Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication

Robert E. Kraut, Mark D. Miller, Jane Siegel
Human Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA, U.S.A. 15213
Robert.Kraut@andrew.cmu.edu
Mark.D.Miller@andrew.cmu.edu
Jane.Siegel@cs.cmu.edu

ABSTRACT
We report an empirical study of people using mobile collaborative systems to support maintenance tasks on a bicycle. Results show that field workers make repairs more quickly and accurately when they have a remote expert helping them. Some pairs were connected by a shared video system, where the video camera focused on the active workspace and they communicated with full duplex audio. For other pairs, either the video was eliminated or the audio was reduced to half duplex (but not both). Pairs' success at collaboration did not vary with the communication technology. However, the manner in which they coordinated advice-giving did vary with the communication technology. In particular, help was more proactive and coordination was less explicit when the pairs had video connections. The results show the value of collaboration, but raise questions about the interaction of communication media and conversational coordination on task performance.

Keywords
Wearable computers, empirical studies, collaborative work, conversation, media effects, vehicle maintenance.

INTRODUCTION
This paper investigates the design and value of collaborative, mobile computer systems to support field repair and maintenance of mechanical devices. When a pilot radios in a possible malfunction in an aircraft, the field crew has only a short time to diagnose and repair the problem. Smailagic and Siewiorek [18] have described mobile maintenance systems, which incorporate computers, on-line repair manuals and diagnostic aids. These systems could give field workers access to schematics, repair histories, instructions, parts lists, and other information about the equipment being repaired. With the addition of telecommunications, they can also provide access to remote experts. Designers expect that these systems will be useful for aircraft and other mechanics as well as emergency personnel and other workers who face unusual field conditions and could be supported by documents, data, and helpers that are not at hand.

Incorporating an interpersonal communication component in what are otherwise lightweight, portable computers makes their design substantially harder. Empirical field research has documented the value of collaboration when workers are troubleshooting and repairing complex equipment like elevators [2], copiers [13], or telephone equipment [14]. Based on this work, we presume that the addition of communication facilities would be valuable, although we know of no empirical tests of this proposition.

Some designers have recommended the use of video as the basis of a shared workspace between a field worker and a remote helper (e.g., [11]). The video could be used to broadcast current conditions in the field to an expert staffing a help desk. For example, an emergency medical field system developed by British Telecom and ABB enables physicians in a hospital remotely to supervise paramedics as they conduct sophisticated procedures on accident victims [1]. Yet, video substantially increases the costs of mobile systems and their technological challenges.

Numerous studies of video telephony for white-collar, office tasks have failed to show that video improves task performance in this domain ([3] [12]; see also [20] for an old but still useful review). The applicability of this earlier research, however, may be low, since most video telephony systems use a "talking heads" model, in which the cameras broadcast pictures of the people in conversation. It is possible that video focused on what a worker is doing rather than on the worker may improve communication, by grounding the conversation in a shared view of the rapidly changing world.

How people ground their conversation and how they exploit features of communication medium to effectively coordinate their speech with each other are

Computer Supported Cooperative Work '96 , Cambridge MA USA
© 1996 ACM 0-89791-765-0/96/11 ..$3.50

important theoretical issues in their own right. In addition, answers to these questions can also provide a theoretically motivated basis for many design decisions. Both Clark [5] and Krauss and Fussell [10] argue that in coordinating speech, conversationalists need to achieve mutual knowledge or common ground. That is, when Person A forms an utterance for Person B, A uses the history of the conversation, their shared context, their prior knowledge of each other, and other information to infer what B would understand. Thus, Adam can say to Eve, "We shouldn't have eaten *that* fruit" because they have a common history that leads Adam to infer that Eve and he will have the same apple in mind, and he knows that Eve will know this as well. Clark and Wilkes-Gibbs [5] argue that when a speaker offers an assertion, both parties to the conversation actively collaborate in making sure that the assertion is understood and that both parties know that it has been understood before the speaker can offer another assertion.

Different communication media have features that change the ease with which conversationalists achieve common ground and the methods that they use to achieve it. Krauss and his colleagues ([8] [9] [10]) have shown that speakers become more efficient in describing objects with repetition, because they can assume that their partner understands the shorthand they are using. This growth in efficiency, however, is disrupted by delays of the sort introduced by satellites or video compression, because with delay a listener cannot easily indicate what he or she has understood. Clark and Brennan [4], in a speculative article, tried to decompose the differences among communication modalities in terms of their features and the costs these features impose on achieving common ground. They noted, for example, several features of face-to-face communication that change the cost of achieving common ground: visual co-presence, which allows a speaker to refer efficiently to objects that are in the speaker's and listener's shared view; sequentiality of message arrival, which allows a speaker to know what a listener has already received before creating a new message; and ephemerality of messages, which means that speakers cannot assume that listeners have access to all the details of a previous message.

We are interested in how these general principles of conversational coordination are deployed in the context of help-giving around a physical repair task. For help to be effective, workers getting the help must receive it when they need it and when their preconditions for understanding it and taking advantage of it have been met. To effectively give help, experts must continually make inferences about the worker's internal cognitive and motivational state and time their contributions to moments of readiness. In repairing a bicycle for example, the task we use in this research, an expert must continually assess whether a worker is having trouble, is finished with a previous subtask, will be able to understand the technical vocabulary that is the efficient shorthand for describing components, and has found the correct parts of the bicycle to apply the recommended procedures.

The cues that allow experts to make these inferences can consist of explicit communication activity on the worker's part or may be side effects of the worker's task performance, with no explicit communicative intent. That is, the worker may explicitly ask for help ("What's the next step"), ask for help using an indirect speech act ("Ok, I finished the first step"), or verbally indicate confusion about the task in such a way that the expert provides unsolicited help ("The wheel won't go in"). Or the expert may infer that the worker is ready for the next step in the procedure by listening to the sound of the tools or by directly observing the worker's performance.

The cues that experts have available are likely to vary with the technology for communication, for two reasons. Consider the effects of video on coordination. When the worker and expert share a visual workspace, the expert can receive feedback from the task itself to precisely time when he gives instructions and which instructions to give. Moreover, it is also likely that workers will change their explicit communication because they are aware that the expert can see their performance. In contrast, when they do not have a shared visual workspace, they need to describe what they are doing, because the expert has no other way of knowing. As a result, we expect workers will be more explicit in describing their state when the experts can not directly observe it, and that experts will be more explicit in asking for information about the workers state when they can't directly observe it.

In the research described here, we examine in detail the way that visual co-presence effects how people coordinate their conversation. By comparing pairs of people who share visual information about a task that one of them is performing with people who do not have this shared visual space, we can examine how people exploit visual co-presence to achieve common ground. Because we are using a new referential communication task, we also included the contrast between full-duplex audio and half-duplex audio as a technological variable, expecting to find as did previous researchers that half-duplex audio degrades the achievement of common ground [8] [9] [10].

In summary, the research we report on has two foci: (1) an applied focus on the development of mobile systems and (2) a theoretical focus on communication and coordination of work.

METHOD

Study Overview

We designed a study where unskilled workers performed three repair tasks on a ten speed bicycle : 1) replacing a seat on the saddle post, 2) adjusting the front brakes, and 3) installing the front derailleur. Workers always had access to an on-line repair manual. They either worked alone or with a remote expert. When they worked with the expert, they were connected or not with a video link between the worker and expert so that the expert could see what the worker was doing. We also varied the quality of audio connection between the worker and expert (either full or half-duplex audio). The experimental design was an incomplete factorial, in which we compared: (1) solo performance to collaborative performance and (2) collaborative pairs with video and full-duplex audio connections to pairs who were connected by full duplex audio without video or to pairs connected with video, but only half-duplex audio. Study participants were randomly assigned to a single treatment.

Apparatus

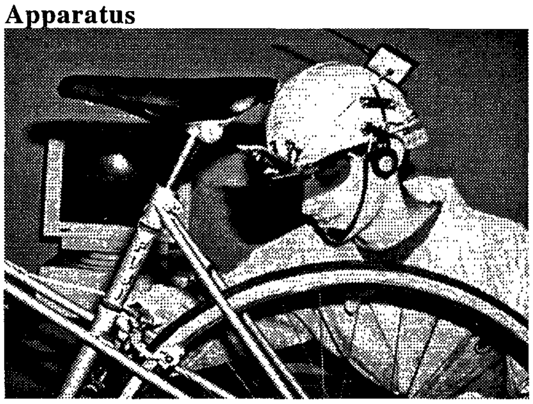

Figure 1: Worker wearing collaborative system. (Photo credit: David Kaplan)

Each worker donned the apparatus shown in Figure 1, a hard-hat on which were mounted various display and audio/video telecommunications devices.

The devices included a sports caster style Radio Shack 49 MHz microphone, headphones, and a tiny Virtual Vision VGA monitor mounted in front of the right eye, with optics that placed the image directly in front of the eye. The visual display had VGA (640x480 pixel resolution) and enabled the worker to view the on-line manual bicycle manual text, parts, and pictures. They also wore a small CCD camera mounted on the brim of the hat above the left eye.

In the video conditions, both the worker and expert could see the output from the camera on their screen and output from a camera focused on the face and upper torso of the remote expert, using Intel's Proshare video conferencing technology[1]. The worker's camera saw approximately what the worker was pointing his head at, which in many cases was what the worker was also attending to. There were problems with the angle of the camera, insufficient resolution for close-up views, variable frame rates (when the worker moved a lot, there was increased delay in video updating), and registration of the camera with what the worker saw. Despite these problems, the simple user interface allowed the expert to keep track of what the worker was doing most of the time. The view from the video conditions is shown in Figure 2.

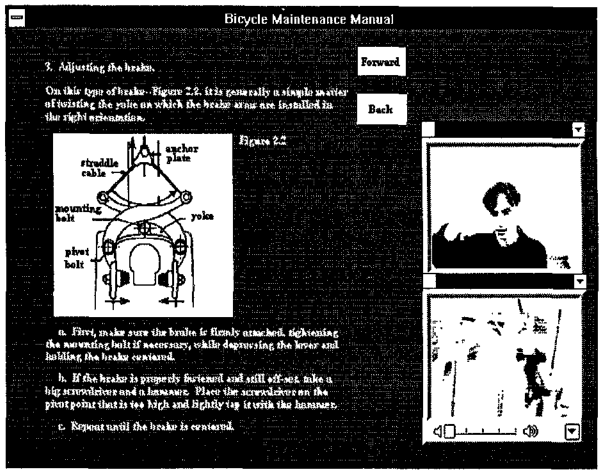

Figure 2. Display for the video conditions

The subjects were tethered by a 20 foot cable that provided power, audio input, and VGA input to the system. Also, they used a remote control/radio controlled trackball to navigate the shared information space with the expert.

The experts were in a separate room in front of a 17 inch VGA monitor and communicated with the worker through an audio link. In each experimental condition, workers and experts had the same view on their displays. The always saw the on-line manual, and both the worker and expert could control the cursor and flip pages.

Subjects

Sixty-four Carnegie Mellon University undergraduate students were recruited to participate in this experiment, of whom sixty completed it. One subject did not finish due to equipment failure and one subject failed a required eye test. Two withdrew from the experiment due to headaches. The students were participating in a subject pool for course credit. Sixty-nine percent of the participants were male. Sixty-three percent owned a bicycle. All subjects also received a

[1] Because Proshare introduces an audio delay, we by-passed Proshare for the audio.

candy bar and competed for a $20 bonus for the subject with the fastest completion time and highest quality task performance.

Two bicycle repair experts also participated in the study. They had both worked in a bicycle repair shop for over a year and were native speakers of English. The experts were tested for their ability to complete the experimental task by themselves and their ability to communicate the appropriate procedures. They were trained for the specific repairs they would advise on during this experiment. They were paid for their participation.

Procedure

Subjects reported to a room to complete consent forms, a battery of tests and questionnaires. Then subjects were taken to the experimental lab and suited up for the task with a jump suit to protect them from dirt. They also put on the hard hat shown in Figure 1 and a lumbar pack containing some of the components and controls. The hard hat was fitted on the subjects' head and adjusted by the experimenter so that it was comfortable and so that the camera tracked the workers gaze. Then the subject was given a brief eye test, reading several lines of text with diminishing font size from the head-mounted monitor. Next, subjects were instructed on how to navigate through the hypertext manual: they used two hypertext buttons, one for forward and one to go back. As soon as the subject could navigate, they were instructed and given a practice task to assemble a small part of a toy car. For the collaborative cases, the expert was introduced prior to the practice task. Subjects were given final instructions and introduced to the tool set they used (a flathead screw driver, crescent adjustable wrench, pliers, hammer, and socket wrenches).

Then the experiment began, with one experimenter in the same room as the worker, behind a computer used for real-time coding of communication behavior. In the solo condition the subject simply did each task. In the collaborative condition, the expert and subject could communicate at will, but the expert had the following rules to follow: (1) answer any question asked, (2) try to give the best answer, (3) if the subject was quiet for one minute, ask if they were doing alright, and (4) offer help or advice if the subject is doing or is about to do something incorrectly.

Following completion of each task, subjects were asked a set of questions by the experimenter about their experience during that task. At the completion of the three tasks, the subjects removed the equipment and returned to the initial room where they took a test of bicycle repair knowledge and completed a questionnaire about their experience. They were debriefed, given a candy bar and dismissed.

Measures

Demographic data for each subject included: age, gender, year in school, major field of study, and grade average. Data about subjects' experience and skill to carry out the repair tasks included: self-reported ratings of computer ability (familiarity, comfort, and average daily use), bicycle experience (ownership, familiarity, comfort, and average daily use), history in specific bicycle repairs (e.g., fixing a flat tire, adjusting the handlebars), as well as their familiarity and comfort with tools. Subjects completed the Eckstrom et al. spatial/visual abilities test [7]. They also completed a ten-item multiple choice instrument to measure bicycle repair knowledge, developed based on information from van der Plas [19] and Sloane [17].

An experimenter coded the worker and expert's speech in real time during the experiment, using MacShapa [15] Video and audio recordings of the sessions were the basis for verbatim transcripts and more detailed, post-experimental coding of the communication.

Post-session, on-line questions asked participants to describe their experiences using the mobile system. Workers rated their perceptions about ease of finding information, seeing the workspace, the ease of hearing the remote expert, being heard by the expert, and being understood by the expert.

Measures of task performance include the number of the tasks that the workers completed, participants' work time for each task they completed, and repair quality of each task they completed. To assess repair quality, both the experimenter and the session expert checked the worker's repair against a checklist, assessing such details as whether the saddle was level to the ground and tight and whether the brake anchor was set correctly.

RESULTS

We present the results in two parts, first examining the effects of collaboration and communication media on measures of task performance and then examining the effects of media on communication patterns among the collaborative pairs. The examination of communication includes data from the real-time coding, more detailed, post-session coding of a single subtask (attaching a brake straddle cable), and several vignettes that illustrate the way that workers and experts coordinated their speech.

Performance

Workers' self-rated mechanical expertise, their gender, and their spatial abilities were associated with their speed and accuracy in performing the repair tasks, and together explain 5% of the variance in the number of tasks completed, 25% of the variance in the log of time to complete the tasks and 20% of the variance in the quality of their repairs. They were included as control variables in the analyses that follow.

Workers do substantially better at performing these repair tasks with collaborative help. Average time to complete the tasks with a remote expert was half as long as in the solo condition (7.5 versus 16.5 minutes respectively; $t(54)=4.54$; $p < .001$). In contrast, being a standard deviation more mechanically skillful than average only reduced completion times by 2.4 minutes ($t(54)=2.87$; $p < .01$). More of the workers who had a remote expert completed all three repair tasks than those who had no expert (93% versus 62%; $p < .01$). Finally, the quality of the repairs they completed was superior when they had a expert than when they worked solo (79% of the quality points for the collaborative condition versus 51% for the solo condition; $t(54)=6.50$; $p < .001$).

While having access to an expert dramatically improved performance, having better tools for communication with the expert did not improve the number of tasks completed, the average time per completed task, or performance quality (for all 3 dependent variables, $F(2, 42) < 1$; $p > .5$). In particular, neither video (the comparison of the full duplex/video condition with the full duplex/no video condition) nor full duplex audio (the comparison of the full duplex/video condition with the half duplex/video condition) helped workers perform more tasks, perform tasks more quickly, or perform them better.

Communication

While technology in this experiment did not change performance, it did influence how workers and experts coordinated their activities. We concentrate here on the effects of video (i.e., the contrast of the full duplex/video condition with the full duplex/no video condition) on experts' ability to give effective help. When video is present, the worker and expert have a similar view of what the worker was doing, on a moment by moment basis. Our goal in this section is to understand how this common view changes coordination of conversation.

Real-time speech coding. To examine how media changes coordination strategies, a single experimenter coded experts' and workers' communication behavior in real-time, during the study, using the following codes:

- Workers' questions about the task or technology.
- Workers' descriptions about the state of the task or technology.
- Experts' questions about the workers' state.
- Experts' help about how to perform the task.
- Worker or Expert acknowledgments of their partner's communication.

Speech act	Video	No Video	*p*
Worker questions	34.8	28.5	.44
Worker state descriptions	40.3	55.2	.02
Worker acknowledgments	32.9	34.6	.72
Expert questions	13.7	16.8	.30
Expert help	111	93.7	.24
Expert acknowledgment	7.67	13.4	.003
Probability of worker question being followed by expert help	.95	.92	.26
Probability of worker description being followed by expert help	.56	.42	.001

Table 1: Speech acts by the presence of video
Note. Real-time coding of three bike repair tasks.

These codes represent over 90% of speech activity during a session. They describe in a very rough way what worker and expert were talking about, but they do not differentiate some interesting speech behaviors (e.g., whether a description or help statement was proactive or a reaction to a prior speech act). Inter-rater reliability analyses of the real-time codes based on recoding of 3 videotapes by five judges show Cohen's kappas in the mid 40s for the set of 6 codes. As one would expect from coding done in real time, judgments are substantially more reliable than chance, but contain a substantial amount of error. The practical consequence of both the small number of differentiations made and their relatively low reliabilities is that we will be able to observe only gross differences in the communication behavior by technology condition. This dataset does not have the power to detect subtle differences among conditions.

Table 1 shows the average number of speech-acts in the full-duplex video and no-video conditions and the probability that difference between conditions could have occurred by chance. A comparison between the two conditions suggests how a shared visual space influences conversation.

As seen in Table 1, in the no-video condition, workers were more explicit in describing their task (i.e., more state descriptions). More of the content of their communication was about their state, e.g., what tool or part they were holding, what they had just completed, and what they were seeing. Second, experts without video were more likely to acknowledge worker comments (e.g., to utter "yes", "uh huh"), as if they needed to be explicit about having heard the workers questions and descriptions.

The bottom two rows in Table 1 show the conditional probability of one speech act following another. The differences in the probability of workers' description being followed by expert help in the video and non-video condtions, shows that experts treated a worker's description of state differently depending on whether video was available. When video was available, experts seem to treat workers' descriptions of state as implicit request for help; they followed with help in 56% of the cases. On the other hand, since the descriptions were more necessary for simply tracking what workers were doing in the no-video condition, they were followed by help in only 42% of the cases. In contrast, an explicit request for help was followed by help in over 92% of cases, and this contingency didn't differ by experimental condition.

Speech coding of the straddle-cable subtask. To examine these phenomena in more detail and to improve the reliability of our coding, we used videotapes and transcripts to code a single subtask—attaching a brake anchor plate to a straddle-cable connecting the two brake pads. As can be seen in Figure 2, workers needed to latch the lip on the back of the anchor plate over the straddle cable. To do this, they needed to create slack in the cable, either by releasing the brake quick release lever (not shown in Figure 2) or by pressing the break pads towards the tire rim.

We divided workers' state descriptions and experts' help, to distinguish reactive cases, in which the speaker was responding to an explicit request for information from a partner, from proactive cases, in which the speaker seemed to initiate the description or help. In addition, we split the experts' help into cases of instruction, in which he was telling the worker what procedural steps to take, from clarifications, in which he attempted to clarify the language he had previously used.

Turn		Speech
1	E.	Then you can hook the plate on. Isn't there hooks on the back of that plate?
2	W.	Hook in the back of the plate?
3	E.	The back of the, that anchor plate is what it's called. That you're holding there in your right hand?"

Table 2. Conversational fragment.

In the example in Table 2, turn number 1 is an instruction, while both sentences in turn number 3 are clarifications, in which the expert informs the worker of the name of the plate that he has previously referred to.

Again we see that people coordinate differently depending on whether they have video present. In the straddle-cable subtask, all of the action is in the expert's behavior. When experts could see the worker, they gave more proactive help (i.e., without the worker explicitly asking for it), presumably because they could see when the worker was finishing a task or when the worker was having trouble. In addition, they were more likely to clarify and elaborate their prior instructions, presumably because they could better understand when the workers were understanding them and when they were not. Finally, they were less likely to offer acknowledgments. This last finding is inconsistent with the data from the real-time coding and deserves futher investigation.

Speech act	**Video**	**No Video**	***p***
Worker questions	2.0	1	.26
Worker proactive state descriptions	2.0	1.1	.70
Worker reactive state descriptions	.60	.7	.92
Worker acknowledgments	4.5	3.4	.35
Expert questions	1.2	.8	.31
Expert proactive help	7.4	3.8	.03
Expert reactive help	1.2	.8	.48
Expert clarifications	1.9	.04	.04
Expert acknowledgments	1.8	2.1	.04

Table 3 Speech acts by the presence of video Note. Coding of the straddle-cable subtask, based on transcripts and video tape.

Vignettes. We can illustrate the ways that people adjust their conversation to the resources they have available by looking at several cases in more detail. Tables 4 through 5 present annotated transcripts from two sessions in which the dyad had video and one session in which they did not. Our annotations are based on multiple viewing of the video and readings of the transcripts. While they are inferences about the workers' and experts' internal states, they represent the informed judgment of two observers who had repeatedly viewed these videotapes. They reflect our assessments of the inferences that the workers and experts must have been making to utter the words they said, in light of their other behaviors.

Transcript	Description
E. If it, if it, uh, won't seem to go on.	Expert visually establishes that workers is having trouble attaching straddle cable.
E. I didn't know this until yesterday. But there's sort of a quick release-mechanism for the anchor plate.	Expert establishes the precondition for giving instruction—term for and location of the anchor plate.
E. It's above it.	Expert visually establishes that worker knows what the quick release mechanism is and where it is.
E. You just kind of, uh, flick it over to the right and it will release a bit of slack.	Expert provides instruction.
E. Yeah.	Worker acknowledges he has understood the instruction. Expert visually establishes that the worker is manipulating the correct component
W. Oh, OK, I got it.	Worker announces he has completed this subtask.

Table 4. Precise timing of instruction on the basis of visual inspection.

Our examination of these video tapes and transcripts strongly suggest that experts were using the knowledge they gained from the video channel to determine what instructions to give the worker and to precisely time them. The experts could deliver appropriate instruction at an appropriate time when they did not have the video, but this required more effort. When they did not have video, both the worker and expert needed to encode in language the worker's state and the state of the task for timing of instructions to be effective. We make no claims that the vignettes we have analyzed are representative of our entire corpus. We present them to give the reader a richer sense of the mechanisms that the dyads use to coordination their speech, and the ways that the video channel influences coordination.

Table 4 is an example of a conversation from the video condition. In it, the expert infers that the worker knows what a quick release mechanism is, because the worker goes directly to it when the expert first mentions it. Thus, the expert relies on the video to infer that the worker understands what the quick release lever is and when the worker locates it on the bike. The expert modifies his instruction on the basis of the video.

In Table 5, the expert uses visual information to make inferences about a worker's approach to the task. Based on the worker's behavior, the expert first believes that the worker has identified the correct problem (increasing slack in the brake cable assembly) and has a workable solution (squeezing the brake pages together). As a result, the expert delays giving his standard instruction about the quick release mechanism. When he notices that the workers is attempting to unscrew the brake cable where it is attached to the hand brake, he gives his instruction about the quick release mechanism, but does so in a mitigated way because the worker hasn't explicitly asked for help and because the expert is proposing a solution that conflicts with the worker's theory that slack can be created by squeezing the brake shoes together. Through the rest of the episode the expert uses a combination of verbal and visual cues to determine how much additional instruction to deliver and when to deliver it

The two preceding conversations, taken from the video condition, contrast with the conversation in Table 6, where no video was available. In Table 6, the expert is more explicit in describing various components, double checking that the worker has understood his description. He withholds instruction until he is sure that the worker has identified the correct parts, but must rely upon language to make that assessment. The worker is more explicit in using language to indicate when she has identified the parts and has completed elements of the subtask. An interesting episode in this conversation occurs when the expert uses the worker's vocabulary to create the link between his technical description and the worker's knowledge ("that anchor plate - that little horseshoe thingy").

Transcript	Description
E. I guess you have to make sure that, in the, in the diagram in the manual.	Expert uses a diagram to identify the components for the worker. He uses the video to see that the worker is having trouble attaching straddle cable.
E. that the anchor plate is in place on the cable. And that's about it.	During this interval, the worker is squeezing the brake shoes together, allowing the expert to infer that a) he has identified the relevant parts and b) has an appropriate theory about how to create slack on the brake cable.
W. OK (3 second pause) OK, I'm just...	Expert can see in the video that the worker's hands are moving away from the brake components to the handlebars and the brake levers.
E If you're having trouble getting the anchor plate on you will find that there's a, a quick-release mechanism in the brake.	Expert gives a mitigated form of instruction.
E. It's just above where the anchor plate is connected to the cable.	Clarification about the location of the quick release lever.
W. Oh, OK, I see.	Worker acknowledges that he has found the quick release lever.
E Yeah.	Expert sees that worker is manipulating the correct component.
W. Oh, OK, I see. Thank you (3 second pause).	Worker indicates that he understands the instruction and has found the correct components.
W. OK.	Worker announces he has completed the subtask.

Table 5. Expert infers a worker's theory of operation based on visual inspection.

DISCUSSION

In summary, this research has shown that collaboration with a remote expert substantially improves a less skilled person's ability to perform maintenance and repair tasks. It also shows that the technology that the collaborators have available to them effects the manner with which they communicate and how they articulate their speech with their task performance. Workers were less explicit in describing the state of the physical world and what they had accomplished when they shared a view of the work environment with their collaborators. When they shared this view, experts were more likely to offer proactive instruction, basing the instruction they delivered and when they delivered it on a combination of the worker's explicit descriptions and their visual inspection of the worker's behavior. When the shared view was available, experts were more likely to treat the workers' explicit description of state as an implicit request for help, which they then provided.

Experts used the shared visual space as a basis for assessing the worker's readiness to receive help as well as for inferring what help the worker needed. They used images from the camera (where it was pointed and what the workers was touching), along with language behavior (e.g., back channel responses such as "ok" and more extended questions or comments that indicated confusion) to infer several aspects of the worker's state. These inferences included a) whether the worker understood the language they were using to describe a repair, b) whether the worker had an appropriate theory about how bicycles operated and how the repair was to be done, c) whether the worker had identified the locations in the physical environment where the tasks were to be done, and d) whether the worker had created the conditions on the bicycle to be ready for the next step in the repair process. The experts also referred to and pointed to drawings in the repair manual to translate between their technical instructions and the physical world in which the repairs needed to be done. While we have shown that collaborators used the video to coordinate their language, we have also shown that they can achieve their collaboration by language alone. We found no evidence that differences in communication technology influenced success in collaboration.

These findings should be viewed as preliminary. Some of our findings are based on real-time coding of a large sample of conversation (about 30 minutes from each of 45 pairs of subjects). This coding, however, was relatively crude, with a small number of categories and only moderate reliability. Other findings are based on more reliable coding, but of only a single subtask. We are in the process of extending these analyses.

Transcript	Description
E. The next thing we got to do uh, ok, hook the brake cable up.	Expert gives overall instruction.
E. The brake cable's unhooked right now. Uhm ... if you look up uh ... ju uh ... to the top part of the brakes here.	Expert tries to figure out if worker knows where the anchor plate is. Worker's lack of responses causes the expert to give more detailed information about the location.
W U-huh	Worker indicates she's located the right spot.
E You see uh this, this loose cable that kind of goes across from this one arm to this other arm.	Expert uses the manual to indicate the correct components.
W. Uh	Worker hasn't identified the location yet.
E. Called the straddle cable	Expert labels the part, for more efficient communication.
W The little tiny one. It's got a little kindy, tiny horse shoe thingy hanging from it	Worker misidentifies part.
E No, that's called, that's the actual cable coming from above.	The expert corrects the worker's misunderstanding of the part.
W. Right	The worker acknowledges that she misunderstood.
E Ok, there's also a second part, which is this cable that just kind of loops around from here to here.	Expert returns to using the manual to indicate the correct part. He moves the cursor over the part in the diagram.
W. Ok. I see it. It's hooked between the two right now.	The worker reports she has identified the part and gives a brief description to show this.
E. Right. Right. It's attached on both ends.	Expert acknowledges that she has identified the correct part, and provides further description to solidly ground that they are referring to the same part.
W: Right.	Worker acknowledges the expert's description.
E: Ok. Uhm, what you want is, you want to hook that uh. . . that anchor plate - that little horseshoe thingy. It. . .	The Expert continues the instruction. He refers to the anchor plate by the previous description given by the Worker.
W: Uh huh.	
E: . . . hooks on the straddle cable	Finishes the instruction.

Table 6. Worker and expert use language to compensate for lack of video.

This research raises both theoretical and engineering questions. The finding that changes in technology didn't lead to changes in collaborative success deserves additional investigation. It is possible that this lack of relationship represents a failure to measure collaborative success sensitively enough to capture potentially subtle effects of technology on outcome. Our failure to replicate earlier studies, which showed that half-duplex audio degrades communication, is consistent with this explanation. Another possibility is that different methods for coordinating conversation are functionally equivalent to each other, at least within limits, because people can modify their behavior to compensate for constraints imposed by technology. This explanation is consistent with the finding that people were more explicit in describing state when they didn't have a video image to show their state to their partner. We need additional analysis to learn how people adapt to these different conversational resources.

While collaborators did take advantage of a shared visual space provided by a video camera, the display itself was not sufficient to substitute for real co-presence. The user interface to the camera—mounting a continuous feed camera on a hat so that head movement produced camera movement—was valuable because it allowed workers to concentrate on their tasks with little regard to controlling the video. Some features of the interface were unsatisfactory, however. Parallax and other misalignment problems meant that the camera sometimes did not show what the worker was attending to (cf. [11]). A camera with a wider field of view, an intelligent camera controller that could keep a predefined object in view, or the expert's control of the camera could all help this problem, but may raise difficulties of their own. Assuming that shared video workspaces prove valuable, designing them so that they are visually adequate, but don't distract from the task at hand will still be challenge.

ACKNOWLEDGEMENTS

This study was conducted with support from the following institutions and individuals: Daimler Benz AG, Advanced Research Projects Agency, The Intel Corporation, National Science Foundation Equipment Grant #9022511, Richard Martin, Dan Siewiorek, Len Bass, Kathleen Carley, Bonnie John, Malcolm Bauer, John Stivorek, Sandra Esch and Tom Pope. Jordan Anderson and David Kaplan were exceptionally helpful in all phases of this research.

REFERENCES

[1] BT Development (1993). Camnet videotape. Suffolk, Great Britain: Author.

[2] Cash, J., O'Neil, J. & Ostrofsky, K. (1991). *Otis Elevator Co.: Managing the Service Force.* Cambridge, MA: Harvard Business Case 9-191-213.

[3] Chapanis, A. (1975) Interactive human communication. *Scientific American*, 232, 36-42.

[4] Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, R. M. Levine, & S. D. Teasley (eds.). *Perspectives on socially shared cognition.* (Pp. 127-149). Washington, DC: American Psychological Association.

[5] Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39

[6] Clark, H. H. (1992) *Arenas of language use.* Chicago: University of Chicago Press.

[7] Eckstrom, R., French, J.W., Harman, H.H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, New Jersey: Educational Testing Service.

[8] Krauss, R. M. & Bricker, P. D. (1966). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society, 41*, 286-292.

[9] Krauss, R. M. & Fussell, S. R. (1990) Mutual knowledge and communicative effectiveness. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual teamwork: Social and technological bases of cooperative work.* Hillsdale, NJ: Lawrence Erlbaum Associates. Pp. 111-145.

[10] Krauss, R.M. & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343-346.

[11] Kuzuoka, H., Kosuge, T., and Tanaka, K.. (1994) GestureCam: A video communication system for sympathetic remote collaboration. *Proceedings of CSCW 94*, Chapel Hill, North Carolina, 35-43.

[12] McCarthy, J.C., Monk, A.F., (in press) Measuring the quality of computer-mediated communication. *Behavior and Information Technology.*

[13] Orr, J. (1989) Sharing knowledge, celebrating identity: War stories and community memory among service technicians. In D. S. Middleton and D. Edwards (Eds.), *Collective Remembering: Memory in Society.* London: Sage Publications (in press).

[14] Sachs, P. (1995). Transforming work: Collaboration, learning, and design. *Communications of the ACM, 38*(9), 36-44.

[15] Sanderson, P, Scott, J., Johnston, T., Mainzer, J, Watanabe, L., & James, J. (1994). MacSHAPA and the enterprise of exploratory sequential data analysis. *International Journal of Human-Computer Studies, 41*, 633-681.

[16] Siegel, J., Kraut, R.E., John, B.E., Carley, K.M. (1995) An Empirical Study of Collaborative Wearable Systems, *CHI Companion 95*, Denver, Colorado, 312-313.

[17] Sloane, E.A. (1991) *Sloane's new bicycle maintenance manual.* Simon & Schuster, New York.

[18] Smailagic, A. and Siewiorek, D. (1993). A case study in embedded systems design: VuMan2 wearable computers. *IEEE Design and Test of Computers,.*

[19] van der Plas, R. (1993) *The bicycle repair book.* San Francisco: Bicycle Books, Inc.

[20] Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin, 84*, 963-976.