LANGUAGE EFFICIENCY AND VISUAL TECHNOLOGY
Minimizing Collaborative Effort with Visual Information

DARREN GERGLE
ROBERT E. KRAUT
SUSAN R. FUSSELL
Carnegie Mellon University

When collaborators work on a physical task, seeing a common workspace transforms their language use and reduces their overall collaborative effort. This article shows how visual information can make communication more efficient. In an experiment, dyads collaborated on building a puzzle. They communicated without a shared visual space, using a shared space featuring immediately updated visual information, and using a shared space featuring delayed visual updating. Having the shared visual space helps collaborators understand the current state of their task and enables them to ground their conversations efficiently, as seen in the ways in which participants adapted their discourse processes to their level of shared visual information. These processes are associated with faster and better task performance. Delaying the visual update reduces benefits and degrades performance. The shared visual space is more useful when tasks are visually complex or when participants have no simple vocabulary for describing their environments.

Keywords: *shared visual space; collaboration; communication; discourse; computer-supported collaborative work*

Consider an architect and client working together side-by-side to discuss architectural plans for a new corporate headquarters. Communication among them does not merely consist of the words they exchange, produced independently and presented for others to hear. Rather, speakers and addressees integrate and take into account what one another can see (Schober, 1993; Schober & Clark, 1989). They

1

notice where the other's attention is focused (Argyle & Cook, 1976; Boyle, Anderson, & Newlands, 1994; Fussell, Setlock & Parker, 2003), point to objects and use deictic references like "that one" and "there" (Barnard, May, & Salber, 1996), demonstrate and manipulate objects (Clark & Krych, 2004), make hand gestures, eye contact, facial expressions, and reference prior discourse and behavioral actions (Clark, 1996). Many of these processes take advantage of shared visual information. Using visual information to infer what another person knows facilitates efficient communication and reduces the ambiguity otherwise associated with particular linguistic expressions.

Shared visual information can be an extremely efficient collaboration mechanism, particularly when behaviors and actions are linguistically complex. It also serves as a precise indicator of comprehension. Finally, it may be used to provide situational awareness in regard to the overall state of a joint task. As pairs attempt to communicate as efficiently as possible, the visual information provided in a shared visual workspace can be used in several ways to minimize the overall level of joint effort required. Although these communicative techniques are often critical to successful interaction in the everyday world, technologies designed to support communication at a distance often fail to support them adequately.

A shared visual space occurs when the architect and client are collocated and gathered around the table, looking at architectural plans. It can also occur through technological mediation, for example, when distant collaborators jointly look at documents on yoked computer screens. In either case, a shared visual space enables people jointly to view approximately the same objects at approximately the same time. In designing a shared visual space technologically, the designers have many choices about how to construct it. For example, they can influence what images are transmitted (the users or the objects), the orientation of the images, refresh rate for the information, or the levels of detail that are transmitted between the communicators. How these decisions are made can be informed by application of the grounding theory of language and communication (Clark & Brennan, 1991). Grounding phenomena shape the language and understandings that communicators exchange. Therefore, language and the results of these understandings provide evidence for the effectiveness of the technology design, as well as for the theory itself.

This article has two major goals. First, it is designed to examine how a shared visual workspace influences communication in a collaborative work task. The second research goal is to examine how one should design a shared visual space to support effective communication.

PREVIOUS WORK

Most of the early research examining the utility of visual information in communication focused on the degree to which collaborators were aware of one another, at the expense of visual information about the objects they discussed. This research tradition is derived from work conducted by the Communications Study Group at British Telecom (Short, Williams, & Christie, 1976) and in Chapanis' lab in the United States (Chapanis, Ochsman, Parrish, & Weeks, 1972). Studies compared dyads performing a referential communication task (i.e., a task where a speaker communicates information about objects, pictures, directions, etc.) using only an audio channel with dyads working face-to-face or using an audio/video connection. This research concluded that visual information was not important for referential communication.

More recent research shifts the focus from a view of the participants' faces to a view of the work area. One line of research using realistic work tasks in this new wave has uniformly found that participants in side-by-side settings, in which they share full views of one another and the workspace, perform better than participants using a variety of communications tools (Fussell et al., in press; Kraut, Fussell, & Siegel, 2003; Nardi et al., 1993).

However, results are mixed when the research uses video to create the shared visual space. For example, Fussell, Kraut, and Siegel (2000) had paired "worker" and "expert" dyads repair a bicycle while conversing side-by-side, using audio plus a head-mounted camera transmitting the worker's view of the bicycle to a remote expert, or via audio only. Pairs were substantially faster when they worked side-by-side than in the audio condition. Although dyads used different techniques to refer to objects in the video-mediated condition than in the audio condition, their overall performance time was no better. In contrast, Fussell, Setlock, and Kraut (2003) found that pairs performed better when they use video tools that provide views of the workspace than when using audio or text-based communication alone.

The differences among video configurations may lead to conflicting results. For example, in Fussell, Setlock, and Kraut (2003), remote communicators could make visible gestures in the video image, whereas in Fussell et al. (2000) they could not. Differences in the quality of the implementation may also account for different effects. For example, in Fussell et al. (2000), technical problems with the field of view, video transmission, and slippage of the camera on the worker's head made the video-mediated shared visual space inadequate. Thus, there is a need for more tightly controlled laboratory studies of shared visual space to complement these previous efforts.

To address these issues, a second line of work has been exploring more stylized communication tasks in tightly controlled laboratory

environments. For example, Clark and Krych (2004) used a stylized communication task in which one participant, a Director, instructed another, a Matcher, on how to construct a simple LEGO form. When the Director could see what the Matcher was doing, the pair was substantially faster, in part because the pair could precisely time their words to the actions they were performing. Although this work provides initial insight into the ways in which shared visual space leads to more efficient conversation, the exact mechanisms by which the improvement occurs are unclear. Consider the nature of a shared visual space when people are working side-by-side: Voice is synchronized to actions, the parties are mobile, both parties can point to objects in space, each party can see both the work area and each other's face and gestures, and each party sees the workspace from a slightly different angle. Which of these features of the side-by-side setting need to be reproduced to re-create the benefits of proximity through technology-mediated communication?

THE CURRENT STUDY

The study reported here uses a new technique to disaggregate the features of a shared space and to observe their effects on performance. In our paradigm, a Helper instructs a Worker in completing an online shape arrangement puzzle. Only the Worker can manipulate the puzzle. The Helper has a model of the completed arrangement of pieces and gives instructions and comments to guide the Worker. They share a visual space consisting of a view of the work area rendered on each of their computer screens. The benefit of this paradigm is that the view presented to the Helper can be any computationally derived transformation of the workspace shown to the Worker. For example, we can manipulate whether the Helper can see the workspace at all, whether the Helper sees the full workspace or only a subset of it, or whether the Helper sees the workspace immediately or after some delay. By using this paradigm we can identify features of the shared visual space that make it valuable.

We applied this paradigm to examine how a shared visual space (whether the Helper could see the shared visual space or not) and one of its attributes (the speed with which the shared visual information is updated) interact with two task attributes (visual complexity and temporal dynamics) to affect communication processes and task performance. We expect that having a shared visual space will be more important for tasks involving difficult-to-describe puzzles or tasks in which the environment changes rapidly. We also expected that delays in updating the shared visual space would degrade its usefulness. Krauss and Bricker (1967) had previously shown that auditory delays as small as 250 msec can affect both communication process and efficiency. Do delays in updating a shared visual space, of the sort pro-

duced by network congestion and video compression, cause similar problems?

IDENTIFYING THE CRITICAL ELEMENTS
OF SHARED VISUAL SPACE

To identify the important elements of a shared visual space, we must first understand how people use specific types of visual evidence for collaborative purposes. Clark and Wilkes-Gibbs (1986) observed that collaborative work occurs at multiple levels simultaneously, although the distinction between levels is not crisp. At the highest level, people collaborate on performing the task. In our experiment, they are jointly solving a puzzle. At a lower level, they use language and other communicative behaviors to coordinate actions in order to perform the task. At yet a lower level, dyads use communicative behaviors to coordinate the language they use. For example, pairs jointly determine the names to apply to pieces in the puzzle or indicate whether they understood a description. Visual evidence can be helpful at each of these levels. It can inform the Helper about the next puzzle action that the Worker needs to perform by giving an up-to-date account of the overall state of the task. It can guide the Helper in planning an instruction by indicating when it should be given and how it should be phrased. Finally, it can provide the Helper with evidence about whether the Worker understood an instruction.

FACILITATING CONVERSATION
AND GROUNDING

A shared visual space may facilitate the communication that surrounds a joint activity. Successful communication relies on mutual knowledge or common ground (Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986): the knowledge, beliefs, understanding, and so on, shared by the speaker and hearer known to be mutually available. Shared visual information helps communicators develop common ground, by giving them evidence from which to infer what others understand at any moment.

Generally, a speaker would not speak in Yiddish unless he thought a partner understood it, would not suggest "pinging the gateway" unless he thought the partner had telecommunications knowledge, nor use a pronoun unless he thought the partner understood the antecedent. Although these inferences about a partner's state of knowledge may be incorrect, they underlie speech production. As a result, throughout a conversation, participants are mutually assessing what each other knows at any moment and then using this knowledge to form their sub-

sequent utterances. Participants are obligated both to assess and give off cues that indicate their understanding. This method of exchanging evidence about understanding over the course of a dialog is referred to as the process of grounding.

Clark and Brennan (1991) hypothesize that different communication media have features that change the cost of grounding. For example, when communicating by electronic mail with large delays between conversational turns, participants cannot simultaneously transmit back channel communications—the "uh-huh", "I see", head nods, and smiles—that signal to one another the degree to which they understand the current utterance. In this research we are interested in how a shared visual space affects grounding. Clark and Brennan (1991) and Kraut, Fussell, Brennan, and Siegel (2002) suggest ways that a shared visual space can be helpful for establishing common ground (see also Brennan, in press; Endsley, 1995).

The principle of least collaborative effort asserts that participants in communication will try to minimize their collaborative effort (i.e., the work that they do from the initiation of each communication contribution to its mutual acceptance) (Clark & Wilkes-Gibbs, 1986). Shared visual information can help reduce collaborative effort at two distinct phases in the communication process: at the *planning* stage and the *acceptance* stage.

Planning takes place when the speaker is forming an utterance; it affects the efficiency of expressions. When describing a puzzle, one of the Helpers' goals is to form expressions that succinctly refer to the puzzle's pieces. If the Helper can see the work area, he can create efficient referring expressions by relying upon what the Worker already sees (e.g., using the phrase "that one" when observing that the Worker is hovering over the correct piece) or anticipating potential ambiguities (e.g., using the phrase "the dark red one" only if he can see that the Worker is likely to be confused by multiple red pieces). If the Helper cannot see the Worker's area, the Helper is likely to provide the wrong amount of information or rely upon the Worker to state explicitly what information she needs. Thus, by the principle of least collaborative effort, we should expect to see shifts in who acknowledges when a task is completed based on the degree of shared visual space.

The *acceptance* stage occurs when the speaker is assessing whether the conversational partner has understood the utterance. It provides comprehension monitoring. According to the collaborative model of conversation, after contributing an utterance to a conversation, a speaker should not move the conversation forward unless speaker and listener believe that the listener has understood the utterance sufficiently (Clark & Marshall, 1981). After giving instructions about a puzzle, seeing the Worker's consequent behavior provides the Helper information about the Worker's comprehension of the instruction. With shared visual space, the Helper can easily recognize when the

Worker is performing an incorrect action, when she appears confused, or did not understand a task. For example, in the present experiment, if a Helper notices that when the Worker puts one piece directly above another in response to the instruction, "put the piece kitty-corner" he can assume that that "kitty-corner" is not part of their shared language. The Helper can easily remedy this mistake by providing a more meaningful directive such as, "Above and to the right so that the corners are touching." Without shared visual space, the Helper needs to make assumptions about what the Worker understood or rely upon the Worker to explicitly state her level of understanding.

Visual information can provide a clearer signal of comprehension than a listener's self-assessment of understanding. If the Helper tells the Worker to "position the piece at 2 o'clock" and he can see the Worker's response, he can tell with certainty that the Worker has understood the instruction. However, if there is no shared visual space, then the Worker must state her understanding, for example, "OK, it's above the last piece," to which the Helper might respond, "Above and to the upper right?" Even at this point, the Helper cannot be certain that they are both speaking about the same piece. In this way visual information can provide a less ambiguous signal of comprehension than can language.

By seeing the partner perform some task, the Helper gets immediate feedback about whether the partner understood a directive. Clark and Krych (2004) demonstrated the *temporal precision* with which speakers use this visual evidence of understanding. For example, when a shared visual space is available, directors change their descriptions and further elaborate mid-sentence in response to their partner's behavior. They use visual information to determine the precise moment at which to disclose new information. Delays of the sort introduced by video compression or network lags are likely to undercut the value of the visual feedback.

Visual feedback, however, may be less necessary if the task is simple enough (e.g., a game of tic-tac-toe in which the pieces and positions are easily described) or if the partners have an efficient, well-practiced, and controlled vocabulary to describe events (e.g., routine communication between pilots and air traffic controllers). In these cases, a shared visual display provides little new information and its value for communicative purposes is diminished.

MAINTAINING AWARENESS
OF TASK STATE

In the previous section, we described how shared visual information can be useful in coordinating language during the planning of utterances that a partner can understand, and in monitoring whether that

partner does understand. Shared visual information can also be valuable for coordinating the task itself. In particular, if collaborators can see the state of the task as it develops, they know what work still needs to be done. This awareness helps them plan how to proceed toward the goal, what instructions they need to give, and how to repair incorrect actions. Shared visual information also provides the ability to monitor specific actions.

Imagine a pair performing a typical referential communication task in which a Helper is instructing a Worker on the order in which to place a set of cards (Isaacs & Clark, 1987). If the Worker places a card to the left when it should have been placed to the right, the Helper can intervene with new instructions if he can see the work area. Otherwise, the Helper must query the Worker on the order of the cards and rely upon the Worker providing an accurate description.

The benefit of the shared visual space should be greater as the task grows more visually complex because the visual complexity introduces more opportunities for task errors, and because the language is less adequate to describe the task state. For example, in the puzzle task used in the present experiment, the puzzles are two-dimensional (with abutting pieces) or three-dimensional (where one piece may overlap and occlude another), with corresponding levels of complexity. In the simple two-dimensional case, the instruction "Put the red piece on top of the blue one" is unambiguous, whereas in the three-dimensional case, the red piece can either overlap the blue piece or be north of it. If the Helper can see the work area, he can intervene to rectify any misinterpretation. He can also see when the Worker is ready for the next instruction.

HYPOTHESES

We can summarize this discussion about the influence a shared visual space on conversational grounding and task awareness in terms of three sets of hypotheses regarding performance in a referential communication task. The first concerns the effect of a shared visual space on task performance as measured by completion time. The second and third address the way visual information changes the content and structure of the communication as the pairs attempt to reduce their joint collaborative effort.

Performance. Because the shared visual space should help participants maintain awareness of what needs to be done in the puzzle and allows them to communicate more efficiently, we expect that it will lead to improved performance.

General Hypothesis 1 (H1): *A collaborative pair will perform a referential communication task more quickly when they have a shared view of the work area.*

When the referential task is more visually complex and involves a rapidly changing environment, language alone becomes less adequate for describing the task state, and the likelihood of errors increases. In these cases, the shared visual space should be more useful, in an interaction effect between the presence of shared visual space the visual complexity of the task.

H1a: *A shared view of the work area will have additional performance benefits when the task is more visually complex.*

We would further expect an interaction between the temporal dynamics of the task objects and the fidelity of the shared visual space.

H1b: *A shared view of the work area will have additional performance benefits when the objects in the task change versus when they are stable.*

However, the shared visual space should be less useful if it is not kept up to date because it will not be synchronized with the state of the task or the language it needs to support. As described by Clark and Krych (2004), spoken language is particularly useful when it can be precisely timed to physical actions and behaviors. Even a small delay in updating the visual space should be enough to disrupt this precision timing and diminish the value of visual information.

H1c: *Delay in transmission will diminish the value of a shared view of the work area.*

Communication efficiency. If a shared visual space allows pairs to communicate with less collaborative effort this should be reflected in the efficiency of a pair's language use, that is the number of words they need to give instructions, refer to objects, or to indicate their state of comprehension.

General Hypothesis 2: *A shared visual space will allow collaborators to communicate more efficiently.*
H2a: *Collaborators will use fewer words to complete their task when they have a shared visual space.*

Even though the shared visual space provides new information to the Helper by allowing him to see the Worker's behavior, we expect that the visual tool will primarily influence the Worker's language efficiency. If the pairs are operating according to the principal of least collaborative effort and the Worker is aware that the Helper can see the

space, then the Worker can let her actions substitute for her words in demonstrating her level of understanding.

> H2b: *A shared visual space should increase the Worker's communicative efficiency more than the Helper's.*

Communication process. To influence communication efficiency, the shared visual space must also affect the strategy collaborators use in forming utterances and indicating their level of understanding. Because the Helper forms his utterances on the basis of intuitive hypotheses regarding what information the Worker needs, providing a shared visual space should allow him to rely on more efficient linguistic shortcuts, such as the use of deictic pronouns and spatial deixis, in the formulation of referential statements. Both of these linguistic forms are ways of verbally referencing (or pointing to) a particular object in the display, or in the case of spatial deixis, the spatial relation between a reference object and a to-be-located object. For example, in the phrase "I want that" (pointing to an object), "that" is a deictic pronoun used to linguistically point to an object. Deictic pronouns are generally efficient, substituting for longer and more linguistically explicit referring expressions. Spatial deictic expressions are an example of longer and more explicit forms. For example, in the expression "It's the one on top of the red block," "on top of" uses the relative spatial position of objects to refer to them. If both Helper and Worker can see the spatial positions of puzzle pieces and know their partner can also see the positions, they should not need elaborated spatial deixis.

> H3a: *A shared visual space should increase collaborators' use of deictic pronouns.*
> H3b: *A shared visual space should decrease collaborators' use of explicit descriptions of spatial position (spatial deixis).*

In addition to the general efficiencies shown in the planning of messages, a shared visual space allows pairs to change their strategies for demonstrating and monitoring comprehension and should also reduce the amount of effort needed to monitor comprehension. With a shared visual space, the Helper can directly observe evidence of the Worker's comprehension. As a result, the Worker need not explicitly state it. On the other hand, with no shared visual space, Workers must frequently indicate verbally whether they have understood utterances.

> H3c: *The shared visual space should decrease the amount of acknowledgements explicitly stated.*

A lack of shared visual space should shift the burden of responsibility for verifying comprehension to the person performing the action. In the puzzle study explored here, this means the Worker will need to take on the responsibility of confirming their actions verbally.

H3d: *A lack of shared visual space should additionally increase the amount of acknowledgements explicitly stated by the Worker.*

METHOD

TASK

We investigated these hypotheses in an experiment that manipulated the fidelity of the shared visual space and attributes of the task. Participant pairs played the role of Helper and Worker in a referential communication task that involved the completion of a geometric puzzle. The goal was for the Worker to arrange puzzle pieces so that they matched the target that the Helper was viewing.

APPARATUS

The Helper and Worker were each seated in front of separate desktop computers with 21-inch color monitors. A divider positioned between the workstations prohibited the participants from seeing one another. This eliminated the pair's ability to use hand gestures, facial expressions, and so on. The Helper and Worker spoke out loud and each speech stream was captured by microphone and integrated with a time-stamped video capture of the displays. The general structure of the Worker's display can be seen in Figure 1a. It contained a staging area, on the right, where eight pieces for the puzzle were stored, and a work area, on the left, where the Worker constructed a four-piece puzzle. The Helper's display is shown in Figure 1b. It contained the target puzzle on the right, representing the goal state. In the left, it showed one of the three views of the Worker's work area, which we describe in more detail below. Pairs were notified before each trial regarding the status of the shared work area for the upcoming trials.

INDEPENDENT VARIABLES

The experimental displays for the Worker and Helper were written as two communicating Visual Basic programs. By constructing the displays computationally, we were able to manipulate the visual space that participants shared and the visual nature of their task in several ways. We manipulated the extent to which participants viewed the same work area (Fidelity of the Visual Space), the adequacy of lexical tokens to describe the puzzle pieces (Color Drift) and the visual complexity of the task itself (Puzzle Difficulty).
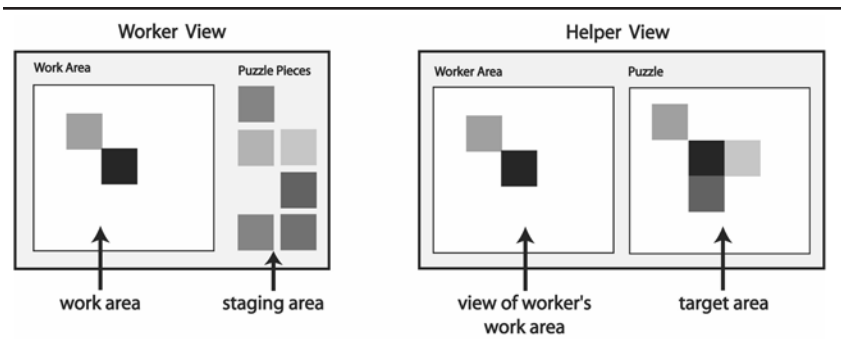
Figure 1a. **The Worker's View (left)**
Figure 1b. **The Helper's View (right)**

Fidelity of the visual space. We varied the degree to which the Helper could see the state of the Worker's puzzle (as reflected back to the Helper's own display). In any trial, the Helper could either see a replication of the Worker's work area with no delay, could see the work area with a 3-second delay, or could not see the work area at all. We call these, respectively, the Immediate, Delayed, and None visual space conditions.

Color drift. We varied the lexicality of the puzzle pieces by manipulating whether the colors of the blocks were static (e.g., red) or constantly cycling (e.g., red to orange to yellow to . . .). In the Stable condition, pieces were chosen randomly for each experimental condition from a palette of easily distinguishable colors (see the staging area in Figure 1). In the Drift condition, each piece slowly changed its color, incrementally cycling through the colors in the color palette. In general, the pieces passed through approximately four to six perceivable color changes every 24 seconds. The pieces changed at a rate of a major color change approximately every five seconds. It took roughly one second of continuous observation to notice whether any given piece was changing color. It should be noted that these values fluctuate somewhat due to the fact that people do not perceive change equally across the color spectrum.

Puzzle difficulty. We varied the difficulty of the puzzles by having configurations where the pieces simply abutted edges (Easy) or overlapped one another (Difficult). In the difficult condition, a piece could overlap either one-quarter or one-half of another piece. The layout algorithm guaranteed that a single piece was never completely occluded.

PARTICIPANTS AND PROCEDURE

Participants consisted of 12 pairs of Carnegie Mellon University undergraduate students, who received $10.00 per person for their participation in the study. The participants were randomly assigned to play the role of Helper or Worker. Color Drift was manipulated between pairs of participants, whereas both Visual Space and Puzzle Difficulty were manipulated within each pair. Each pair participated in six experimental conditions, once in each Visual Space (3) × Puzzle Difficulty (2) combination, counter-balanced. Pairs solved four puzzles within each experimental condition.

MEASURES

Task Performance Measure

The participants were instructed to complete the task as quickly as possible, so task performance was the time it took to complete the puzzle. Custom software logged and time-stamped all mouse events. Puzzle completion times were extracted from the logs by calculating the time between when both partners pressed buttons indicating they were ready to proceed with the next trial to the time the Helper pressed a button indicating the trial was successfully completed. Overall, the vast majority of the puzzles were solved correctly so differences in error rates were a less useful indicator of task performance.

Conversational Coding

To investigate the relationship between the shared visual space and dialogue we employed a coding scheme to identify the speaker (Helper or Worker) and the primary purpose of each utterance and action (see Table 1). The method was modified from the coding scheme described in Kraut et al. (2003). The typical cycle of performing this task involved the Helper describing one of the puzzle pieces, waiting until he was convinced that the Worker had identified the correct piece, and then telling the Worker its position in the work area. When he was convinced the piece was placed correctly he would describe the next piece. This would be repeated until the puzzle was completed.

In this report, we are especially interested in the language efficiency and manner in which participants referred to the objects in the puzzle, described the spatial positions of those objects, and how they verified that they were manipulating the correct pieces and positioning them correctly. To examine these issues in detail, we conducted our analyses using the categories presented in Table 1. In particular, the *reference* and *position* categories represent the substantive task communica-

Table 1
Types of Utterances Coded

Utterance Types	
Referents	References to and attempts to describe a specific piece (e.g., "Take the red one").
Referential context	Information providing the context for identifying a specific piece (e.g., "What colors do you have available?").
Position	Attempts to describe the position of a single specific piece (e.g., "Put that one in the upper right corner").
Positional context	Description of several pieces together (e.g., "The last three blocks should form a triangle like shape").
Acknowledgements of understanding	Responses to statements confirming an understanding (e.g., back-channel responses, "mmm-hmm").
Acknowledgements of behavior	Acknowledgements directly following a behavior indicating whether a partner had made a correct or incorrect move.

Table 2
Types of Deixis Coded

Deictic Expressions	
Deictic pronoun	Utterances that use the deictic pronouns "this," "that," "there," and related terms.
Spatial deictic	Utterances that refer to terms using spatial position, such as "above," "below," "in front of," "on top of," "next to," "behind," "right," "left," "up," "down," "touching."

tion. When spoken by the Helper, they were often instructions telling the Worker what to do. When spoken by the Worker, they were often attempts to clarify an instruction or verify that she had understood it correctly. The *acknowledgement* categories were brief exchanges asserting that the Worker had understood an instruction or performed it correctly.

We also assessed efficiency of communication by examining the use of deictic pronouns and spatial deictic expressions. Table 2 presents the types of deixis coded for in this analysis.

Two independent coders classified a 12% sample of utterances until they reached 90% agreement on all categories. They then each coded different transcripts, periodically coding a common transcript to ensure that the categories they used did not drift during the duration of the coding. Agreement remained high throughout.

STATISTICAL ANALYSIS

Each analysis is a repeated measures analysis of variance in which Block (combination of conditions 1-6), trial (1-4), Puzzle Difficulty (Easy or Hard) and Visual Space (Immediate, Delayed, None) were repeated,

and Color Drift (Stable or Drift) was a between-pair factor. We included 2-way and 3-way interactions in the analysis. Because each pair participated in 24 trials (6 conditions by 4 trials per condition), observations within a pair were not independent of each other. Pairs, nested within Color Drift, were modeled as a random effect. Our analysis of performance uses time to complete a puzzle, recorded as seconds, as the dependent variable. When we conducted the analysis of conversational efficiency, we included the number of words as the dependent variable and time to complete the task as a covariate. The analysis for conversational content included the number of referents, position statements, acknowledgements, and deictic expressions, with both time and number of words as covariates.

Our interest in this study is on the impact of the fidelity of a shared visual space on task performance, conversational efficiency, and conversational tactics. Although our analysis was a full factorial analysis of co-variance, with 3-way interactions, for reasons of space in this article, we focus on the influence of Visual Space and its interactions with Puzzle Difficulty, Color Drift, and Speaker Role.

RESULTS

MANIPULATIONS CHECKS

The manipulation of puzzle difficulty had a significant impact on the speed with which the pairs solved the puzzles. The pairs were faster when the pieces simply abutted edges [LS Mean (and standard error) = 62.5 (3.8)] than when they were difficult and the pieces overlapped [70.0s (4.3)], t (258) = 2.40, p = .017.[1] The manipulation of color drift also had a significant impact on performance speed. The pairs were significantly faster in trials where the colors were stable [LS Mean = 54.4s (5.3)] than when they were drifting [78.0s (5.3)], t (10) = 3.19, p = .009.

TASK PERFORMANCE

This experiment was designed to examine the impact of the fidelity of shared visual space on performance for different types of tasks. Consistent with *General Hypothesis 1*, the results show that a shared view of the work area benefited performance. The pairs were about a third quicker at solving the puzzles in the Immediate Shared Visual Space than in either the Delayed Shared Visual Space condition, t (258) = 4.57, p < .001, or the No Shared Visual Space condition, t (258) = 6.61, p < .001 [LS Means (se): Immediate = 52.3s (4.2); Delayed = 69.6s (4.5); None = 76.7s (4.4)]. However, consistent with *Hypothesis 1c*, delays in
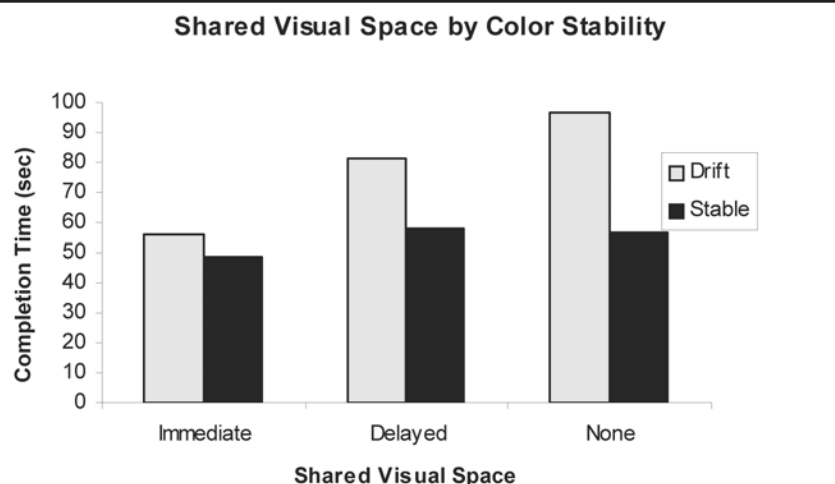
Shared Visual Space by Color Stability

Figure 2. **Effect of Shared Visual Space and Color Drift on Performance Time**

updating the shared view reduced its benefits. Indeed, the 3-second delay eliminated its benefit completely; the delayed shared view was no better than no shared view at all.

Consistent with *Hypothesis 1b*, the Visual Space × Color Drift interaction demonstrates that having a shared view of the work area had greatest benefit in the drift condition, when the objects being discussed were lexically unstable and difficult to describe (see Figure 2) interaction F (2, 258) = 11.41; $p < .001$). Decomposition of this interaction reveals that the Immediate Shared Visual Space condition led to substantially faster completion than the No Shared Visual Space condition when colors were changing than when they were stable, interaction t (258) = 4.33, $p < .001$. Similarly, the Immediate Shared Visual Space condition was faster than the Delayed Shared Visual Space condition when the colors were drifting than when they were stable, interaction t (258) = 2.19, $p = .03$ (see Figure 2.). Phrased another way, a shared view of the work area was less beneficial when words themselves could easily describe the objects (e.g., they could be called by concise color terms such as red, blue, or aqua). Because people precisely time their utterances in the grounding process (Clark & Krych, 2004), temporal synchrony matters a great deal.

It is instructive that the Visual Space × Puzzle Difficulty interaction, although in the hypothesized direction, was not statistically significant, F (2, 258) = 1.01, $p = .37$. Visual complexity itself did not raise the value of a shared view of the work area. Thus, we found no statistical support for *Hypothesis 1a*. It was primarily when the task was dynamic and the environment was changing that the display was most beneficial.

The next stage of analysis explored the way in which the language between the Helper and the Worker varied when the shared visual space was perturbed.

Communication Efficiency

We explored the rate at which the pairs produced words (in the log scale) in order to examine the efficiency with which they communicated. We examined word rate (the number of words, controlling for time) to test this prediction. The ANOVA model for the word rate analyses was similar to that for examining task performance, with a few exceptions. It included the speaker's role as a factor in the design (Helper or Worker) and used time to complete the task as a covariate. Because none of the three-way interactions were significant, with the exception of Block $\times$ Visual Space $\times$ Speaker Role, they were removed from the model in subsequent analyses.

Consistent with *General Hypothesis 2* and *Hypothesis 2a*, the pairs produced more efficient speech when they had higher fidelity shared visual space. They used fewer words to solve the puzzles, controlling for time, as the shared visual space was more accurate [LS Means (se): Immediate = 2.97 (.14) words (nLog) per puzzle; Delayed = 3.40 (.15); None = 3.81 (.15)]. The Immediate Shared Visual Space condition was more communicatively efficient than both the Delayed Shared Visual Space condition, $t(110) = -2.55$, $p = .01$, and the No Shared Visual Space condition, $t(110) = -4.84$, $p < .001$. In turn, the Delayed Shared Visual Space condition was more efficient than the No Shared Visual Space condition, $t(110) = 5.78$, $p = .017$.

An examination of the Shared Visual Space $\times$ Speaker Role interaction depicted in Figure 3 reveals that the fidelity of the shared visual space influenced the Workers' efficiency more than the Helper's, interaction $F(2, 110) = 10.81$, $p < .001$. Because the Workers could always see the work area, changes in Workers' behavior reflected their accommodation to differences in the Helpers' view of the workspace. This provided support for *Hypothesis 2b*.

Communication Process

We expected that the shared visual space would be useful in allowing the pairs to monitor the state of the task. When the workspace was present, the Helper could monitor the Worker's progress and issue corrections. However, when the shared space was not visible, the responsibility of communicating the task state shifted to the Worker. One of the ways this shift in responsibility might be seen is in the issuance of acknowledgements. We examined two types of acknowledgements. Acknowledgements of behavior examine the use of acknowledgements in response to behaviors or physical actions. Acknowledgements of
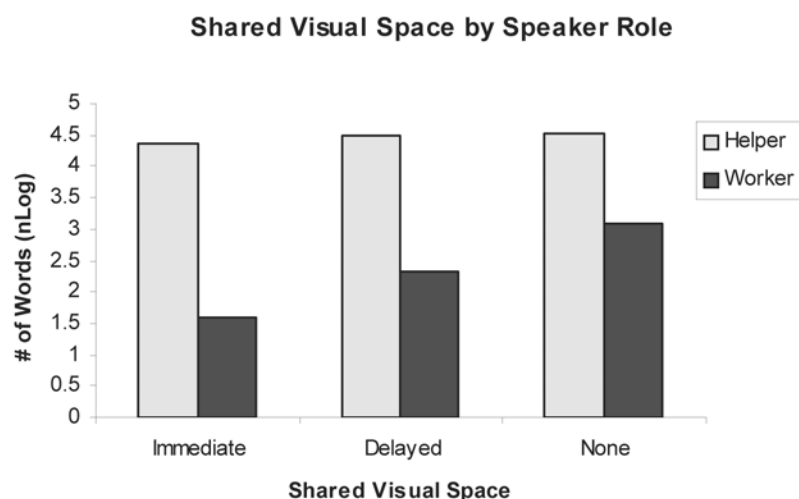
Shared Visual Space by Speaker Role

Figure 3. **Effect of Shared Visual Space and Speaker Role on Word Rate**

understanding look at the use of acknowledgements in response to statements or questions. The models used for the content-count analyses were similar to the ANOVA model for examining word rate, but they included the number of words as a covariate. This allows us to view the values discussed here as proportions of overall word production. These analyses allow us to investigate in further detail, changes in the language structures.

ACKNOWLEDGEMENTS OF BEHAVIOR

Table 3 demonstrates a typical example of how the pairs acknowledge behaviors with and without a shared visual space. Consistent with *Hypothesis 3d*, the Workers took over the responsibility for assessing and communicating the state of the task when the Helpers did not have up-to-date visual information. When the pair had no shared visual space, the Worker had to indicate explicitly whether she understood an instruction and performed it correctly by reporting on the current task state (e.g., "OK, so it's like [on the] side of it and you see half of the red block"). The Helper then confirmed this understanding with the phrase, "Right of the red, yeah."

In contrast, when the shared space was available, the Helper could visually confirm that the Worker understood the instruction (e.g., with the statement, "Yeah. All right, that's good") without the Worker explaining.

Consistent with *Hypothesis 3d*, statistical analyses supported the shift in responsibilities. In the Immediate Shared Space condition,

Table 3
Shifts in Responsibility in Assessing and Communicating Correctness of Performance

Immediate Shared Visual Space	No Shared Visual Space
H: The right hand, the top right hand corner of the blue block touches the bottom left hand corner of the first orange block.	*H:* And that's gonna be on top of the red one but only the right side of the red is going to be showing.
W: [Positioned piece correctly]	*W:* [Positioned piece correctly]
W: Like that?	*H:* You know what I mean?
H: Yeah.	*W:* OK, so it's like . . .
H: All right, that's good.	*H:* Oh, like, put it on the left side of the red.
H: Right of the red, yeah.	*W:* . . . side of it and you see half of the red block.
	W: OK.

the Helper issued nearly as many behavioral acknowledgements as the Worker. That is, the Helper was as likely to tell the Worker that she had positioned a piece correctly as the reverse. However, when the shared visual space was limited, the Workers increased their production of acknowledgements (see Figure 4), interaction $F(2,105) = 33.56$, $p < .001$. Workers told Helpers about their success in following instructions. This Shared Visual Space × Speaker Role interaction is stronger when comparing the Immediate and No Shared Visual Space conditions, $t(105) = 8.10$, $p < .001$, than in comparing the Immediate and Delayed conditions, $t(105) = 2.49$, $p = .014$. *Hypothesis 3c* was not supported for acknowledgements of behavior. Although responsibilities for acknowledging correct behavior shifted across the shared visual space conditions, the total rate did not change.

ACKNOWLEDGEMENTS OF UNDERSTANDING

Another way in which the pairs use visual information is to support the grounding process. When the shared visual space is available, it is more efficient and easier for the pairs to follow a cycle of the Helper giving instruction and the Working performing actions. They can reserve speech for interrupting when things go wrong. There is little need for the Workers to state their understanding of instructions explicitly, becuase the Helpers can infer understanding by observing whether Workers performed correctly. However, when the fidelity of the space decreases, the Workers must be more explicit in communicating their understanding.

Consistent with *Hypothesis 3c*, the pairs were most explicit in stating their understanding when they had no shared visual space, $F(2, 105) = 12.43$, $p < .001$. They used acknowledgements of understanding more when they had no shared visual display than when it was avail-
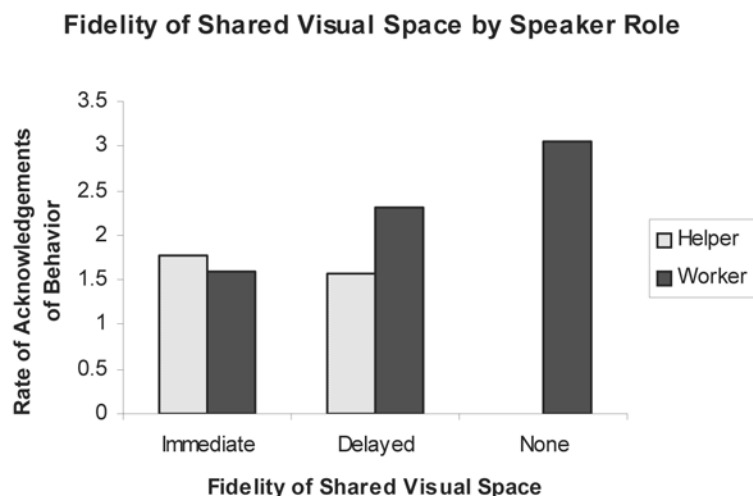
Figure 4. **Effect of Shared Visual Space and Speaker Role on the Production of Acknowledgements of Behavior**

able, t (105) = 4.59, p < .001, or when it was delayed, t (105) = 4.10, p < .001. However, there was little difference between having an immediate display and having a delayed one, t (105) = .57, p = .57, [LS Means (se): Immediate = 1.30 (.27); Delayed = 1.51 (.27); None = 3.11 (.29)].

The Shared Visual Space × Speaker Role interaction demonstrates further support for *Hypothesis 3d*. Workers were more explicit in stating their understanding when the shared visual space was of lower fidelity (see Figure 5) interaction F (2, 105) = 8.66, p < .001, whereas the Helpers behavior did not change much with variations in the shared visual space.

The Shared Visual Space × Color Drift interaction showed an additional increase in the use of acknowledgements of understanding when the colors were drifting than when they were stable, interaction F (2, 105) = 5.30, p = .006.

DEICTIC EXPRESSIONS

Deictic pronouns. Because the task in this study required the pairs to identify specific objects and then place them in a spatial arrangement, we expected that they would prefer to use shorthand references to objects as opposed to lengthy verbal descriptions when they could. Consistent with *Hypothesis 3a* pairs used differing rates of deictic pronouns, F (2, 105) = 5.47, p = .006. They used more in the Immediate condition than in the No Shared Visual Space Condition, t (105) = 3.31, p = .001. However, although the difference between the Immediate and
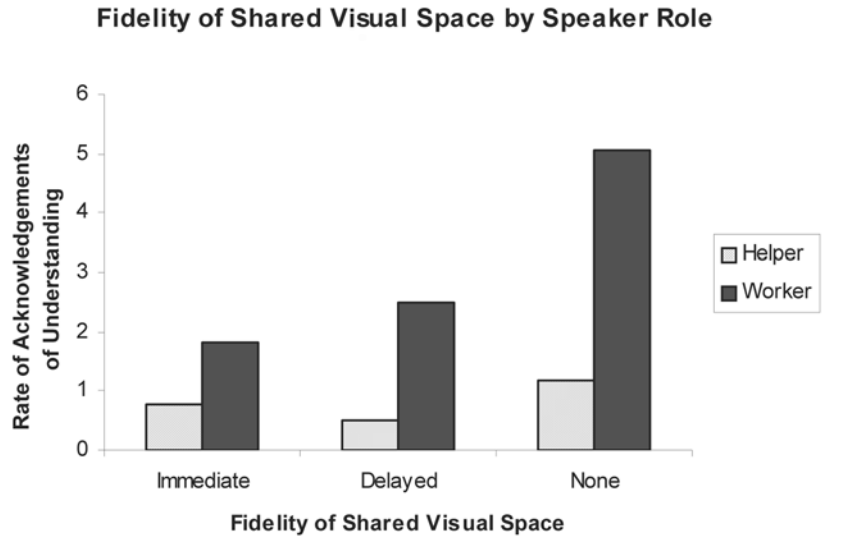
Fidelity of Shared Visual Space by Speaker Role

Figure 5. **Effect of Shared Visual Space and Speaker Role on the Production of Acknowledgements of Understanding**

Table 4
Use of Deictic Pronouns With and Without Shared Visual Space

Immediate Shared Visual Space	No Shared Visual Space
H: And that over . . . put that on top of the red one.	*H:* The bright blue's, the bright blue's, um, bottom left corner touches the bright red's upper right corner.

Delayed conditions was in the expected direction, it was not significant, t (105) = 1.71, p = .09 [LS Means (se): Immediate = 1.50 (.20); Delayed = 1.01 (.21); None = 0.512 (.22)].

Spatial deixis. Spatial deixis is the term we use for attempts to refer to an object by describing its position in relation to others, in phrases such as "next to," "below," or "in front of." Spatial descriptions are expensive. They are less efficient than a simple noun phrase (e.g., "the blue one") or a deictic pronoun (e.g., "that one"). If pairs are trying to minimize collaborative effort they should use spatial deixis less with a high fidelity shared visual space that is immediately available. Analyses showed a trend for the pairs to use differing proportions of spatial deixis depending on the fidelity of the shared visual space. Although the overall F-test did not reach statistical significance, F (2, 105) = 2.67, p = .074), pair-wise comparisons revealed that the Pairs tended to use spatial deixis more in the Delayed than in the Immediate Shared Vis-

ual Space condition, t (105) = 2.26, p = .02. However, the difference between the No Shared Visual Space and the Immediate Shared Visual Space did not reach significance, t (105) = 1.58, p = .11 [LS Means (se): Immediate = 2.82 (.29); Delayed = 3.64 (.30); None = 3.41 (.31)].

The shared visual space had less of an impact on spatial deixis when the colors were stable, interaction, F (2, 105) = 3.21, p = .04, and when the puzzle configurations were easy, interaction, F (2, 105) = 3.65, p = .03. Thus, if the task was linguistically or spatially difficult, the absence of a shared visual space caused subjects to resort to costly spatial description to solve it.

There was also a trend for the shared visual space to affect the Helpers' use of spatial deixis more than the Workers'. Although the overall F-test did not reach statistical significance, interaction F (2, 105) = 2.15, p = .12, pair-wise comparisons indicated that the Helpers used spatial deixis more when the fidelity of the display was decreased, whereas the Workers tended to produce a consistent number of spatial deixis per puzzle regardless of the view. This interaction was significant for the comparison between the Immediate and Delay conditions, t (105) = –2.01, p < .05, however it failed to reach significance for the comparison between the Immediate and No Shared Visual Space conditions, t (105) = –1.58, p = .12.

DISCUSSION

Communication media influence how well people collaborate. In this study, we found broad support for Clark's thesis that common ground is crucially important for conversation and specific support for Clark and Brennan's (1991) hypothesis that different communication features change the cost of achieving common ground. In particular, we examined the value of shared visual space as it pertains to conversational grounding and task awareness.

FACILITATING CONVERSATIONAL GROUNDING

The research shows that collaborative pairs can perform accurately and more quickly when they have a shared view of a common work area. The shared visual space improved task performance and conversational efficiency. Delay in updating the visual information diminished the benefits of having a shared visual space in most dimensions.

There are two major ways that the shared view of the work area improved performance by allowing Helpers to accurately ground their instructions. First, the shared work view allowed Helpers to use more efficient referring expressions to describe objects and positions in the work area. Seeing the Workers' behavior allowed Helpers to use deictic pronouns and other compact expressions instead of longer noun-

phrases to refer to elements in the puzzle. In addition, Helpers could see directly when their partners were ready for the next instruction, reducing the time between their instructions. Similarly, Workers, knowing that their partners could see their moves, could ask for confirmation with compact expressions such as "Like that?," rather than verbally describing the new state of the puzzle.

The second way that the shared visual space improved task performance was that it made conversational grounding more accurate and efficient. The shared visual space provided an important resource that allowed participants to comprehend the degree to which their partners understood an utterance. In particular, when the Helpers could see the Workers' behavior, they used this information to infer whether the Worker understood the current instruction. We observed that when Helpers saw that their partner made a correct move following an instruction, they cut short their descriptions and did not elaborate, but instead continued to the next instruction. In contrast, if they observed that their partner made an error, they would provide more detail, to describe a puzzle piece or its position.

This reasoning is consistent with the finding that Helpers used explicit descriptions of spatial positions (i.e., spatial deixis) less frequently in the Shared Visual Space than in the No Shared Visual Space condition. When the Helper could see the Workers' behavior, the Worker's placement of a piece in the correct place was immediate, costless evidence that they understood an instruction. Therefore, they could curtail their spatial description. However, without this evidence, the Helpers continued to elaborate the spatial description until they got explicit confirmation from the Workers about understanding.

The data presented here are broadly consistent with a cooperative model of communication. In particular, Workers adapted their communication and behavior to compensate for what the Helper could or could not see. It is important to note that in this experimental design the Worker's view of the workspace was always the same whether or not the Helper could see it. If Workers were using a purely egocentric approach to communication they would not change their communication behavior in response to variations in the shared visual space because their view of the space never changed. Instead, they changed their communicative behavior in response to what their partner could see. When the Helper could not see the work area, Workers used more words to complete the task, were more likely to describe the work area after they made moves, and were more likely to indicate explicitly whether they understood an instruction.

The results are consistent with Clark and Brennan's (1991) framework for analyzing the costs and benefits of different communication technologies. When media provide visual information about what the Worker is doing, the ability of Workers to ground their utterances via actions reduces their need to provide verbal indicators of comprehen-

sion. Instead, they let their actions speak for themselves and demonstrate their understanding of the Helpers' instructions. Elsewhere, we have used sequential analysis techniques to examine this issue in more detail (see Gergle, Kraut, & Fussell, in press). In particular, sequential analyses show that Helper's instructions were more likely to be followed by the Worker's movement of a puzzle piece in the Shared Visual Space than the No Shared Visual Space condition. In contrast, a Helper's instructions were more likely to be followed by a Worker's acknowledgement of understanding in the No Shared Visual Space than the Shared Visual Space condition.

These results, like others in this issue, show that people try to compensate for limitations in the communication technologies available to them. However, these compensations often fall short with regard to communication efficiency. For example, as previously discussed, when Workers believe that their partners cannot see their behavior, they are more explicit in indicating their level of comprehension. Yet, acknowledgements of understanding can be inaccurate. As any teacher knows, students can think they understand an instruction without really doing so. When Helpers could view the Workers' behavior, they got more accurate information about Workers' level of understanding, untainted by the Workers' self-assessments.

MAINTAINING TASK AWARENESS

We extended the work of Clark and Brennan (1991) by illustrating how features of the task interact with features of the communication setting to influence the grounding process. In this work, the value of a shared visual space depended on the task being performed. The shared visual space helped performance and conversational efficiency more when the tasks were dynamic (i.e., in the Color Drift condition). In other research we have shown that a shared visual space is more valuable when objects being discussed do not have common English names, and when verbal communication channels between partners is degraded (Gergle, Millen, Kraut & Fussell, 2004; Kraut, Gergle, & Fussell, 2002).

The interactions between the fidelity of shared visual space and the features of the task demonstrate the importance of understanding task characteristics when determining the value of a shared visual space. Our results suggest that the utility of a shared visual space depends in part on the visual complexity of the task. In dynamic settings or ones with many objects in a variety of spatial relationships to one another (e.g., for distributed medical teams, aircraft repair), visual space may be particularly important. For less complex visual tasks, especially those in which objects and spatial relationships are static and easily lexicalized, an audio-only connection may suffice. These findings help to rectify the disparity between early and more

recent research on the value of visual information in distributed communication.

In this study, task objects changed rapidly in the drift condition, and when they did, temporal delays in visual information rate had a significant negative impact on communication and performance. We would expect these results to generalize to other settings with rapidly changing events, such as an operating room. Temporal delays may be less problematic when task objects are relatively static, as they might be in an architectural design task.

Further work is necessary to understand the impact of other task attributes (e.g., size and number of task objects, types of task actions) on the use of shared visual space. Continuing an empirical investigation of shared visual space may provide us with a better understanding of the ways in which we can improve existing technologies and may also provide direction for the development of new technologies to improve distance collaboration.

LIMITATIONS OF THE STUDY

The stylized task used in this research is both a strength and a weakness of the study. It allowed us to examine basic principles required for successful collaborative interaction in a shared visual environment and provided a glimpse of the mechanisms and features through which a shared visual space improves performance. However, it does so at the cost of realism and generalizability.

Another potential limitation to this study is the discrete way we manipulated the fidelity of the shared visual space. We included three conditions: no shared visual space, a shared space with a 3-second delay, and an immediate visual space. The 3-second delay was unrealistically high for many users of today's technologies. Other research manipulating delay as a continuous variable is needed to gain more insight into the specific point at which a temporal breakdown occurs.

CONCLUSION

We have argued that shared visual space is essential for complex collaborative visual problem solving because it facilitates the ability of the pairs to maintain awareness of the task state, helps them to reduce errors and ambiguities when the environment is visually complex, and facilitates grounding and communication by allowing the use of efficient language and a method for monitoring comprehension. The effects of new communication technology are not superficial, and their developers should not be guided by surface characteristics. By considering the ways that technologies, and the task we attempt with their aid, interact with, modify, and rely on language, greater strides can be

made in understanding and design. Moreover, these developments illuminate basic principles of conversation and social psychology in profound ways, bringing into focus not only technological but traditional communication processes.

NOTE

1. Because of missing data, the independent variables were not completely orthogonal. Therefore, we used Least Squared Means (LS Means) to compare experimental conditions. When calculating the means for an experimental condition, LS Means control for the value of the other independent variables.

REFERENCES

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Barnard, P., May, J., & Salber, D. (1996). Deixis and points of view in media spaces: An empirical gesture. *Behaviour and Information Technology, 15*, 37-50.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language & Speech*, *37*, 1-20.

Brennan, S. E. (in press). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *World situated language use: Psycholinguistic, linguistic and computational perspectives on bridging the product and action traditions*. Cambridge, MA: MIT Press.

Chapanis, A., Ochsman, R. B., Parrish, R. N., & Weeks, G. D. (1972). Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors, 14*, 487-509.

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., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory & Language*, *50*, 62-81.

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

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

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors Special Issue: Situation Awareness*, *37*, 32-64.

Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of the Conference on Computer-Supported Work (CSCW)*, *2*(3), 21-30.

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 Conference on Computer-Human Interaction (CHI)*, *5*(1), 513-520.

Fussell, S. R., Setlock, L. D., & Parker, E. M. (2003). Where do helpers look? Gaze targets during collaborative physical tasks. *Proceedings of the Conference on Human Factors in Computing Systems (CHI) (Extended Abstracts)*, *5*(1), 768-769.

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E. M., & Kramer, A. (in press). Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*.

Gergle, D., Kraut, R. E., & Fussell, S. R. (in press). Communicating with action. *Proceedings of the Conference on Computer-Supported Work (CSCW)*.

Gergle, D., Millen, D., Kraut, R., & Fussell, S. (2004). Persistence matters: Making the most of chat in tightly-coupled work. *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, *6*(1), 431-438.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*, 26-37.

Krauss, R. M., & Bricker, P. D. (1967). Effects of transmission delay on the efficiency of verbal communication. *Journal of Acoustical Society of America*, *41*, 286-292.

Kraut, R. E., Fussell, S. R., Brennan, S. E., & Siegel, J. (2002). Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In P. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 137-162). Cambridge, MA: MIT Press.

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 Conference on Computer-Supported Work (CSCW)*, *4*(3), 31-40.

Nardi, B. A., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S., & Sclabassi, R. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, & T. White (Eds.), *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (pp. 327-334). New York: ACM Press.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1-24.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211-232.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.

Darren Gergle (M.S., University of Michigan) is a doctoral student in the Human Computer Interaction Institute at Carnegie Mellon University. His research interests include small group communication in face-to-face and mediated environments, and the design and study of tools to support distributed collaboration.

Robert Kraut (Ph.D., Yale) is the Herbert A. Simon Professor of Human Computer Interaction and Social Psychology at Carnegie Mellon University. He conducts research on the design and impact of computer-mediated communication systems.

Susan R. Fussell (Ph.D., Columbia University) is a research scientist in the Human Computer Interaction Institute at Carnegie Mellon University. Her research interests include interpersonal communication in face-to-face and computer-mediated contexts, online communities, and the dynamics of collaboration in work teams and organizations.