The DrawStream Station
or The AVCs of Video Cocktail Napkins

Steve Harrison
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, California 94304 USA
harrison@parc.xerox.com

Scott Minneman
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, California 94304 USA
minneman@parc.xerox.com

Joshua Marinacci
Realm Information Technologies
5555 Oakbrook Parkway
Norwood, Georgia 30093 USA
joshy@mindspring.com

Abstract

This paper reports on the development of a multimedia system to support collaborative design processes. The DrawStream Station was developed through a cycle of observation of real work, identification of problem characteristics and system requirements, development of the technology, and observation of the technology in use. The system is a combination of a high quality multimedia infrastructure and a particular form of "Space Above / Space Below" interface merging real and virtual representations. The control interface mixes a variety of temporal and logical representations to display a threaded conversation. We report on the use of the system as a way to understand the dynamics of genre systems. Two nascent genres are identified: Asynchronous Video Conversations and Video Cocktail Napkins.

1. Introduction

We develop new technology through a cycle of observation of real work, identification of problem space, establishment of system requirements, development of the technology, and observation of the technology in use [3, 6]. Observing media in use means observing the kinds of uses it gets put to, that is, observing *genres*.

Six or seven years ago, no one had heard of "homepages," now the term is understood by even casual computer users. The media of the internet has led to the creation of new forms of content. This social process of the emergence of genres seems to happen in most every media. The development of media therefore implies the seeding of new genres.

As developers of new media and new media applications, we are curious about the process and are looking at the possibilities that these socio-technical systems can be, to some degree, designed. We will report on one new media project, the DrawStream Station (DS S), and its use, as a way to understand the dynamics of genre systems.

First, lets look at the design of the technology, starting with a few observations about the activities of the application area, architectural and engineering design.

2. The Problem Space: Process Ephemera

The DrawStream Station was developed with an eye toward supporting salient aspects of design: primarily, supporting the large quantities of *process ephemera* and conversations that employ process ephemera.

The majority of documents used during a design process are "process ephemera." *Process ephemera* are representations, in a heterogeny of media, that come and go over the course of a project.[1] At one moment, designers might be working with sketches and a rough model, at another, it might be product samples and presentation drawings. While much work is done on the drafting board and computer, some of the material is around the room: drawings are piled on desks and in flat files, sketches and photos may be tacked to walls, and copies of specifications from old projects often lie piled within easy reach [3].

These materials are ephemeral in that they are useful in the moment. A classic example is the cocktail napkin sketch—not necessarily much to look at, but possibly pivotal for illustrating some elusive point in the course of a conversation. One often-missed detail of this example is the extent to which the hasty marks on that flimsy paper are a small fragment of a larger whole. The context of that drawing—the place where the sketch was made, the conversation around it, the gestures over it, the facial expressions about it—are where the real value lies. In fact, these sketches are often meaningless devoid of that conversational context [11].

The materials are also ephemeral in their relation to the project and to other representations in it: a couple of days might be spent putting together a study model, referring to sketches and early plan layouts. When complete, it might be looked at intensely, then tacked up on the wall along with photos cut out of magazines and photographs of the site. At a presentation to the client, someone might grab it again to explain some point that a drawing doesn't effectively convey.

2.1 System Requirements

How might we include these ephemera when designers are remote from one another in space or time? How can we integrate ephemera from the physical world with those of the virtual world? The system must provide a means to include the ephemera in remote collaborations, to record the interactions with the ephemera and conversations when people are talking about the ephemera, to index to these key events, and to replay the recordings so that they can be built upon in the same way that the ephemera are built upon over the course of design work.

[1] Process ephemera are an aspect of the documents, not a genre in themselves. The various forms of these ephemera, on the other hand—sketches, models, and the like—*are* genres found in design work, but they too are *constituents* (genres nest) of the conversational phenomena we are centrally concerned with here.

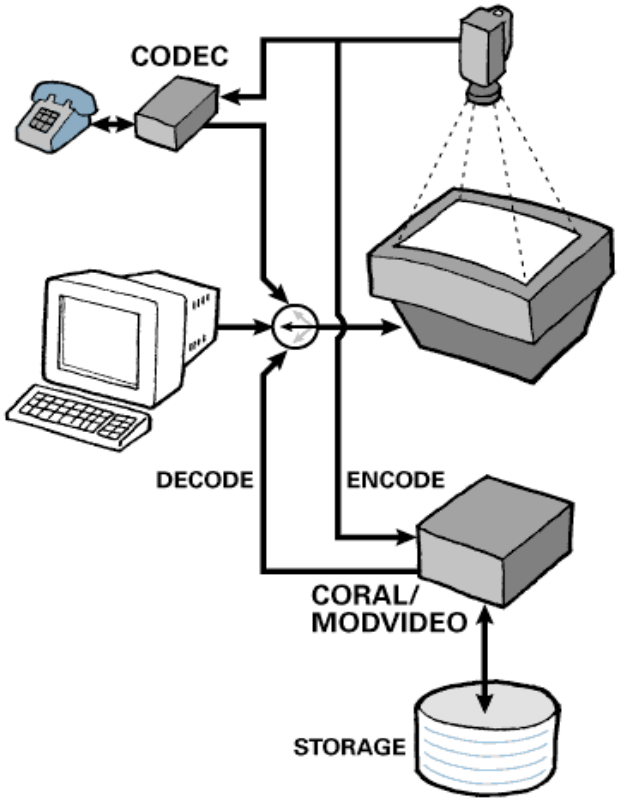

Figure 1: The DrawStream Station. The user's work surface is the region between the camera and the upward facing digital video monitor.

3. The System

To address these requirements, the DrawStream Station consists of a unique camera and monitor work surface and a high-quality indexed multimedia infiastructure. The station is a downward facing camera looking at an upward facing monitor. The upward facing monitor is the user's work surface (see Figure 1).

3.1 SASB Interfaces

The camera/monitor pair is a Space Above / Space Below" (or "SASB" interface). *SASB's are electronically connected work surfaces that augment conversations by supporting conversational hand gestures. Images that represent real objects appear approximately life-size.*

Real and virtual objects and actions must work seamlessly together; they do so by having the *real space above* and the *virtual space below* have a relationship to one another. The DS S is not the first to use a SASB interface. Its design is informed by a number of SASB's that have addressed a variety of problem spaces, most notably, remote collaboration and moving between real and virtual worlds.

The DrawStream Station is based upon the VideoDraw remote collaboration SASB. VideoDraw [12] used downward facing cameras cross-connected to upward facing monitors to create a shared work surface which presents both gestures and sketches to multiple users.

Ishii's ClearBoard [4] extended the VideoDraw idea by including the facial view of users. This optimized communication channels and reduced the number of cameras and monitors needed to simulate face-to-face communication over a shared works surface. In both VideoDraw and ClearBoard, both the representations and the hands that animate them are carried into the surface of the display; while VideoDraw supports the use of tracing paper on its surface, it would occlude the image of users in ClearBoard.

The DigitalDesk [13] is a SASB computer interface to allow work on and with paper at desks. A downward facing video camera is coupled with a video projector. Paper objects can be brought into play, elements extracted into computing systems, and virtual elements projected onto the paper surface to create a buffer between the two worlds.

The DigitalDesk has (in part) been the progenitor for a few shared computational work surface projects that try to bridge between computation and the physical world. Working with Fitzmaurice and Buxton at the University of Toronto, Ishii built the "Active Desk" as part of their work on graspable user interfaces[2] [2]. Apple ATG did something similar that featured the ability to rotate images for parties standing in different orientations[3]; and most recently, GMD/Darmstadt's InteracTable features the "Passage" device which, like Ishii's Bricks (and more recently, his Triangles), uses abstract physical icons for virtual documents [10]. All of these ideas share the notion that there is some representation "below" that is manipulated by hands working on real things "above" and that the appearances in both realms are related in scale and dynamics.

3.2 Multimedia Collaborative Systems

The other part of the DS S is a facility that can distribute, store, index and retrieve the streams that represent the activities at a SASB interface. Video has, from its infancy, been seen as a platform for shared activity across space and more recently, across time. Schooler [9], among others, lays out the space of collaborative tools; video systems are the obvious first choice to support the interface.

The problem with video has always been to align the visual fields of the medium with the visual space of activities. Issacs and Tang [5] focus on the difficulties of making that leap. In this, the application and the multimedia infiastructure need close coupling.

PARC's Coral multi-media infiastructure does the distributed indexing and our ModVideo digital video infiastructure stores and replays the video segments. This infiastructure has been reported previously [7, 8]. Briefly, ModVideo is a motion JPEG real time encoder and decoder that uses RTP and IP to transport the streams over an ATM network. This is a network appliance that produces digital video streams of a high enough quality that fine detail of small sketches and quick gestures are captured and

[2] Since interacting with (or through) these systems does not depend on holding or manipulating an object, graspable user interfaces are a kind of manipulator for SASBs., but not the same thing as a SASB.

[3] While well-known in the field, the project was not reported in publicly-accessible documentation.

reproduced accurately. ModVideo also handles the CD quality audio necessary to support the conversational part of the system.

3.3 The Control Interface

The control interface mixes a variety of temporal and logical representations to display a threaded conversation. The design permits the AVCs to be added to and be replayed separately from the DrawStream Station. Currently, the control interface is outboard from the DrawStream Station on a separate workstation.

In addition to recording and playback controls, the interface organizes conversational fragments (see Figure 2). The dominant (topmost) bar in the interface is the first recording. Responses to points made in it are represented as the "flags" hanging off of it. This is not a traditional timeline: whilst bar length corresponds to segment duration, it does not represent *when* it was created. Playback can traverse the commentary tree so that all opinions can be heard and layers of images can be seen in context. Playback can also be by time sequence, speaker, or topic.

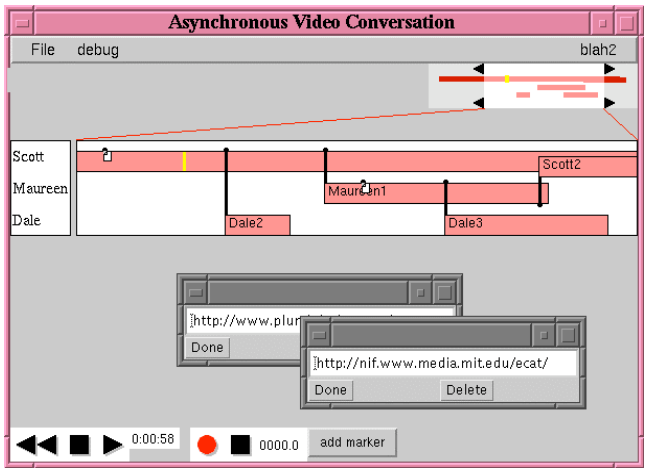

Figure 2: Control Interface window.

Also visible in this example are some textual note markers and a "radar" view of the commentary tree. The textual notes are placed, like the commentary flags, on a segment of video and pop-up when clicked on or when playback crosses their reference marker. In the figure, the notes are URLs that the participants were discussing. The radar view in the top right selects the viewable portion of the body of recorded material.

3.4 Surface Sources

The DrawStream Station sits in a larger context; users can gesture and mark over images from a variety of sources. The electronic work surface can *display a remote user in real time* (that is, *be* a shared work surface), *display a computational screen* (such as a web browser or a CAD system), or *re-play recorded digital video material* (see Figure 1). Note that while the work surface can play out images from various sources, the recording is only made of the appearance of the work surface and objects above it.[4]

[4] Coral is capable of managing recording and playback of many types of time-based streams. It is possible for us to record more streams than the video and audio (e.g., direct capture of the computational system or replayable interactions with applications as was reported in [Minneman, et al., 1995]). As we acquire more experience with the use of DSS, we may fold in more of the capabilities of Coral, but for now, we are keeping things simple.

4. In Use: PARC / MIT Design Studio

To better explain how the DrawStream Station works, let us take a look at the observational study of its use that closed this round of the development cycle. The DS S was used in a one semester architectural design class at MIT, partly taught by the authors, working from their Lab at PARC. This setting was selected because architectural school projects have many of the same traits as commercial design practice. It was also chosen since it is a relatively forgiving setting (as opposed to a commercial setting that *must* be operational 24 hours a day, 7 days a week) and part of the MIT educational agenda is to prepare students for future practice.

The DrawStream Station was part of a suite of tools being used in the course and needed to interoperate with other media: video teleconferencing, CAD systems, and the paper ephemera of the studio. The meetings were presentations, discussions, and most importantly, design crits.

Crits are the central pedagogy of architectural education. The crit is both feedback about the design and designing with the student. In this classic master-pupil situation, a critic (usually the course faculty member) will sit with the student, review progress on assigned design projects, critique presented work, suggest alternatives that the student may not have explored, and even insert him/herself into the process, drawing over the student work

4.1 Remote Desk Crits

During the project, DrawStream Station was used in myriad ways in design crits. It worked with the existing process ephemera of the studio and created its own process ephemera in the form of recordings and indices; this new media ephemera were themselves new forms (more on the new genres appears in Section 5).

The source images displayed on the surface of the DrawStream Station were of models, sketches, and renderings (and, occasionally, faces). The images came from three sources: the *transmitted* video image from MIT, a computer screen that *was scan converted* (and zoom/pan capable) for display on the DrawStream Station, or *replayed* from the ModVideo digital video store.

PARC critics gestured over the images so that the MIT students could understand what was being referred to and get feedback about how well the critics understood the idea. PARC users could also offer suggestions about modifying the design or how the students might alter the presentation in progress so as to better convey whatever aspect of the design under discussion.

PARC critics drew over the images so that explicit changes could be described and to reinforce gestures made over the images. Drawing was done by both *marking directly on the screen surface* using erasable markers and by using *tracing paper unrolled on the surface*. The latter is a traditional medium in architecture; it partially obscures the underlying image and highlights the new marks made on it. As in the traditional paper-over-paper method, it also permits removal and repositioning.

Just as it is possible to gesture and mark on present images, it is also possible to gesture at and mark on past ones. While the DrawStream Station supports recording of marking and gesturing

over previous recordings of gesturing and marking, the studio project did not present many opportunities to use this multi-layered capability.

4.2 An Example

Let us now follow the cycle of use from the setting in which a recording is made to the place it was replayed. We will look at the recording and re-use of two fragments of conversations.

Late in the project, one team is wrestling with the scope of their project and how its components fit together (see Figure 3). Their project is a modular corridor for the automatic display of information. This group works occasionally together, occasionally apart.

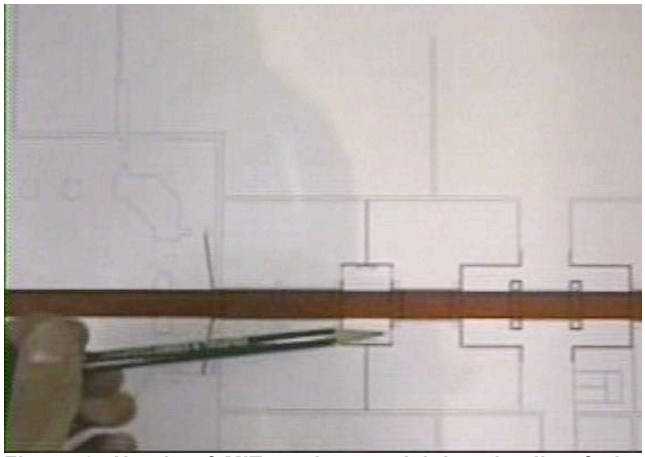

Figure 3: Hands of MIT student explaining details of plan diagram of new hallway.

"We have up here the dome at Lobby 7. Here is the beginning of the Infinite Corridor, right here. So, essentially, this long access that we're running And at different locations we have these cross-corridors coming in."

Figure 4: Later in same conversation, talking about a model.

"So this being our sort of introduced tube-thing and we're looking at this potential cross-corridor location. What we're looking at is providing moments that bump-out from the main corridor here. By looking at these rib elements which fit within this tray, these panel pieces right here which then plug in. This begins to setup a rhythm. We're looking at this in several ways: one is contructability and in terms of system distribution where we saw the ribs providing structural support as well as service supporting. So in this instance, power would be fed into the ribs and then be distributed to the panels so that the ribs were critical to holding it up as well providing necessary services."

At this meeting, one of the three students on this team meets with both PARC critics. Other members of the team are not present in the studio and therefore, cannot overhear the discussion. He talks through the scheme for the modular corridor using a plan drawing of the space and a variety of large corrugated cardboard models. During the conversation he focuses attention of the critic by gesturing over the plan drawings with his hands and a pencil and by animating various pieces of the model (see Figure 4).

Figure 5: talking with second student, hands of PARC critic gesturing over image from previous session.

A few days later, another member of the same project, who had participated in the previous discussion, had some related concerns. Again, there is only one of the team present in the studio. Recordings from the previous discussion with the other student were replayed by the PARC critics and were visible at both ends of the link (see Figure 5). This led to a discussion of the critics' understandings of what had been presented. In the figure, note the use of the fingers to express an interlocking feature; this gesture reproduces an action shown with the models.

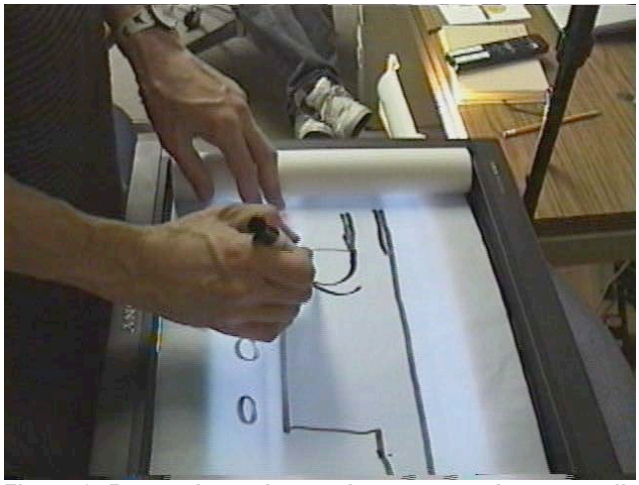

Figure 6: Researcher using tracing paper to draw over discussion of plan diagram of new hallway replayed from digital file store.

This led to a discussion of this particular team member's understanding of her teammate's scheme for fitting the new modules into the existing structure. The image showing the plan was recalled and the suggested approach drawn on tracing paper layered onto the DS S (see Figure 6).

Critic: "Could you back that up? So, if I understand this, this part and this part are going to be much simpler somehow. How is this going to be simpler?"

Student: "move down that thickness."

Critic: "So this is somehow thinner?"

Student: "Thinner, right. So that red line that goes down that plan will actually be imbedded."

Critic: "Imbedded how?"

Student: "the existing floor."

Critic: "Oh, I missed that part."

5. New Genres

Observing this novel medium in this setting we noted the emergence of two new genres.

Genres are intrinsically socio-technical systems. That is, the conventions of a genre are understood by both authors and readers. For example, in the genre of newspapers, the stories on the first page above the fold are understood to be the most important. Knowing this is important to the editors, authors, layout artists, and readers of newspapers, although the degree to which those populations are conscious of the properties of genre varies widely (with readers' knowledge probably being most tacit).

5.1 Video Cocktail Napkins

The cocktail napkin has long been used as a metaphor for a kind of casual creative conversation, usually around sketching. The multimedia recordings in this project became a kind of cocktail napkin.

There are key elements of the genre of a video cocktail napkin: a recording made of a conversation over the playback of another video stream, the source video is common ground in the conversation, the content of the source is based on shared experience of the parties to the conversation, the content depends on there being conversation about it, and the conversation depends on its presence. This is a DrawStream-independent description; a video cocktail napkin on the DrawStream Station has the added element of the content being communicated by the inclusion of hand gestures, hand-sized objects, and marking.

In the example of use, we saw how the playback of the first student's comments to the second were used to elicit an explanation of the changes that had occurred in their team's design. Both the critic and the student knew the context of original fragment, although only one of them had been present during its recording. The fragment is unintelligible to outsiders—and certainly devoid of the richness that the parties read into it.

There is a possible confusion in that the DS S station was used itself for sketching like a cocktail napkin; the point to this is not that the video could be sketched on, but that the playback was talked over and explicated through the conversation.

5.2 Asynchronous Video Conversations

The other genre takes advantage of the DrawStream Station's indexing of recordings to create *asynchronous video conversations (AVCs)*. AVCs are a sort of multimedia threaded conversation. By juxtaposing fragments of conversations created in everyday work situations, a rich interplay of discourse can be created that spans a number of temporal settings. AVCs are a process manager for the improvisational acts around process ephemera.

This genre is a sequence of video cocktail napkins where one or more parties to the conversation are not temporally present. There are 3 key elements of the genre of AVCs: *the video is common ground, conversation depends on the presence of other video recordings to give it context,* and *the content is based on shared experience.* .As in the previous genre, an AVC on the DS S adds the *content being communicated by the inclusion of hand gestures, hand-sized objects, and marking.*

AVCs differ from threaded textual discussions in a number of ways. The primary difference is that the form and content more rigidly directs the subsequent direction of the conversation. The form—indexed audio and/or video records—must be experienced in some temporal form and is therefore, more experientially sequential than blocks of text. The content—images which can be layered—provides the basis for subsequent dialog and therefore comes as built-in context. For a more complete discussion, see Churchill and Minneman [1].

6. Conclusions

Our conclusions fall into two areas: those having to do with multimedia systems, and those that concern the emergent genres.

6.1 Design of Multimedia Systems

The primary observation of the design of the DS S is that the SASB interface directly affects the content of the multimedia and the audio and video quality requirements. In turn, the overlaying quality of the content determines the fundamental requirements for the indexing structure and the control interface. That is, *the content largely drives the form of the technology*.

6.2 Genres and Document Types

At a different level, the DS S demonstrates how the workspace of process ephemera can be turned into a piece of process ephemera. Video cocktail napkins and asynchronous video conversations are both process ephemera and a record of process ephemera. That is, they are more than props in the play of events; they are explicit representations of conversations that form the events. The next step in the design of the underlying media is, then, an understanding of the distinction between the two.

The two document types we report on here are the result of the authors' long research into collaborative tools for distributed design. However, they need to be out in the world so that the social system around the particular genre can be further understood. We are confident that this exploratory method provides a fruitful point of departure for refining the medium and a illustrates a likely first set of uses that will arise when the system is deployed in other places.

7. References

1. Churchill, E. and Minneman, S., "What Was That Again? Conversations Using Asynchronous Multimedia Discourses," FX/PAL Technical Report, (1998).
2. Fitzmaurice, G., Ishii, H., and Buxton, W. "Bricks: Laying the Foundations for Graspable User Interfaces" in *Human Factors in Computing Systems. The Proceedings of CHI '95*. Addison-Wesley Publishing, Reading, MA (1995) pp. 442-449.
3. Harrison, S. and Minneman, S., "Studying Collaborative Design to Build Design Tools" in M. Tan and R. The (eds) *The Global Design Studio. Proceedings of the Sixth International Conference on Computer-Aided Architectural Design*. Center for Advanced Studies in Architecture, National University of Singapore (1995) pp 687-698.
4. Ishii, H., Kobayshi, M., and Grudin, J., "Integration of Interpersonal Space and Shared Workspace: Clearboard Design and Experiments" in *ACM Transactions On Information Systems*, Association for Computing Machinery, New York. vol 11, no. 4, (1993) pp. 349-375.
5. Issacs, E and Tang, J, E. What Video Can and Cannot Do for Collaboration: a Case Study" in *Multimedia Systems*. vol. 2, no. 2, Springer International Publishing, Heidelberg (1996) pp. 63-73.
6. Minneman, S L *The Social Construction of a Technical Reality: empirical studies of group engineering design practice*, Xerox PARC Technical Report #SSL-91-22 (Ph.D. Dissertation, Stanford University), Palo Alto, CA (1991).
7. Minneman, S., Harrison, S., Janssen, W., Kurtenbach, G., Moran, T., and Smith, I. A confederation of tools for capturing and accessing collaborative activity. *Proceedings of ACM Multimedia 95* , November 5-9. Reading, MA: Addison-Wesley, (1995) pp 523-534.
8. Moran, T., Chiu, P., Harrison, S., Kurtenbach, G., Minneman, S., and van Melle, W., Evolutionary Engagement in an Ongoing Collaborative Work Process: A Case Study. *Proceedings of ACM CSCW 96*. November 17-20,. Reading, MA: Addison-Wesley (1996) pp 150-159.
9. Schooler, E. "Conferencing and Collaborative Computing" in *Multimedia Systems*. Vol. 4, number 5. Springer International Publishing, Heidelberg (1996) pp 210-225.
10. Streitz, NA, Geibler J, and Holmer, T, "Roomware for Cooperative Buildings: Integrated Design of Architectural Spaces and Information Spaces" in *Cooperative Buildings – Integrating Information, Organization, and Architecture. Proceedings of CoBuild98*. Lecture notes in Computer Science, Vol 1370. Springer: Heidelberg, (1998) pp 4-21.
11. Tang, J. C., *Listing, Drawing, and Gesturing in Design: A Study of the Use of Shared Workspaces by Design Teams*, Xerox PARC Technical Report #SSL-89-3 (Ph.D. Dissertation, Stanford University), Palo Alto, CA (1989).
12. Tang, J. C., and Minneman, S. L., "VideoDraw: A video interface for collaborative drawing" in *Proceedings of the CHI 90 Conference on Human Factors in Computing Systems,* Seattle, WA (1990).
13. Wellner, P., "Interacting with Paper on the DigitalDesk" in Communications of the ACM. Association for Computing Machinery, New York. vol. 36, no. 7 (1993). pp. 86-89.