

Ubiquitous Media and the Active Office

Bill Buxton

University of Toronto & Xerox PARC[\[1\]](#)

UbiMedia = UbiComp + UbiVid

In 1991, Mark Weiser published an article that outlined Xerox PARC's vision of the next generation of computation (Weiser, 1991). He referred to this model as *Ubiquitous Computing*, or *UbiComp*. In what follows, we introduce a complimentary component of Weiser's story: what we call Ubiquitous Video, or *UbiVid*.

The groundwork for UbiVid was laid by research into in "media spaces" (Gaver *et al.*, 1992; Mantei *et al.*, 1991; Stults, 1986; Bly, Harrison & Irwin, 1993). The ideas that we discuss build upon this work. We argue that UbiComp and UbiVid are two sides to the same story. Together, they make up something that may best be called *Ubiquitous Media*, or *UbiMedia*. Our belief is that this notion of Ubiquitous Media provides a useful model for conceiving of future systems and their usage models.

This paper is based on research undertaken at Xerox PARC, Rank Xerox EuroPARC and the Ontario Telepresence Project. Many of the ideas discussed have been implemented or prototyped. Many more have not. Our purpose here, however, is not to report on research *per se*. Rather, our intent is to convey the model of future computation that lies behind much of our work. Our hope is to aid communication and provide the basis for future discussion.

UbiComp: a Brief Review

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 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. Central to UbiComp is a break from the "Henry Ford" model of computation in which, can be paraphrased asfsim:

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.

Technology Warms Up

We can most easily place Weiser's model 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 next generation of computation was available in rooms outside of computer centres; however, these had to have special electrical cabling and air conditioning. Therefore, computation was still restricted to special "computer rooms."

In the next generation of heating system, we moved to Franklin stoves and even to 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, this is not unlike the next generation of computation, (the generation in which we are now living), where we have access to distributed computation everywhere, 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 next (today's) 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 actually 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 one tailor the climate to their individual preference. This is the heating equivalent of UbiComp: the service is ubiquitous, yet the delivery is invisible. In this mature phase, the technology is seamlessly integrated into the architecture of the workplace.

Thus, 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."

UbiVid

UbiVid is the video compliment 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. Generally, the video interactions are confined to this single camera-monitor pair.

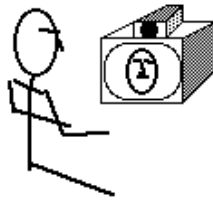


Figure 1: A Typical Desktop Video Conferencing Configuration

Conferencing is typically channeled through a video camera on top of a monitor on the user's desktop.

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 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: *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.*

Example: Visitor vs. Office Mate

Figure 2 illustrates two scenarios for how an office might be laid out. The first (a), shows an office with a desk holding a single video monitor and a visitor's chair. This is typical of the layout that one would find in most "mediaspace" environments (including our own). This layout, however, violates design Principles 1 & 2. All video transactions occur on a single monitor in a fixed position. This not only causes contention when there are overlapping demands for services (such as when someone wants to conference while I am watching a video). It also means that location/function relationships cannot be exploited, as we shall see below.

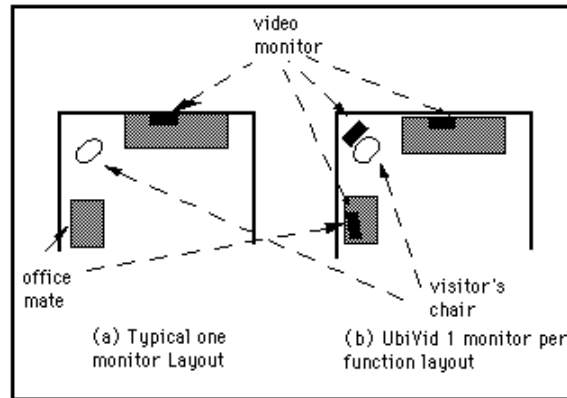


Figure 2: Desktop Video vs UbiVid

Figure 2(a) shows channeling all video interactions through a single camera/monitor pair. Figure 2(b) illustrates the UbiVid approach of distributing functionality. For simplicity, only monitors (shown in black) are illustrated.

The second layout (b) captures the character of UbiVid. The video monitor on the desk is for "reading" video documents and doing "up-close" work with a remote colleague. The monitor for a video "visitor" is behind the visitor's chair. Furthermore, a monitor for a virtual office mate is off to the side where an office mate's desk would be.

For simplicity, the figure shows video monitors only. In each case, there would be a camera and loudspeaker paired with each monitor, to ensure reciprocity. In addition, one would typically also have one or more computers (which may or may not use the same monitor(s) as the video).

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 3.[\[2\]](#) The intent here is to enable remote participants to "take their place at the table."

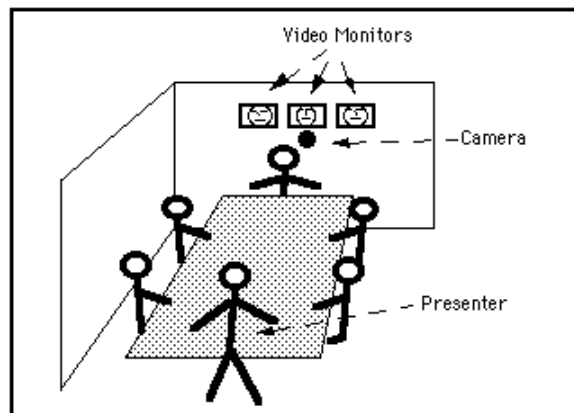


Figure 3: Back-to-Front Videoconferencing

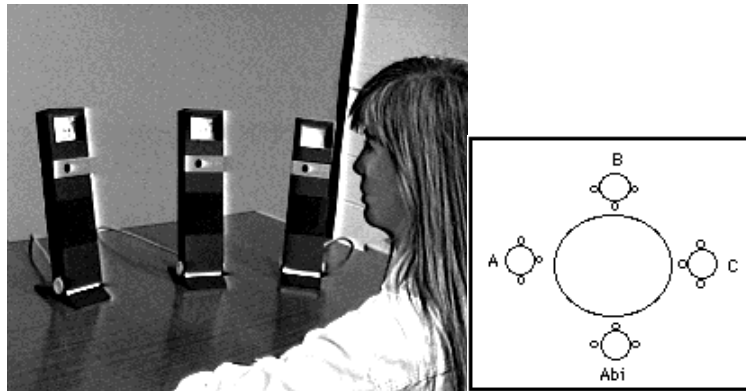
Remote attendees to a meeting take their place at the table by means of video monitors mounted on the back wall. They see through the adjacent camera, hear via a microphone, and speak through their monitor's loudspeaker. The presenter uses the same conventional skills in interacting with those attending 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 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 life time 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 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."^[3] This is illustrated in Figure 4.



(a) (b)

Figure 4: Using video "surrogates" to support a 4-way video conference

Figure (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). Sitting in front of the desk is a colleague, Abi Sellen. Each unit provides a unique view of her for one of the remote participants, and provides her a unique view of them. The spatial relationship of the participants is illustrated by the "round-table" in the right-hand figure. Hence, person A, B and C appear to Abi on the Hydra units to her left, centre and right, respectively. Likewise, person A sees her to their right, and B to their left, etc.

Collectively, the units shown in the figure mean that Abi has three monitors, cameras and speakers on her desk. Yet, the combined footprint is less than that of her telephone. These Hydra units represent a good example of transparency through ubiquity. This is because each provides a distinct point source for the voice of each remote participant. As a result, the basis for supporting parallel conversations is provided. This showed up in a formal study which 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.

Our next example pushes even harder on 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 location and distance in this space, but social graces are determined by how we *move* in this space, such as in approaching one another, or in taking leave. 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, thereby violating normal social behaviour.

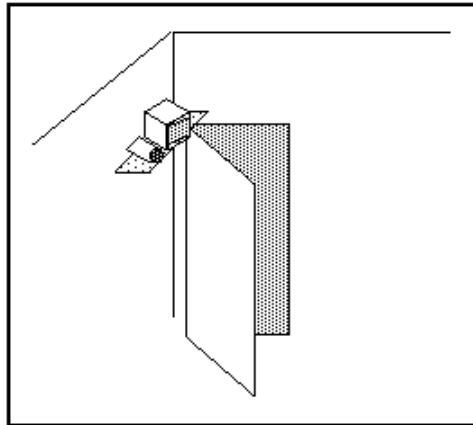


Figure 5: Maintaining social distance

In establishing contact, one appears by the door and has a from-the-door view via the camera, monitor and speaker mounted at that location. The social graces of approach are preserved, and the same social conventions are used for both physical and electronic visitors.

Figure 5 above shows a UbiVid approach to 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 appear in the monitor by the door, I hear you from the speaker by the door, and you see me from a wide angle low-resolution camera by the door. 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 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" chair, i.e., the visitor's monitor seen in Figure 2. On the other hand, on glancing in, you may well see that I am busy and choose to come back later (physically or electronically). This design serves both parties in the interaction. The visitor is saved the potential embarrassment of intruding in the middle of something, and the integrity of the personal space of the occupant is preserved.

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. That same door controls my accessibility to physical visitors. If it is open, you are welcome to "pop in." If it is ajar, you can peep in and determine if I am busy. You will probably knock if you want to enter. If it is closed, you will 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 6 represents the interface, 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.

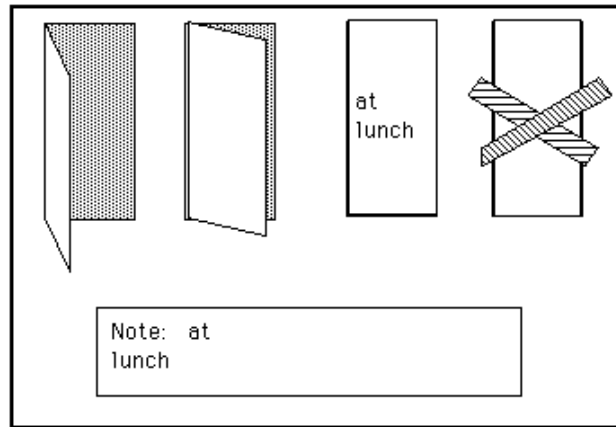


Figure 6: Using "door state" to specify accessibility

The figure illustrates a technique for users in a media space to control their own accessibility following the same approach used in physical space: by the state of their door. Each one of the selectable door states allows a different level of accessibility. The specified door state is visible to potential callers who, likewise, know the implied permissions by analogy with the physical world.

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*.

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). One action and one protocol controls all.[\[4\]](#)

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

Consider the UbiVid equivalent to sitting across the desk from one another, as illustrated in Figure 7. Here, through rear projection, the remote participant appears life-size across the desk. What we are trying to capture in this example is where two people are working together on a joint project, such as a drawing or budget, which is on the desktop. In implementing this example, a number of significant points arise.

First, it is not like watching TV. Because of 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* (see also Ishii, Kobayashi & Grudin, 1992) gives us the sense that they are looking onto our own desktop. Their gaze traverses the distance onto our shared workspace, thereby strengthening the sense of *Telepresence*.

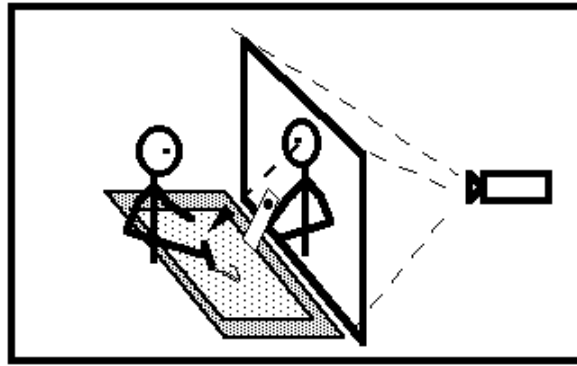


Figure 7: Face-to-Face

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.

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.*

Active Sensing and the Active Office: UbiComp Meets UbiVid

The earlier example of having the computer sense the state of our physical door breaks set with conventional practice, yet is a natural outgrowth of both UbiComp and UbiVid. It bridges the gap between human-human and human-computer interaction.

Observation: *A door is just as legitimate 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. We want to explore the degree to which sensed potentials can be mapped into system control signals. 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 since 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 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 switched in the computer, AI notwithstanding?

In addition to door sensors, motion sensors, and the like, the technologies of UbiVid 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).*

My desk-top 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 to 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. Central to his 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.

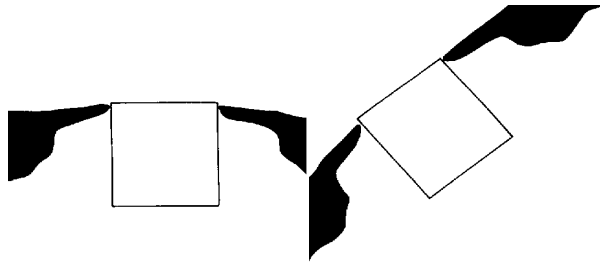


Figure 8: Myron Krueger's *Videodesk*:

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 reorienting a square.

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 artifacts 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 (1992), illustrated in Figure 9. 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 desk. 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.

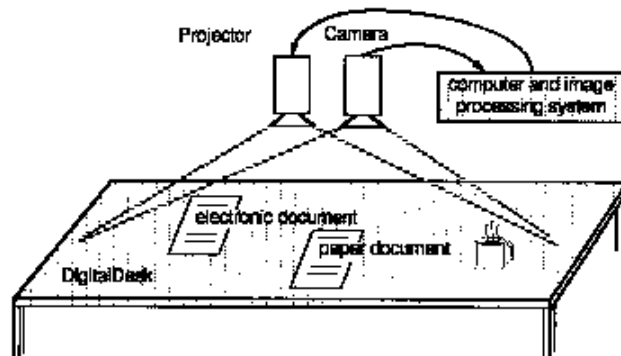


Figure 9: The Digital Desk (Wellner, 1992)

With this system, electronic documents are projected onto the desktop. Similarly, a camera enables the computer to see what is on the desktop. It can see documents, and "read" them using optical character recognition (OCR) techniques. It can also "See" the user's hands and recognize gestures, such as pointing, selection, and activating graphical "buttons" of devices projected onto the desk surface.

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 the Ubiquitous Media approach 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" *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

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. In addition, I would like to thank Barbara Whitmer, who made many useful 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.
- Fields, C.I. (1983). Virtual space teleconference system. *United States Patent 4,400,724*, August 23, 1983.
- Gaver, W., Moran, T., MacLean, A., Löfstrand, L., Dourish, P., Carter, K. & Buxton, W. (1992). Realizing a video environment: EuroPARC's RAVE System. *Proceedings of CHI '92*, 27-35.
- Ishii, H., Kobayashi, M. & Grudin, J. (1992). Integration of inter-personal space and shared workspace: Clarboard 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.
- 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*, 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. (1992). The DigitalDesk Calculator: Tangible manipulation on a desktop display. *Proceedings of the Fourth Annual Symposium on User Interface Software and Technology*, 27-33.

[1] Author's current address: W. Buxton, Head, User Interface Research, Alias | Wavefront Inc., 110 Richmond St. E., Toronto, Ontario, Canada M5C 1P1. Fax: 416-861-8802. Email: buxton@aw.sgi.com.

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

[3] 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).

[4] In reality, it is probably wrong to hard-wire such protocols into a system. The meaning of door state is culture specific, for example. As the ability of a system to sense the context within which it is to react increases, 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.