

Supporting Distributed Groups with a Montage of Lightweight Interactions

John C. Tang, Ellen A. Isaacs, and Monica Rua

SunSoft, Inc.
2550 Garcia Avenue
Mountain View, California 94043
(415)336-1636, (415)336-1167, (415)336-2037
tang@eng.sun.com, isaacs@eng.sun.com, rua@eng.sun.com

ABSTRACT

The Montage prototype provides lightweight audio-video glances among distributed collaborators and integrates other applications for coordinating future contact. We studied a distributed group across three conditions: before installing Montage, with Montage, and after removing Montage. We collected quantitative measures of usage as well as videotape and user perception data. We found that the group used Montage glances for short, lightweight interactions that were like face-to-face conversations in many respects. Yet like the phone, Montage offered convenient access to other people without leaving the office. Most glances revealed that the person was not available, so it was important to integrate other tools for coordinating future interaction. Montage did not appear to displace the use of e-mail, voice-mail, or scheduled meetings.

KEYWORDS: awareness, media space, informal communication, video, remote collaboration.

USING VIDEO TO PROVIDE A SENSE OF PROXIMITY

Many work groups are becoming distributed across different geographical sites. These groups can include workers from different company sites as well as telecommuters who work from home. There is growing interest in providing these groups with a sense of cohesion and proximity that can be lost when group members are dispersed among different locations and time zones. In particular, team members who are co-located can easily initiate spontaneous, lightweight interactions. Previous research [Whittaker et al., 1994; Kraut et al., 1990] found this informal awareness to be an important aspect of collaborative work activity.

Connecting distributed group members with audio-video links is an intuitive approach to providing a sense of proximity. A few research prototypes (e.g., Cruiser [Root,

1988], Portholes [Dourish & Bly, 1992]) have demonstrated various ways of applying video to support informal awareness. Fish et al. [1993] evaluated the use of Cruiser in a realistic work setting, but we are still early in our understanding of how actual work groups react to such technology. Studying how audio-video links are used in actual, distributed work settings would help improve the design of this technology and increase our understanding of the needs of distributed workers.

We began exploring this issue as the result of our previous research on a desktop video conferencing (DVC) prototype [Tang & Isaacs, 1993]. That prototype enabled audio, video, and shared drawing connections among computer desktops. The interface for making connections followed the telephone model, where a person explicitly answers a call request before audio-video connections are made. One finding from our study of the DVC prototype in real use was that only 43% of the attempts to call someone were answered and turned into desktop video conferences. Most of the call requests were not answered because the recipient was not in the office. This statistic suggested that an application that supports audio-video interactions should also support the *pre-interaction* coordination that allows remote collaborators to find opportune times to contact each other.

Users of our previous DVC prototype commented that using the telephone model to request a conference felt rather heavyweight. Consequently, they tended to use desktop conferences for longer interactions, averaging 17 minutes and 11 seconds in duration. This observation suggested the need to provide the ability to have lightweight communications for distributed groups. By lightweight communication, we mean impromptu interactions that are quick and easy to initiate and tend to be short and informal.

Building on our previous experiences, we developed a prototype called Montage, which attempts to provide a sense of *teleproximity* 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 for coordinating future interactions. We

deployed this prototype in a working group that was distributed among different locations to observe how real users would react to this technology. This paper describes what we learned from the participants' usage of and reactions to Montage.

THE MONTAGE PROTOTYPE

Although the Montage prototype is more fully described by Tang and Rua [1994], we briefly describe it here to provide a context for this study. 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 it is not a good time to interact (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.

In Montage, a user typically selects the name of a person they wish to glance from a menu (see Figure 1). Within a few seconds, a sound notifies the recipient of the onset of a glance and video windows fade-in on both users' screens (see Figure 2). The fade-in effect provides a graceful approach for the people involved in a glance.¹ 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.

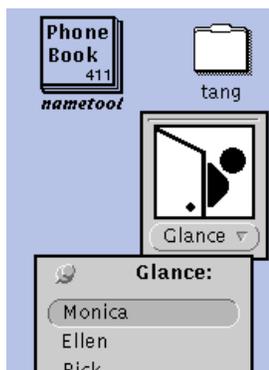


Figure 1. John initiates a glance at Monica by selecting her name from the Montage menu.

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 a more extended interaction, either one can move into a full-featured desktop video conference by press-

1. In an earlier version of Montage, the glance windows popped up without the fade-in effect, and people found the abruptness of the approach disconcerting.



Figure 2. 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.

ing the Visit button. A visit offers enlarged video windows (256 x 240 pixels) and access to tools for sharing bitmap graphics (ShowMe Whiteboard™) and short text messages (Stickup notes). Glances and visits are ended by pressing a button that immediately dismisses the video window.

If the glance shows that the person is not available, the buttons along the bottom of the glance window (see Figure 2) provide quick access to browse her on-line calendar, send her a short text note (Stickup), or send her an e-mail message. The on-line calendar and e-mail functionality are adaptations of existing tools widely used on the Solaris® platform. 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 3). 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.

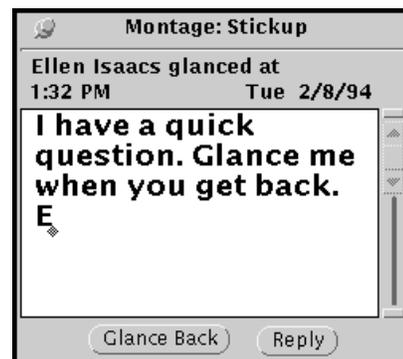


Figure 3. A Stickup from Ellen. Note the **Glance Back** and **Reply** buttons that quickly initiate a glance or Stickup back to the person who posted the Stickup.

Since Montage allows audio-video connections with any other user, it is important to enable users to protect their privacy. Montage addresses this issue by building on existing

social mechanisms for protecting personal privacy. Because all Montage glances are reciprocal, users can see if anyone is glancing at 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.

Montage also offers modes that convey different levels of accessibility. If a user sets Montage to 'do not disturb,' anyone who tries to glance will see an image indicating that the person does not want to be disturbed. However, this mode still offers the option of 'glancing in' to negotiate an interruption, similar to peeking through the glass window of a closed door and knocking to interrupt. Other modes offered are: 'locked' (no interruptions allowed), 'out of the office', and 'other', which provides a blank image on which users can type a message. Users can type a message over any of the access mode images (e.g., one can type "I'll be back at 3pm" over the empty office image).

Montage runs on the Sun SPARCstation™ line of computer workstations running the Solaris 2 UNIX operating system. Each workstation must be equipped with a video camera and a board that encodes the video signal using the CellB video compression algorithm. The video image is transmitted at a four frame per second refresh rate. The workstations must be connected by a conventional Ethernet network. External speakers are typically added to provide higher quality audio.

STUDYING MONTAGE IN USE

To learn how people would use Montage in their daily work, we deployed the prototype in an existing working group. We selected a group of ten people (six men, four women) who developed strategies and procedures for enabling a computer operating system to run on a variety of hardware platforms. The group had several interesting characteristics that tested the utility of Montage. The group was:

- distributed among three buildings on a campus site — five were on two floors in one building, two were in a neighboring building, and three were in a building almost 1/2-mile away;
- cross-functional — there were three engineers, three marketing account managers, a test manager, a legal representative, a project coordinator, and a manager;
- flexible in work schedule — one engineer worked four days a week on average, and one engineer worked at home, coming into the office only one day a week;
- unaffiliated with the Montage project — although part of the same company, they were not involved in the design or development of Montage and did not have much of a vested interest in it.

We studied the group's communication patterns during the four weeks before we installed Montage (pre-Montage condition), 12 weeks with Montage installed (Montage condi-

tion), and four weeks after we removed Montage (post-Montage condition). To determine how Montage affected their work communication, we collected:

- logs of their use of Montage,
- logs of their use of the company voice-mail system,
- copies of all e-mail sent within the team, and
- logs of appointments scheduled in their on-line calendars that involved other team members.

We would have liked to collect data on phone calls among the team, since the availability of Montage might be expected to affect their phone usage. However, the company phone system could not provide that information automatically. Similarly, we would have liked to collect data on their face-to-face meeting activity, but we found no reliable way to gather that information. Previous experience with self-reports of meeting activity [Tang & Isaacs, 1993] suggested that this measure is not accurate enough to show reliable changes in patterns.

In each of the three conditions, we videotaped samples of the group's work activity by leaving a video camera running in individual offices throughout a day. In all, we collected 13 days of activity among six of the team members, including two team members who were videotaped in each of the three conditions. We also administered surveys to the team to gather their perceptions of their work activity and their reactions to Montage. Surveys were collected before starting the study, just before deploying Montage, after the group had been using it for three weeks, just before removing Montage, and after the study had ended.

Real world, real data, real pain

Our approach to studying technology is to observe its use in real-world activity as it naturally occurs, without trying to control that activity in any way. This commitment to studying naturalistic activity often trades off against gathering data that cleanly distinguishes between experimental conditions. Several factors in the data collection process prevented us from getting data that would distinguish among the three conditions as well as we had hoped. Although these factors limit some of the claims we can make on the quantitative data, we accept them as a characteristic of studying real world activity rather than behavior isolated in a laboratory setting.

The most disappointing aspect of the data is that the Montage prototype experienced performance and reliability problems during the study. The effectiveness of Montage depended on quickly setting up glances among team members. However, the average length of time between selecting a person's name and starting to see a glance was about 11 seconds, and could range as long as 50 seconds. Such a long wait tarnished the notion of using Montage for quick glances, and the users sometimes used the phone instead.

Despite substantial testing, the reliability of the prototype was also disappointing. The prototype would not only quit unexpectedly, but would often require help from the development team to get it running again. This was inconvenient to the users, and also prevented them from using Montage while they waited for us to restart it. Another technical problem sometimes caused the prototype to hang indefinitely when a user tried to make a glance. This problem further eroded the notion of using Montage for lightweight glances.

These performance and reliability problems became so troublesome that Montage became practically unusable during the Montage condition. We decided to re-engineer the prototype and deploy a revised version during the seventh week of the Montage condition. The revised version greatly reduced (but did not completely eliminate) these problems for the remainder of the Montage condition.

Because we were using open microphones and speakers, there were considerable problems with the audio channel. It took a fair amount of adjustment to avoid acoustic feedback and echo in a glance, and some tuning was needed for each person glanced. Occasionally during long glances, the audio cut out or the video froze. Users reported that they sometimes used the phone while in a glance to work around these problems.

Despite the prototype's problems, the users continued to use Montage, and no one stopped using it during the Montage condition. Even though the revised version of Montage greatly improved its usability, their disappointing experiences with the initial prototype probably biased their impressions of Montage. In the survey responses, most users reported that they would have used Montage more if it had been more robust. Thus, the data underestimate the usage potential of Montage, but they still provide viable information about how the users responded to the concepts of Montage.

Independent of Montage, there were also a few factors that affected the work activity observed during the study. A few personnel changes occurred during the study. The group's manager changed two weeks before the study started and the project coordinator returned from a leave of absence just when Montage was installed, replacing a temporary substitute. Three weeks before the end of the Montage condition, two people left the company. While such changes were not unheard of in their work environment, accommodating these changes had an indeterminate effect on the group's activity during the study.

Throughout the study, the nature and pace of their work varied as they met deadlines and moved to different phases of their project. The timing of the study also included some substantial holiday disruptions. The company's 11-day Christmas holiday occurred during the pre-Montage condi-

tion, and the Montage condition was interrupted by two three-day weekend national holidays, one of which was combined with a two-day off-site group meeting.

There were also some errors in the data collection that caused us to lose very small amounts of the voice-mail, e-mail, and Montage usage data during short periods of time. To compensate for the lost data, and because there were varying numbers of people in the office on any given day, most of the frequency data is reported on a daily basis as a total number of events divided by the number of people participating (i.e., in the office and correctly logging data) for that day.

MONTAGE USAGE

The quantitative data were analyzed for statistically significant differences among the three conditions. The videotaped samples of actual work activity and the group's survey responses also helped us understand their communication activity and how it changed with Montage. Taken together, these data gave us a variety of views on the frequency of Montage glances, the nature of the group's interactions through Montage, the effect of Montage on their overall communication activity, and its effect on their privacy.

Many glance attempts, much fewer actual interactions

In the 12 weeks of Montage usage, there were 1188 attempts to glance someone in the group through Montage. After accounting for days when individuals were not in the office, this usage corresponded to an average of 2.9 glance attempts per person per work day. Of the 1188 attempted glances, only 302 (25%) were acknowledged by either person enabling the audio. The other 886 (75%) glance attempts were unacknowledged — neither party enabled the audio for verbal interaction.

Figure 4 plots the number of acknowledged glances as a proportion of total glance attempts (averaged by the number of people per day) for the Montage condition. Besides showing how few glance attempts were acknowledged, the graph also indicates the usage of Montage over time. A noticeable novelty effect seems to have subsided by the fourth week of the Montage condition.

Why so many unacknowledged glances? The most likely reason for most of the unacknowledged glance attempts is that the recipient was not in the office at the time of the glance. Although the usage logs cannot indicate the reason a glance was not acknowledged, they do show that 38% of the unacknowledged glances occurred when the recipient had enabled a screen lock program, which some users routinely invoked when they intended to be out of the office for a long time.

Glancing at someone who was not running Montage at the time or was experiencing technical problems with Montage also resulted in an unacknowledged glance. Additionally,

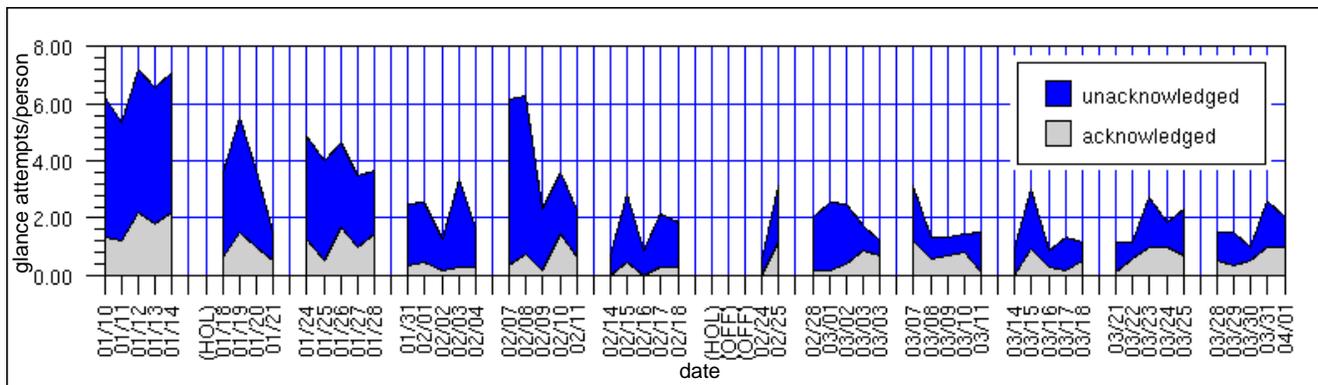


Figure 4. Total number of glance attempts averaged by the number of people in the office each day during the Montage condition. The proportion of glance attempts that were acknowledged by enabling the audio is shown in the lower shading of the graph.

the usage logs showed 43 attempts to glance at someone in an access mode that did not immediately accept glances (e.g., do not disturb, out of the office). Of course, users could intentionally ignore a glance because they were otherwise occupied in the office. Although the usage logs cannot indicate these intentionally ignored glances, we observed several instances in the videotapes of selected office activity, and we expect that they account for a small percentage of the unacknowledged glances.

This large percentage of unacknowledged glances is consistent with other research on communication patterns. Rice & Shook [1990] and Whittaker et al. [1994] reported that the majority of phone calls do not connect with their intended recipients. In our study of the previous DVC prototype [Tang & Isaacs, 1993], we found that 57% of the attempts to conference did not result in an interaction. These findings emphasize the importance of supporting the *pre-interaction* coordination and negotiation that is often necessary to find an opportune time to contact others. Finding that someone is unavailable can still be productive if you can easily determine when you can interact in the near future or if you can exchange the information using a different channel.

Low but valuable use of Stickups. Montage integrates access to the on-line calendar, Stickups, and e-mail to support pre-interaction coordination. We hoped that when a glance showed that a person was unavailable for an interaction, these utilities would make it easy to either find a convenient time to re-establish contact or convey the information in another mode. For the entire group over the 12 weeks of the study, the Montage logs recorded:

- 77 instances of posting a Stickup,
- 20 instances of browsing a calendar, and
- 16 instances of sending e-mail.

These numbers seemed very low, especially compared with the large number of unacknowledged glances (886) when they might have been useful. We were particularly disappointed that Stickups were not used more often. However,

from the users' perceptions collected in the surveys, eight of the ten users commented that they especially liked Stickups. One user commented:

I really like the Stickup feature. It's easy, convenient, and very useful. The tool lets us behave as we normally would (i.e., stick a postit to someone's machine). Stickup makes it easier (no running around) and is also more reliable (doesn't fall off like paper notes do).

The users' enthusiasm for Stickups suggested reinterpreting the relatively low amount of usage in evaluating their utility. Especially when introducing a new tool, it is hard to gauge how much usage indicates a useful feature; some valuable resources (e.g., phone book, dictionary) are used infrequently. The data also showed that Stickups were used to follow up on missed contact attempts. Of the 77 Stickups that were posted, 22 were used to glance back to the person who sent the Stickup and 25 were used to post a Stickup reply. Although usage was relatively low, Stickups did seem to encourage people to establish contact and send short messages to those who were not available.

The amount of calendar browsing and e-mail messaging through Montage was even lower, and some users remarked that it was just as easy to use the stand-alone calendar and mail applications that were usually open on their computer desktop. Because the users were already familiar with these applications, most had established deeply ingrained habits in using them. The videotapes showed some examples of glances when users opened their desktop applications rather than accessing them through Montage. Although users did sometimes browse calendars and send e-mail in conjunction with Montage glances, the prototype's design did not appear to integrate these tools effectively.

Overall patterns in glance attempts. Group members made an average of 2.9 glance attempts per person per work day. This is comparable with our previous DVC prototype and with Cruiser, which had a similar video glance mechanism. Users of the DVC prototype placed an average of 1.4

conference requests per person per work day [Tang & Isaacs, 1993]. In a study of Cruiser, Fish et al. [1993] reported an average of 2.7 call attempts per person per workday. Fish et al. compare it with a company average of 5.4 telephone calls placed per user per day. A higher usage of phone calls is to be expected, since it is a much more pervasively deployed communication medium.

Not surprisingly, the data show that two-thirds of the Montage glance attempts were between people in different buildings. Over one-quarter of the glance attempts were between people on different floors of the same building. Most of the glances also seemed to cross work functions (e.g., engineer to account manager, project coordinator to group manager). For the most part, however, the function groups were physically located together, so the patterns of cross-building and cross-function contact are interrelated.

Individuals' usage of Montage varied considerably. The frequency distribution of glance attempts *initiated* by each group member showed that one user (the project coordinator) initiated a high number of glances (291), most initiated a moderate amount (between 79 and 174), and two users initiated relatively few (23 and 20). However, the frequency distribution of glance attempts *received* was more balanced among the group members. Eight users received between 97 and 202 glances, and the other two received fewer (63 and 24). In general, the amount of glances that a person initiated did not correlate strongly with the amount received. For example, one person rarely initiated glances (23), but was the fifth most frequent recipient of glances (127). Even though she did not initiate many glances, she indicated in the surveys that she found Montage useful, apparently from the glances that she received through Montage.

Interactions in Montage were short and spontaneous

Acknowledging a Montage glance by enabling the audio resulted in an interaction in which the users could both see and hear each other. The average duration of an interaction (the time between enabling the audio and ending the interaction) was 3:09 (minutes:seconds). The durations ranged from 0:02 to 47:10, and the median was 1:08. The distribution of Montage interaction durations is graphed in Figure 5.

These data indicate that participants used Montage mostly for short interactions, but they had some extended interactions as well. The glances that were captured on videotape reveal that a typical short interaction consisted of a simple question and a short discussion of the answer. Longer interactions involved addressing several different issues or exploring an issue that required retrieving and analyzing more information.

By comparison, the median duration of Cruiser calls was reported to be 27 seconds, and the median duration of conferences in our previous DVC prototype was 8:55. Cruiser calls may have been shorter because those users often used it

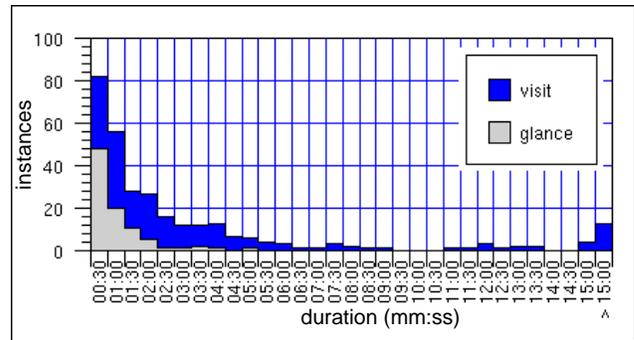


Figure 5. Distribution of glance and visit durations, graphed according to half-minute intervals.

to check whether someone was in before setting up a longer interaction. The previous DVC prototype offered much of the same capability as Montage, but it had an interface modeled after the phone. The data suggest that the Montage interface encouraged more frequent and shorter connections than our previous DVC prototype.

Distinguishing glances from visits. The design of Montage offered two levels of audio-video interactions: glances with small video windows and access to coordination tools, and visits with large video windows and access to ShowMe Whiteboard and Stickups. Of the 302 interactions during which audio was enabled, 90 (30%) stayed in the small video windows of a glance until it ended. The other 212 (70%) resulted in a visit, either by first enabling a glance and transforming it into a visit, or by going directly into a visit.

Figure 5 also shows the distribution of glance and visit durations. Glance duration was measured from the time the audio was enabled until the glance window was dismissed. Visit duration was measured from the time audio was enabled until the larger visit window was dismissed, including any time spent in a glance before entering the visit. The average duration of glances was 0:44, with the longest glance lasting 4:59. The median glance duration was 26 seconds, indicating that most of the glances were very short. The average visit lasted 4:10, which was considerably longer than the average glance. Visits ranged in duration from 0:04 to 47:10, with a median duration of 1:57. The distribution shows that most visits were also short with a sparsely scattered tail of longer interactions.

Thus the data show some evidence that, overall, glances were used for shorter interactions and visits for longer interactions. However, some users commented that they routinely went into a visit because the glance windows were too small to provide enough information. Both the logs and the videotapes confirmed that this was a common usage pattern. The utility of the glance windows may have been diminished by the rather low quality video images available in this prototype. The user interface of Montage made it easy to move into a visit, so even those who did not use glances were not hindered by the interface.

Perceptions of Montage interactions. In the surveys, users reported that they tended to use Montage as the need arose, rather than waiting for enough issues to accumulate to justify the effort of physically visiting someone. One user commented about Montage:

I think it also helped efficiency, since I would contact someone right when I came across the need for information. Previously, I would collect up a number of issues and walk over to the other building and do all my business there at one time. I often forgot about things when I put them off.

The users also reported that their communication improved. One user commented: “I think decision making was faster using Montage. Certainly the e-mail process took too long in that regard.” Another said: “I had meetings I wouldn’t have otherwise had (because I was too lazy to walk to another building). Made impromptu meetings much easier. As a result, communication was improved.” After we removed Montage, half the team reported that they were walking and driving around more to visit people. Unfortunately, we do not have any reliable measures to corroborate these perceptions.

These comments indicate that the participants felt comfortable using Montage for some interactions that they traditionally reserved for face-to-face visits rather than the phone. It is surprising that, having learned the benefit of resolving issues quickly using Montage, once it was removed they did not continue to do so over the phone. The main difference between Montage and the phone is the visual access provided by the video and the approach interface. It appears that these differences encouraged the users to think of Montage interactions more like physical visits than phone calls. On the other hand, survey responses indicated that the users tended to prefer face-to-face visits over Montage for private or sensitive conversations.

Low usage of ShowMe Whiteboard. From a visit, Montage also integrated access to the ShowMe Whiteboard shared drawing application, which enabled the two users to share and mark over an image on their screens. The Montage logs showed 11 instances of launching ShowMe Whiteboard from within a Montage visit. Unfortunately, starting up and connecting ShowMe Whiteboard from the Montage prototype was unacceptably slow, often taking over a half minute. We suspect that this slow performance discouraged more frequent use of this feature.

Despite the very small amount of usage, responses from the surveys indicated that three of the ten users especially liked having access to ShowMe Whiteboard. Most of the glances captured on videotape revealed instances when it would have been useful to quickly share a view of something on their computer screens. For example, users looked at the same e-mail message together or synchronized their views of an application as one person directed the other on how to

use it. Although the slow start-up performance of ShowMe Whiteboard from Montage discouraged people from using it, we found evidence for the need to have quick access to shared drawing and viewing.

Other communication activity was mostly unchanged

We expected that the introduction of Montage would affect the group’s use of other communication tools. It seemed possible that the group would leave less voice-mail or have fewer scheduled meetings, for example. However, when we did an analysis of variance to compare their use of other communication methods before, during and after using Montage, we found that only their use of voice-mail changed significantly ($F(2, 88) = 11.01, p < .001$). Table 1 shows that this change was due to a drop in voice-mail usage after we removed Montage. It is unclear whether this change in voice-mail was related to the removal of Montage or other factors in the group’s work.

	Pre-Montage	Montage	Post-Montage
Voice-mail	5.7	6.0	4.1
E-mail	5.4	6.2	6.4
Meetings	3.1	3.2	2.2
People/mtg	2.9	2.8	2.7

Table 1. Average frequency of communications per person per day and average number of people per meeting across the three conditions. Only voice-mail showed a significant change.

We found no evidence that the number of e-mail messages per person per day changed over the three conditions ($F(2,93) = 2.01, ns$). This result is in contrast to our previous study, in which we found a drop in e-mail usage when the group was equipped with the DVC prototype [Tang & Isaacs, 1993]. One reason that this group may not have changed its e-mail usage is that they used e-mail to document their activities. We also found no differences across conditions in the average number of meetings scheduled per day ($F(2,91) = 2.94, ns$), or in the average number of people attending each meeting ($F(2, 278) = 0.24, ns$).

These results roughly corresponded with the participants’ perceptions of their activity. When we asked them at the end of the Montage condition whether their use of other communication media had changed, six of the ten said they thought e-mail frequency had not changed and seven said scheduled meetings had not changed (the remainder said activity had declined). However, six said they thought voice-mail had declined and one said it had increased, when in fact it had not changed between the pre-Montage and Montage conditions ($t(69) = 0.46, n.s.$). Of course, there is some possibility that voice-mail among group members had declined but that the drop did not significantly affect their overall voice-mail activity.

One might expect that Montage would be most likely to affect the frequency of phone calls and face-to-face visits, but as explained, we were unable to collect those data. Still, it is notable that by the end of the Montage condition, most participants perceived a drop in these two activities. Eight said that face-to-face visits had declined and seven said that phone usage had decreased.

Thus, introducing Montage did not completely displace the use of their existing collection of tools. Rather, the group incorporated Montage along with their other tools, using each medium to its advantage. Even when they might have discussed something over Montage, for example, they still might have followed up with an e-mail message to keep everyone informed. This finding reinforces the importance of integrating novel communication tools with existing ones, in this case by having them easily accessible on their workstation.

Privacy concerns

When we interviewed the group to learn about their initial reactions to the prospect of using Montage, quite a few were concerned about the effect it would have on their privacy. One person almost declined to participate for this reason. Three weeks after Montage was installed, we asked the group members in the surveys whether they felt Montage had reduced their privacy. Only the person who had been initially hesitant to participate felt his privacy had been reduced, specifically during meetings with other individuals in his office. Another user who had initially been particularly concerned said the reduction in privacy was “not as much as I was worried about.”

One of the ways that Montage enabled users to maintain their privacy was through the access modes (e.g., do not disturb, out of the office). The usage logs indicate that over the 12 week period, the group members changed their access mode 175 times, an average of 19.4 times per person. Group members’ usage varied quite a bit. Five of them only changed their access mode between 0 and 9 times, whereas the other five members did so between 17 and 44 times. The person who had been most concerned about privacy changed his accessibility 35 times.

Not surprisingly, about half the time (84 times) the group members set their access to ‘available.’ That is, after they had changed their access, they usually reset it to become ‘available’ once again. Another 56 times, they set access to ‘out of the office’ or ‘other,’ and wrote in a short note to indicate where they were (usually at a meeting, at a specific event, at lunch, or gone home for the day). Another 32 times, they set their access to ‘do not disturb.’ Users locked their Montage only three times. It was also possible to reduce access by turning off the camera, locking the screen or quitting Montage for privacy reasons, but we were unable to log these events.

Despite the relatively frequent use of access modes, some users indicated that they did not always remember to use them. One user wrote, “as a manager, I have one-on-one and other very private meetings. But unfortunately, being human I forget to turn the camera off most of the time.” Nonetheless, it is interesting that when users remembered to use access modes, it was usually because they wanted to provide *more* information about their whereabouts.

A few also mentioned some ways in which the openness of Montage changed their behavior. One user commented that, “a glance via Montage seems more of an interruption than someone walking down the hall and peeping in. Without interrupting my work space, people can look into my office and see that I’m busy.” Two users said they felt more obligated to respond to Montage glances than to other interruptions. And finally, a few mentioned that they thought the phone was more private than Montage because it allowed them to speak directly into the handset rather than talking at their computer (usually at a higher volume). One user reported, “If the conversation was sensitive or about someone, I’d talk over the phone.”

COMPARING MONTAGE WITH OTHER INTERACTIONS

Within the videotape data, we captured 30 Montage glance attempts (15 of which were acknowledged), 48 face-to-face visits, and 83 phone calls. The videotape data, along with the group’s survey responses, provided some information about the differences between Montage interactions, face-to-face visits and phone calls.

The value of video in Montage interactions

One of our goals was to understand the role of video in Montage interactions. Given the expense of providing video, it is important to understand its benefits in supporting communication among a distributed group.

We saw several examples of the subtle benefits of the video channel to convey non-verbal cues. Users were able to detect cues from body language and facial gestures, and used those cues to work through disagreement and sensitive issues. We refer to previous studies of video conferencing that have discussed these benefits [Isaacs & Tang, 1993; Heath & Luff, 1992] for more elaboration. These non-verbal cues seem to have played a major role in giving a natural and familiar feel to the group’s interactions in Montage.

The users also remarked that the video provided an ongoing indication of each participant’s attention. Since Montage users are each located in their own office, they have the potential to engage in many other activities and distractions. Some users commented that it can be annoying in phone conversations when they can hear that their partner is attending to other things. During face-to-face visits, participants are less likely to engage in other activities unless they are doing so to signal a desire to end the interaction. When

interruptions do occur in face-to-face interactions (e.g., a visitor drops by, the phone rings), they are typically self-evident from the visual cues, making it easier for those involved to manage the interruption. The videotapes show some of these characteristics of visual access in action in Montage interactions.

Managing “visual interruptions.” The video channel enabled participants to interpret an interruption without an explicit explanation. For example, in a Montage visit captured on videotape between team members B and J, the Montage video view revealed that a visitor was interrupting J. J made a series of facial gestures toward her door, each lasting less than three seconds. B responded at first by stretching out his utterance and then pausing briefly, but then B continued with his remarks. In this instance, B could tell from his video view of J that she was handling a brief interruption, but that the interruption did not warrant a break in the conversation. Conversely, B’s video presence in J’s office probably kept the interruption to a minimum because it indicated to J that B wanted to continue the conversation. The video channel enabled J to effortlessly negotiate what we call a “visual interruption.”

Visual interruptions are similar to verbal interruptions, which occur when a third party breaks into an ongoing conversation with a short interaction. Such interruptions are typically unremarkable because they do not derail the flow of the main conversation [Grosz & Sidner, 1986]. Visual interruptions are accomplished through gestures rather than words. Verbal and visual interruptions occur naturally in face-to-face interaction, but visual interruptions can be problematic over the phone because not everyone can see the visual cues of an interruption occurring. Stopping to provide an explicit verbal explanation (e.g., “just a second, someone just came by”) can be more disruptive than handling the interruption visually.

The interaction between B and J illustrated that participants in a Montage glance can interpret interruptions that occur in each others’ offices. The videotapes also captured instances of using Montage to accomplish a visual interruption, such as glancing at someone who was talking on the phone at the time. The glance made it clear that the person was busy, and it was sometimes used to convey a gesture to indicate that they had seen the glance. Meanwhile, the phone conversation continued without disruption. Being aware of the glance attempt usually prompted a follow-up glance after the phone call was completed. One user reported another type of example, which occurred when he glanced to see if someone was ready to go to lunch: “He was on the phone. If I’d only called, I would have gotten voice mail. Instead I got a sign-language ‘one minute’ response.” In this case, the purpose of the glance was completed without verbal interaction and again without disrupting the phone conversation.

The fact that we saw several examples of visual interruptions in such a small sample of videotaped activity suggests that interruptions in the office can be fairly common. The video channel gives users an additional resource to handle these interruptions.

The drawbacks of video. Of course, visual access does have its disadvantages. Users remarked that they were used to doing other work while talking on the phone and they felt restrained from doing so when glanced. In this case, users liked video when it commanded more attention from their conversation partners, but they did not like it when it forced them to pay more attention to their partners. Another problem is that visitors may not realize when a glance is in progress, especially when the computer screen is not visible from the office doorway. Users commented that they felt awkward “talking to their computer,” especially when people passing their office could not see that they were interacting with someone through audio-video connections on their computer.

Montage interactions have a distinct structure

The videotape data allowed us to compare the structure of interactions in Montage with those over the phone and face-to-face. Some of these structural differences suggest that audio-video connections share some aspects of face-to-face interaction and some of phone calls. These differences indicate that audio-video links offer a distinct communication medium between face-to-face visits and the phone, with its own advantages and disadvantages.

Comparing the duration of interactions. Using the interactions captured on videotape, we compared the average duration of Montage glances, Montage visits, phone calls, and face-to-face visits. However, there is some uncertainty in this comparison because the Montage interactions included only those people within the work group, while the phone calls and face-to-face visits included anyone who contacted them in their offices.

The average duration of the various modes of interaction were: 0:44 for Montage glances, 4:10 for Montage visits, 4:58 for phone calls, and 6:39 face-to-face visits. Montage glances and visits may have been shorter than face-to-face visits because participants used Montage for issues as they arose, rather than waiting for enough issues to accumulate to justify a physical visit. Interactions in Montage also seemed more task-focused and included less small talk. Alternatively, users might have perceived Montage as somewhat fragile (because of its reliability problems), or the poor audio might have encouraged them to complete their interactions in Montage as quickly as possible.

The structure of openings and closings. Opening and closing conversations are accomplished by well-recognized rituals [Clark, 1985]. Comparing openings and closings in Montage interactions, phone calls, and face-to-face visits

illustrates some of the structural differences among the interactions.

Clark identifies three stages to openings:

- contact initiation — mutually recognizing an intent to communicate,
- greetings — acknowledging each other’s acquaintance, and
- topic initiation — introducing the first topic.

Likewise, there are three stages to closings:

- topic termination — mutually recognizing the lack of further topics,
- leave-taking — confirming the conversation end, and
- contact termination — disabling the means of contact.

Our comparison focuses on the first two opening stages of contact initiation and greetings, and the last two closing stages of leave-taking and contact termination. The openings and closings of all interactions that were captured on videotape were transcribed for comparative analysis.

In face-to-face visits in an office setting, contact initiation is established by approaching an office and establishing eye contact with the occupant. Greetings (e.g., “Hi Ellen, how are you?”) typically confirm that you know each other, although this stage is often skipped or abbreviated among people who are familiar with each other because seeing each other usually establishes mutual acquaintance [Clark, 1985]. In closing a conversation, leave-taking (e.g., “Goodbye”) is also often abbreviated or done non-verbally (e.g., walking toward the door). Whittaker et al. [1994] found that informal interactions in an office setting tend not to include formal greetings or goodbyes. Terminating contact is typically accomplished by leaving the office, although this process usually allows some time to restart the conversation if the need arises. In fact, the closing process can be fluidly accomplished by standing up to indicate that you have finished the topic, gradually moving toward the door to confirm the end of the conversation, and walking away to end the contact.

In a phone call, contact initiation is accomplished when the caller dials a number and the recipient answers the phone. In office settings, people often combine answering the phone with a greeting (e.g., “Hello, this is John”), to which the caller usually responds with her identity (e.g., “Hi John, this is Ellen”). Leave-taking is typically marked by at least one exchange of goodbyes, and hanging up the phone terminates the contact. We observed that phone closings often exhibited more than one exchange of leave-taking (e.g., “OK, thanks”, “OK”, “Goodbye”, “Goodbye”), perhaps because hanging up the phone is so explicit and irreversible.

In Montage, contact initiation is started by selecting a person to glance, and once the glance fades in, either person can

respond by enabling the audio. The usage logs show that enabling the audio is fairly evenly distributed between the person glancing (52%) and the person being glanced (48%). This is contrasted with the phone, where the person being called must answer the phone before contact is established. As in face-to-face interactions, greetings in Montage are usually abbreviated or skipped because both parties can see each other. Unlike face-to-face interactions, closings in Montage usually included some formal exchange of goodbyes. These explicit goodbyes probably occur because contact is terminated by pushing a button on the user interface, which immediately breaks the audio-video connection. Like the phone, terminating contact in Montage is abrupt and irreversible.

We noticed one problem caused by these explicit endings in a videotaped glance between J and S. J prematurely ended the glance, just as S started saying something else. J had to glance back to S to complete the interaction. This problem suggested redesigning the Montage interface so that closings in Montage interactions could more closely follow face-to-face closings. Perhaps ending a glance or visit could stop the audio and cause the video window to fade out over a few seconds, during which either party could reactivate the glance by pressing a button.

We observed a problem in Montage greetings that was caused by a limitation in the technology. Glances are made by connecting audio and video between the machines as quickly as possible, but those connections were not guaranteed to be exactly simultaneous. We observed that there could be a second or two when one party could hear and see before the other. By nature, greetings tend to be exchanged in pairs [Schegloff, 1972], and in Montage they were often very short (e.g., “Hi”, “Hi”). These non-simultaneous connections sometimes confused the greeting ritual. Person A would say “Hi”, but person B would not hear it because the audio had not been connected. B would then say “Hi” and wait for A’s completion of the greeting exchange. Meanwhile, A, thinking the greeting was already completed, would wait for B to initiate the first topic. This awkwardness was usually repaired easily, but it indicates that reciprocal audio-video connections should be made simultaneously.

Comparing with caller-id phones. The company phone system offers some users a phone with a display screen that identifies the caller’s name if the call originated from within the company (caller-id phones). Eight of the ten group members had caller-id phones. Caller-id capability offered some but not all of the advantages of Montage (presumably at less expense).

Because caller-id phones revealed the caller’s identity, greetings could be shortened from the usual phone greeting (e.g., “Hello, this is John”) to a more abbreviated form typical of face-to-face and Montage interactions (e.g., “Hi Ellen”). The videotapes captured one example in which

caller-id even enabled the receiver to infer the topic of the call. The person answered a call from her manager by saying: “I’m doing it right now!”

However, since caller-id is based on a database of names and numbers, it does not offer the same degree of positive identification as seeing an image of the person. If the caller used someone else’s phone, some amusing mis-identifications can occur (e.g., “Hi Monica. Oh, it’s Ellen?! The phone says you’re Monica”). The videotapes captured one example when J received a call while she was on another line. She checked the identity of the incoming call but, for some reason, it provided an incorrect name. When she finished the first call, she called back the person indicated by caller-id, and the resulting conversation was confused and somewhat embarrassing.

Of course, caller-id does not offer visual access and so precludes the handling of visual interruptions and gestural cues discussed earlier. It is important to note that the users had caller-id phones prior to this study of Montage. Thus, the group’s perceptions of Montage’s benefits compared to the phone already took into account caller-id capabilities.

The interaction context. In contrast to face-to-face visits, both participants in Montage are typically located in their own offices. This is often considered an advantage in phone calls because both people can access their own office resources and be available for urgent interruptions. The videotapes confirm that many glances involved accessing these resources (e.g. documents, calendars, artifacts). Since Montage appeared on their computer screen, it was particularly convenient to interact with other computer-based resources while involved in a glance.

Although Montage does provide visual access to the participants, it is more limited than the shared visual space during face-to-face contact. The video cameras typically gave a view of the other person’s head and shoulders. Although the users could easily redirect their cameras to show other people or objects in the office, they did so only occasionally. Many hand gestures were missed unless they were brought into camera view. Furthermore, users typically had to stay seated in front of their computers to be seen. This limitation reduces the ability to use certain postural cues, for example standing up to initiate the end of a conversation.

Although Montage was designed to support connections between only two workstations, the videotapes captured several instances when the participants included a visitor who happened to be in one person’s office when a glance arrived. Sometimes being able to see visitors through the video prompted their inclusion in the interaction. More often, their participation was encouraged by the “open-air” audio channels — the visitor could hear the conversation in the glance and join in with minimal effort. This arrangement is in contrast to phone conversations, where the phone

handset keeps the conversation private unless the call is switched to the speakerphone.

Montage, like face-to-face interaction, allows others to join an interaction through its open video and audio channels. Whittaker et al. [1994] found that many dyadic interactions in the office expand to include more people. Reports from the surveys indicated that users had a need for three-way Montage interactions, which the prototype did not support. Users worked around this limitation by having two people gather in one office and glance at the third, or using the phone to add the third party. A system like Montage should support at least three- and four-way connections.

GLANCING AT WHAT WE HAVE LEARNED

This study of Montage in a real work setting has taught us much about how users respond to such a tool and the potential role it can play within a distributed group. We also learned more about the issues involved in conducting such studies and the design of effective networked multimedia applications.

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 allowed rich interactions and also discouraged attending to other distractions while in an interaction. Although the visual access helped people interpret those interruptions and transitions that did occur, it did not provide the full range of flexibility possible in face-to-face interactions. Like both the phone and face-to-face settings, only a small portion of their attempts to reach each other actually resulted in an interaction.

This characterization, along with users’ enthusiasm for Montage, indicate that tools that provide lightweight audio-video connections can be effectively integrated into existing work settings in which group members are distributed. Not only can they provide a sense of awareness, but they seem to encourage people to contact each other just when an issue arises rather than waiting until they physically visit each other. People still rely on their existing communication mechanisms, but they have more flexibility in using the right medium for the task.

Methodologically, this study underscored the value of collecting a variety of perspectives on the work activity being studied. By combining quantitative measures, users’ perceptions, and analyses of videotaped samples of activity, we learned about reactions to Montage at many different levels. Some of our conclusions were limited because we did not have access to some key measures of activity, such as the frequency of phone calls and impromptu face-to-face visits. However, to the extent that some objective measures of

human activity are inherently inaccessible, it is important to invest the time and effort to gather qualitative measures of work activity to get a complete picture of users' reactions.

Our experience with Montage raised several technical requirements for providing effective interactive multimedia applications. Audio connections must be free of feedback and annoying echo. Connections among workstations must be quick to initiate, simultaneous, and capable of sustaining the large amounts of data generated by sharing audio, video and computer applications. Capabilities for shared drawing and viewing must be quickly and easily available. Mechanisms and policies for integrating these multimedia interactions with other forms of communication (e.g., notes, phone) are needed. Mechanisms and interfaces must be developed to support the n-way interactions that often occur in groups.

Perhaps even more challenging are the many social issues raised by this technology. How does the functionality of providing lightweight connections among people scale up in large organizations? What social mechanisms need to be developed to respond to this new communication channel? As this technology develops, we must continue to explore these questions by studying how real groups use interactive multimedia.

ACKNOWLEDGEMENTS

We especially thank the ten users who allowed us to study their work activity while we deployed the Montage prototype. We also acknowledge the help of the rest of the Collaborative Computing (COCO) group within SunSoft, especially Alan Ruberg for implementing all of the digital video effects and helping support the Montage prototype during the study. Sun, SPARCstation, Solaris, and ShowMe Whiteboard are trademarks or registered trademarks of Sun Microsystems, Inc.

REFERENCES

Clark, Herbert H., "Chapter 18: Language Use and Language Users", *The Handbook of Social Psychology*, Gardner Lindzey and Elliot Aronson (Eds.), New York: Harper and Row, 1985, pp. 179-229.

Dourish, Paul and Sara Bly, "Portholes: Supporting Awareness in a Distributed Work Group", *Proceedings of the Conference on Computer Human Interaction (CHI) '92*, Monterey, CA, May 1992, pp. 541-547.

Fish, Robert S., Robert E. Kraut, Robert W. Root, Ronald E. Rice, "Video Informal Communication", *Communications of the ACM*, Vol. 36, No. 1, January 1993, pp. 48-61.

Grosz, Barbara J. and C. L. Sidner, "Attention, intentions,

and the structure of discourse", *Computational Linguistics*, 12, 1986, pp. 175-204.

Heath, Christian and Paul Luff, "Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction", *Human-Computer Interaction*, Vol. 7, 1992, pp. 315-346.

Isaacs, Ellen, A. and John C. Tang, "What Video Can and Can't Do for Collaboration: A Case Study", *Proceedings ACM Multimedia '93*, August 1993, Anaheim, CA, pp. 199-206.

Kraut, Robert E., Carmen Egido, and Jolene Galegher, "Patterns of contact and Communication in Scientific Research Collaboration", *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, Jolene Galegher, Robert E. Kraut, Carmen Egido (Eds.), Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1990, pp. 149-171.

Rice, Ronald E. and Douglas E. Shook, "Voice Messaging, Coordination, and Communication", *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, Jolene Galegher, Robert E. Kraut, Carmen Egido (Eds.), Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1990, pp. 327-350.

Root, Robert W., "Design of a Multi-Media Vehicle for Social Browsing", *Proceedings of the Conference on Computer-Supported Cooperative Work*, Portland, OR, September 1988, pp. 25-38.

Schegloff, Emanuel A., "Sequencing in Conversational Openings", *Directions in Sociolinguistics: The Ethnography of Communication*, J. J. Gumperz & D. H. Hymes (Eds.), New York: Holt, Rinehart, & Winston, 1972, pp. 346-380.

Tang, John C. and Ellen A. Isaacs, "Why Do Users Like Video? Studies of Multimedia-Supported Collaboration", *Computer Supported Cooperative Work: An International Journal*, Vol. 1, Issue 3, 1993, pp. 163-196.

Tang, John C. and Monica Rua, "Montage: Providing Teleproximity for Distributed Groups", *Proceedings of the Conference on Computer Human Interaction (CHI) '94*, Boston, MA, April 1994, pp. 37-43.

Whittaker, Steve, David Frohlich, and Owen Daly-Jones, "Informal workplace communication: What is it like and how might we support it?", *Proceedings of the Conference on Computer Human Interaction (CHI) '94*, Boston, MA, April 1994, pp. 131-137.