Chapter 15:

The Future and Emerging Potential

Force Feedback

- IPO Trackball (Keyson,D., 1996) note claim that eye-hand reaction time is about 175 ms. while force-motor reaction times are normally about 50 ms faster (i.e. 125 ms). Also, has pressure sensing for state)-1 transitions. Als. Note that if add rotation, operating graphical potentiometer by rotation means no state1 transition.
- Beck & Stumpe
- Pantograph
- Minsky (Margaret)

Akamatsu, M., Sato, S. & MacKenzie, I.S. (1994Multimodal mouse: A mouse-type device with tactile and force display. *Presence* 3(1), 73-80.

- SensAble: phantom, MIT
- Cybernet Systems, Ann Arbour, Mi
- Immersion Corp., Santa Clara, Ca

Ramstein, C., & Hayward, V. (1994).

Massie, T. (1998). A tangible goal for 3D modeling. *IEEE Computer Graphics and Applications*, 18(3), May/June 1988, 62-65.

Brooks, F.P. Jr., Ouh-Young, M, Batter, J. & Kilpatrick, P.J. (1990).
Hayward, V. (in press). Towards a seven axis haptic device.
Iwata, H. (1990). Artificial reality with force-feedback: development of desktop virtual space with compact master manipulator, *Computer Graphics* 24(3), Proceedings of SIGGRAPH '90, 165-170.

Schmult, B. & Jebens, R. (1993). Application Areas for a Force-Feedback Joystick, *Proceedings of the ASME Winter Annual Meeting: Advances in Robotics, Mechatronics and Haptic Interfaces,* 47-54.
Schmult, B. & Jebens, R. (1993). A High Performance Force Feedback Joystick, *Proceedings of Virtual Reality Systems '93*, 123-129.

Tangible Interfaces

- George's stuff
- Hirochi

- Continue from Hinkley
- Wacom paint
- Epistemic vs pragmatic action

Eye Input:Eye Input:

- Hutchinson, T., White, K., Martin, W., Reichert, K. & Frey, L. (1989)

- Jacob, R. (1991). The use of eye movements in human-computer interaction techniques: What you look at is what you get, *ACM Transactions on Information Systems,* 9(3), 152-169.

Reactive Environments and Augmented Reality

Introduction

With multimedia and desktop videoconferencing, computer are begining to be equipped with a richer set of input technologies. Some are there originally to support human-human interaction, such as video cameras. However, an important observation is that any device that can be used for human-human communication is also a candidate for human-computer interaction.. 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.*

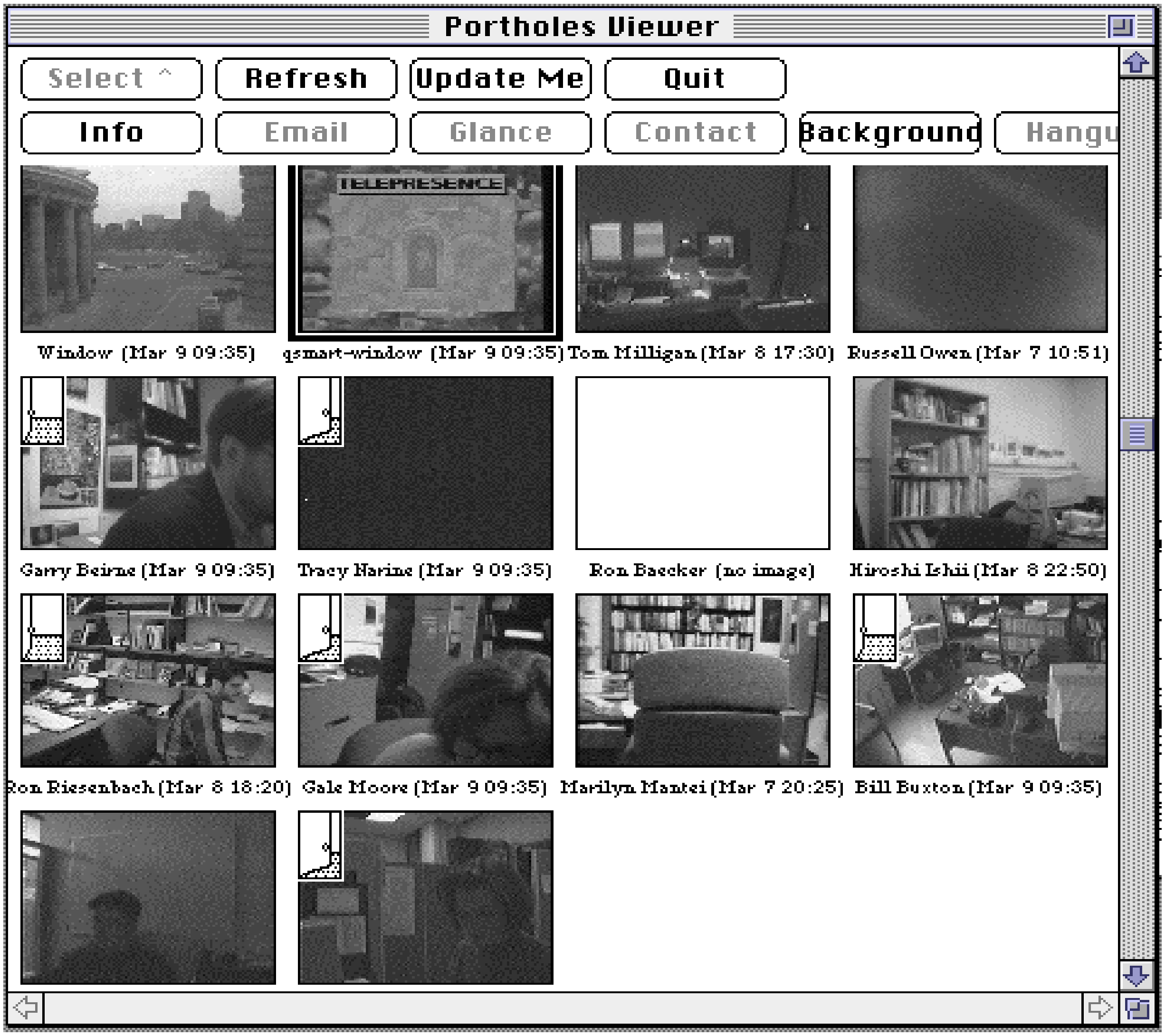

Figure 15: Portholes: Peripheral Awareness of One's Social Network

The original Portholes system was a joint development of Xerox PARC and Rank Xerox EuroPARC. Every 5 minutes, in the case of the Telepresence Project, a snapshot of each member of the workgroup is distributed to all other members of the group. In the Telepresence implementation, this is accompanied by an icon of that member's door icon (see Figure 11). 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.

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).*

Krueger (1983, 1991) has demonstrated how video signals of the user can be effectively used in human-computer interaction. 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 to track the position of the hands over the desk. This is illustrated in Figure 16, which illustrates 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.

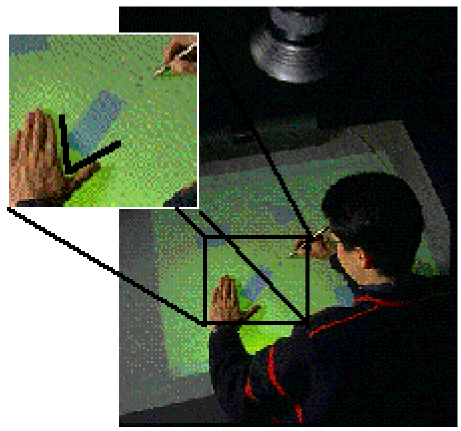

Figure 16: Using the Hand as Input

In this prototype system developed by Yuyan Liu, a video camera mounted above the Active Desk captures the position and orientation of the operator's left hand. It also determines the angle between the thumb and forefinger. The user is able to grasp and interact with 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 that people do: by watching the hands or body.

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 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 determine — within 5 minutes of resolution — 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 could look up at Portholes to see if you are there. If so, I could double click on your image to assert a connection. Otherwise, I could instruct the system that I want to talk to you. In the background, while I get on with other work, it could monitor the state of your office and alert me when you appear to be in *and* (by virtue of your door state) when you are available. The benefit of such a utility increases dramatically when it is a conference call that one wants to set up.

The Digital Desk

The information store of most of us consists of two solitudes: the information that is electronic form and that which is not (such as all the paper in our filing cabinet.) As was discussed earlier, it is important to be able to maintain a shared space of task as well as person. And, yes, video can transmit images of documents and other nonelectronic artifacts, while the computer can distribute the electronic ones. But the combined forces of the technologies can do even better than that, as has been shown by the *Digital Desk* of Wellner (1991).

With this system, the video camera enables the computer to "see" what is on the desktop. Like the systems of Krueger, it 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 Wellner's working prototype, for example, the camera was used to scan alphanumeric data to which optical character recognition techniques are applied, thereby enabling the computer to "read" what is on the desk. Wellner's system is an excellent example of the concepts of Ubiquitous Media in action.

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.

Specifying door state using the mechanism illustrated in Figure 11 preserves the protocols of the physical world *by metaphor*; however, it fails to comply fully with the design principal of using the same mechanism in both the electronic and the physical domain. The reason is that while the protocols are *parallel,* they are not *one.* One still has to maintain two systems: the physical door and the logical one, as represented in the computer application.

Using the *physical* door to control both means that accessibility for both electronic and physical visitors are handled by the same mechanism. Hence (naturally subject to the ability to override defaults), closing my physical door is be sensed by the computer and prevents people from entering physically or electronically (by phone or by video). One action and one protocol controls all.[512]

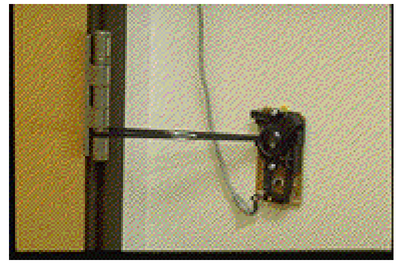

Figure 17: The "Door Mouse"
This is a mechanism for instrumenting the door such that its current state can be monitored by my computer. It was designed and implemented by Andrea Leganchuk. It consists of a Macintosh computer's mouse, with the cover removed, screwed onto the wall. A belt couples the door hinge to one of the mouse's shaft encoders. The mouse is then connected to the computer.

Such a system was implemented in a number of rooms in our lab by a student, Andrea Leganchuk. Her simple but elegant solution is illustrated in Figure 17.

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 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?*

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 of 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 of 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 has to 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 8), 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, in press).

---

[512] 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.