interaction in a Mediaspace.  Consider the impact of electronically sitting across the desk from one another, as illustrated in Figure 1 compared to Figure 10.  In Figure 10, through rear projection, the remote participant appears life-size across the desk.  In this case, we are using essentially the same configuration as we saw in Figure 6; however, in this case the large display is showing the image of the remote person in a 1-on-1 conversation. I am captured by the Hydra camera, but the large display replaces the Hydra monitor. A number of significant points arise from this example.

[3] In reality, it is not quite this simple.  The meaning of door state is culture specific, so there are larger issues that need to be considered in implementing this feature for it to work.  As the ability increases of a system to sense the context within which it is to react, so must the quality and flexibility of the tools for user tailoring of those actions.  The examples that we give are to establish another way of thinking about systems.  They are not intended to provide some dogma as to specific designs.

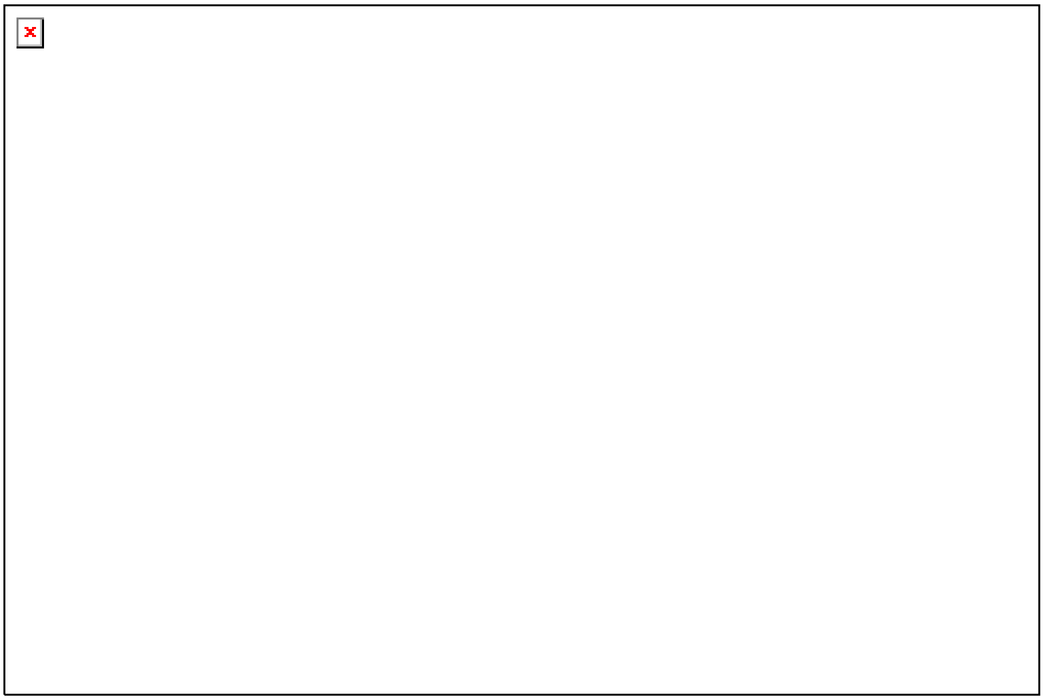

**Figure 10: Face-to-Face.**

*In this scenario, each participant has a computerized desktop on which the same information is displayed. The intention is to capture the essence of working across the desk from one-another. Each sees the remote participant life-size. The video camera (from a Hydra unit) is unobtrusive on the desk. Participants interact with the computer using a stylus. When one participant looks down to their desktop, their eyes seem to project into the space of the other, thereby strengthening the sense of telepresence. While there is a considerable amount of technology involved, it is integrated into the architectural ecology. What one gets is lots of service and lots of space, not lots of gear and appliances.*

First, it is not like watching TV. Due to the scale of the image, the borders of the screen are out of our main cone of vision. 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, and perhaps most important, 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 feeling that they are looking onto our own desktop. Their gaze traverses the distance onto our shared workspace, thereby strengthening the sense of telepresence. What is central to this example is the contrast between the simplicity and naturalness of the environment and the potency of its functionality. In keeping with the principle of invisibility, a powerful, non-intrusive work situation has been created.

> ***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. That is the difference between the ecological design of Ubiquitous Media and the design of appliances.*

The relevance of the above to future design is all the greater given that inexpensive large format displays will be one of the most "visible" new technologies emerging over the next few years.

---

# Proximal Sensing and Reactive Environments

The example of having the computer sense the state of our physical door, illustrated in Figure 9, breaks with conventional practice. Yet it is perfectly in keeping with our earlier example of the supermarket door. What is interesting is that the design bridges the gap between human-human and human-computer interaction.

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

The ability to make computers more "aware" of their surroundings is an important part of our work ((Buxton, 1995; Cooperstock, Fels, Buxton & Smith, 1997). We want to explore the degree to which sensed potentials can be mapped into system control signals. Doing so is similar to what happens in remote sensing, where sensors on satellites collect information about the ecology of Earth. Since it is the same thing, just a little closer, what we describe in this section could perhaps best be described as *proximal sensing.*

Consider this: a computer is made up of thousands of switches, yet AI notwithstanding, a motion-sensing light switch is smarter than any of them because it has the ability to sense motion and turn a light on when someone is present.

> *When you walk up to your computer, does the screen saver stop and the working windows reveal themselves? Does it even know if your 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?*

In addition to door sensors, motion sensors, and the like, all of the mediaspace technologies expand the potential for interaction in the UbiComp environment. The same cameras that I use for video conferencing can give my computer "eyes." The same microphone through which I speak to my colleagues can also provide my computer with an "ear." The displays on which I view my video may also display data, and vice versa: when the world is digital, video and data are one.

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

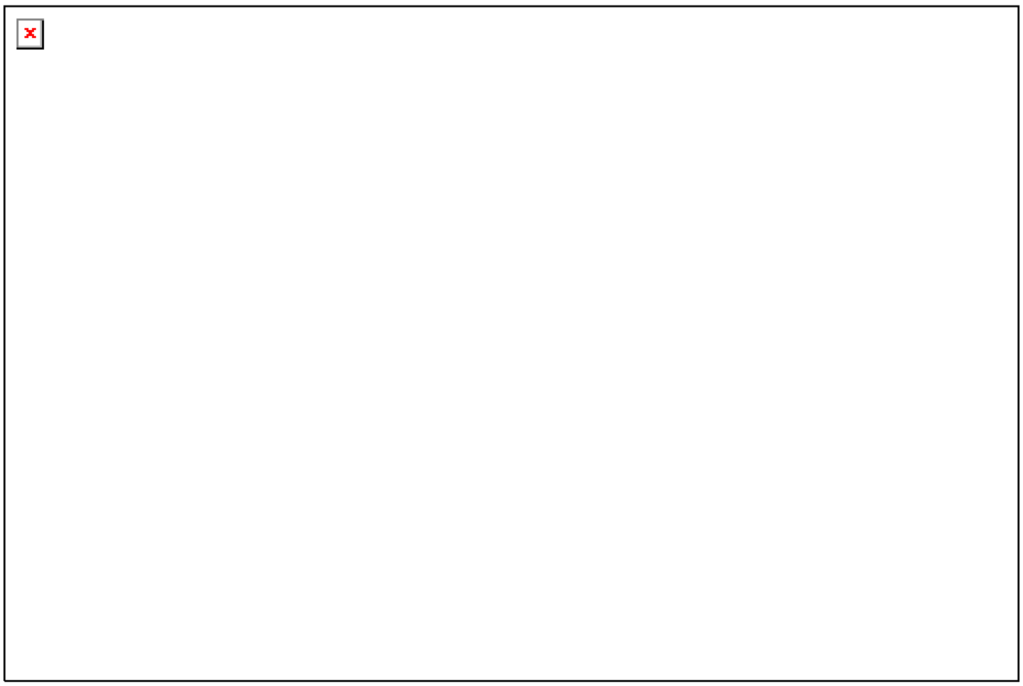

**Figure 11: Myron Krueger's Videodesk.**

*The user's hands are "seen" by the computer and superimposed on the display. The system can recognize each hand, its position and its shape (open, closed, pointing, etc.). Based on this, one can manipulate objects in the scene. In this case, a user is manipulating an oval type shape. Not only are both hands able to be used without holding any devices or wearing any intrusive gloves, the tip of the thumb and index finger of each hand is also being used to control the shape, thereby giving four times the amount of control that would be available using a mouse, for example.*

My desktop camera could sense if I am at my desk. If I am not, but the door-way camera senses that I am in the room, then the computer could switch from visual output to audio output in communicating with me. Also, since it is analyzing the input to the microphone (through simple signal detection), it knows if I am speaking or not. If so, it will wait until I am finished so that it doesn't interrupt.

This expanded repertoire of technologies can lay the basis for a far more seamless interface between the physical and electronic worlds. Krueger (1983, 1991) has shown that video cameras can be effective input devices for controlling computer systems. An example of his work is illustrated in Figure 11.

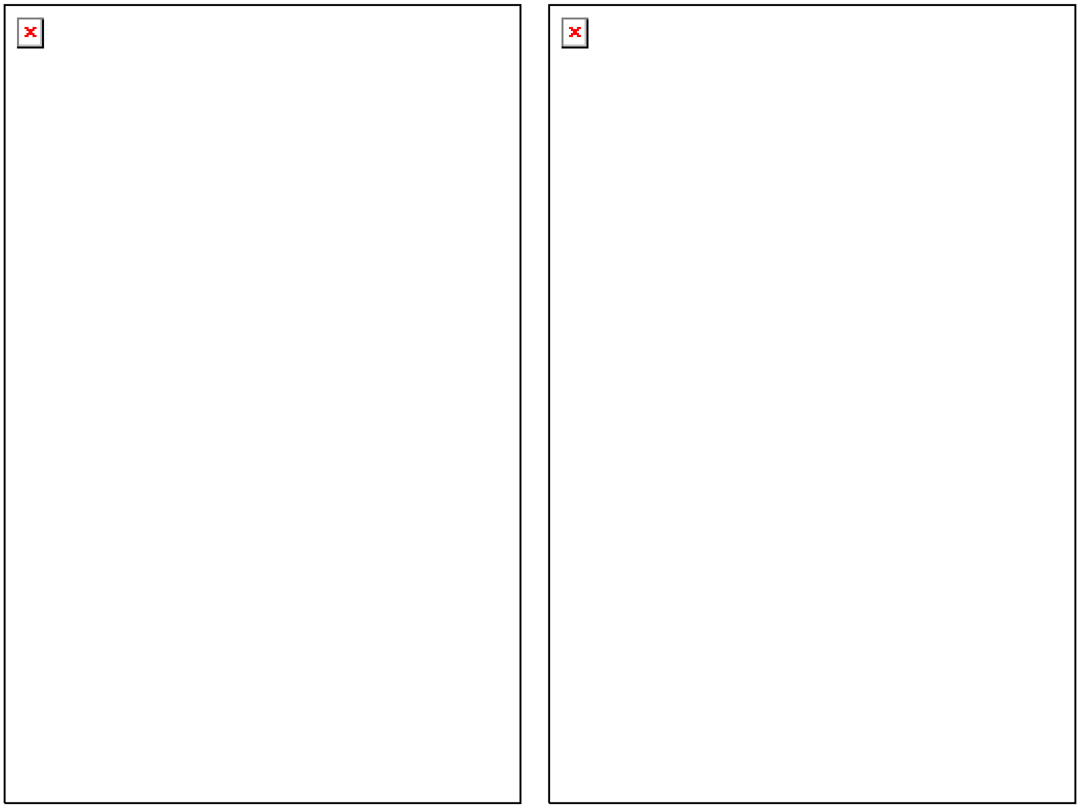

**Figure 12: Hand Tracking and Pose Recognition**

*The image on the left shows the "Active Desk", a rear projection work surface in the format of a traditional drafting table. The main mode of interaction is with a stylus, as the rear projection surface is actually a high-resolution translucent digitizing tablet. In the image, the desk is being used with a Hydra unit for collaborative design. On the right, a camera mounted above the desk is tracking the position of the left hand of the user relative to the graphical objects displayed on the surface. The camera also recognizes the hand pose - actually the angle between the thumb and the forefinger. In the style of Krueger, this enables the left hand to manipulate graphical objects, in a manner that compliments the use of the stylus in the right hand.*

Central to this approach (as opposed to that commonly seen in virtual reality systems) is that it is non intrusive. One need not wear any special gloves or sensors. The system sees and understands hand gesture much in the same way that people do: by watching the hands.

We have taken some of the ideas from Krueger's Videodesk, and done a simple proof-of-concept of how this type of video input can be used to augment interaction with desk-size displays, and in combination with more conventional input devices. This is illustrated in Figure 12.

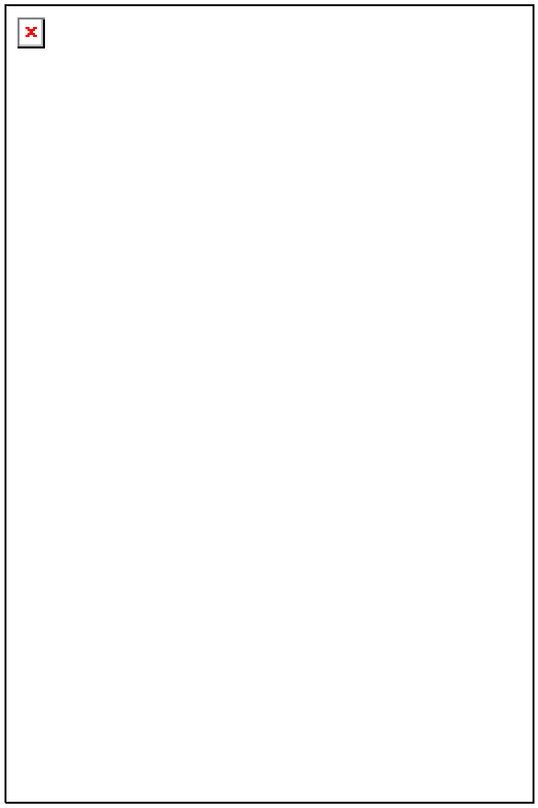

**Figure 13: The Digital Desk**

*With this system, the physical desktop can "hold" both paper documents, and electronic documents and applications that are projected onto it from a computer. In this image, the document on the left is paper, while to the right, a projected digital calculator is shown. The desk surface is also "seen" by two cameras. One looks at the whole surface, and looks at the location and the pose of the hands. Here it sees that the right hand is over the paper document, and in a pointing pose. The finger is pointing at a specific number on the paper document. The second camera can "read" what is written in the highlighted area around the right hand, and the number 4834 that the user is pointing at on the paper document is entered into the calculator where it can be manipulated by tapping on the projected calculator keys using the same finger that is being used for pointing.*

Here we see how a camera is being used as input to the computer, enabling the left hand of the user to manipulate graphical (virtual) objects projected onto the work surface, in a complimentary manner to the manipulations being done with the conventional stylus, held in the right hand.

However, it is not just the link between human and machine that these technologies facilitate. It is also the provision of a more seamless link between the artefacts of the physical and electronic worlds. As technologies become more "intimate," or close to the person, they will increasingly have to provide a bridge between these two worlds. Small portable tab-sized computers may more resemble a camera than a calculator, for example. One of the best examples of using these media to provide such a bridge is the Digital Desk of Wellner (1991), illustrated in Figure 13. This system goes beyond both desktop computers and the desktop metaphor. In this case, the desktop is the computer.

As shown in the figure, there is a projector and a camera mounted over the desk. The former projects the computer's display onto the desktop. The camera enables the computer to "see" what is on the desktop. Hence, electronic documents can be projected, as can active widgets such as a calculator or a browser. And, like the Krueger example, the camera enables the computer to see the actions of the hands on the desk, and to use this as input. It also enables the computer to "see" documents and objects on the desktop. Here again the potential exists for recognition. In the working prototype, for example, the camera can be used to scan alphanumeric data to which optical character recognition techniques are applied, thereby enabling the computer to "read" what is on the desk.

# 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, as described, represents a break from previous practice. It is a shift in design attitude that builds upon existing skills on the part of the user, especially those pertaining to their understanding of space/time/distance relationships. It is a mature approach to design that breaks out of the "solution-in-a-box" 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 improve, reconcile and balance skills in technology with those in design and social science. Throughout, we must keep in mind that humans come to our systems with a broad repertoire of skills and knowledge. Any design that does not reflect this with a high degree of fidelity demonstrates a lack of respect, and is unacceptable.

# Acknowledgments

The ideas developed in this essay have evolved over countless discussions with colleagues at Rank Xerox EuroPARC, Xerox PARC and the Ontario Telepresence Project. To all of those who have helped make these such stimulating environments, I am very grateful. I would like to especially acknowledge the contributions of Abi Sellen, Sara Bly, Steve Harrison, Mark Weiser, Brigitta Jordan and Bill Gaver. Finally, I would like to acknowledge the contribution of my wife, Elizabeth Russ, who made many helpful comments on the manuscript.

The research discussed in this paper has been supported by the Ontario Telepresence Project, Xerox PARC 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. Earlier version appears in Proceedings of Groupware '91, Amsterdam, Oct. 29, 1991, 27-36.

Buxton, W. (1995). Integrating the Periphery and Context: A New Model of Telematics. *Proceedings of Graphics Interface '95*, 239-246.

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.

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. & Sheasby, M. (1997). Interfaces for Multiparty Videoconferencing. In K. Finn, A. Sellen & S. Wilber (Eds.). *Video Mediated Communication.* Hillsdale, N.J.: Erlbaum, 385-400.

Cooperstock, J., Fels, S., Buxton, W. & Smith, K.C. (1997). Reactive environments: Throwing away your keyboard and mouse, *Communications of the ACM*, 40(9), 65-73.

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, 1983.

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*, *ACM Conference on Human Factors in Software* 27-35.

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, ACM Conference on Human Factors in Software*, 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, ACM Conference on Human Factors in Software*, 49-59.

Sellen, A., Buxton, W. & Arnott, J. (1992). Using Spatial Cues to Improve Videoconferencing. *Proceedings of CHI '92, ACM Conference on Human Factors in Software*, 651-652. Also videotape in CHI '92 Video Proceedings.

Stults, R. (1986). Media Space. *Systems Concepts Lab Technical Report*. Palo Alto, CA: Xerox PARC.

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. *Communications of the ACM*, 36(7).