

Studying video-based collaboration in context: From small workgroups to large organizations

Ellen A. Isaacs and John C. Tang

SunSoft, Inc., 2550 Garcia Ave., Mountain View, CA 94043
ellen.isaacs@sun.com; john.tang@sun.com

To be published in: K. Finn, A. Sellen, & S. Wilbur, *Video-Mediated Communication*, Mahwah, NJ: Lawrence Erlbaum.

Abstract

Over the past few years, our group has developed and studied a variety of video-based prototypes to support remote collaboration. We started with a basic desktop video conferencing system with a shared whiteboard and found that it was effective in supporting a distributed group's conversations, but that many of their attempts to contact each other were unsuccessful. We went on to build another desktop video conferencing prototype called Montage, which focuses on awareness and helps people find good times to interact by integrating other coordination and collaboration tools. We also built Forum, which allows people to give live, interactive video-based presentations to distributed audiences. We describe the methods we used to study the long-term use of these prototypes among existing distributed groups and the lessons we learned by doing so. Finally, we discuss the roles that video plays in supporting distributed collaboration and we enumerate a list of design principles for those who wish to use video to support such activity.

Keywords: video-mediated communication, desktop video conferencing, awareness, distributed presentations, distance learning, distributed collaboration, use studies, design methodology.

Introduction

Lydia is developing some marketing materials for an upcoming customer presentation and she decides to show her latest idea to her colleague Sonia for feedback. From her computer workstation, she initiates a “glance” into Sonia’s office, 500 miles away. As the image fades in, she sees that Sonia is away from her desk. She checks Sonia’s on-line calendar and sees that she is in a meeting for another hour, so she leaves a Stickup “electronic note” asking Sonia to glance her back when she gets in. A few minutes later, a notice pops up on her screen indicating that the CEO’s quarterly talk is about to begin. She “attends” the talk by opening an application, and she is presented with a video window of the CEO, who will be broadcasting from the corporate headquarters, along with the presentation slides and a list of others starting to join. After five minutes, over 600 people have joined the talk. Lydia looks through the slides to preview the CEO’s remarks. She sees that the CEO will not address certain issues that are important to her. She decides she will ask about it if no one else does, though she will probably keep the question anonymous. Perusing through the list of names and faces, Lydia notices that her colleague Bjorn is attending and she sends him a note asking if he wants to have lunch after the talk. He replies quickly saying he’s free and they make arrangements.

This scenario illustrates the range of tasks and group sizes our Collaborative Computing group has been trying to support through the use of video. For example, the “glance” tool (which we call Montage) supports one-on-one interactions and attempts to help people find opportune times to interact. It is integrated with other on-line communication and coordination tools (e.g. calendar and Stickup electronic notes) to support the “pre-interaction coordination” that is often needed to set up an interaction. The video presentation tool (Forum) supports one-to-many

presentations among potentially hundreds of people. It also supports questions from the audience and passing notes among the audience to retain some of the interactive richness that makes presentations collaborative events.

Studying Collaboration

Over the past five years, our group has developed a variety of video-based prototypes to support remote collaboration. Our approach begins by studying collaborative work activity to identify how the existing technology does and does not meet the collaborators' needs. These studies guide the design of prototypes that help address those needs and in particular enable more natural collaboration across distance. Once we have developed a functioning prototype, we deploy it into real use over a long period of time and study how people react to it. We take a combination of quantitative and qualitative measures of users' activities, and we compare the groups' behavior using the prototype to their behavior without it so that we can learn how the prototype affects their ability to work together.

Although our work shares much in common with other chapters in this book, we have made some choices that have enabled us to explore new issues. For example, nearly all the other experimental video systems aim to support small groups of people with relatively informal interactions, yet Forum supports large communities by broadcasting formal presentations. Most projects, including Cruiser at Bellcore (Fish, Kraut, Root & Rice, 1993), CAVECAT and Hydra at Toronto (Mantei et al., 1991, Sellen, 1992), the Media Spaces at PARC (Bly, et al., 1993) and at EuroPARC (Gaver, et al.), the research at University of Michigan (Olson, Olson & Meader, 1995), and others have focused on analog video, whereas we have used digital video. The latter is currently of lower quality (e.g., 4-10 frames per second, grainy resolution), but it allows us to better integrate our prototypes with other applications and to use digital video effects when appropriate. So far our work has used a "connection-based" model of communication, in that contact is explicitly initiated by one person, rather than leaving audio-video links on all the time. Other projects, such as the Media Spaces, have explored such continuous connections that enable extremely short, spontaneous interactions. Finally, while many other use studies¹

involved the developers of the system or other early adopters, we have tested our prototypes with people who were not affiliated with the prototype development to better approximate the response of non-technical users.

This chapter reviews the three video-based prototypes we have developed to support remote collaboration. We started with a basic desktop video conferencing (DVC) prototype, which gave small groups of about 5-10 people the means to see each other and share documents. Studies of the DVC prototype in use indicated that people often had problems finding opportune times to make contact. This finding prompted us to design Montage, which made it easier to start interactions or coordinate future times to interact. In parallel, we worked on Forum, which enabled people to give interactive presentations to hundreds of people watching from their workstations. The following sections describe these projects and summarize lessons learned. In each case, we describe our techniques for studying the use of the prototype, in part to demonstrate how we reached our conclusions and in part to illustrate techniques we believe are effective for studying the social implications of introducing new technology. We conclude with some lessons learned about the use of video, the design of video-based applications, and our methodology for studying collaborative activity.

Desktop Video Conferencing Prototype

Goals

We began our research on the support of remote collaboration by studying how people used existing technologies to collaborate over long distances (Tang & Isaacs, 1993). We surveyed users of our company's video teleconference rooms to learn how that system was or was not meeting the needs of remote collaborators. The biggest problem identified was availability; the teleconference rooms were often booked weeks in advance. Respondents also complained about audio quality and delay in hearing the remote collaborators. When asked which additional video conferencing capabilities they desired, users most often requested a shared drawing space.

1. We use the term "use studies" rather than the more common term "user studies" because we study the *use* of the technology in context rather than the *users* of the technology.

We also studied the work of a four-person team that was split between East and West Coast sites of the United States, about 3,000 miles apart. We studied their work activity in three settings: face-to-face meetings, video teleconference room meetings, and phone conferences. By studying videotape records of the team's activity, we found that the main problem in using the video teleconference rooms was the delay in transmitting the audio from one site to the other. This delay disrupted many mechanisms of natural conversation (e.g., interrupting a speaker, completing sentences, timing jokes). Eventually, the team elected to turn off the audio of the video teleconference system and use speakerphone audio through the phone system. They strongly preferred this arrangement because of the negligible latency in phone audio, even though it disrupted the synchrony between the audio and video (audio arrived before the accompanying video).

Observations from the survey and the preliminary use study suggested that users would benefit from a system that provided video conferences on demand, with minimal audio latency, and a shared drawing space. Thus, we designed a desktop video conferencing (DVC) prototype (Tang & Isaacs, 1993) that provided real-time audio and video links and a shared drawing program. The DVC prototype used digital video and audio transmitted over standard computer networks. This DVC prototype enabled us to explore the technical issues of using digital audio and video and introduced us to the interface and usability issues of desktop video conferencing.

Design

The user interface for establishing and managing desktop video conferences was modeled after the process of placing a phone call. The interface, shown in Figure 1, allowed users to specify the recipient of a conference request. When a request was made, a copy of the interface appeared on the screens of all the specified users, accompanied by an audio alert. A shared message area allowed users to type text messages to negotiate their entry into the conference.

Users could also select which collaborative tools would be used in the conference. The tool supported multipoint (up to three-way) conferences, but once a conference was established no one else could join.

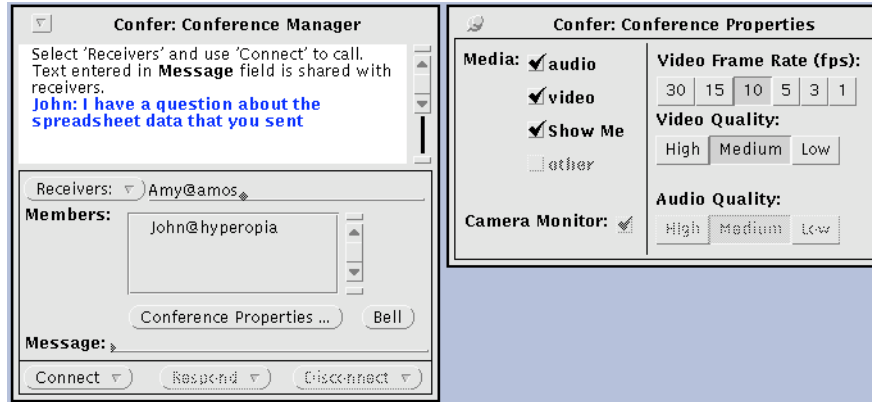


FIGURE 1. John uses the conference manager application to request a conference with Amy.

Figure 2 shows a screen image of a typical desktop video conference. For a 2-way conference, each user's screen displayed:

- a video window of a remote collaborator,
- a preview window of the video being sent to the remote collaborator, and
- a shared markup and drawing program (called Show Me) for drawing, typing, pointing, and erasing over shared bit-map images.

Show Me allowed users to create shared free-hand graphics and to grab bitmap images from their screens and share them with the other conference members.

When designing the DVC, we made sure to minimize the audio transmission delay. We did so by designing the infrastructure for making connections to handle the video and audio data streams separately, which enabled us to give priority to the audio transmission during periods of heavy network usage. At these times, the quality and/or latency of the video was degraded.



FIGURE 2. The desktop video conferencing prototype consists of the Show Me shared drawing tool, receive video window of remote user, and preview video window of outgoing video signal.

Study

To understand how people would use the DVC prototype, we conducted a study of distributed collaboration. We studied a 5-person team that was distributed among three locations: two buildings on a campus site on the West Coast, and another building on the East Coast. We observed their collaborative work under three conditions:

- *pre-DVC* – using conventional tools (phone, e-mail, video conference rooms, etc.)
- *full-DVC* – adding the DVC prototype (audio, video, Show Me)
- *DVC-minus-video* – subtracting video from the DVC prototype (audio and Show Me only)

We studied the team for three weeks in the pre-DVC condition, six weeks in the full-DVC condition, and four weeks in the DVC-minus-video condition. This team had previously been located together in neighboring cubicles at one site, so they were particularly aware of things that became difficult in their distributed locations.

We used a variety of observational methods to get different perspectives on their work activity. We monitored the team's observable collaborative activity, including the number and duration of phone calls, their usage of electronic mail, the frequency of their face-to-face contact, and their usage of the DVC prototype. Additionally, selected samples of collaborative activity in the three conditions were recorded on videotape. These tapes were analyzed by a multi-disciplinary group in the tradition of interaction analysis (Brun-Cottan & Wall, 1995). Furthermore, at various stages during the study, we interviewed each team member to gather their perceptions about their work activity. Details of the study are described in Tang and Isaacs (1993).

Lessons Learned

Although this study was not specifically intended to measure the value of video, one finding that clearly emerged is that video was very important to the users. The group used the prototype when video was provided, but their use declined dramatically when the video was removed. They commented that the main reason they stopped using the prototype without video is that its audio was worse in quality than the phone and exhibited an annoying echo and a noticeable transmission delay. This pattern indicates that the main benefit the prototype offered was through the video. In fact, the users were willing to endure the poor audio to gain the advantages of the video.

Qualitative analyses of the videotapes of their use of the prototype helped identify why video was valuable. The visual access provided cues that facilitated the mechanics of turn-taking and the interpretation of gestures, facial expressions, and pauses. We observed that non-verbal cues were especially important for signalling disagreement and handling sensitive issues. (See Isaacs and Tang, 1993, for a further discussion.) Furthermore, the users commented that the video capability made their interactions generally more satisfying (see also Rudman, et al., this volume). This support for interactional mechanisms makes video-mediated communications more efficient, effortless, and effective. A richer communication channel affords greater mutual understanding among the participants, and we would expect it to help improve the quality of their collaborative work in the long term.

One related issue is the role of eye contact in video-mediated communication. While direct eye contact is expected in face-to-face meetings, conventional desktop video conferencing configurations can provide only near eye contact by positioning the lens of the camera as close as possible to the video window of the remote collaborator. In our DVC setup, this arrangement gave each collaborator a clear sense of their partners' direction of gaze, known as gaze awareness (Ishii & Kobayashi, 1992). All the team members initially remarked that their inability to establish direct eye contact felt strange. However, we found considerable evidence in the videotapes that the collaborators were able to make use of gaze awareness in their interactions. For example, if someone paused and looked upward, their partner could infer that they were searching for the right thing to say. In another example, one collaborator expressed disagreement by avoiding looking at his partner until they moved onto another topic.

Only 43% of the call attempts turned into desktop video conferences. Most of the call requests were not answered because the person being called was not in the office. This statistic suggested that an application supporting audio-video interactions should also help users find a good time to interact. The usage logs also showed that desktop conferences tended to be relatively long, with a median duration of 8 minutes and 55 seconds. In the interviews, users commented that the interface for requesting a conference felt heavyweight, and they tended to use the phone for shorter interactions.

Our experience with desktop video conferencing also revealed some important ways in which it is different from other forms of interaction. In desktop video conferences, all participants were located in their offices where each person had access to his or her resources and distractions (e.g., phone calls, e-mail arrivals, visitors). Thus, it was not unusual for people to read e-mail or take phone calls during a conference. Furthermore, these interruptions were managed without causing confusion or offense because the aural and visual cues enabled remote collaborators to interpret what was happening when one person temporarily stopped participating. Thus, desktop video conferencing was a medium for focused interaction (like a phone call or meeting), but also one that tolerated long periods of independent work.

One implication of this observation is that desktop video conferencing is a distinct collaboration setting that has its own characteristics and limitations. As such, it is not intended to replace other forms of interaction (as some marketing promises might suggest). Our data showed that there were no statistically significant decreases in the amount of phone or face-to-face meeting activity when the full DVC prototype was available compared to the other two conditions (Tang & Isaacs, 1993). Desktop video conferencing offers a communication choice that complements, rather than replaces, face-to-face, phone, e-mail, and other interactions.

In summary, the quantitative measures of DVC prototype use helped us detect that people stopped using the prototype once we removed the video. Interviews with the users confirmed that they did so in fact *because of* the lack of video. Qualitative analyses of the videotaped activity revealed ways in which video helped them accomplish and enrich their interactions, which helped explain why they found the video-mediated interactions so satisfying. The studies also revealed some design implications that led to our next project.

Montage

Goals

Our experiences with the DVC prototype prompted us to take a broader perspective on how audio-video connections could be used to support remote collaboration. We realized that it is important to support the process of finding an opportune time to interact. To understand this problem better, we interviewed a range of people in the U.S. (including those who spent a portion of their work time physically separated from their work group) to explore how people want to be aware of and accessible to their colleagues. These interviews confirmed the need to help people find good times to make contact. People commented that they wanted help finding people who were not in their office or the ability to “leave a note on their chair” to set up a future contact. When presented with the idea of using audio-video connections to see when people were available, interviewees expressed strong concerns about preserving their privacy.

Design

Guided by these interviews, our previous experiences with the DVC prototype, and lessons learned from other video conferencing efforts (e.g., Dourish & Bly, 1993; Fish et al., 1993), we developed a prototype called Montage (Tang & Rua, 1994), which tries to provide a sense of proximity for distributed groups. It does so by providing an easy way to make audio-video connections between computer desktops and by integrating other communication applications.

Montage uses momentary, reciprocal *glances* among networked, media-equipped workstations to make it easy to peek into someone's office. It is modeled on the process of walking down a hallway to visit a colleague in her office. If you peek in and see that she is not available (e.g., not in the office, busy on the phone), you might pass by the door without stopping. If you find her in, you might pause at the doorway to indicate what you want to discuss before entering and settling in for a discussion. By basing Montage on the hallway model, we hoped to provide a familiar way of increasing the accessibility of colleagues without disrupting their privacy.

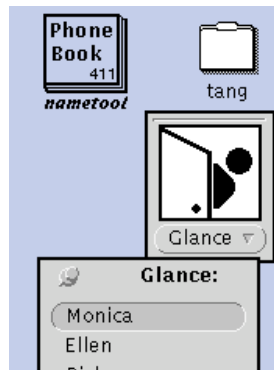


FIGURE 3. John initiates a glance at Monica by selecting her name from the Montage menu.

In Montage, a user typically selects the name of a person they wish to glance from a menu (Figure 3). Within a few seconds, a sound notifies the recipient of the onset of a glance and video windows fade-in on both users' screens. The fade-in effect provides a graceful approach for the people involved in a glance. Either party can *acknowledge* the glance by pressing the audio button to open an audio channel. If neither party enables the audio channel, the glance fades away after 8 seconds. Once either person presses the audio button, a two-way audio-video

connection is established. The relatively small (128 x 120 pixels) video windows of the glance are intended to support short, lightweight interactions. If participants want to have an extended interaction, either one can initiate a full-featured desktop video conference by pressing the **Visit** button. A visit offers enlarged video windows (256 x 240 pixels) and access to tools for sharing bitmap graphics (ShowMe Whiteboard™ —a product version of the shared drawing tool from the DVC prototype) and short text messages (Stickup notes). Glances and visits are ended by pressing a button that immediately dismisses the video window.

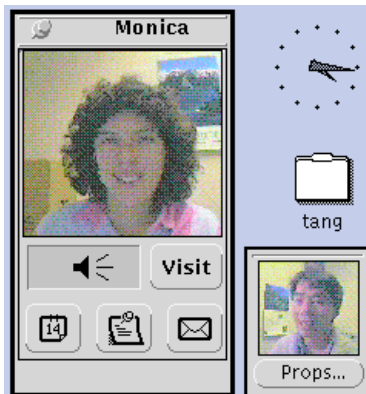


FIGURE 4. After John glances Monica, a small window appears on his screen providing a view into her office. At the same time, he sees a preview of his own image.

If the glance shows that the person is not available, the buttons along the bottom of the glance window (see Figure 4) provide quick access to browse her on-line calendar, send her a “Stickup” note, or send her an e-mail message. The on-line calendar and e-mail functionality are adaptations of existing tools widely used in our company. We developed Stickup, which enables users to type a text note that appears in a popup window on the recipient’s screen (shown in Figure 5). Stickups also include a **Glance Back** button that quickly starts a Montage

glance back to the person who posted the Stickup, and a **Reply** button that opens a Stickup to post back. By integrating quick access to these other communication tools, we hoped that Montage would help coordinate opportune times to make future contact.

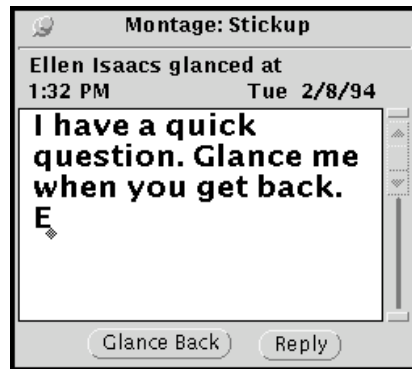


FIGURE 5. A Stickup from Ellen. Note the **Glance Back** and **Reply** buttons that quickly initiate a glance or Stickup back to the person who posted the Stickup.

Since Montage allows audio-video connections with any other user, it is important to enable users to protect their privacy. Montage addresses this issue in part by building on existing social mechanisms for protecting privacy. Because all Montage glances are reciprocal, users can see if anyone is glancing them. Just as it is considered rude to stand outside someone's door and stare in, it is equally impolite to do so through Montage, which provides the aural and visual cues to make such eavesdropping obvious. This symmetry enables users to socially negotiate their privacy. In addition, Montage offers a 'do not disturb' mode that blocks incoming glances.

Study

To learn how people would use Montage, we deployed the prototype in an existing working group (Tang, Isaacs & Rua, 1994). We selected a group of ten people distributed among three buildings on a campus site. The group was multi-disciplinary (including marketers, engineers, a manager, and a project coordinator), and it included people who worked part time or telecommuted. As with our other studies, we chose a group that had not been involved in the design or implementation of Montage.

We studied the group's communication patterns for four weeks before we installed Montage, 12 weeks while they had Montage, and four weeks after we removed Montage. To determine how Montage affected their communication, we collected logs of their use of Montage, logs of their voice-mail system use, copies of all e-mail sent within the team, and logs of appointments scheduled in their on-line calendars. Unfortunately, we were unable to collect reliable information about their phone calls and face-to-face meetings. In each of the conditions, we videotaped samples of the group's work activity by leaving a video camera running in individual offices throughout a day. We administered surveys to the team during the study to gather their perceptions of their work activity and their reactions to Montage.

Lessons Learned

Even more than with the DVC prototype, the logs of Montage use demonstrated how frequently attempts to contact someone were unsuccessful. The Montage logs showed that on average, users attempted to glance others 2.9 times per day, but that 75% of glance attempts were not acknowledged (i.e., neither party enabled the audio). This high rate of unacknowledged glances underscores the importance of helping people find opportune times to make contact. Despite the likelihood that a glance would not immediately turn into an interaction, people continued to use Montage. This continued use suggested that glancing was sufficiently lightweight and that it provided valuable help in coordinating future contact.

The logs showed further evidence of the lightweight nature of interactions in Montage. Of the acknowledged glances, resulting interactions tended to be relatively short, with a median duration of 1 minute 8 seconds. This median compares to 8 minutes 55 seconds in our previous DVC prototype, which suggests that Montage glances were used for shorter, more lightweight interactions. In the interviews, participants indicated that they tended to use Montage for small issues just as they arose. Without Montage, they either handled the issue themselves or waited to contact someone until a few such issues had accumulated.

However, the quantitative data demonstrated that the Montage features for coordinating future contact were not extensively used. There were 886 unacknowledged glance attempts, when the user might be expected to use the other communications applications integrated with Montage. However, the logs showed that people posted Stickups only 77 times, they browsed calendars only 20 times, and they sent email only 16 times. These results surprised us, especially since the users' perceptions collected in the surveys told a different story. Eight of the ten users said they especially liked Stickups and found them to be very useful. We speculate that Stickup use may have been low because of the impromptu nature of the contact; if the person wasn't there when contacted, the issue would be handled without their input.

In the interviews, users were generally enthusiastic about Montage and the visual access provided by the video. Analyzing the videotapes identified specific ways in which the video was useful. The tapes provided more examples of the subtle benefits of using the video to convey non-verbal cues during an interaction, similar to what we saw with the DVC prototype. In addition, video was also used in Montage to interpret people's availability and willingness to interact. The information in the video window enabled the person being glanced to identify who was requesting their attention and to convey whether they welcomed an interruption. Furthermore, when a glance revealed that someone was on the phone or occupied with a visitor, the participants often used non-verbal signals to set up future contact (e.g., a look and a hand gesture to indicate 'I see you, I'll glance you back'). Other times, they used visual cues to interrupt the activity gracefully. Thus, in addition to the advantages of "talking heads" video, we also observed the benefits of "silent heads" video in leading up to an interaction.

Both quantitative and qualitative measures indicated that Montage provides a communication medium that is between face-to-face visits and the phone. Like the phone, it provides quick access to people who are located elsewhere, and allows both participants to remain in their own offices with access to their own resources. Like face-to-face interactions, the video channel in Montage allows rich interactions and facilitates more frequent, shorter interactions that addressed specific issues just as they arise.

Forum

Goals

After focusing exclusively on small group coordination, we decided to explore the use of networked video and audio to provide a sense of community in a large organization.

Presentation and training sessions are often used to communicate information to large groups of people. These sessions help large groups create common knowledge and shared experiences and they help reinforce the organization's culture.

As organizations become distributed, they have to work harder to create these shared experiences. We felt that it could be useful to enable people from different locations to attend presentations from their computer desktops. Presenters could reach more of their intended audience, and individuals could attend more presentations that interested them. This idea led to the development of Forum, a tool that enables distributed video-based presentations.

Before designing Forum, we observed face-to-face presentations and interviewed people who gave or attended many presentations. In doing so, we became especially aware of the interactive nature of presentations. Speakers rely heavily on feedback from the audience in the form of questions and non-verbal cues. Audiences also pick up information from each other by chatting among themselves and by seeing how others react to the speaker's comments. In designing Forum, we tried to find ways to build in tools that enabled interaction between the audience and speaker and among the audience. We also tried to provide a basic level of awareness so that participants would know who else was attending a talk.

Design

Forum is a distributed application that enables speakers to broadcast talks over a network and enables audience members to participate in the talks from their workstations. Speakers sit in front of a media-equipped workstation that lets them control the display of their slides, manage their interaction with the audience, and see a list of audience members. Audience members receive live audio and video of the speaker as well as the slides and slide annotations. Audience members can interact with the speaker in three ways: they can speak to the presenter, they can

“vote” anonymously on an issue raised by the speaker, and they can send in written comments. Since they are in a multitasking environment, they can switch their attention between the Forum talk and other applications on their desktop or other activities in their office.



FIGURE 6. Forum’s audience interface. Audience members watch the video and interact with the speaker in the upper left window. Here, Ellen Isaacs is asking a question as three others wait in line to speak; the results of an earlier poll appear in the Poll Meter. Audience members view the slides in the lower window, using the thumbnails to view slides independently of the speaker. The Audience Window shows a list of attendees and allows audience members to send each other text messages.

The audience's interface is shown in Figure 6. The main window in the upper left shows video of the speaker, and the controls below it manage the audio parameters. The control panel to the right of the video provides three mechanisms to interact with the speaker: spoken questions, polls, and written comments. An audience member who wants to speak gets in the queue by clicking on the button at the bottom of the window. When the speaker calls on her, she presses and holds down the **Speak** button. Everyone can hear her speak and they can see her picture with her name above it. Speakers use the poll meter in the upper right to ask the audience a question. To vote on a poll, an audience member clicks on the **Yes** or **No** option and the bar chart changes accordingly. To submit a written comment, the user clicks on the **Comments** button, types a comment into the popup window that appears, and sends it to the speaker.

Audience members can also find out who else is watching a presentation by clicking on the **Audience** button, which brings up the Audience window, shown to the right. Users can click on a name to see that person's icon, their location and phone number. They can send that person a short message by clicking on the **Message** button, which brings up a small window pre-addressed to that recipient. When they send the message, it pops up on the other person's screen in a small window with a **Reply** button. Finally, audience members see the speaker's current slide in the slides window, and they can click on a thumbnail to view a different slide at any time. Users can see the speaker's annotations and they can make their own private annotations

Study

We studied the use of Forum using both informal and formal approaches. The informal testing went on throughout the process of developing Forum. On a weekly basis, we asked people around the company to give talks over Forum on a topic of their choice. The talks were attended by a small community of people who were willing to test Forum and, for any given talk, were interested in the topic. During each talk, we videotaped the speaker and at least one audience member. We sent out periodic questionnaires to the audience members, and we interviewed the speakers and some audience members. During that period, we made many

refinements to the functionality and interface, many to enable smoother interactions. We also learned a great deal about both speakers' and audiences' experiences (Isaacs, Morris & Rodriguez, 1994).

After a series of design iterations based on informal testing, we ran a more rigorous test to compare Forum presentations with those given in a local setting (Isaacs, Morris, Rodriguez & Tang, 1995). In this study, seven talks were given once over Forum and once in a local setting. The talks ranged in topic and style (e.g. lecture, informal talk, discussion session). For each talk, we videotaped the speaker and an audience member, sent questionnaires to all audience members and the speaker, interviewed the speakers, and logged Forum user activity.

Lessons Learned

Much of what we learned about video stems from the fact that video is used asymmetrically in Forum. Although the audience members can see the speaker, the speaker cannot see the audience. We chose this design to minimize network bandwidth and because very few audience members had video equipment. In the future, these limitations will disappear, but our goal was to design something that could be used with current technology.

Our most striking finding was that audiences were extremely enthusiastic about Forum and in most cases preferred it to local talks, whereas speakers found it less rewarding to give a talk over Forum than in a face-to-face setting. Although other factors contributed to this difference, video played an important role. The questionnaire data and interviews indicated that speakers' biggest complaint was that they could not see the audience. They found it difficult to gauge the audience's level of interest, degree of understanding, and general attitude toward their material. One speaker said he liked trying out Forum because it was an interesting technology, but "insofar as a way to actually communicate the content, it was less fun, because of the inability to judge audience response and to get to know any members of the audience." Unless they worked hard to draw out the audience through the other interaction channels, speakers found it difficult to adjust and respond to the audience. Even the static image that appeared when an audience member asked a question was a welcome visual cue for speakers.

On the other hand, audiences could see the speakers and felt more connected to the speakers than vice versa. We can tell from analyzing videotapes that the video played an important role in helping audience members focus on the presentation. Still, few audience members explicitly mentioned the value of video in their questionnaires. Instead, when asked about the value of the visual information, most audience members mentioned that the slides were helpful and that they liked it when speakers showed videotapes or demos through the video channel. It appears that audiences, who had video of the speaker, took that information for granted, whereas speakers, who didn't have video of the audience, felt limited by its absence.

A second major finding was that more people attended Forum talks than local talks (for the same talk), by more than a 2 to 1 margin. An average of 141 people attended Forum talks, compared with 60 for local talks. As a result, Forum speakers received greater exposure than they would have otherwise. By interviewing speakers, we learned of many cases when speakers were approached long after their talk by people who recognized them from their Forum talk. We suspect that audiences felt as if they knew the speaker more than they might at a large presentation because each person had a close-up image of the speaker, who appeared to be talking directly to them.

The video channel was also used to transmit visual information other than the speaker's head and shoulders. A number of speakers played videotapes during Forum talks, which appeared in the same video region of the interface, temporarily replacing the speaker's image. The speaker could be heard while the video was playing. In addition, two speakers showed demos during their talks, pointing the camera at an object and manipulating it as they described how it worked. In these cases, the video channel was used to provide visual information about objects, rather than to support face-to-face interaction (Nardi, et al., this volume).

These observations indicate that video plays an important role in establishing relationships among the participants and in demonstrating objects. However, we once again found that video is most effective when combined with other interaction tools. Clearly, the speaker's audio was the most important channel for the audience; it was not uncommon for people to listen to a talk while doing other work, focusing on the video when they wanted to pay closer attention. We

know from the videotapes and speakers' interviews that the audio questions from the audience also provided the richest source of information about the audience. The importance of audio is highlighted by one of the audience's frustrations, which was their inability to express laughter or applause. The "press to talk" audio model was required to avoid audio echo problems, but it prevented audiences from giving ongoing audio feedback to the speaker. Some Forum audience members even sent in written comments to speakers at the end of their talks explicitly praising the talk. Since they could not applaud, they had to find another way to show appreciation.

One of the more surprising findings was how effectively the poll provided a sense of the audience. The poll is a very simple device that gives anonymous, yes/no information about the audience, but it was because of this simplicity that it became so useful. Speakers could ask frequent questions to get a feel for the audience's attitude and the audience could easily and anonymously convey their opinion. The poll served to keep the audience involved and the speaker connected to the audience. The effectiveness of the poll was a good reminder that interactivity can take many forms; the key is to provide the right tool for the right situation. In some cases, video provides the right information, in others a simple yes/no or text-based interaction tool fills the need.

The Role of Video

Based on our work with these prototypes, we have come to appreciate the variety of ways in which video supports and enhances collaboration. There is no one right use of video. Each is appropriate for different tasks and settings, and in many cases video serves multiple purposes.

Enhancing the users' experience

Perhaps our most obvious finding is that people like to see each other when they interact (Gale, 1990, Rudman, this volume). Regardless of any cognitive benefit video may provide, people like having it, whether they are in one-on-one interactions or watching lectures. We saw from the DVC study that people stopped using the prototype when the video was removed. Later we

saw that people contacted each other for impromptu interactions over Montage more often than they did when they had to use the phone. In Forum, the video is an important reason why audiences had such an enthusiastic response to the technology.

Interpreting visual information in interactions

Video helps people interpret the many subtle visual cues that accompany interactions (Isaacs & Tang, 1993, Rudman, this volume). In one-on-one interactions, people use video to help time their contributions and interpret each others' attitudes. We saw cases when people used gaze awareness to indicate that they disagreed with a speaker. A smile along with a sarcastic remark helped defuse the comment. Gestures were used to enhance descriptions. In addition, video opened up room for more casual and less focused conversations. If someone paused to consider an issue, the other person understood that they were not simply being unresponsive. If a visitor dropped by or a person became distracted, the other person could easily understand what was happening. As a result, we, like others, saw cases of longer "office share" connections with intermittent focused interactions (Bly, 1993; Fish, et al, 1993; Mantei, et al., 1991).

In multi-party interactions, video also helps people manage not just when to speak but who will speak next. It also helps a speaker get a sense of the groups' reaction and to adjust as they speak. The lack of this video feedback in the one-to-many situation of Forum clearly disrupted speakers. They had few cues of the audience's understanding of, agreement with, or appreciation for their presentation, and so could not adjust their remarks accordingly. They ended up less satisfied with the experience relative to a face-to-face presentation.

Enabling distributed conversations that would not happen otherwise

Because video enables the interpretation of subtle visual feedback, it opens the possibility of having conversations about sensitive or private issues that people are reluctant to conduct over the telephone. Although people prefer face-to-face settings for such delicate discussions, we found that people were willing to hold them over video but not over the phone. During one conversation, two people turned off the video camera that was observing their activity for our study purposes because they thought the topic was too sensitive for us to record, but apparently

it was not too sensitive to discuss over a video link. When people work across distant locations, they rarely have the opportunity to have face-to-face conversations, so the video enables them to talk about issues they simply would ignore if only the phone were available. This is a subtle issue that affects groups that work together over long periods of time; conflict is bound to emerge and it must to be handled well to keep the group functioning productively. This effect has not been explicitly or extensively studied, so it remains an open question how far we can generalize from our findings.

Awareness

Video plays a critical role in providing awareness. From the DVC prototype, we learned that interactions are greatly facilitated if people know when others are around and whether they are available to talk. Video is perfectly suited to providing these cues. The video glance mechanism of Montage enabled people to interpret effortlessly whether someone was available to interact or whether they should try again later. Through video it is easy to tell if a person is on the phone, busy with a visitor, engrossed in some work, or not there. If someone is glanced when they are busy, they can look toward the video and recognize the glance, perhaps using non-verbal cues to indicate that they will get back to the person later. Of course, awareness usually trades off against privacy, so it is important to enable users to control others' access to them.

Our experience with Forum expanded our understanding of the role of awareness. Among large groups, awareness involves knowing not only who is around and their level of activity, but the size of the group and their level of responsiveness. An awareness of others present helps participants in large groups frame their contributions, and it sets up later interactions. An awareness of the size of the audience shapes participants' style of interaction, and knowledge of participants' level of responsiveness influences everyone's interpretations of the event. The role of video in enabling awareness has been noted for some time (see Isaacs et al., this volume).

Providing identity and recognition

Forum also demonstrated the importance of video for establishing the identity of a person, especially when the participants have never seen each other. When people first meet, they feel they know each other better if they have previously seen each other than if they have only heard each other. In the case of Forum, where speakers' images were broadcast to people distributed across many locations, the video provided valuable recognition to the speaker. Others see that person later and identify them, and perhaps approach them for an interaction. This role of video has not been discussed much in the video conferencing literature, but we note that television has demonstrated convincingly the power of video to enable widespread recognition.

Creating a focus

We were reminded by Forum of the simple point that video helps provide a point of focus. We noted that people often did other lightweight work while attending Forum talks, but when they wanted to pay closer attention, they watched the video, even though the "talking head" image provided relatively little dynamic information. In other cases, speakers used the video to show an object, which also gave the audience a focus and a shared understanding of the visual material. Nardi, et al. (this volume) further discussed this use of "video as data."

Design of Video

During the course of our work, we also have learned some design guidelines for using video in applications for face-to-face interaction, which we describe here. We qualify these comments with a reminder that we have worked with low-quality digital video (4-10 frames per second, low resolution), which may reduce the visibility of subtle social cues. Previous work indicates that very high quality video only marginally improves the quality of work (Olson, et al., 1995), but that it does affect the mechanics of conversation and users' satisfaction with the experience (O'Conaill, et al, 1993).

Audio latency matters more than audio-video synchronization for real-time interaction

In our DVC pre-study, we first learned the importance of minimizing audio latency when supporting live interaction, a finding that has been confirmed by others (Kurita, Iai, & Kitawaki, 1993; O’Conaill, Whittaker, & Wilbur, 1993). When it takes longer than about 400 milliseconds between the time when one person says something and the other person hears it, conversations become far less interactive. We found this effect even with constant delay from a switched system. Since it is difficult to time a response, people can no longer rapidly exchange utterances. Since humor heavily depends on timing, it starts to disappear. Since conflict management depends on quick responses to feedback, people back away from controversial issues. As a result, conversations with delayed audio are not only frustrating for participants, but they consist of extended monologues about straightforward topics with little humor (Krauss & Bricker, 1967; O’Conaill, Whittaker, & Wilbur, 1993). If a design tradeoff must be made, it is best to sacrifice audio-video synchrony to enable low-latency audio. Although it is also disturbing to see someone’s mouth moving after hearing their utterance, people seem to be able to adjust to this much more effectively than they do to audio delays. Of course, when video and audio are used in non-interactive contexts, audio and video should be synchronized.

Video is often more effective when combined with other means for interaction

Although video can be used alone in some cases, it is usually more effective when combined with other media. Audio is the obvious complement to video, but other types of media are often useful. When people interact, they often want to show each other things and write things down so they can be saved for later use. It is useful to enable them to share graphics, text, and applications. In Forum, we provided a “poll mechanism” to enable the relatively common technique of surveying the audience. Even if video of the audience had been provided, it would have been too unruly to manage this process. The lesson is that one should consider providing mechanisms to support visual behavior that cannot be accomplished comfortably through video.

Very short connection times are critical for supporting lightweight interactions

To effectively support lightweight interaction, audio and video connections among computers must be established very quickly. When using the phone, people often expect to make contact within three rings (about 10 seconds). Our experience with the DVC prototype indicated that longer delays caused users to resort to the phone for short interactions. Even with Montage, it took an average of 11 seconds for a glance to appear after it was initiated. This delay sometimes detracted from the goal of supporting lightweight interactions. While responsive performance is always an issue with applications, it is especially important with desktop conferencing. If connections take longer (or even feel like they take longer) than a few seconds, users will use another mechanism or use video conferencing for more formal, extended interactions.

Provide a sense of approach when establishing video connections

When video is used to connect people, designers should let people prepare for the pending connection by providing some warning. In the physical world, we can often hear people approach, which smooths the transition from the previous activity to the interaction. The designers of Cruiser found that people were disturbed when large video images of others' faces suddenly appeared on their monitor, even though they were aware that this could happen (Fish et al., 1993), whereas the RAVE system found audio cues to be helpful (Gaver et al., 1992). Based on these observations, we designed Montage glances to fade in rather than pop up and to be preceded by an audio cue. This approach felt more comfortable and natural to users.

Provide ways for people to protect their privacy

When people are told about applications that allow them to contact people over video, many react by expressing concern about their privacy. Although most users do not often take advantage of privacy controls, we have found that they will not even experiment with the system without the possibility of blocking access. Therefore, it is very important for the adoption of the technology to provide reasonable privacy controls. It would be a mistake to try to convince someone that they don't need access control because most people don't use it. They "use" it to feel comfortable that the system will let them control their privacy.

The Role of Methodology

Our understanding of our prototypes is shaped by the type of use studies we conducted and the measures we took. As we have shown, we believe it is important to use a combination of approaches and measures when collecting and analyzing data. Each method provides useful information, but when different ones are combined, a fuller picture emerges of the effect of technology on people.

In particular, we prefer to combine both quantitative and qualitative measures. Quantitative results, such as usage statistics or frequency counts of specific behaviors, can identify reliable patterns in the data, which can be used to confirm or deny theories about user behavior. But quantitative information often provides only broad descriptions of phenomena. To understand how and why the behavior takes shape, we use qualitative measures, such as descriptions of specific events detected in the videotapes or comments from user interviews. It is also important to distinguish objective and subjective data. For example, it is helpful to use videotapes to count the occurrences of a certain behavior (an objective measure), but asking users why they chose that behavior (a subjective measure) helps fill out our understanding.

During the Forum study, we used findings from different sources to motivate investigations with other types of data. For example, we used the videotapes and the logs to count the number of questions asked in local and Forum presentations (a quantitative, objective measure of interactivity). However, we noticed from the tapes that the local talks “felt” more interactive (a qualitative analysis), and the questionnaires indicated that local audiences thought the questions were handled better (a quantitative measure of a subjective reaction). We returned to the videotapes and noticed that questioners in local talks seemed to ask more follow-up questions, which created a dialogue between the speaker and the audience member. From this qualitative analysis, we then counted the number of follow-up questions in Forum and local talks (quantitative measure) and found that indeed they did occur significantly more often in local talks. Finally, the interviews confirmed that this type of dialog contributed to the speakers’

satisfaction with the experience. By combining measures, we not only know that local talks had more speaker-audience interactions, but we know the form they took and their effect on the participants.

In some cases, we noticed a conflict between the objective and subjective findings. In the Montage study, early interviews indicated that people considered protecting their privacy to be critical, which led us to build in a “do not disturb” mechanism to block video access. Once Montage was deployed, the logs showed that most people rarely used the privacy control. Interviews confirmed that they felt comfortable with the privacy offered. By combining these subjective and objective data, we determined that user acceptance depends in part on making privacy controls available, even though in practice most people do not actively use them. The behavioral measure would lead us to conclude that access control is not a necessary feature because it is rarely used, but the subjective reactions help us understand its importance to users.

In general, we have found that objective performance or usage data do not fully reveal the effect of a technology on users. We do not always know what aspects to measure to predict users’ reactions (e.g. we may measure efficiency when users are more concerned with the quality of social interaction), and some aspects are difficult to measure (e.g. whether people feel comfortable using video to discuss sensitive matters). Since we do not understand when and why objective and subjective measures are not well correlated, we should explore both aspects of technology to fully understand its implications for use.

Conclusion

We have reported on use studies of a variety of prototypes that use video to support collaborative work. Our studies not only confirmed that users like video, but also uncovered concrete information about the value of video in supporting interaction. In many ways, our work builds on the emerging body of work that conflicts with the earlier literature that did not find any demonstrable value of video. Studies by Ochsman and Chapanis (1974) and Gale

(1990) found no significant differences in comparisons of collaborative activity with and without video. There are two fundamental reasons why our studies found evidence for the value of video where earlier studies did not.

First, our studies involved prototypes whose design was shaped by an understanding of the needs of remote collaborators. Our question has not been “Is video useful?”, but rather “How can video be usefully integrated into people's work practice?” As is often the case, it is not raw technology that is useful, but rather its design into artifacts that fit into users' work practice. By doing studies of users' work activity with existing technology, we come to understand people's needs well enough to integrate video with other technology to develop useful designs.

Secondly, our use studies combined methodologies and observed the prototypes in the context of real work activity. By combining quantitative measures of user behavior (computer logs, behavior frequency counts from videotape data, questionnaires) with qualitative analyses (descriptions of videotaped activity, interviews, essay questions in surveys) we could appreciate a variety of perspectives on people's use of the technology. Also, studying people using technology in the context of their real work over time can reveal patterns of use that may be missed in laboratory studies, which often must use relatively contrived short-term tasks, sometimes with groups that would not otherwise work together. Those studies enable researchers to isolate specific causes of behaviors, whereas field studies reveal realistic responses among representative groups of users doing naturalistic tasks.

We are encouraged that so much innovative research is being conducted, both in the lab and in the field, in the spirit of understanding users' goals and needs. Taken together, all this research will help us design video-based technology that not only supports people's tasks but expands their ability to work with a wider range of people who may be located around the globe.

References

Bly, S.A., Harrison, S.R., & Irwin, S. (1993). Media spaces: Bringing People Together in a Video, Audio, and Computing Environment, *Communications of the ACM*, 36:1, 28-47.

Brun-Cottan, F., & Wall, P. (1995). Using Video to Re-Present the User, *Communications of the ACM*, 38:5, 61-71.

Dourish, P., & Bly, S. (1992). Portholes: Supporting Awareness in a Distributed Work Group. *Proceedings of the Conference on Computer Human Interaction*, (541-546). Monterey, CA: ACM Press.

Fish, R.S., Kraut, R.E., Root, R. W., & Rice, R. E. (1993). Video as a Technology for Informal Communication, *Communications of the ACM*, 36:1, 48-61.

Gale, S. (1990). Human aspects of interactive multimedia communication, *Interacting with Computers*, 2:2, 175-189.

Gaver, W., Moran, T., MacLean, A., Lovstrand, L., Dourish, P., Carter, K., & Buxton, W. (1992). Realizing a video environment: EuroParc's RAVE system. *Proceedings of the Conference on Computer-Human Interaction*, (27-35). Monterey, CA: ACM Press.

Isaacs, E.A., Morris, T., & Rodriguez, T.K. (1994). A Forum for Supporting Interactive Presentations to Distributed Audiences, *Proceedings of the Conference on Computer-Supported Cooperative Work*, (405-416). Chapel Hill, NC: ACM Press.

Isaacs, E.A., Morris, T., Rodriguez, T.K., & Tang, J.T. (1995). A Comparison of Face-to-face and Distributed Presentations, *Proceedings of the Conference on Computer-Human Interaction*, (354-361). Denver: ACM Press.

Isaacs, E.A., & Tang, J.C. (1994). What Video Can and Cannot Do for Collaboration: A Case Study, *Multimedia Systems*, 2, 63-73.

Ishii, H., & Kobayashi, M. (1992). ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact, *Proceedings of the Conference on Computer Human Interaction*, (525-532). Monterey, CA: ACM Press.

Krauss R.M., & Bricker, P.D. (1967). Effects of Transmission Delay and Access Delay on the Efficiency of Verbal Communication, *Journal of the Acoustic Society of America*, 41, 286-292.

- Mantei, M. M., Baecker, R. M., Sellen, A. J., Buxton, W. A.S., Milligan, T., & Wellman, B. (1991). Experiences in the Use of a Media Space, *Proceedings of the Conference on Computer Human Interaction*, (203-208). New Orleans: ACM Press.
- Ochsman, R.B., & Chapanis, A. (1974). The Effects of 10 Communication Modes on the Behavior of Teams During Co-operative Problem-solving, *International Journal of Man [sic]-Machine Studies*, 6, 579-619.
- O’Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication, *Human-Computer Interaction*, 8, 389-428.
- Olson, J.S., Olson, G.M., & Meader, D.K. (1995). What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work?, *Proceedings of the Conference on Computer Human Interaction*, (362-368). Denver: ACM Press.
- Sellen, A. (1992). Speech Patterns in Video-Mediated Conversations. *Proceedings of the Conference on Computer-Human Interaction*, (49-59). Monterey, CA: ACM Press.
- Tang, J. C., & Isaacs, E. (1993). Why Do Users Like Video? Studies of Multimedia Supported Collaboration, *CSCW: An International Journal*, 1:3, 163-196.
- Tang, J.C., Isaacs, E.A., & Rua, M. (1994). Supporting Distributed Groups with a Montage of Lightweight Interactions, *Proceedings of the Conference on Computer-Supported Cooperative Work*, (23-34). Chapel Hill, NC: ACM Press.
- Tang, J.C., & Rua, M. (1994). Montage: Providing Teleproximity for Distributed Groups, *Proceedings of the Conference on Computer-Human Interaction*, (37-43). Boston: ACM Press.

