Integrating the Periphery and Context: A New Model of Telematics

Bill Buxton
University of Toronto & Alias Research Inc.
c/o Alias Research
110 Richmond St. E.
Toronto, Ontario
Canada M5C 1P1
buxton@alias.com

Abstract

From many directions, we hear about a new era of telematics, emerging from a convergence of telecommunications, computer and consumer electronics technologies, and its potential for social benefit. However, of the resources invested in developing telematics applications, the majority are spent on so-called (but unproven) 'killer' applications, such as home shopping and video on demand. Our belief is that the real promise of telem!tics, and the opportunities for realizing social benefits, lie in the symbiotic relationships among a suite of applications, rather than the value of any one killer' application.

We introduce a new model directed at supporting a holistic way of thinking about such systems, a prime property of which is the consideration it gives to both foreground and background interactions.

Keywords: videoconferencing, CSCW, telecommunications, electronic highway, technology convergence

Introduction

From many directions we hear about a new era of telematics which emerging out of a convergence of telecommunications, computer and consumer electronics technologies. From most quarters, great things are promised, and one of the main points of agreement seems to be that it is the applications that will determine the true value of the end result. Certainly, applications are important. But relatively few resources (human or financial) are being invested in their study or development. Of these, the majority are being allocated towards hoped-for "killer" applications, such as home shopping and video on demand.

There are two main concerns that arise from this. First is the limited scope of the applications that are being investigated. Second is our belief that the real value lies in the symbiotic relationships among a suite of applications, rather than the value of any one "killer" application. Hence, approaches that are limited to individual applications, or applications in isolation, run the risk of missing the target. If true, then the consequence is that a far more holistic approach must be taken.

In what follows, we introduce a new model which is directed at supporting just such a holistic way of thinking about such systems. Our hope is that it is simple enough to be understood and used, yet rich enough to be useful.

Our belief is that these emerging technologies have real potential for social benefit. If this model makes some contribution to the realization of this potential, then it will have served its purpose.

Basic Framework

Our model is based on the simple 2 x 2 matrix illustrated in Figure 1. The two dimensions are the *ground*, as represented by the columns, and *object of communication*, as represented by the rows. The dimension of "ground" is the one most needing clarification. By this we do *not* mean synchronous/asynchronous. That is covered in the previous complimentary taxonomy. What we mean by *Foreground* are activities which are in the fore of human consciousness. Intentional activities, such as speaking on the telephone, or typing into a computer.

By *Background*, we mean tasks that take place in the periphery, "behind" those in the foreground. Examples would include being aware of someone in the next office typing, or the light in your kitchen going on automatically when you enter it, as opposed to you manually flicking the switch (which would be a foreground intentional act). The notion of background, as used in this essay, relates directly to the *context* of interaction, a topic discussed at length in Moran (1994).

Foreground / Background

Human - Human	conversation, telephone video conf.	"Portholes"
Human - Computer	GUI's	smart house technology

Figure 1: The Basic Model

The two rows concern who or what the user is communicating with: another human, or a computer. *Human-Human* communication in the foreground might be a simple telephone call. (In this row, we actually mean technology-mediated human-human interaction.) The second row, *Human-Computer* communication, can be characterized by interactions with a computer using a graphical user interface (GUI).

From this simple model, we can already derive some valuable observations. First, while it is easy to populate the left column, this is much less true with the right one. Second, nearly all work in technology mediated human-human and human-computer interaction falls within the left column. However, a fundamental belief of our work is that the real "sweet spots" of future applications lie on the right hand side. This requires further explanation.

Communication in the Background

One could argue that insofar as supporting human-human interaction, telephones and videoconferencing do a reasonably good job. We can hold fairly rich conversations, see each other, judge moods, etc. So why is there still such a sense of distance between people, despite such technology? Our belief is that this is due to the fact that such technologies do not share some of the key affordances that occur naturally when people work in close physical proximity (Gaver, 1992). Regardless of the fidelity of the videophone, I still have no sense of who is in when. I can't "bump into" people in the hall, know who is available and who is busy, or take advantage of synergistic opportunities when just the right combination of people happen to be at the water cooler at a particular time. Yet, in shared physical space, all of these are commonly available almost effortlessly in the background, due to our awareness of the periphery.

This, as well as previous work on informal communication (Kraut & Egido, 1988; Whittaker, Frohlich & Daly-Jones, 1994) suggests that what is required is a "social prosthesis" which provides a sense of the periphery over distance. Used in combination with existing (and new) methods for supporting foreground conversation (such as the "Mediaspace" described by Bly, Harrison & Irwin, 1993), we believe this will achieve a significant improvement in the sense of copresence, or "telepresence" (Buxton, 1992).

One example of such a technology is referred to in the upper right corner of Figure 1, the *Portholes* system developed by Xerox PARC and Rank Xerox EuroPARC (Dourish & Bly, 1992). Portholes is a system which takes video "snapshots" of members of a community every 5 minutes, and circulates them to the computer screens of the members of that same community, as shown in Figure 2. Hence, all members have an increased awareness of who is in, what they are doing and if they might be available. They also provide a means of combating the all too human tendency towards "out of sight, out of mind." All members of the community have a visual presence, regardless of actual geographical location.

Our claim is that Portholes is an excellent example of a background "awareness server," of which there are many others.

Likewise, along the Human-Computer interaction dimension, there are also background technologies. The example cited in the bottom right quadrant of the figure is "smart house" technology. These are technologies that turn down the heat on weekends, automatically water your plants, close blinds, turn on lights, etc., under computer control.

Seamless Transitions Between Quadrants

At this stage, we could expand upon these examples, and further populate these two (admittedly sparse) right-hand quadrants. However, while we think this is worthwhile, and is something which we fully intend to do as part of our research, we believe that there is a larger and more important point to be made.

The real power of this model comes not from merely populating the individual quadrants, *but by providing the means to make transitions seamlessly from quadrant to quadrant.* This is illustrated by the arrows in the version of the model shown in Figure 3.

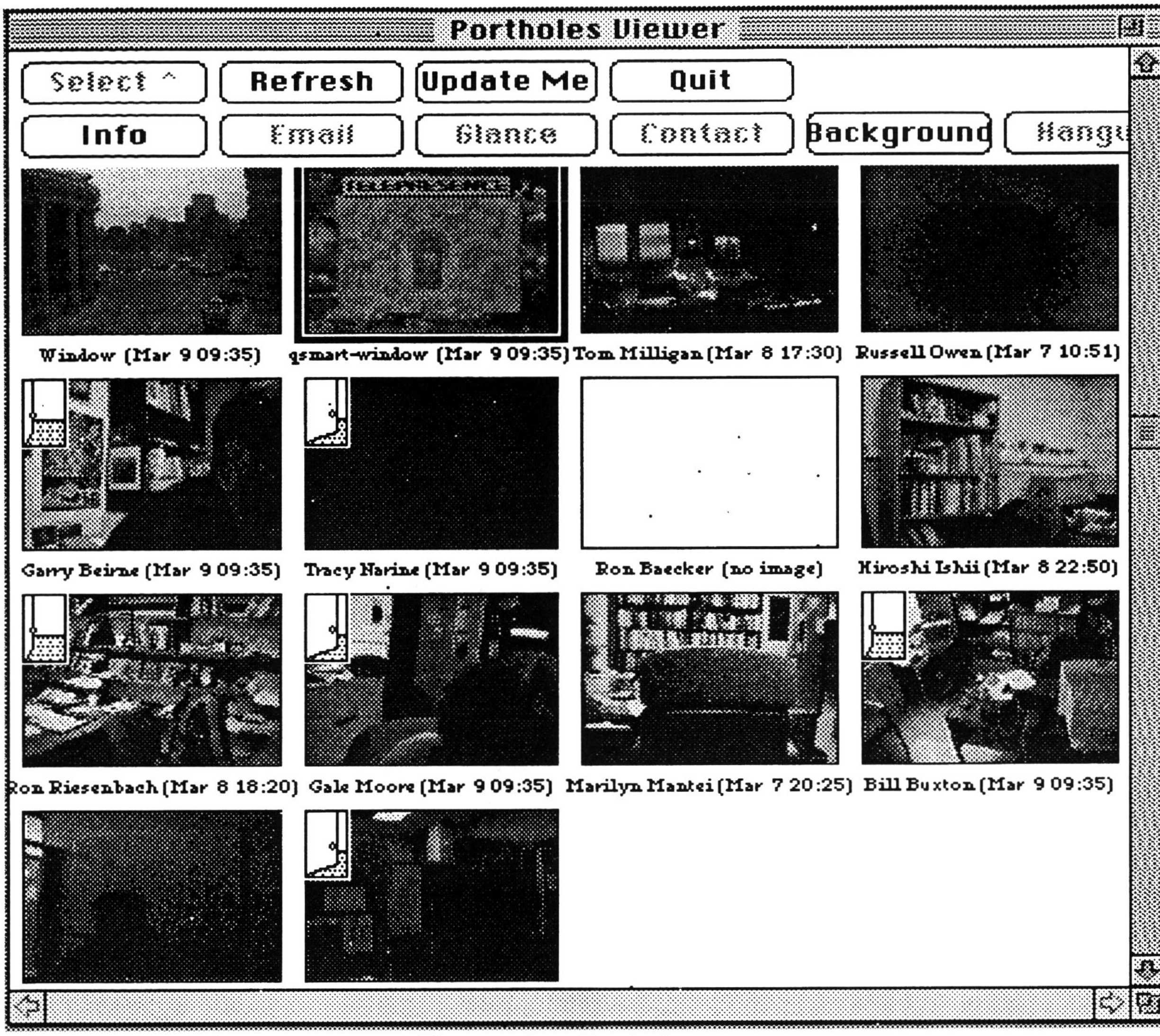

Figure 2: Portholes

Let us illustrate this point with an example that will relate to a problem familiar to many: trying to arrange a conference call among a number of colleagues, all of whom are busy, hard to reach, and at different sites.

Using our tools, the user would glance at their portholes window to determine if the people seemed to be available. If so, they would use portholes itself to contact them and the problem is solved — we would have made a transition from the top right to the top left quadrant (via the bottom left, when interacting with portholes).

However, what if the more typical case were true: nobody appeared to be available. In this case we instruct a process on our machine to let us know when the parties are available. This is done by simply selecting the appropriate people by pointing at their portholes images, and selecting an operator, such as "set up videoconference when available." Moving to the bottom right quadrant, in the background - while you resume other work - the process "looks" at the incoming portholes images, looking for any changes. Through simple image processing it can detect comings and goings in the remote offices. When all parties appear to be available, the process initiates a foreground dialogue with the user, suggesting that now might

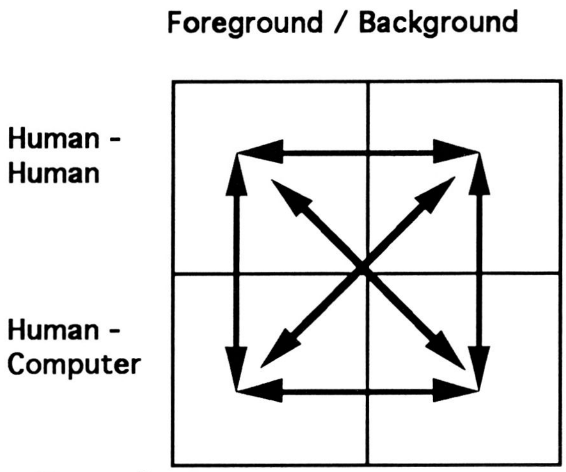

Figure 3: Seamless Transitions Between Quadrants

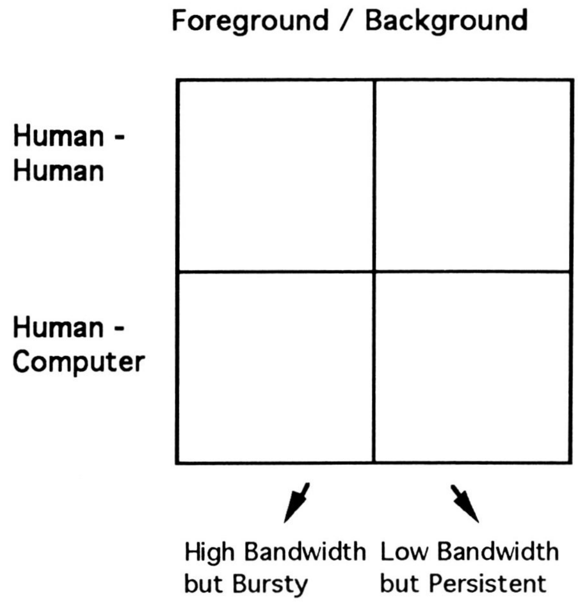

Figure 4: Bandwidth and Persistence

be an opportune time for the meeting. If so, the user initiates the meeting, and the conversation begins. In a seamless manner, one has moved counter-clockwise from the top right to the top left quadrant. High value and functionality is obtained with minimal complexity for the user. A prosthesis which makes up for many of the problems of distance is provided.

Our belief is that this is just one example of many, and that the architecture which we are pursuing affords exploring such synergies in an effective and coherent manner. As well, the program which we have described, and a number of examples which follow, are simple examples of what many call *intelligent agents* (Riecken, 1994, for a recent collection of papers on the topic). Our agenda here is less to discuss agents *per se*, than to set the work in context with other software and usage.

Telephony and Bandwidth on Demand

An interesting point of the previous scenario was that it was an example of "video on demand." Yet, this is a form of video on demand which is used rather differently than is usually discussed - that is, calling up videos. There are subtle differences that emerge from the example that make us rethink our notions about telephony. The basis for the change is highlighted in Figure 4.

What we have added here are labels that characterize the bandwidth of the two columns. It is our claim that generally speaking, activities in the left column are high bandwidth, but bursty, whereas those in the right are relatively low bandwidth, but persistent. For example, a videophone call is high bandwidth, but we may only make 5 calls a day. On the other hand, distributing Portholes images is persistent, going on constantly in the background. However, the bandwidth required to distribute images is relatively low. Viewed in the context of seamlessly moving from quadrant to quadrant, what we have is a means of capturing the notion of "bandwidth on demand." Furthermore, the model which emerges from this approach is in may ways richer than those commonly used, such as video on demand.

Finally, note that there are some subtle but important economic and technical issues that emerge from this seemingly simple example. Note that a large percentage of telephone traffic consists of incomplete calls. This consumes switch capacity and bandwidth, yet there is no cost-recovery mechanism (i.e., we don't get charged for incomplete calls, such as when nobody is home, or the line is busy). But notice two things from the example. First, mechanisms like Portholes and other background processes have the potential to greatly reduce such non-billable traffic. Second, these very same processes are of value to the consumer. Hence, they not only have the potential to reduce non-billable consumption of resources, they are of sufficient value that the user will be willing to pay for them.

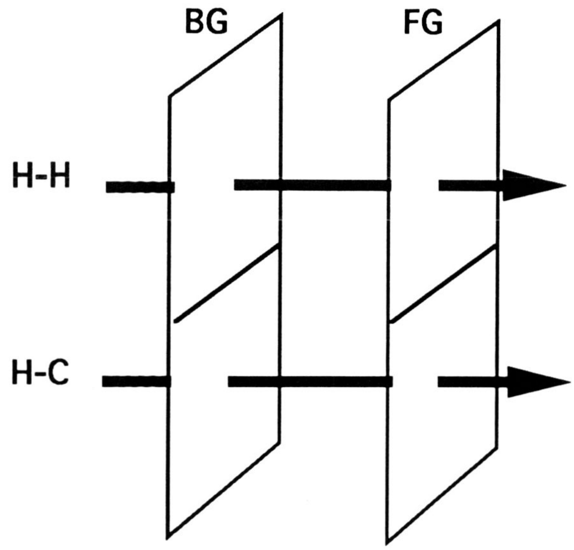

Figure 5: Context-Sensitive Interaction

Context-Sensitive Interaction

Up until now, we have addressed transitions *between* the foreground and background. We now progress to consider the case where foreground and background work in concert. In human-human communication, this is the norm. For example, your reaction to my calling out "Duck!" will differ if we are hunting in a marsh with shotguns, versus standing on a golf course fairway. Our argument is that a significant amount of the complexity in humans dealing with technology is due to having to explicitly maintain state (or context) as a foreground activity. This is in contrast with most everyday communication, where state maintenance is in the background, or periphery. Our argument is that we can significantly reduce the complexity of working in the emerging technological world if, likewise, maintenance of state is pushed into the background. This situation is illustrated in Fig. 5.

Walking into a room and having the lights automatically go on is a simple example of how complexity can be reduced by background sensing of the environment of action. While the person is responsible for the foreground action (walking into a room), the responsibility (load) associated with the background action (turning the lights on) is assumed by the system and the associated motion sensor driven switch.

While turning on the lights in a room is normally a fairly low-load task, consider the following analogous task from the world of portable computing. I have three different offices, each with its own network and laser printer. When I plug in, my laptop computer, it has no idea which network it is connected to. Consequently, if I want to print a document, I must explicitly specify the laser printer for that site. This is over head that the user should not have to deal with. It should be taken care of in the background by the system itself.

Photography offers another example – one which leads us to a practical application of this approach to design. As a start, consider my first camera. It had two controls: the button to snap a picture and the handle for advancing the film. While rather inflexible, it was easy to use. It was "point and shoot," and then roll the film. My next camera was a marvel of technology. I could adjust it in any number of ways and get virtually any effect that I could imagine. The only problem was, without a Ph.D. in "Camera Arts", the likelihood of getting anything usable, much less what I wanted, was low. My current camera, however, has the best of both worlds. It has all of the controls that an expert could dream of. But for mere mortals, good – even excellent – results can be obtained by the simple old "point-and-shoot" approach. And you don't even have to remember to roll the film yourself.

All of this is achieved because, in the background, the camera senses all kinds of things: what the subject is, its distance from the camera and the illumination of the subject and the surround. All of this is integrated with knowledge about the kind of film so that the camera can automatically set its configuration to a state appropriate for the foreground action of the photographer. Like the light example, the only cognitive burden on the user is that imposed by the foreground action. Context-sensitive system state is maintained automatically the background.

Now consider the increasingly common task of scanning documents into computers. Does this not have a strong resemblance to my experience with my second camera? The task normally involves setting a relatively large number of controls, and several attempts are generally required before achieving acceptable results. This is especially true when dealing with complex documents, such as those consisting of a colour photograph, text and artwork all on the same page.

Here is a prime candidate for applying precisely the same kind of automation that we see in modern cameras. Like the camera, a "smart" scanner is able to sense relevant properties of its "subject" and adapt its processing accordingly. In this case, however, it is going to base its decisions on a knowledge of document morphology and properties of the page being scanned. But the process is essentially the same. The user just places the document in the scanner ("points") and pushes as single button ("shoots"). All of the parameters that determine the state of the scanner are set automatically, thereby reducing the cognitive burden of system operation to that of the simplest copier.

And what does the user get for this single button push? The contrast is set automatically and adaptively on different parts of the page, the colour photo on the hypothetical page is scanned at 24 bits/pixel, the gray-scale image at 8 bits/pixel, and the line art and text at 1 bit/pixel. This alone results in higher quality, less complexity and significant reduction in file size. (Remember, the norm in this example would be to scan the whole page at the "worst case" condition: 24 bits/pixel overall.) But we get even more. Having isolated the different morphological parts of the page, we can compress each with the algorithm most suited to the type. Hence, quality and storage are even further optimized automatically.

As our last example, let us push to a scenario a bit further from mainstream computation. You are in a videoconference with a colleague. You each have the requisite camera, monitor, microphone and speaker, all of which are connected to your coder/decoder (CODEC). This is illustrated in Figure 6.[1]

Now consider what would be required if in the midst of your conversation, it was suggested that you record a part of the meeting. Assuming that you have a VCR, this should be fairly simple, shouldn't it? After all, you know how to put a tape into a VCR and hit record. But it is not. The wiring plan is shown in Figure 7. This makes working a manual camera seem simple by comparison!

In order to record the conference, both the incoming and outgoing video signals need to be combined using a "picture-in-picture" (PIP device) before being fed into the VCR. Likewise, both the incoming and outgoing audio signals must be mixed together before recording. Assuming all of this was cabled together correctly and the conference recorded successfully. Now consider what would happen if your remote colleague asked to see the recorded segment. This would require a complete reconfiguration of the A/V gear, in a manner such as shown in Figure 8.

This configuration lets both sides hear and see the video, as well as talk to each other over the soundtrack, while it is playing. It also lets me, the presenter, see you in a small window on the screen while you watch the tape.

The first observation is that working at the cable level, a fair amount of knowledge is required to set up any of these configurations. While more common than it should, users should never have to work at this level. Consequently, current design practice would be to have some kind of preset button for each configuration. In this case, recording or playing back a tape would be a two step operation: selecting the appropriate preset, then operating the VCR. A variation would be to have the preset also control the VCR, thereby reducing the transaction to one step.

The third approach is the one which most follows from our context sensitive approach. In this case, the user just puts the tape in the VCR and hits the "record" or "playback" buttons. The state of the VCR is monitored by the computer, as is the state of the room (i.e., the fact that the user is in a videoconference). Consequently, the computer automatically reconfigures the A/V network in the background, based on the system state, or context, in which the user's foreground action takes place.

There are some subtle but important distinctions between this third solution and the more conventional second approach. This is not a computer-controlled VCR. Rather, it is a VCR controlled computer. Users interacts directly with the components in the workplace with which they are familiar. In this case, it is a VCR. The interface is decentralized and builds upon existing skills. By decentralizing, we move away from the "super appliance" approach which is prevalent today, in which all of our interactions with the electronic domain are channeled through one of two overworked and less than appropriate appliances: the television and the computer. This centralized approach will not scale up as the range of applications expands.

[1] If you do not understand the diagrams in the next three figures, do not worry. You should not have to deal with this level of detail and complexity to fulfill simple reasonable desires or requests of a system. The figures highlight the problem. Having all of this taken care of in the background is the solution.

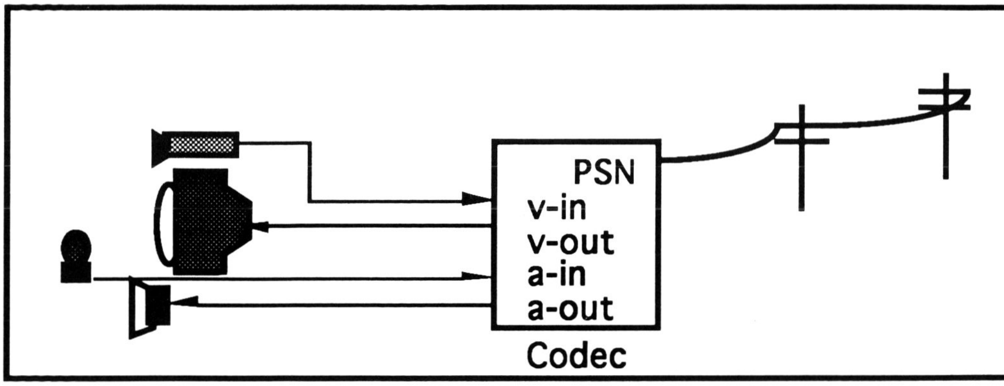

Figure 6: A Simple Videoconferencing Configuration

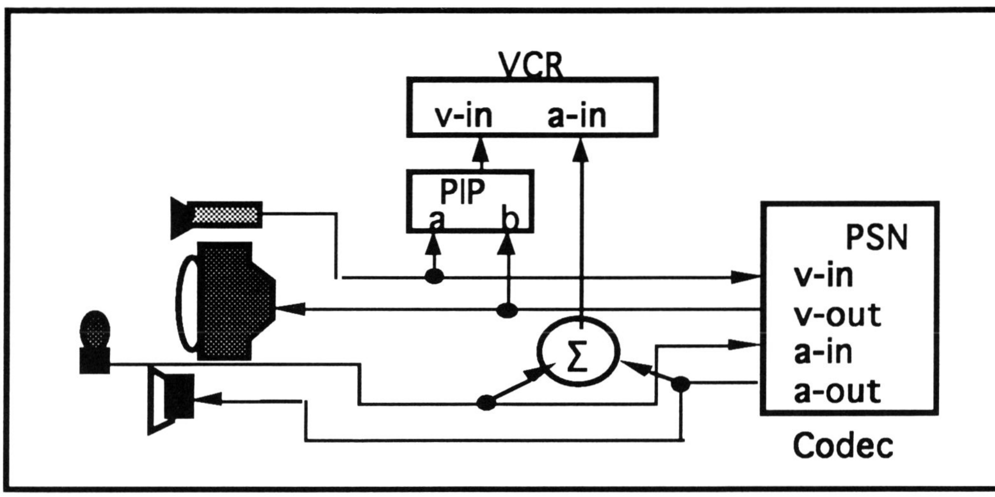

Figure 7: Recording a Video Conference

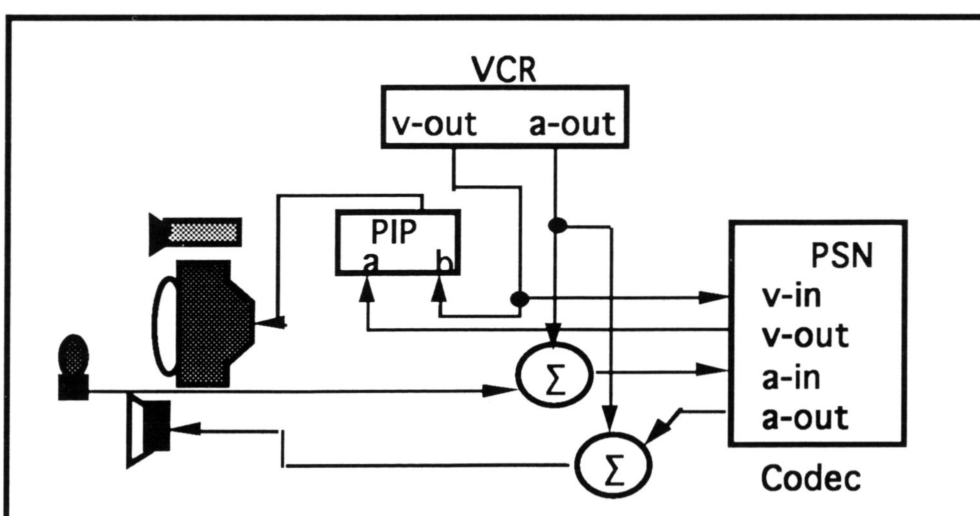

Figure 8: Playing back a tape in a videoconference

Systems that do make use of context and knowledge of the domain to reduce the load on the user are now practical and are beginning to appear commercially. *Easy Scan*, a new system from Xerox Corp. has precisely the properties described in the scanning example, above. Work in our lab, described in Cooperstock, Tanikoshi, Beirne, Narine and Buxton (1995) addresses the issues discussed in the conference room example. Our belief is that this class of architecture will play an increasing role in future systems design. Our hope is that our arguments and examples will accelerate the process whereby this comes to pass.

Summary and Conclusions

From the telecommunications perspective, what we have introduced is *a usage based model* which argues strongly that a traditional telephony model (i.e., foreground calls, video conferencing, etc.) is not adequate to support *telepresence* applications (including telework, distance education, etc.). At the same time, it has presented some structure around which we can organize our thinking about such matters, and assist in our planning.

What is equally important is the point that this model has emerged from a methodology based on placing the emphasis on *usage* not technology. As such, it provides some motivation to redistribute resources in such a way as to put more emphasis in this area in the future.

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.

Cooperstock, J.R., Tanikoshi, K., Beirne, G., Narine, T., Buxton, W. (1995). Evolution of a Reactive Environment. To appear in the *Proceedings of CHI'95*, May, 1995.

Dourish, P. & Bly, S. (1992). Portholes: Supporting Awareness in a Distributed Work Group. *Proceedings of CHI'92*, 541-547.

Elrod, S., Hall, G., Costanza, R., Dixon, M. & Des Rivieres, J. (1993) Responsive office environments. *Communications of the ACM*, 36(7), 84-85.

Gaver, W. (1992). The affordances of Mediaspaces for collaboration, *Proceedings of CSCW'92*, 17-24.

Kraut, R. & Egido, C. (1988). Patterns in contact and communication in scientific collaboration. *Proceedings of CSCW '88*, 1-12.

Moran, T.(ed.)(1994). Special issue on context in design, *Human-Computer Interaction*, 9(1).

Riecken, D.(ed.)(1994). Special issue: Intelligent agents, *Communications of the ACM*, 37(7), July 1994.

Whittaker, S., Frohlich, D. & Daly-Jones, O. (1994). Informal workplace communication: what is it like and how might we support it?, *Proceedings of CHI'94*, 131-137.