HUMAN–COMPUTER INTERACTION, 2004, Volume 19, pp. 273–309

# Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks

**Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer,** and **Adam D. I. Kramer**
*Carnegie Mellon University*

## ABSTRACT

This article considers tools to support remote gesture in video systems being used to complete collaborative physical tasks—tasks in which two or more individuals work together manipulating three-dimensional objects in the real world. We first discuss the process of conversational grounding during collaborative physical tasks, particularly the role of two types of gestures in the grounding process: *pointing gestures,* which are used to refer to task objects and locations, and *rep-*

**Susan Fussell** is a social and cognitive psychologist with interests in face-to-face and computer-mediated communication; she is a Research Scientist in the Human Computer Interaction Institute at Carnegie Mellon University. **Leslie D. Setlock** is a researcher with interests in computer-mediated communication; she is a Research Associate in the Human Computer Interaction Institute at Carnegie Mellon University. **Jie Yang** is a computer scientist with interests in multimedia communication; he is a Senior Systems Scientist in the Human Computer Interaction Institute and the Language Technologies Institute at Carnegie Mellon University. **Jiazhi Ou** is a computer scientist with interests in multimodal communications; he is a doctoral candidate in the Language Technologies Institute at Carnegie Mellon University. **Elizabeth Mauer** is a researcher with interests in human–computer interaction; she is currently a Human Systems Engineer at Aptima Corporation. **Adam D. I. Kramer** is a social psychologist with interests in human–computer interaction and decision making; he is a Research Associate in the Human Computer Interaction Institute at Carnegie Mellon University.

## CONTENTS

*resentational gestures*, which are used to represent the form of task objects and the nature of actions to be used with those objects. We then consider ways in which both pointing and representational gestures can be instantiated in systems for remote collaboration on physical tasks. We present the results of two studies that use a "surrogate" approach to remote gesture, in which images are intended to express the meaning of gestures through visible embodiments, rather than direct views of the hands. In Study 1, we compare performance with a cursor-based

pointing device that allows remote partners to point to objects in a video feed of the work area to performance side-by-side or with the video system alone. In Study 2, we compare performance with two variations of a pen-based drawing tool that allows for both pointing and representational gestures to performance with video alone. The results suggest that simple surrogate gesture tools can be used to convey gestures from remote sites, but that the tools need to be able to convey representational as well as pointing gestures to be effective. The results further suggest that an automatic erasure function, in which drawings disappear a few seconds after they were created, is more beneficial for collaboration than tools requiring manual erasure. We conclude with a discussion of the theoretical and practical implications of the results, as well as several areas for future research.

## 1. INTRODUCTION

As the workforce becomes increasingly distributed across space and time and is increasingly mobile, the need to collaborate with remote partners to accomplish collaborative tasks has risen substantially. Most current systems for remote collaboration (e.g., desktop video conferencing, electronic mail, audio teleconferencing), however, are designed to support group activities that can be performed without reference to the external spatial environment (e.g., decision making). Development of systems to support collaborative physical tasks has been much slower. By “collaborative physical tasks” we mean tasks in which two or more individuals work together to perform actions on concrete objects in the three-dimensional (3D) world. For example, a remote expert might guide a Worker’s performance in emergency repairs to an aircraft, a group of students might collaborate to build a science project, or a medical team might work together to perform surgery. Such tasks play an important role in many domains, including education, design, industry, and medicine. Because the expertise to perform collaborative physical tasks is often distributed across locations, there is growing demand for technologies to support their remote accomplishment.

Observational studies of physical collaboration suggest that people’s speech and actions are intricately related to the position and dynamics of objects, other people, and ongoing activities in the environment (e.g., Flor, 1998; Ford, 1999; Goodwin, 1996; Kuzuoka & Shoji, 1994; Tang, 1991). Conversations during collaborative physical tasks typically focus on the identification of target objects, descriptions of actions to be performed on those targets, and confirmation that the actions have been performed successfully. During the course of the task, the objects themselves may undergo changes in state as people perform actions upon them (e.g., a piece of complex equipment may undergo repair) or as the result of outside forces (e.g., a patient might start hemorrhaging).

As they speak, collaborators on physical tasks use gestures to clarify or enhance their messages. Pointing gestures are used to refer to task objects and locations (e.g., “that piece goes over there”). Representational gestures, such as hand shapes and movements, are used to represent the form of task objects and the nature of actions to be performed on those objects (Bekker, Olson, & Olson, 1995; McNeill, 1992). For example, a speaker might say, “turn the screw,” while using his or her hands to indicate the direction to turn it. In face-to-face settings, people can make full use of both pointing and representational gestures because they share a physical environment that includes both task objects and other participants. However, in most systems to support remote collaboration, participants have limited ability to gesture. Even systems that specifically provide views of the work area (e.g., Fussell, Kraut, & Siegel, 2000; Fussell, Setlock, & Kraut, 2003; Kraut, Fussell, & Siegel, 2003) typically only allow for pointing by collaborators co-located with task objects, not for remote participants. When views of remote participants’ bodies are available, spatial relationships among people and objects are typically not preserved as they must be for pointing gestures, and the field of view is often not wide enough or of sufficient resolution to show a full range of representational hand gestures. Collaborators often express frustration with this situation, saying, for example, “If I could just point to it, its right there,” or “if only I could show you how to do it,” when struggling to provide instructions for their partners.

In this article, we describe two studies of tools we have developed that provide remote collaborators with the ability to make certain types of gestures. Both tools make use of the concept of overlaying images—either a cursor pointer or pen-based drawings—on a live video feed from a workspace. The goal of both tools was to facilitate communication and performance of distributed teams by allowing remotely located participants to gesture in the workspace. In the remainder of this article, we first describe the theoretical framework guiding our work. Then, we present two laboratory studies that tested the value of our gesture tools for collaboration on physical tasks. We conclude with a discussion of the theoretical and practical implications of our findings and some suggestions for future areas of work.

## 1.1. Collaboration on Physical Tasks

Collaborative physical tasks are tasks in which people work together to manipulate objects in the 3D world. These tasks can vary along a number of dimensions, including number of participants, temporal dynamics, type and size of objects, and the like. Our focus in this article is on what we call “instructional" collaborative physical tasks, in which one person (whom we call the "Worker") directly manipulates objects and tools under the guidance of another person (whom we call the "Helper") who provides instructions but does not physically manipulate objects. The Helper may be in the same physical location as the Worker or at a distance, connected by communications media including audio and video conferencing systems. The relationship between participants is thus similar to a teacher guiding a student's lab project, a remote surgeon guiding an operation, or a remote expert guiding machinery repair.

## 1.2. Conversational Grounding in Collaborative Physical Tasks

Providing instruction during collaborative physical tasks requires complex coordination (Clark, 1996; Kraut et al., 2003): A Helper must determine what instructions are needed and when, how to phrase their messages of assistance such that their partner understands them, and whether the message has been understood as intended and the task is proceeding appropriately. The *conversational grounding* framework proposed by Clark and his colleagues (e.g., Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986) provides a useful theoretical foundation for understanding the relationships among verbal and nonverbal communication, actions, and environment in collaborative physical tasks.

Research has shown that interpersonal communication is demonstrably more efficient when people share greater amounts of *common ground*—mutual knowledge, beliefs, goals, attitudes, and so on. People may have common ground prior to an interaction if they are members of the same group or population (e.g., Fussell & Krauss, 1992). In addition, they construct and expand their common ground over the course of the interaction on the basis of *linguistic co-presence* (because they are privy to the same utterances) and *physical co-presence* (because they inhabit the same physical setting; Clark & Marshall, 1981). The term *grounding* refers to the interactive process by which communicators exchange evidence about what they do or do not understand over the course of a conversation, as they build up common ground (Clark & Brennan, 1991).

In a series of prior studies (Fussell et al., 2000; Fussell, Setlock, & Kraut, 2003; Kraut et al., 2003; Kraut, Gergle, & Fussell, 2002; Kraut, Miller, & Siegel, 1996; see also Bolt, 1980; Emmorey & Casey, 2001), we have found that conversational grounding during collaborative physical tasks tends to follow a predictable pattern: First, collaborators come to mutual agreement upon or "ground" the objects to be manipulated using one or more referential expressions. Next, they provide instructions for procedures to be performed on those objects. Finally, they check task status to ensure that the actions have had the desired effect. Figure 1 shows a sample dialogue with the segments of the grounding se-

***Figure 1.* Example of the structure of conversational grounding during a bicycle repair task. (The Worker is fixing the bicycle; the Helper is providing guidance.)**

| | Sample Dialogue | Phase of Task Performance |
|---|---|---|
| Helper: | Now these fork looking things down here. | Object identification |
| Worker: | uh huh | |
| Helper: | Those should go on the wheel axle inside of the nuts on the axle. | Procedural statement |
| Worker: | Ok | |
| Helper: | Are they on ok? | Monitor comprehension, task status |
| Worker: | Yep, all set. | |

quence indicated. This material is adapted from conversations between a Helper and a Worker attempting to repair a bicycle (Fussell et al., 2000). In this example, each phase of the task was grounded immediately with a Worker acknowledgement ("uh huh," "ok"). In other cases, each phase of the task dialogue may include clarification and other sub-sequences of dialogue before grounding is established (Jefferson, 1972; Sacks, Schegloff, & Jefferson, 1974).

Our research suggests that grounding sequences are more efficient when people are co-located as opposed to when they are linked by audio or video technologies, even when those video technologies provide views of the workspace (Fussell et al., 2000; Fussell, Setlock & Kraut, 2003; Kraut et al., 1996; Kraut, et al., 2003). We hypothesized that, in part, this greater efficiency of co-located partners stems from their ability to make full use of both pointing and representational gestures to ground their conversations. In the next sections of the article, we focus on the role of gesture in conversations during collaborative physical tasks and on the design of systems to support gesture capabilities in video-mediated collaboration.

## 1.3. The Role of Gesture in Conversational Grounding

Our previous studies suggest that grounding during collaborative physical tasks occurs through gestures and actions in addition to speech, and that this use of gesture facilitates task performance. As they talk, people use several types of gestures to clarify or enhance their messages (e.g., Bekker et al., 1995; McNeill, 1992). *Pointing gestures* are used to refer to task objects and locations. *Representational* or *iconic gestures*, such as hand shapes and hand movements, are used to represent the form of task objects and the nature of actions to be used with those objects, respectively. Although the classification of gestures differs somewhat between systems (e.g., Efron, 1941; Ekman & Friesen, 1969; Kendon, 1972; McNeill, 1992), all make distinctions between pointing and representational hand movements. (In this article we do not consider other

*Figure 2.* **Definitions and possible functions of gestures used in collaborative physical tasks.**

| Type of Gesture | Definition | Possible Functions |
|---|---|---|
| Deictic (Pointing) | Orienting a finger or hand toward a point in the environment | Reference to objects and locations |
| Concrete representational | | |
| Iconic representations | Forming hands to show what a piece looks like, or to show how two pieces should be positioned relative to one another | Reference to objects, procedural instructions (particularly orientation), descriptions of task status |
| Spatial/Distance | Indicating through use of one or both hands how far apart two objects should be or how far to move a given object | Procedural instructions, descriptions of task status |
| Kinetic/Motion | Demonstrating through use of hands what action should be performed on a task object | Procedural instructions |

categories of hand gestures, such as *beats* marking speech tempo, that are viewed as inherently noncommunicative, cf. McNeill, 1992).

The four types of gestures we focus on in our research are shown in Figure 2. *Pointing (deictic) gestures* are motions in which a person uses his or her hands (typically one finger extended and the others curled inward) to indicate a person, object, or location within the shared physical environment. They often co-occur with deictic verbal expressions such as *this, that,* and *there.* For example, a speaker might say, "take *that piece* and put it *there*," while using a pointed index finger to indicate the intended piece and target location. Pointing gestures provide a quick and efficient way to indicate objects and locations that would otherwise require lengthy verbal descriptions (e.g., Bauer, Kortuem, & Segall, 1999; Fussell et al., 2000; Karsenty, 1999). In a study of collaborative bicycle repair (Fussell et al., 2000), for example, we found that when participants worked side-by-side, and thus both gestures and task objects were visually shared, participants used more pointing gestures and deictic expressions to refer to task objects and this use of deictic expressions was associated with shorter, more efficient referring expressions and faster task performance than when pairs were linked by audio-only or audio-video connections.

In addition to pointing gestures, there are several types of representational gestures that may facilitate conversational grounding. Representational gestures are those that bear a relationship to the corresponding speech stream—for example, a speaker might hold his or her hands out straight while saying, "the flat piece." We focus on *concrete* representational gestures that may encode details of spatial relationships, shapes, and actions pertinent to collaborative physical tasks, in contrast to *abstract* representational gestures that may have a more indirect or metaphorical relationship to speech content (McNeill, 1992). Prior research suggests that representational gestures are common in physical tasks. Tang and Leifer (1988), for example, found that approximately 35% of the gestures produced by a sample of design teams served to either convey or represent ideas (see also Bekker et al., 1995; Emmorey & Casey, 2001). Concrete representational gestures may facilitate conversational grounding in collaborative physical tasks by allowing speakers to communicate multiple pieces of information about the task simultaneously (Clark, 1996; McNeill, 1992).

Three types of representational gestures—iconic representations, spatial gestures, and kinetic gestures—appear to play a critical role in task-oriented dialogues (e.g., Bekker et al., 1995; Bolt, 1980; Tang & Leifer, 1988). *Iconic representations* use the shape of the hand(s) to project an image of what a particular piece might look like. For example, a speaker might form a circle with his or her thumb and forefinger to indicate a round object. Iconic gestures may also be used to demonstrate the relationship between two or more objects, as, for instance, when a speaker places one straightened hand against the middle of the other to form a "T" shape.

*Spatial (distance) gestures* involve placing two fingers or both hands a certain distance apart, where the distance can literally reflect the actual physical distance between two objects. For example, a person might say, "move it up this much" while positioning his or her hands apart the specific distance intended. Bekker et al. (1995) found that design team members used spatial gestures to indicate distance between people and objects in the hypothetical systems they were designing. We have observed Helpers in bicycle and aircraft repair tasks use similar spatial gestures (Fussell et al., 2000; Kraut et al., 1996; Siegel, Kraut, John, & Carley, 1995).

*Kinetic (motion) gestures* are those in which the speaker uses the tempo and motion of the hand(s) to specify manner of motion (McNeill, 1992). Bekker et al. (1995) found that design team members used these gestures to illustrate how users would interact with the systems they were designing. Cassell (1998) similarly described how an instructor used his hands to specify the details of how a caulking gun should be used as he verbally describes the process.

## 1.4. Implementing Gesture in Video Systems for Remote Collaboration

In face-to-face collaboration on physical tasks, people can readily combine speech and gesture because they share the same physical space. They can monitor one another's hands and jointly observe task objects and the environment. Designing systems to provide remote support for gesture is complicated by the different visual requirements for pointing and representational gestures. Pointing gestures require a view of the pointing device (typically a hand), the target object or location, and the relationship between the two; representational gestures require a view of the speaker's hands (or a surrogate). Most current technologies to support gesture either enable pointing or show a view of the speaker's hands, but not both. The few exceptions, such as the ClearBoard system (Ishii, Kobayashi, & Gruden, 1993), require expensive, specialized equipment that makes their use impractical for most collaborative work.

Systems such as ClearBoard use what we call a "direct view" approach to remote gesture, in which they attempt to retain people's natural modes of hand gesturing. In contrast, what we will call "surrogate" approaches to remote gesture assume that the communicative intent of a gesture can be expressed through alternative means that do not show a speaker's hands. These types of systems are sometimes said to incorporate visible *embodiments* of gesture, rather than the natural gestures per se (e.g., Gutwin & Penner, 2002). A familiar example of the surrogate approach is the laser pointer often used in lectures to large audiences. Kuzuoka and colleagues have developed a series of elaborate laser pointer systems suitable for collaborative physical tasks (e.g., Kuzuoka, Kosuge, & Tanaka, 1994; Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000). In the current research, we explore surrogate approaches to remote gesture that provide substitutes for both pointing and representational gestures. Our goal is to allow people to make these gestures in ways that, although not entirely natural, are readily integrated with the rest of their conversation and task performance.

Two types of devices that are especially promising as gesture surrogates in collaborative physical tasks are cursor pointers and pen-based drawing tools. Cursors have a long history in HCI as a collaborative tool (see, e.g., Greenberg, Gutwin, & Roseman, 1996). Mutually visible cursors have been implemented for telepointing in shared Web pages (Greenberg & Roseman, 1996), in collaborative whiteboard systems such as Colab (Stefik, Foster, Bobrow, Kahn, Lanning, & Suchman, 1987), and many other domains. In general, the impermanence of cursor marks as well as their small size makes them more appropriate for pointing gestures than for representational gestures (although there are some exceptions, such as Gutwin and Penner, 2002).

Pen-based systems that permit mutual sharing of sketches, diagrams, and handwritten text in addition to pointing may be more appropriate for showing the full range of gestures. Shared drawing systems such as Tivoli (Pedersen, McCall, Moran, & Halasz, 1993), Conversation Board (Brinck & Gomez, 1992), and DOLPHIN (Streitz, Geissler, Haake, & Hol, 1994) allow participants to use styli to mark on the shared drawing. Although many drawing tools were designed primarily to support the creation of shared images (e.g., architectural diagrams or design sketches), studies of their use have observed the importance of the tools for gestural communication. In We-Met (Wolf, Rhyne, & Briggs, 1992; Wolf & Rhyne, 1993), a pen-based shared drawing tool, 15% of people's speaking turns included gestures. Circles, arrows, and similar marks were used for indicating parts of the shared drawing, and two-way arrows were used to express more complex relationships. With the Commune system (Bly & Minneman, 1990), in which pens can be used for both drawing (when pressed down on the drawing surface) and gesturing (when not pressed down), nearly half of all users' pen activity was for gesturing purposes.

Surprisingly, cursor and pen-based tools have rarely been combined with live video feeds. One exception is the VideoDraw system (Tang & Minneman, 1991), in which video feeds of a collaborative drawing tablet allowed collaborators to view one another's hand gestures overtop of the drawings they construct. In this system, natural hand gestures are combined with pen-based drawing. Roussel (2001) similarly combined camera shots of hand gestures over live video feeds. Drawing over video is also common in television sports broadcasting, although it is primarily used on replays rather than live video. In the next section, we describe the current studies that combine either cursor pointing or pen-based drawing with live video to provide support for remote gesture.

## 1.5. Overview of the Current Studies

In the current studies, we examine two tools that combine embodiments of gesture with live video feeds from the workspace. One system uses a mouse-based system that supports remote pointing gestures only; the second uses a pen-based system that supports remote drawing of both pointing and representational gestures. In both cases, the Helper's gestures are displayed on a monitor in front of the Worker's workspace. Both tools have the benefits of being inexpensive, easy to use, and readily incorporated into most video conferencing systems. Unlike fancier laser pointing systems, however, our tools have the cost of requiring Workers to map the gestures over video displayed on their computer monitors to the actual objects and locations in the workspace. They also have the cost that gestures must be embodied in cursor or pen movements, rather than performed directly with the hands. Despite these potential costs, we hypothesized that our drawing-over-video tools would improve performance over a video-only system. We test these hypotheses within the context of an instructional collaborative physical task, in which the Worker builds a large toy robot under the guidance of a co-located or remote Helper.

## 2. STUDY 1: CURSOR POINTING

Previous research has indicated that remote collaborators on physical tasks would benefit from being able to point to objects and locations in the workspace. In fact, people devise novel strategies for pointing when technologies make it difficult to do so through natural hand gestures. For example, Workers use head-worn cameras to point to task objects during a collaborative bicycle repair task (Fussell et al., 2000), people hold objects up to cameras that would not otherwise be visible (Fussell, Setlock, & Kraut, 2003; Tang & Minneman, 1991), and they point with their heads when their hands are not available (Emmorey & Casey, 2001). Although Workers can adapt video technologies to enable the use of pointing gestures, the additional time required for such maneuvering hinders conversational grounding and task performance, making overall performance times longer than when partners are face-to-face. Furthermore, such technological adaptation is of no assistance to remote Helpers—they still have no way to point to objects in the Workers' workspace. These results strongly suggest that there is a need for implementing pointing capabilities in video systems to support collaborative physical tasks.

Support for remote pointing is particularly difficult because of the need to maintain spatial relationships between communicators and objects in the environment. One approach to this problem has been to user laser pointers that allow remote partners to point beams of light directly at objects in the workspace (e.g., Kuzuoka et al., 1994, 2000). For example, in the GestureMan system (Kuzuoka et al., 2000), a remote partner manipulates a robot with attached laser pointer. These systems have several practical limitations, particularly their expense and the specialized equipment required, making them unlikely candidates for widespread adoption.

In this study, we implemented and evaluated a simple cursor pointing tool. The tool combines a video camera oriented at the workspace ("scene camera") with a cursor pointer. The scene camera sends a live feed of the workspace to a remote Helper, who can overlay the cursor on top of this feed to point to objects and locations. The resulting combination of video feed and cursor overlay is displayed on a large monitor in front of the Workers' workspace. This system has the benefits of being inexpensive, easy to use, and easily implemented in most any video conferencing system, but has the cost of requiring Workers to map the view from the video feed on the monitor to the actual objects and locations in the workspace. We hypothesized that despite this limitation, the cursor pointer would improve communication and performance over a video-only system. However, we anticipated that communication and performance would be best in a side-by-side condition, in which Helpers and Workers were co-located. In addition to testing these hypotheses, we examined the targets of participants pointing gestures to better understand the role of pointing in task performance.

## 2.1. Method

### Design

We used a within-subjects design in which pairs completed three robot assembly tasks, one in each of three media conditions: side-by-side, video only, and video plus cursor pointer. One partner was randomly assigned to the "Helper" role and was responsible for providing instructions; the other partner was assigned to the "Worker" role and was responsible for actually building the robot. Trials, tasks, and media conditions were counterbalanced.

### Participants

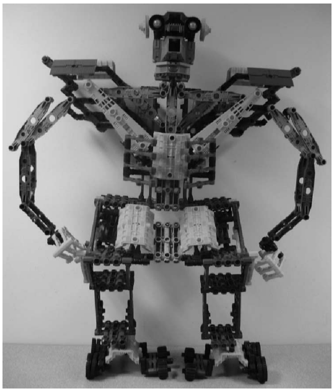
Participants consisted of 48 pairs of undergraduate students at Carnegie Mellon University (52% male), ranging in age from 19 to 55 years ($M = 24.38$). Seventy-three percent reported having prior experience constructing objects from kits. Each participant received $15 for their participation; each pair also competed for a $25 bonus to the fastest and most accurate pair in the study.

### Materials

The Robotix Vox Centurion robot kit (Figure 3) was used as the basis for the instructional tasks. The kit contains a variety of black, orange, purple, and gray pieces ranging from approximately .5 in. to 7 in. (12.7 mm – 177.8 mm) long. When fully assembled, the robot stands 3.5 × 2.5 ft (1.07 × .76 m) wide. We identified three tasks of similar difficulty (left arm, right ankle, and right hand), each of which took less than 10 min on average to perform. An instruction manual with bulleted items that was created in PowerPoint® outlining the steps to be completed and printed for use in the side-by-side condition.

Three sets of online surveys were created and then implemented in HTML and Microsoft Access for online presentation and automatic response recording. A *pretest questionnaire* collected basic demographic information (e.g., gender, age). A *post-task questionnaire*, administered after each task, asked questions about the success of each collaboration (e.g., "I am confident we completed this task correctly"). Responses were made on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The survey also included questions tailored specifically for the participant's role. Helpers indicated agreement with statements such as "I could tell when my partner needed assistance," and "It was easy for me to point out objects in the shared workspace." These questions were

***Figure 3.*** **The large toy robot used in our experiments.**



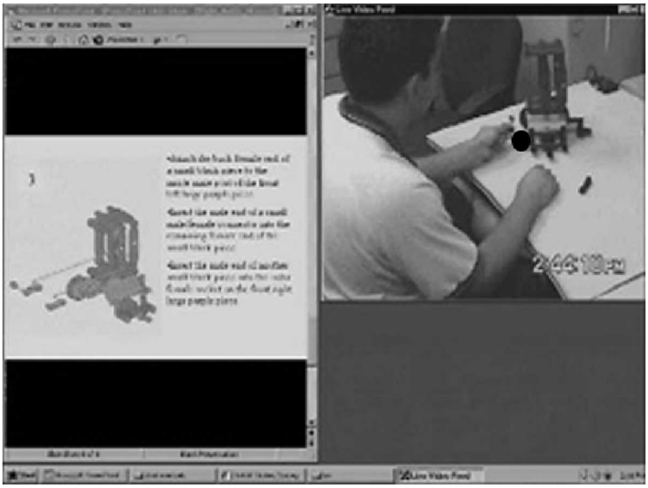
rephrased for Workers (e.g., "My partner could tell when I needed his or her assistance"). Helpers also rated the extent to which they relied on different resources (the manual, previous experience doing the task, ability to see what partner was doing, cursor pointer, and partner's requests for help) as they assisted their partner, on a scale of 1 (*not at all*) to 5 (*extensively*).

A *final questionnaire*, presented upon completion of all tasks, asked broader questions about the technology and collaboration. Two versions of the final questionnaires were created, one for each experimental role. The Helper survey included questions about the success of the collaboration (e.g., "My partner and I worked well together on these tasks") and importance of visual information (e.g., "It was important to me to be able to see what my partner was doing"). Helpers also rated the similarity between each technology and face-to-face communication and rated the usefulness of specific features of the technology, including the cursor pointer. Ratings were made on 5-point scales. The Worker version included questions about the success of the collaboration but no ratings of specific technologies.

## Equipment

A Sony Handycam Hi 8 video camera was positioned 5 ft (1.52 m) behind and to the right of the Worker and showed a view of the Worker's hands, robot pieces, and part of the robot being completed. Helpers saw this view in the upper right of their computer screen; Workers saw the view on a large monitor in front of their workspace. In the pointing condition, Helpers could press a mouse button to create a pink circle on the video feed. The pink circle appeared in the same location on the Workers' monitor (Figure 4). An AverMedia AverKey 300 Gold was used to merge the video feeds on the Helper's PC. The

*Figure 4.* **Helper's view with the manual on left and video with cursor overlayed on right (Study 1).**



output was sent to a Panasonic DV-VCR (Model No. AG-2000P) for recording. Two Samson MR1 microreceivers received audio between the two rooms. The audio feeds were input into the DV recorder.

### Procedure

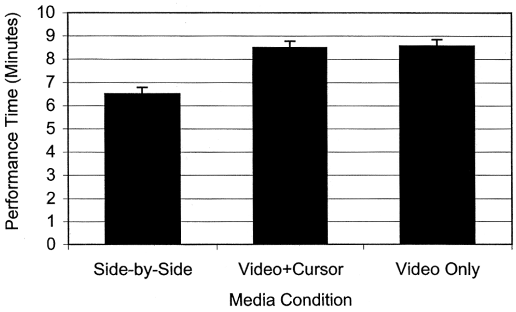
Workers and Helpers were given an overview of their roles in the experiment—to build the robot and to instruct, respectively—and completed consent forms and pretests. They were then shown the robot and the communications technologies they would be using. The Helper was further instructed on use of the online manual and cursor pointing device.

Pairs exchanged small talk to familiarize themselves with the equipment and then began their series of three trials. Participants were told what technology would be available to the Helper prior to each trial. Upon completion of the task, or after a 10-min period, the work was halted and participants completed posttask questionnaires. They then moved on to the next task. After all tasks were done, they completed the final questionnaire.

## 2.2. Results

Survey and performance results were analyzed in 3 (trial) × 3 (task) × 3 (media condition) repeated measures analyses of variance (ANOVAs). We

*Figure 5.* **Task performance time by media condition (Study 1).**



first focus on pairs' performance across media conditions; then, we discuss the questionnaire data followed by a more detailed analysis of cursor usage.

## Performance

As shown in Figure 5, performance was fastest and most accurate in the side-by-side condition. Contrary to our hypotheses, however, adding a cursor pointer to the video system was not sufficient to improve performance over that in the video-only condition. Analysis of performance times indicated main effects of trial, $F(2, 70) = 11.73$, $p < .0001$; task, $F(2, 70) = 27.73$, $p < .0001$; and media condition, $F(2, 70) = 22.92$, $p < .0001$. Performance was significantly faster in the side-by-side condition ($M = 6.50$ min, $SD = 1.99$) compared to the two other conditions—for cursor, $t(70) = -5.78$, $p < .0001$ ($M = 8.49$, $SD = 1.98$); for scene camera only, $t(70) = -5.94$, $p < .0001$ ($M = 8.57$, $SD = 2.00$)—but there was no difference between the two video conditions, $t(70) = -.21$, *ns*.

## Questionnaire Results

***Coordination.*** Posttask ratings for pairs' ability to provide assistance, how well they worked together on the task, their confidence that they had performed the task correctly, and four other coordination-related questions were highly correlated. Scores on these seven scales were averaged into one coordination scale ($\alpha = .88$). Coordination scores were highly negatively correlated with performance time, $r(144) = -.62$, $p < .001$, and positively correlated with completion of the task, $r(144) = .63$, $p < .001$. Pairs rated their coordination highest in the side-by-side condition ($M = 3.93$, $SD = .62$), intermediate in the video + cursor condition ($M = 3.49$, $SD = .62$), and lowest in the video-only condition ($M = 3.26$, $SD = .62$). A 3 (trial) × 3 (task) × 3 (media condition) repeated measures ANOVA indicated significant main effects for trial, *F*(2, 70) = 4.86, *p* = .01; task, *F*(2, 70) = 28.57, *p* < .0001; and media condition, *F*(2, 70) = 23.54, *p* < .001; and a significant interaction between trial and task, *F*(4, 70) = 2.49, *p* < .05. Post hoc comparisons indicated that the coordination was rated significantly higher in the side-by-side condition than in the video + cursor or video-only conditions, *t*(70) = 4.48 and 6.74, respectively, *p* s < .0001, and higher in the video + cursor than the video-only condition, *t*(70) = 2.31, *p* < .05.

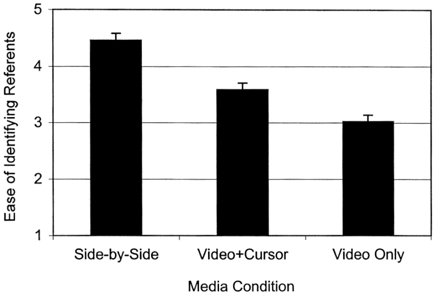
***Ease of Identifying Referents.*** Pairs indicated that they could identify objects in the workspace best when side-by-side (*M* = 4.46, *SD* = .83, on a scale of 1 [*very difficult*] to 5 [*very easy*]), and better with the cursor pointer (*M* = 3.49, *SD* = .83) than with video alone (*M*= 3.26, *SD* = .83; see Figure 6). Rated ability to refer to objects was highly correlated with mean coordination scores, *r*(144) = .68, *p* < .001; negatively correlated with performance time, *r*(144) = –.46, *p* < .001; and positively correlated with task completion, *r*(144) = .44, *p* < .001. A 3 (trial) × 3 (task) × 3 (media condition) repeated measures ANOVA indicated a significant main effects for trial, *F*(2, 70) = 3.26, *p* < .05; task, *F*(2, 70) = 13.20, *p* < .0001; and media condition, *F*(2, 70) = 46.88, *p* < .0001; but no significant interactions. posthoc tests showed a significant difference between all three conditions: for side-by-side versus video + cursor, *t*(70) = 5.90, *p* < .0001; for side-by-side versus video-only, *t*(70) = 9.60, *p* < .0001; and for video + cursor versus video-only, *t*(70) = 3.81, *p* < .001.

### Analysis of Cursor Usage

The videotapes from 36 of the participants were of sufficiently good quality that we were able to code their use of the cursor pointer. Cursor use was coded into one of four categories: pointing to objects, pointing to locations, showing motion by moving the cursor, and other. Two coders each coded half of the videos, plus an overlapping sample of five clips (approximately 200 cursor usages) for reliability. Reliability of counts of the numbers of cursor usages per participant falling into each category was excellent ($\alpha = .97$).

Overall, Helpers used the cursor from 0 to 72 times during a single task (*M* = 32.56, *SD* = 21.02). The cursor was used fairly evenly across the two pointing targets. Participants used the tool an average of 15.36 (*SD* = 13.15) times to point out an object, 15.42 (*SD* = 9.34) times to point out a location, and 1.61 (*SD* = 2.75) times to demonstrate motion. The latter mean was due primarily to a single Helper who used the cursor 9 times to demonstrate movement; the vast majority of Helpers never used it for this purpose.

*Figure 6.* **Ease of referent identification by media condition in Study 1 (1 = very difficult; 5 = very easy).**



To investigate the effects of cursor usage on task performance, ratings of coordination, and ratings of ability to point at objects, we first calculated the rate of cursor use per second by dividing total cursor use by the total time required to complete the task. Higher rates of cursor usage were significantly correlated with faster performance times, $r(34) = -.36, p < .05$, and nonsignificantly but in the predicted direction with coordination ratings, $r(34) = .26, p = .13$.

## 2.3. Discussion

In summary, participants reported finding value to the cursor pointing device, but the tool did not improve performance times over the scene camera alone. Higher rates of cursor use were, however, correlated with faster performance within the cursor + video condition, suggesting that the tool had some benefit for performance.

One possibility is that the cursor tool was too limited in functionality, in that it supported pointing but not representational hand gestures. Representational gestures showing orientation, movement, and the like may be crucial to the instructional phase of the grounding process outlined in Figure 1. Although a few Helpers used the cursor dynamically to indicate direction of movement, the vast majority did not use it for this purpose. In Study 2, we test a system that allows for both pointing and representational gestures.

# 3. STUDY 2: DRAWING GESTURES

Study 1 demonstrated that a cursor pointing tool was perceived as valuable by collaborators for referring to task objects and locations, but the tool did not improve performance over the scene camera alone. One possible reason for the lack of performance effects can be seen by referring back to the sample conversation in Figure 1: Identification of objects and locations is only one of several stages in grounding task instructions. Another important phase is that of explaining procedures. It is possible that the cursor facilitated faster object identification times, but that this process is such a small percent of the overall grounding time that no effects on overall task performance times are visible.

In Study 2, we assess the value of a drawing tool that can be used for pointing, for drawing representational gestures, and for making sketches. In the DOVE (Drawing over Video Environment) system, the workspace is visually shared through video cameras and is equipped with tablet PCs, desktop PCs, or other handheld devices. Real-time video streams from the camera(s) are sent to collaborators' computing devices. A Helper can make freehand drawings and pen-based gestures on the touch-sensitive screen of a computing device, overlaid on the video stream, just like using a real pen on a piece of paper in a face-to-face setting. The results are observable by all collaborators on their own monitors. The technical details of DOVE are presented in Ou, Fussell, Chen, Setlock, and Yang (2003).

We compare communication and performance with two versions of the DOVE system—automatic versus manual erasure of drawings—to communication and performance with a scene camera alone. As Wolf and Rhyne (1993) observed, drawing persistence has both pros and cons. When the partner is not directing his or her attention to the spot where the drawing is being made, persistence allows them to view the drawing at a delay. At the same time, drawing persistence and manual erase functions may detract from smooth, natural conversation. To investigate the issue of drawing persistence in the context of collaborative physical tasks, we devised two versions of our system. In the auto-erase version, drawings disappeared after 3 secs, much in the way that hand gestures are gone once completed. In the manual-erase version, drawings remained on the screen until the Helper pressed one of several "erase" buttons with the pen ("erase all," "erase most recent"). With manual erase, our software functions more like a drawing tool, in that series of gestures can be combined to convey complex ideas.

We hypothesized that the drawing tool would benefit both communication and performance in our collaborative robot construction task by aiding both the reference identification and the procedural instruction phases of the conversational grounding process. We further hypothesized that the manual erase version of the tool would be more beneficial than the auto-erase version because of the potential to combine drawings to convey complex meanings. In addition to these explicit hypotheses, we also investigated the types of drawings participants made as they provided their instructions.

## 3.1. Method

### Design

Pairs of undergraduate students completed three robot assembly tasks (e.g., left arm, right foot) under three different media conditions: (a) *Video Only*: Helper could view the output of the camera focused on the Worker's task environment, but could not manipulate the video feed; (b) *DOVE + Manual Erase*: Helper could draw on the video feed but had to manually erase the gestures; and (c) *DOVE + Auto-Erase:* Helper could draw on the video feed and the gestures disappeared after 3 sec. Trials, tasks, and media conditions were counterbalanced.

### Participants

Twenty-eight pairs of undergraduate students served as participants (55.2% male), ranging in age from 18 to 44 years ($M = 22.48$). Seventy-eight percent had prior experience building objects from kits. They each received $15 for their participation, with a chance to win an additional $25 by being the fastest and most accurate pair to complete the task.

### Materials

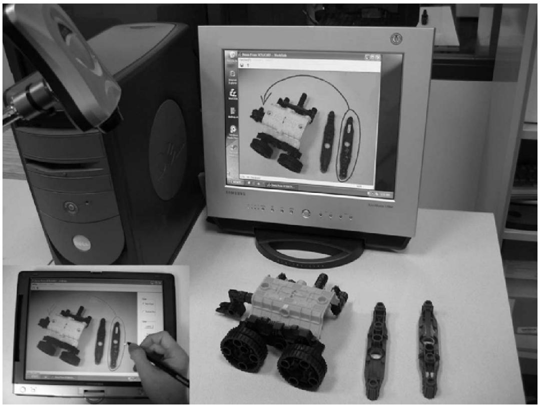
The Robotix Vox Centurion robot shown in Figure 3 was used as the basis for three tasks of equivalent difficulty (robot left arm, right ankle, and left foot). As in Study 1, Helpers were provided with PowerPoint instruction manuals outlining the steps to be completed for each task.

Participants completed the preliminary, posttask, and final questionnaires described under Study 1. In addition to the previous questions, the Helper version of the final questionnaire asked respondents to rate the relative value of the auto-erase and manual-erase versions of the software on a scale of 1 (*strongly prefer automatic erasure of gestures*) to 5 (*strongly prefer manual erasure of gestures*).

### Equipment

The DOVE drawing tool software was installed on a Toshiba Protégé 3505 Tablet PC, with a Mobile Pentium III 1.33 GHz CPU, 496 MB RAM, and 12-in. (304.8 mm) monitor, running Windows XP Tablet edition. An Intellinet Network ENC-001A IP camera, installed 30 in. (762 mm) back and to the right of the Worker's task space, served as the scene camera. It showed a 27 × 31 in. (685.8 mm × 787.4 mm) block of the work area. The feed from the IP camera

***Figure 7.*** **Close-up of the DOVE drawing tool on the Helper's tablet PC (left front insert) and on the Worker's monitor (right).**



was input into the tablet PC, where the Helper could then draw upon it using a pen (see Figure 7). In the manual-erase version of DOVE, Helpers pressed a key at the bottom right of the screen to erase their drawings. In the auto-erase version, the drawings disappeared after 3 sec. This time was chosen as the erasure point for this study because pretesting suggested that this was the time required for Workers to notice and view the drawings on their monitor.

The output from the tablet PC consisted of a video feed with the drawing gestures overlaid, and was distributed via a Lynksys BEFW 1154 wireless local network to a 17 in. (431.8 mm) flat screen monitor in front of the Worker's task area and to the Experimenter's workstation for recording. An AverKey Media iMicro and an AverKey Media 300 Gold were used to convert the video signals for input into the experimenter's workstation. The video images were 6 × 4.5 in. (152.4 × 114.3 mm) on the Helper's Tablet PC, and 6.5 × 5 in. (165.1 × 127 mm) on the Worker's station. A Sony WCS-999 wireless microphone system was used to transmit audio. The audio feeds were input into the DV recorder along with the output from the Helper's PC with instruction manual, the Helper's tablet PC, and the Worker's monitor.

## Procedure

Participants were randomly assigned to the Helper or Worker role. They were situated in the same room, approximately 11 ft (3.35 m) apart, with a 3′6″ × 6′10″ (1.07 × 2.03 m) barrier between them. Participants were provided with an overview of the study and then signed consent forms and filled out the preliminary questionnaire. Next, pairs were shown the technology used in the study, the Helpers received practice using the DOVE drawing tool, and both partners were shown what the output of DOVE looked like on the Worker's monitor.

Pairs then performed the three robot assembly tasks. Helpers constructed each part of the robot prior to providing their instructions to Workers, to familiarize them with the task. At the end of each task, or after a 10 min maximum, they completed posttask questionnaire. Upon completion of all three tasks, they completed the post experimental questionnaire, and were debriefed. Sessions took approximately 60 min. The contents of the Helper's monitor, the drawing tool, and the Worker's monitor were combined with audio recordings for later transcription and analysis.

## 3.2. Results

We discuss the findings in four sections. First, we present overall performance times and error rates; then, we present the analysis of posttask and final questionnaires followed by an examination of conversational efficiency across the three conditions. Finally, we present results of a more extensive examination of the types of drawings participants made during their tasks.

### Performance

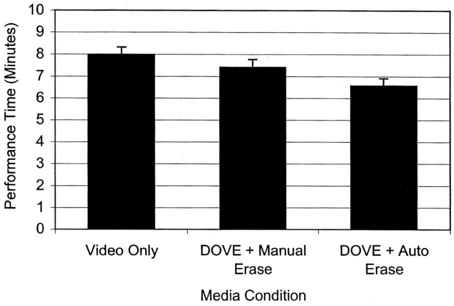
As hypothesized, pairs completed the task faster using the DOVE drawing tool than when using the video camera alone (see Figure 8). Performance times were analyzed in a 3 (trial) × 3 (task) × 3 (media condition) ANOVA. Results indicated significant effects of task, $F(2, 29) = 34.93$, $p < .0001$; and media condition, $F(2, 29) = 8.24$, $p = .002$. There was no main effect of order and no significant interactions. posthoc comparisons of conditions indicated that the auto-erase condition ($M = 6.57$, $SD = 1.79$) was significantly faster than both the manual-erase condition, $t(29) = -2.28$, $p < .05$ ($M = 7.41$, $SD = 1.90$), and the video-only condition, $t(29) = -4.04$, $p < .001$ ($M = 7.99$, $SD = 1.80$). The difference between the manual-erase and video-only conditions was in the expected direction but nonsignificant, $t(29) = -1.56$, $p = .13$.

### Questionnaire Results

Helpers and Workers responded to a series of questions at the end of each trial. Because their responses were highly correlated, we averaged across both experimental roles to obtain a value for each question for each pair.

***Coordination.*** As in Study 1, posttask responses to seven coordination-related questions were highly correlated and averaged into a single coordination scale ($\alpha = .77$). Coordination scores were highly negatively corre-

***Figure 8.*** **Task performance time by media condition (Study 2).**

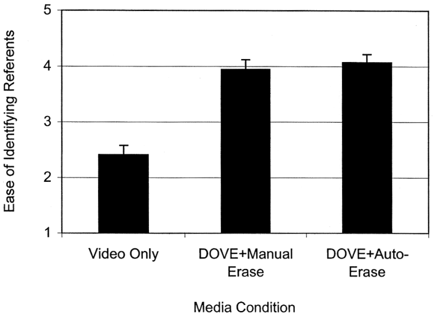


lated with performance time, *r*(82) = –.61, $p < .001$, and positively correlated with completion of the task, *r*(82) = .45, $p < .001$. Coordination was highest in the auto-erase condition ($M$ = 3.63, $SD$ = .53), intermediate in the manual-erase condition ($M$= 3.45, $SD$= .58), and lowest in the video-only condition ($M$= 3.13, $SD$= .58). A 3 (trial) × 3 (task) × 3 (media condition) repeated measures ANOVA indicated significant main effects for task, *F*(2, 29) = 4.59, $p < .05$; and media condition, *F*(2, 29) = 7.02, $p < .005$; and a significant three-way interaction, *F*(8, 29) = .05, $p = .05$. Post hoc comparisons indicated that the video-only condition differed significantly from the manual-erase condition, *t*(29) = –2.20, $p < .05$, and from the auto-erase condition, *t*(29) = –3.73, $p < .001$, but that the two erasure conditions did not differ significantly from one another.

***Ease of Identifying Referents.*** Pairs indicated that they could identify objects in the workspace better when the Helper used the drawing tools (Figure 9). Mean ratings were 4.07 ($SD$= .81), 3.95 ($SD$= .88) and 2.41 ($SD$= .88) for the automatic-erase, manual-erase, and video-only conditions, respectively. Rated ability to refer to objects was positively correlated with mean coordination scores, *r*(83) = .55, $p < .001$, and task completion, *r*(82) = .23, $p < .05$, and negatively correlated with performance time, *r*(82) = –.34, $p < .002$. A 3 (trial) × 3 (task) × 3 (media condition) repeated measures ANOVA indicated a significant main effect for media condition, *F*(2, 29) = 31.77, $p < .0001$, but no other main effects or interactions. Post-hoc tests showed a significant difference between the video-only and manual-erase conditions, *t*(29) = –6.49, $p < .0001$, and between the video-only and auto-erase conditions, *t*(29) = –7.33, $p < .0001$, but no difference between the erasure conditions.

***Final Questionnaire.*** On the final questionnaire, answered after all three tasks were completed, participants rated the value of different features of the

***Figure 9.*** **Ease of referent identification by media condition in Study 2 where the scale ranges from 1 (*very difficult*) to 5 (*very easy*).**



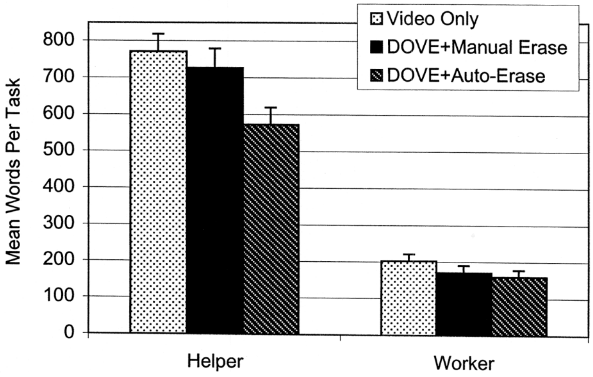
technology. Both Helpers and Workers rated both versions of the drawing software as helpful for their collaboration, but there were no differences between ratings of the two types of erasure (for Helpers, *M*= 4.14 for auto-erase and 4.11 for manual-erase; for Workers, *M*= 3.88 for auto-erase and 4.00 for manual-erase). Helpers also rated their preference for the two types of erasure for pointing, explaining procedures, and overall, using a scale of 1 (*strongly prefer auto-erase*) to 5 (*strongly prefer manual-erase*). Means were 2.43, 2.86, and 2.61 for pointing, explaining, and overall, respectively, showing a slight preference for the auto-erase mode.

Helpers were also asked to rate the potential usefulness of a number of new features for the technology on a scale of 1 (*not at all useful*) to 5 (*extremely useful*). As we have found in previous work (Fussell, Setlock, & Kraut, 2003), Helpers saw value in a shared manual that both they and their partners can view (*M*= 3.93, *SD*= 1.04). Helpers also rated the capability to manipulate the camera (zoom in and out, change camera orientation) as a potentially useful feature (*M*= 3.46, *SD*= 1.29). Having a laser pointer to point directly to task objects and locations was viewed as potentially useful (*M*= 3.19, *SD*= 1.39). Contrary to expectations, Helpers saw less value in gesture recognition software that would normalize their gestures (*M*= 2.21, *SD* = 1.03) or in a function that would allow the drawings to remain on the screen for a longer period of time (*M*= 2.39, *SD*= .99).

## Efficiency of Communication

We hypothesized that gains in performance and coordination with DOVE would stem at least in part from increased efficiency of communication. To investigate this hypothesis, we counted the total number of words used by

***Figure 10.*** **Mean words per task by media condition and participant role (Study 2).**



Helpers and Workers on each task. As expected, Helpers used fewer words to complete the task with DOVE than with video alone (Figure 10).

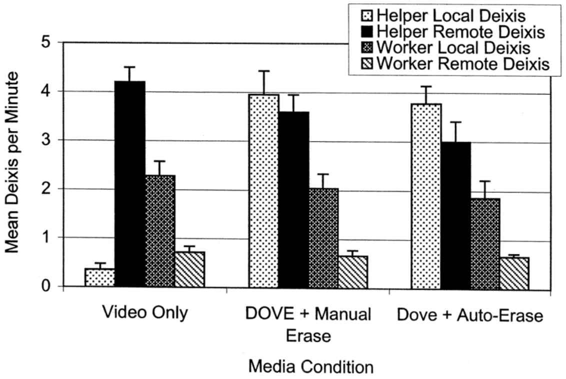
We analyzed words per task in 3 (trial) × 3 (task) × 3 (media condition) repeated measures ANOVAs. For Workers, there were no significant effects. For Helpers, there were effects for task, $F(2, 17) = 27.69$, $p < .0001$, and media condition, $F(2, 17) = 22.01$, $p < .0001$. Post hoc comparisons indicated significant differences between video-only and manual-erase, $t(17) = 6.50$, $p < .05$, and between video-only and auto-erase, $t(17) = 3.91$, $p < .001$, but not between manual-erase and auto-erase.

***"Local" Versus "Remote" Language.*** Previous studies have shown that shared visual space increases the use of deictic pronouns over lengthier descriptions of objects and locations and that this increase in deixis is associated with greater conversational efficiency (Kraut et al., 2003). We coded participants' language for the presence of six deictic terms: *this, these, here, that, those,* and *there*. Factor analysis revealed two dimensions: use of "local" terms that suggest being in the workspace (*this*, *these*, and *here*) and use of "remote" terms that suggest being at a distance (*that*, *those*, *there*).

We calculated deixis per minute by participant and condition (Figure 11). Workers used mostly local deixis and their rates were consistent across media. Helpers, however, showed a shift in their use of local deixis as a function of media: with video only, they used predominantly remote deixis, but with DOVE they used more local than remote deixis. Furthermore, rates of Helper local deixis were significantly correlated with self-reported ease of identifying objects in the workspace, $r(64) = .52$, $p < .001$.

For Helper local deixis, a 3 (trial) × 3 (task) × 3 (media condition) repeated measures ANOVA indicated significant main effects for task, $F(2, 17) = 12.21$,

*Figure 11.* **Local and remote deixis per minute by media condition and participant role. (Study 2).**



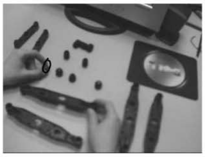
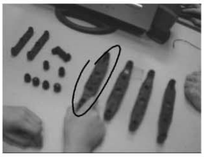
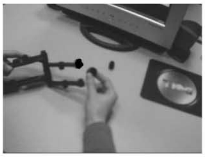
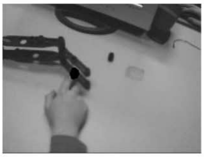
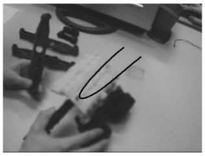
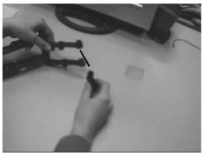
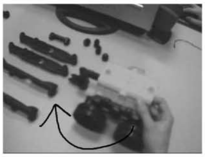
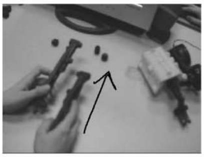
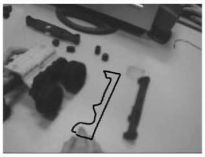
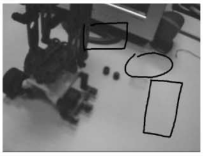
*p* < .001, and media condition, *F*(2, 17) = 99.78, *p* < .0001, and a Trial × Media Condition interaction, *F*(4, 17) = 2.90, *p* = .05. Post hoc comparisons indicated significant differences between video-only and manual-erase, *t*(17) = –10.97, *p* < .0001, and between video-only and auto-erase, *t*(17) = –13.01, *p* < .0001, but not between erasure conditions.

## Analysis of Drawings

To understand better how Helpers used drawings to facilitate task instructions, we coded each drawing in terms of its function in the interaction. A preliminary examination of the video tapes indicated that the drawings fell into five functional categories: pointing out task objects, pointing out locations, showing angles of insertion, showing the orientation of the object being constructed, and actual drawings or sketches. Examples of each category are shown in Figure 12, along with samples of drawings in that category. Occurrences and durations of each type of drawing were coded using a key-press software coding system we developed for video coding. Due to problems with the quality of recordings of a few sessions, coding was performed on observations from 24 of the 28 pairs. We computed reliability by comparing the mean scores by trial for each pair for each drawing category. Agreement was excellent ($\alpha$ = .98).

Overall, participants made between 2 and 62 drawings per trial, with a mean of 25.96, when the drawing tool was available (see Figure 13). The most frequent type of drawing was pointers to locations (*M* = 11.75), followed by pointers to objects (*M* = 7.69). Drawings indicating the angle of insertion of one piece into another, indicating the orientation of individual

*Figure 12.* **Functional categories used to classify participants' drawings in Study 2, along with examples of variations.**

| Category | Examples |
|---|---|
| 1A. Pointing to task objects | <br> |
| 1B. Pointing to target locations | <br> |
| 2. Indicating angle of insertion | <br> |
| 3. Indicating orientation | <br> |
| 4. Drawing Sketches | <br> |

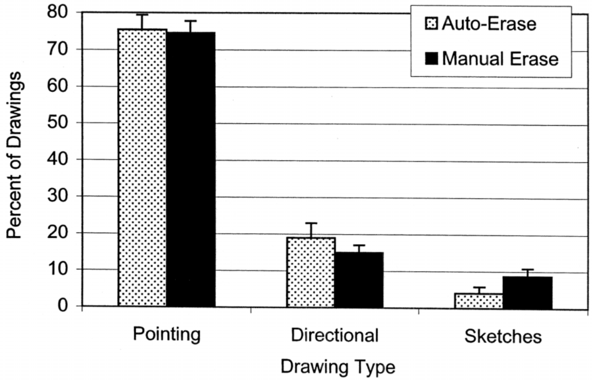
pieces or the task as a whole, or diagramming what pieces or the construction should look like were less frequent (*M*s = 2.71, 1.85 and 1.44, respectively). Repeated measures ANOVAs as well as chi-square tests indicated no significant differences between rates of drawing in each category, not even for sketches, which we had hypothesized would be more common in the manual-erase condition.

To better understand the roles different types of drawn gestures in the robot construction task, we grouped the two pointing categories (objects, locations) and the two directional categories (angle of insertion, orientation) together into "pointing" and "directional" drawings respectively. Figure 14 shows the percentages of pointing, orientation, and sketch drawings by erasure condition. As can be seen, pointing toward objects or locations made up

***Figure 13.*** **Mean frequencies of drawings per task by category and erasure condition (Study 2).**

| | Auto-Erase Condition | | Manual Erase Condition | | Overall | |
|---|---|---|---|---|---|---|
| **Type of Drawing** | **M** | **SD** | **M** | **SD** | **M** | **SD** |
| Pointing to objects | 7.92 | 6.19 | 7.46 | 5.60 | 7.69 | 5.85 |
| Pointing to locations | 11.13 | 8.11 | 12.38 | 7.91 | 11.75 | 7.95 |
| Indicating angle of insertion | 3.00 | 3.16 | 2.42 | 2.26 | 2.71 | 2.74 |
| ndicating orientation | 1.92 | 1.89 | 1.79 | 1.89 | 1.85 | 1.87 |
| Drawing sketches | 0.96 | 1.92 | 1.92 | 3.45 | 1.44 | 2.81 |
| Other | 0.42 | 0.83 | 0.63 | 1.41 | 0.52 | 1.15 |
| Total drawings | 25.33 | 17.20 | 26.58 | 14.35 | 25.96 | 15.68 |
| *N* | 24 | | 24 | | 48 | |

***Figure 14.*** **Percentage of pointing, directional, and sketch drawings by erasure condition (Study 2).**



the preponderance of drawings, comprising a mean of 75% of the total (*SD*= 18) and there was little difference between the two erasure conditions. Directional drawings comprised less than a fifth of all drawings (*M*= 16%, *SD*= 18) and although the percentage was somewhat larger in the auto-erase condition (19% vs. 15%), this difference was not significant. Sketches of pieces or how pieces should be combined were a relatively small percentage of all drawings (*M*= 6.4%, *SD*= 9.7). Sketches comprised a larger proportion of drawings in the manual-erase condition (*M*= 8.4%) than in the auto-erase condition (M = 4.1%), but this difference was not significant.

### 3.3. Discussion

The results of Study 2 demonstrate that a pen-based drawing tool can facilitate task communication and performance on our collaborative robot construction task, particularly when the drawings are automatically erased after several seconds. Performance times with the auto-erase version of DOVE were nearly identical to those reported for pairs working side by side in Study 1 and Fussell, Setlock, and Kraut (2003). In addition, conversations using DOVE in auto-erase mode were more efficient than those using manual-erase mode or video-only.

The findings suggest that pointing gestures were most prevalent. Helpers made heavy use of the drawing tool for indicating objects and locations as they provided their instructions. Helpers also used the tool in ways that parallel certain types of concrete representational gestures—they indicated angles of insertion (a type of iconic representation) and they used arrows to show direction and angle of motion. We return to this issue in the general discussion. Contrary to our expectations, sketches of what specific pieces or combinations of pieces looked like were relatively rare. In addition, we observed only two uses of the tool for writing directly over the video feed (labeling directions or pieces), although it could have been used more often for this purpose in the manual-erase condition. We suspect that the nature of the task, particularly the fact that the robot parts often lacked conventional names, is one reason why participants did not write on the video feed.

## 4. GENERAL DISCUSSION

Taken together, our studies suggest that simple tools can be used to convey gestures remotely in collaborative physical tasks, but that these tools need to be able to convey representational as well as pointing gestures. A cursor pointer alone was of no benefit over a video-only connection (Study 1), but a pen-based drawing tool led to significant improvements in performance times over video alone (Study 2). Furthermore, when the tool includes an automatic erase function, in which drawings disappear after a few seconds just as hand gestures disappear when they are completed, performance was better than when participants had to manually erase their drawings. This was true despite the fact that the manual-erase condition allowed participants to make complex drawings that combined different drawing strokes.

We believe the findings provide strong support for the value of gesture surrogates as a technique for implementing remote gesture in video systems supporting collaborative physical tasks. In particular, the DOVE pen-based system evaluated in Study 2 appears to provide an alternative means for creating both pointing and representational gestures that participants can use as readily as natural hand gestures. Performance times with the auto-erase version of DOVE, in particular, are close to the same as those in the side-by-side conditions of Study 1 and Fussell, Setlock, and Kraut (2003). It is important to note that both of the surrogates we examined (cursors and pen-based drawing) involve activities that are already natural to participants in other contexts (computer work, paper and pencil sketching). This may be one reason why they could so ready adapt to the use of drawing as a manner of gesture.

The monitor located in front of the Worker's task space likewise serves as a surrogate for the actual view of the space. We had anticipated that Workers might find the need to align the camera view with their own view of the workspace potentially problematic, but no signs of difficulty were observed during the sessions or reported in the questionnaires. Thus, it appears that using a separate monitor for remote gesture rather than overlaying these gestures on the actual workspace (as, for example, is done in Kuzuoka's laser pointing systems; Kuzuoka et al., 1994, 2000) may be sufficient for this type of collaborative task. It is certainly possible, however, to adapt the system to work with a head-worn camera (e.g., Fussell et al., 2000; Fussell, Setlock, & Kraut, 2003; Kraut et al., 1996), such that the video feed of the Helper's gesture is aligned completely with the corresponding object–task in the Worker's visual field.

In the remainder of this discussion, we first discuss the relationships between surrogate gestures created with our tools and task communication and performance; then, we briefly describe several of the limitations of this work; finally, we conclude with directions for future research.

## 4.1. Language and Surrogate Gestures

Our word count analyses in Study 2 demonstrate that participants' language while using the drawing tool was strikingly efficient. Examples of comparable instructions from the Scene Camera and DOVE + auto-erase conditions are shown in Figure 15. As can be seen in the examples on the right side of the figure, messages using the drawing tool incorporated deictic pronouns such as "this" and "that" in ways undistinguishable from the use of deixis in the side-by-side conditions of our and others' previous studies (e.g., Bauer et al., 1999; Fussell et al., 2000; Karsenty, 1999; Kraut et al., 2003). Furthermore, as shown in the left side of this figure, and likewise consistent with our earlier studies, Workers in the scene camera condition could use deictic pronouns because they knew the Helper could view their activities on the video feed, but Helpers had to use lengthier verbal expressions to denote objects and locations. The findings thus suggest that, as we had suspected, pairs' performance in video-only conditions suffers at least in part because of the asymmetry between Workers' and Helpers' ability to point within the shared visual

*Figure 15.* **Examples of instructions from the video-only (left) and DOVE plus auto-erase (right) conditions for matched steps in the procedure.**

| **Video Only** | | **DOVE + Auto-Erase** | |
|---|---|---|---|
| Helper: | And then you're going to take the little piece–the little gray piece with the black knobs in it, you're going to attach it to those–the holes below your finger. Does that make sense? | Helper: | All right, and then take the black medium one and then connect them like that. |
| Worker: | These two holes? | | |
| Helper: | Yes. | | |
| Worker: | Okay. Is that right? | | |
| Helper: | Yes. | | |
| Helper: | Um, actually, flip your main piece, the big piece, uh, the other way, the other way, so that the wheels are away from you. The big wheels are away from you, yes like that okay. | Helper: | Okay, so first turn that 90 degrees in that direction. |

*Note.* Underlined text corresponds to the drawing of gestures

field. Eliminating this asymmetry through the use of surrogate gesture techniques greatly enhances performance.

A puzzling finding concerns the large percentage of pointing gestures found in Study 2. If pointing gestures do comprise nearly 70% of all gestures in remote instruction-giving, then it is curious that we did not find performance improvements with the cursor pointer in Study 1. One possibility is that representational gestures describing how to insert or manipulate the pieces comprised a small but crucial component of conversational grounding. Angles and directions are especially difficult to communicate verbally (Fussell et al., 2000; Fussell, Setlock, & Kraut, 2003). Another possibility is that the manner of pointing in Study 2 (often drawing a circle, oval, or rectangle around the intended piece) was easier for Helpers to use and/or for Workers to identify on the monitor. However, Helpers used the cursor and drawing tools for pointing with similar frequencies, and participants in Study 1 never mentioned experiencing problems with the cursor pointer. For these reasons, we conclude that it is the ability to draw representational gestures in addition to pointing that is the critical difference between the two tools.

A comparison of the hand gesture categories presented in Figure 2 with the drawing coding scheme in Figure 12 suggests that drawn gestures are not always readily characterized as pointing, iconic, spatial, and kinetic gestures. Sketches are clearly iconic, similar to using a finger to draw what a piece looks like in space. Gestures showing direction of rotation are clearly kinetic. But some drawings appear to be hybrids, encompassing more than one of the categories in Figure 2. For example, one Helper drew a line between two parts of the robot to show the way in which a straight piece connects to two larger, parallel pieces. Such a drawing can be considered iconic, in that it represents the image of the smaller piece attached to the larger one; it can be considered spatial or orientational; and it can be considered an instance of pointing. In fact, this type of drawing, which occurred frequently in the corpus, often co-occurred with the expression "there," suggesting that it is intended as a pointing gesture.

The differences between our drawing coding system, developed on the basis of what was observed in the corpus, and the more conventional hand gesture system in Figure 2 arise, we suspect, from the inherent differences between hand gesturing and drawing. Hand gestures can involve either one or both hands, and various fingers, depending upon the type of gesture. Drawing a gesture with our system, in contrast, always involves a single pen tip, although it can be used in various ways to represent different types of gestures. Hand gestures also involve 3D space, whereas the drawing tool does not. Pointing with a hand is not static, as in Study 1. The hand moves out from the body, signifying movement to the destination, as was possible to represent only in Study 2.

A fuller understanding of the role of drawing gestures in task performance will require more detailed measures of Workers' cognitive activities. Our data do not allow us to answer questions, for example, about whether Workers always examined the drawings on the monitor in front of them, or what behaviors they subsequently performed. In face-to-face conversations, listeners do not always fixate speakers' gestures (Gullberg, 2003; Gullberg & Holmvquist, 1999). The effects of the drawing tool on performance times strongly suggests, however, that Workers were looking at these gestures. In addition, most studies of representational gestures use an experimental paradigm in which a speaker narrates a story to a passive listener (e.g., Alibali, Health, & Meyers, 2001; McNeill, 1992). The frequency, forms, and interactional functions of gestures may differ substantially in collaborative physical tasks, in which listeners must actively manipulate objects or the environment in response to speaker's messages. To address these issues, we are extending our previous research on the Helper's visual attention during collaborative physical tasks (Fussell, Setlock, & Parker, 2003) to the study of the Worker's visual attention.

## 4.2. Limitations of the Current Studies

Like any experimental paradigm, our choices of tool, testing conditions, and task have implications for the generalizability of the findings. First, we specified default values for the tools (e.g., cursor pointer size and color, pen-based tool width and color) based on pretesting. It is possible that different values for these features could impact collaboration. Furthermore, in Study 2 we specified erasure conditions in advance, rather than allowing participants to switch between automatic and manual erase as their needs shifted. In some task situations, it may be desirable to place the pen in manual-erase mode to construct complex diagrams but to use auto-erase the remainder of the time. Our pretesting also led us to select a 3-sec interval for the auto-erase condition; it is certainly possible that other intervals would lead to even better performance.

Second, our surrogate gesture systems, particularly the pen-based system in Study 2, were tested under better than normal network conditions. The wireless network we set up between the IP camera, Worker computer, and Helper tablet PC minimized the effects of jitter or delay on communication and performance. Gutwin (2001), for instance, showed that delay and jitter can disrupt communication and performance using collaborative cursor pointers. We recently completed a study of nine pairs using DOVE with a 0-, 1-, and 2-sec delay and found no effects of delay on performance. However, it is possible that delay would affect a more time-critical task.

Third, we examined our gesture tools in the context of a single type of task. As noted earlier, collaborative physical tasks vary along a number of dimensions including the nature of the task, the number of participants, and the roles of the collaborators. Further evaluation of the drawing tool will need to systematically manipulate these factors to identify the tasks for which it is most beneficial. For example, the robot construction task is characterized by pieces that vary in color and range from a half-inch to several inches in size. Whether the drawing tool will remain valuable for tasks involving much smaller or less identifiable pieces remains to be determined.

In addition, the robot task is a dyadic instructional collaborative physical task, in which the roles of the participants are clearly distinguished. A number of real-world collaborative physical tasks, such as telemedical applications, can involve multiple experts at multiple remote sites working together to provide instructions to a medical team. Minneman and Bly (1991) found that the Commune collaborative drawing tool extended well from dyadic to triadic teams, but this must be tested empirically for our drawing tool. The robot task is also asymmetrical, in that only the remote partner needs to gesture on the video feed. Other tasks, such as collaborative design, may require a system in which both parties can gesture over the same video feed. We are currently generalizing the system to allow Workers to draw on the shared video feed and to support drawings by multiple remote participants.

### 4.3. Future Directions

Although we have demonstrated the value of a simple drawing tool for remote gesture in collaborative physical tasks, there are several key areas for future work: extending the tool functionality, extending the video capability of the system, and expanding the range of tasks.

With respect to tool functionality, we are extending the drawing system to incorporate additional features. One feature, a gesture recognition module, has already been built and tested for accuracy (Ou et al., 2003). This module takes user input that approximates common shapes (e.g., circles, rectangles) and normalizes them. Whether this normalization will enhance interpersonal communication remains to be tested. Notably, participants in Study 2 did not foresee value to this functionality in their postexperimental questionnaires, and our observations of experimental sessions gave no sign that Workers had trouble understanding the freehand drawings. For other tasks, however, such as the drawing of complex diagrams, gesture normalization may be very useful. We are also exploring the dual use of the pen for both interpersonal and human–computer communication. In our latest DOVE system, pen-based input can be used to control camera functions such as pan and zoom in addition to drawing on the video. We are currently assessing the value of this system for collaboration in a controlled laboratory study. We are also considering a feature that allows users to save drawings, as has been implemented in collaborative drawing tools such as Commune (Bly & Minneman, 1990). It is not clear, however, that saved drawings would have as much value in a setting in which the environment is constantly changing.

A second direction for future work involves extending the video capabilities of the system. In Study 2, and in our previous work (Fussell, Setlock, & Kraut, 2003), Helpers indicated that they would value a video feed from an overhead camera. An important consideration is the effects of the camera position and orientation on Workers' ability to make the appropriate correspondence between what they view on their monitor and their view of the workspace. In the current system, the Worker's view and the camera view are closely aligned and problems of interpretation were never observed. If the monitor shows an overhead view or shifts views depending on Helpers' camera manipulations, problems of interpretation may increase. A second problem concerns the presentation of multiple views. In several previous studies, problems in establishing joint focus of attention between collaborators were observed when more than one view was present (Fussell, Setlock, & Kraut, 2003) or when people could switch among views (Gaver, Sellen, Heath, & Luff, 1993).

A final area for future research is to investigate the value of the pen-based gesture surrogate in a variety of task domains. We are currently examining the effects of the task objects themselves, particularly their size and differentiability, on task performance with and without the drawing tool. Future studies will be needed to examine the value of pen-based drawing for tasks with different numbers of participants and different role structures, and in settings in which both Helpers and Workers need to gesture via drawings in a shared video space.

## 5. CONCLUSION

In this article, we have explored two tools for remote gesture during collaborative physical tasks. Both tools take a surrogate approach to remote gesture, in which a cursor (Study 1) or a pen-based drawing tool (Study 2) is used in lieu of the hands to make gestures overlayed on a video feed of the work environment. This type of system is much less expensive and easier to implement than alternative approaches that attempt to convey actual hand gestures. The results demonstrate that a simple cursor pointing tool is not sufficient for remote collaboration on physical tasks, but that the DOVE pen-based drawing tool, which allows for a range of pointing and representational gestures, can lead to communication and performance virtually identical to that found in side-by-side collaborations.

## NOTES

*Background.* A portion of Study 1 was presented as a short paper at the CHI 2003 Conference on Human Factors in Computing Systems, Fort Lauderdale, Florida, April 2003.

***Acknowledgments.*** We thank William Kunz for developing the pointing software, Asa Sherrill for developing the key-press video coding system, and Katherine Bankson, Angela Ho, Marina Kobayashi, Kathy Lau, Katherine Waite, and Jim Zheng for running participants and preparing data for analysis. We also thank three anonymous reviewers for their helpful comments on the article.

***Support.*** This research was funded by National Science Foundation Grants 9980013 and 0208903. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

***Authors' Present Addresses.*** Susan R. Fussell, Leslie D. Setlock, Jie Yang, and Adam Kramer, Human–Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: {sfussell, lsetlo, yang, adk}@cs.cmu.edu. Jiazhi Ou, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: jiazhiou@cmu.edu. Elizabeth Mauer, Aptima Inc., 12 Gill Street, Suite 1400, Woburn, MA 01801. E-mail: lmauer@aptima.com.

***HCI Editorial Record.*** First manuscript received July 14, 2003. Accepted by Sharon Oviatt. Final manuscript received April 6, 2004. —*Editor*

## REFERENCES

Alibali, M. W., Health, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44,* 169–188.

Bauer, M., Kortuem, G., & Segall, Z. (1999). “Where are you pointing at?” A study of remote collaboration in a wearable video conference system. *Proceedings of the Third International Symposium on Wearable Computers (ISWC’99),* 151–158. Piscataway, NJ: IEEE Press.

Bekker, M. M., Olson, J. S., & Olson, G. M. (1995). Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. *Proceedings of the DIS 1995 Conference on Designing Interactive Systems,* 157–166. New York: ACM.

Bly, S. A., & Minneman, S. L. (1990). Commune: A shared drawing surface. *Proceedings of the OIS 1990 Conference on Office Information Systems,* 184–192. New York: ACM.

Bolt, R. (1980). “Put-that-there”: Voice and gesture at the graphics interface. In *Proceedings of the SIGGRAPH 1980 Conference on Computer Graphics and Interactive Techniques,* 262–270. New York: ACM.

Brinck, T., & Gomez, L. M. (1992). A collaborative medium for the support of conversational props. *Proceedings of the CSCW 1992 Conference on Computer Supported Cooperative Work,* 171–178. New York: ACM.

Cassell, J. (1998). A framework for gesture generation and interpretation. In R. Cipolla & A. Pentland (Eds.), *Computer vision in human–machine interaction* (pp. 191–215). Cambridge, UK: Cambridge University Press.

Clark, H. H. (1996). *Using language.* Cambridge, UK: Cambridge University Press.

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: APA.

Clark, H. H., & Marshall, C. E. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.

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

Efron, D. (1941). *Gesture and environment.* New York: King’s Crown Press.

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavioral categories: Origins, usage, and coding. *Semiotica*, *1*, 49–98.

Emmorey, K., & Casey, S. (2001). Gesture, thought and spatial language. *Gesture, 13,* 35–50.

Flor, N. V. (1998). Side-by-side collaboration: A case study. *International Journal of Human–Computer Studies, 49*, 201–222.

Ford, C. E. (1999). Collaborative construction of task activity: Coordinating multiple resources in a high school physics lab. *Research on Language and Social Interaction, 32,* 369–408.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers’ assumptions about what others know. *Journal of Personality and Social Psychology*, *62*, 378–391.

Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of the CSCW 2000 Conference on Computer Supported Cooperative Work,* 21–30. New York: ACM.

Fussell, S. R., Setlock, L. D., & Kraut, R. E. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of the CHI 2003 Conference on Human Factors in Computing Systems,* 513–520. New York: ACM.

Fussell, S. R., Setlock, L. S., & Parker, E. M. (2003). Where do helpers look? Gaze targets during collaborative physical tasks. *Extended Abstracts of the CHI 2003 Conference on Human Factors in Computing Systems,* 768–769. New York: ACM.

Gaver, W., Sellen, A., Heath, C., & Luff, P. (1993). One is not enough: Multiple views in a media space. *Proceedings of the Interchi 1993 International Conference on Human Factors in Computing Systems,* 335–341. New York: ACM.

Goodwin, C. (1996). Professional vision. *American Anthropologist, 96,* 606–633.

Greenberg, S., Gutwin, C., & Roseman, M. (1996). Semantic telepointers for groupware. *Proceedings OzCHI '96 Australian Conference on Computer–Human Interaction,* 54–61.

Greenberg, S., & Roseman, M. (1996). GroupWeb: A WWW browser as real time groupware. *Companion Proceedings of the CHI 1996 Conference on Human Factors in Computing Systems,* 271–272. New York: ACM.

Gullberg, M. (2003). Eye movements and gestures in human face-to-face interaction. In J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eyes: Cognitive and applied aspects of eye movements* (pp. 685–703). Oxford, England: Elsevier Science.

Gullberg, M., & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition, 7,* 35–63.

Gutwin, C. (2001). The effects of network delays on group work in real-time groupware. *Proceedings of the ECSCW 2001 European Conference on Computer Supported Cooperative Work,* 299–318. Bonn, Germany: Kluver.

Gutwin, C., & Penner, R. (2002). Improving interpretation of remote gestures with telepointer traces. *Proceedings of the CSCW 2002 Conference on Computer Supported Cooperative Work,* 49–57. New York: ACM.

Ishii, H., Kobayashi, M., & Grudin, J. (1993). Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems, 11,* 349–375.

Jefferson, G. (1972). Side sequences. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 294–338). New York: Free Press.

Karsenty, L. (1999). Cooperative work and shared visual context: An empirical study of comprehension problems in side-by-side and remote help dialogues. *Human–Computer Interaction, 14,* 283–315.

Kendon, A. (1972). Some relationships between body motion and speech. In A. R. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–210). Elmsford, NY: Pergamon.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human–Computer Interaction, 18,* 13–49.

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *Proceedings of the CSCW 2002 Conference on Computer Supported Cooperative Work,* 31–40. New York: ACM.

Kraut, R. E., Miller, M. D., & Siegel, J. (1996). Collaboration in performance of physical tasks: Effects on outcomes and communication. *Proceedings of the CSCW 1996 Conference on Computer Supported Cooperative Work,* 57–66. New York: ACM.

Kuzuoka, H., Kosuge, T., & Tanaka, K. (1994). GestureCam: A video communication system for sympathetic remote collaboration. *Proceedings of the CSCW 1994 Conference on Computer Supported Cooperative Work,* 35–43. New York: ACM.

Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., & Mitsuishi, M. (2000). GestureMan: A mobile robot that embodies a remote instructor's actions. *Proceedings of the CSCW 2000 Conference on Computer Supported Cooperative Work,* 155–162. New York: ACM.

Kuzuoka, H., & Shoji, H. (1994). Results of observational studies of workspace collaboration. *Electronics and Communication in Japan, 77*, 58–68.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

Minneman, S. L., & Bly, S. A. (1991). Managing a trios: A study of a multi-user drawing tool in distributed design work. *Proceedings of the CHI 1991 Conference on Human Factors in Computing Systems,* 217–224. New York: ACM.

Ou, J., Fussell, S. R., Chen, X., Setlock, L. D., & Yang, J. (2003). Gestural communication over video stream: Supporting multimodal interaction for remote collaborative physical tasks. *Proceedings of ICMI 2003 5th International Conference on Multimodal Interfaces,* 242–249. New York: ACM.

Pedersen, E. R., McCall, K., Moran, T. P., & Halasz, F. G. (1993). Tivoli: An electronic whiteboard for informal workgroup meetings. *Proceedings of InterCHI 93,* 391–398. New York: ACM.

Roussel, N. (2001). Exploring new uses of video with videoSpace. In R. Little & L. Nigay (Eds.), *Proceedings of EHCI 01, the 8th IFIP International Conference on Engineering for Human–Computer Interaction,* 73–90. Heildelberg: Springer.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.

Siegel, J., Kraut, R. E., John, B. E., & Carley, K. M. (1995). An empirical study of collaborative wearable computer systems. *Conference Companion to the CHI 1995 Conference on Human Factors in Computing Systems,* 312–313. New York: ACM.

Stefik, M., Foster, G., Bobrow, D., Kahn, K., Lanning, S., & Suchman, L. (1987). Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. *Communications of the ACM, 30,* 32–47.

Streitz, N. A., Geissler, J., Haake, J. M., & Hol, J. (1994). DOLPHIN: Integrated meeting support across local and remote desktop environments and liveboards. *Proceedings of the CSCW 1994 Conference on Computer Supported Cooperative Work,* 345–358. New York: ACM.

Tang, J. C. (1991). Findings from observational studies of collaborative work. *International Journal of Man–Machine Studies, 34*, 143–160.

Tang, J. C., & Leifer, L. J. (1988). A framework for understanding the workspace activity of design teams. *Proceedings of the CSCW 1988 Conference on Computer Supported Cooperative Work,* 244–249. New York: ACM.

Tang, J. C., & Minneman, S. L. (1991). VideoDraw: A video interface for collaborative drawing. *ACM Transactions on Information Systems, 9,* 170–184.

Wolf, C. G., & Rhyne, J. R. (1993). Gesturing with shared drawing tools. *Adjunct proceedings of the Conference on Human Factors in Computing Systems (Interact93)*, 137–138. New York: ACM.

Wolf, C. G., Rhyne, J. R., & Briggs, L. K. (1992). Communication and information retrieval with a pen-based meeting support tool. *Proceedings of the CSCW 1992 Conference on Computer Supported Cooperative Work,* 322–329. New York: ACM.