The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence

Robert E. Kraut, Darren Gergle, Susan R. Fussell
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
+1 412 268-7694
robert.kraut@cmu.edu; dgergle+@cs.cmu.edu; susan.fussell@cmu.edu

ABSTRACT
A shared visual workspace is one where multiple people can see the same objects at roughly the same time. We present findings from an experiment investigating the effects of shared visual space on a collaborative puzzle task. We show that having the shared visual space helps collaborators understand the current state of their task and enables them to communicate and ground their conversations efficiently. These processes are associated with faster and better task performance. Delaying the visual update in the space reduces benefits and degrades performance. The shared visual space is more useful when tasks are visually complex or when actors have no simple vocabulary for describing their world. We find evidence for the ways in which participants adapt their discourse processes to their level of shared visual information.

Keywords
Shared visual space, computer-supported collaborative work, conversational analysis, empirical studies, language and communication

INTRODUCTION
Distance collaboration is an increasingly common occurrence, driven by the global nature of business and enabled by improvements in computing and communication technologies. Technologies that provide visual information to people collaborating at a distance have been available for decades. Early research on media effects of video suggested that adding an audio channel to any other medium substantially improved communication, but that adding video to the audio provided little benefit (see Williams [18] for a review). Recent research is starting to identify the conditions under which visual information is valuable. This work shows that in some cases, having a shared visual environment improves communication [16,14,17,11,2,5,9]. Much of this work suggests that the benefit of visual information comes from allowing collaborators to share the work area rather than from seeing one another.

Despite these breakthroughs demonstrating the benefits of a shared visual workspace, we still do not have a complete understanding of the mechanisms and features through which a shared visual workspace improves performance. In addition, we do not have a good grasp on the types of tasks for which a shared visual workspace is most useful.

Answering these questions will help to provide a scientific foundation to design research in CSCW. Imagine the challenge of designing systems to allow an anesthesiologist to monitor surgery at a distance (e.g., see Nardi et al. [14]), to allow a master mechanic at a factory to help a less experienced mechanic repair an aircraft engine at a remote airport (e.g., see Fussell, Kraut & Siegel [9]), or to permit an instructor to provide guidance to a distributed classroom working on a physics problem. A shared visual workspace is likely to be useful in each of these situations. However, there is currently no principled way to determine the requirements of such systems – whether it is high-resolution displays, wide-angle fields of view, synchronization between the audio and video, or a host of other design features.

Previous Work
Much research on the utility of shared visual information compared pairs of subjects performing a referential communication task using only an audio channel with pairs working face-to-face or using an audio/video connection. Pairs could see each other, but not the objects they worked on. This research tradition is derived from work conducted by the Communications Study Group at British Telecom [15] and in Alphonse Chapanis' lab in the United States [4].

More recent research shifts the focus from a view of the participants' faces to the work area. Studies in this new wave differ primarily on how realistic or stylized the task is. For example, Clark [5] used a stylized task, in which a

CSCW'02, November 16—20, 2002, New Orleans, Louisiana, USA.

Director instructed 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 this 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; 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 create artificial proximity?

Several studies have tried to disentangle which features of shared visual space help influence its value by comparing different video conferencing configurations to a side-by-side condition. Each video conferencing configuration makes available a subset of the visual cues present when people are collocated [6,13], and thus allows us to test the value of these specific visual cues for communication and performance. For example, Fussell, Kraut and Siegel [9] had dyads repair a bicycle while conversing side-by-side, using a head-mounted camera, or via audio only. Pairs were substantially faster when they worked side-by-side. Some of their speed-up occurred because they could more efficiently refer to parts of the bike when they could see the work area, while pairs in the audio-only condition spent more time coordinating their messages and acknowledging their partners' messages. Contrary to expectations, performance in the video-mediated condition was no better than in the audio-only condition.

While these tasks have the benefit of being relatively realistic, this realism often leads to a lack of control over experimental conditions. In the Fussell et al. study [9], for example, technical problems with the camera, slippage of the camera on the worker's head, and other technical problems in the video condition make it difficult to identify the reasons why performance was worse than in the side-by-side condition. Thus, there is a need for more tightly controlled laboratory studies of shared visual space to complement these studies.

Research by Brennan and Lockbridge [3] experimentally manipulated features of the shared visual space under more rigid laboratory control. In their experiment, Directors instructed Matchers to place a series of picture cards in a specific order. The Director could see the staging area (where the Matcher had the inventory of cards), the work area (where the Matcher placed the cards), or the Matcher's face. They also varied the nature of the task, by making some trials involve difficult to describe baskets or easy to describe animals. Their research showed that seeing the staging or work area improved performance by enabling more rapid entrainment on pair-specific referring expressions, but that seeing the partner's face had little benefit.

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 guided a Worker in completing a simple online jigsaw puzzle. They shared a visual space consisting of a view of the work area rendered on each of their computer screens. The benefit of such a setup is that the view the Helper sees can be any computationally derived transformation of the workspace the Worker sees.

We applied this technique to examine the impact of temporal delay in the shared visual space on task performance and communication. Krauss and Bricker [12] showed that relatively small auditory delays impede effective communication. Do delays in updating a shared visual space, of the sort that network congestion and video compression might bring about, cause similar problems? We also examined how two task attributes—visual complexity and temporal dynamics—interact with shared visual space. For example, shared visual space may be more important for tasks involving many small pieces or pieces arranged in difficult-to-describe configurations. Shared visual space may likewise be more important for tasks in which the environment is rapidly changing, such as a hospital operating room.

Identifying the Critical Elements of Shared Visual Space

In order 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 [8] observes that collaborative work occurs at multiple levels simultaneously, although the distinction between levels is not crisp. At one level, people collaborate to perform a task; in this paper, they are jointly constructing a puzzle. At a lower level, they use language (and other communicative behaviors) to coordinate actions in order to perform the task. For example, they reach agreements about the names to apply to objects they can jointly see. Visual evidence can be helpful at each of these levels. Visual information can give collaborators an up-to-date view of the state of the task. Additionally, it provides evidence about a partner's level of understanding of the language that is being used for coordination.

Maintaining Awareness of Task State

Shared visual information is important for maintaining an *awareness* of the current state of the collaborative task in relation to an end goal. This awareness helps a pair plan how to proceed towards the goal, what instructions need be given, and how to repair incorrect actions. Shared visual information provides the ability to monitor specific actions.

Imagine a pair performing a typical referential communication task [10] in which a Helper is instructing a

Worker on the order in which to place a set of cards. If the Worker places a card to the left when it should have been to the right, the Helper can intervene with new instructions if they 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 the language is less adequate to describe the task state. For example, in the jigsaw 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). In the simple two-dimensional case, the instruction "Put the red piece on top of the blue one" is unambiguous, while in the three dimensional case, the red piece can either overlap or be north of the blue piece. If the Helper can see the work area, he can intervene to rectify any misinterpretation.

Facilitating Conversation and Grounding

A shared visual space may also facilitate the *communication* that surrounds a joint activity. Communication rests on a foundation of information of which participants are mutually aware, termed mutual knowledge or common ground [7,8]. Generally, a speaker would not speak in Yiddish unless he thought a partner understood it, would not use a technical term like multiplexor 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 subsequent utterances. Participants are obligated both to assess and give off cues that indicate their understanding.

Grounding refers to the interactive process by which people exchange evidence about things they understand over the course of a dialog. Clark and Brennan [6] 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 give off back channel communications—the "uh-huhs", "I sees", head nods, and smiles—that signal to the speaker the degree to which they understand the current utterance.

In this paper we are interested in how a shared visual space facilitates grounding. Clark and Brennan [6] and Kraut, Fussell, Brennan and Siegel [13] suggest ways that a shared visual space can be helpful for establishing common ground.

1) Creating efficient messages

People do not like to work more than they have to and a shared visual environment may reduce the amount of effort a pair jointly needs to expend to describe some state. This is the basis for the principle of least collaborative effort [8]. If the goal of a pair of speakers is to reduce their overall effort in grounding, we should expect to see changes in the way the pairs communicate based on the fidelity of the shared visual space.

For example, if the Helper cannot see the Worker's area, it will be up to the Worker to determine that the current task is complete. This requires the Worker to ask several questions before concluding that the task is done. However, if the Helper can see the work, then it should be more efficient for the Helper to simply say, "You got it", and proceed with the next directive. 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.

2) Monitoring comprehension

Seeing the performance of a partner will provide some knowledge of the partner's level of comprehension. In a shared visual space, one can more easily recognize when an individual is performing an incorrect action or when they are confused or do not understand a task.

For example, in the present experiment, shared visual information might provide a basis for the pair to better coordinate their language. If a Helper notices that when they say, "put the piece kitty-corner" the Worker simply tilts the piece, it can be assumed that that "kitty-corner" is not part of their shared language. Thus, when there is a shared visual space, the Helper can easily remedy this mistake by providing a more meaningful directive.

By seeing the partner perform some task, the Helper gets immediate feedback about whether the partner understood a directive. If the visual feedback were delayed (e.g., as caused by video compression or network lags), the value of the visual information may be diminished. Delay in updating the display may diminish the value of the visual information both for the ability to compare the current work state to the goal state (as described in the earlier scenario) and for the team's ability to coordinate language (as discussed here).

Visual feedback, however, may be less necessary if the task is simple enough (e.g., a game of tic-tac-toe) or if the pair has an efficient, well-practiced vocabulary to describe events (e.g., routine communication between pilot and air traffic controllers). In these cases, a shared visual display provides little new information.

Hypotheses

The ability of a shared visual space to influence both the maintenance and creation of awareness along with facilitating conversation and grounding, as described above,

leads to the formation of several hypotheses regarding the performance of the pairs in the referential task explored in this study.

Since the shared visual space provides additional awareness of the task state we would expect to see increased performance when it is available:

Hypothesis 1: A collaborative pair will perform their task more quickly when they have a shared view of the work area.

Increased visual complexity and a changing task environment increase the potential for errors and make language less adequate for describing state. In addition, a shared visual space may support grounding and aid in monitoring and comprehension. For these reasons, the following benefits are expected:

Hypothesis 2: A shared view of a work area will have more benefit when the task is visually complex.

Hypothesis 3: A shared view of the work area will have more benefit when the task environment is rapidly changing.

While we may expect to see such advantages of the shared view space, any disruption to the immediacy of the visual information may reduce the value of a shared visual space.

Hypothesis 4: Delay in transmission will diminish the value of a shared view of the work area.

In addition, we may expect to see changes in the ways the pairs use language to perform the task. In following the principle of least collaborative effort, we would expect pairs to change the structure of their discourse in ways that lead to the least amount of effort for the pair as a whole.

Hypothesis 5: To minimize collaborative effort, pairs will change the structure of their communication to be more efficient.

METHOD

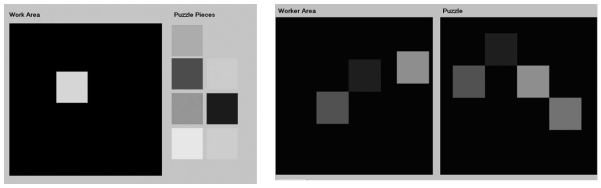
Participant pairs played the role of Helper and Worker in a referential communication task. The Helper instructed the Worker in completing a simple jigsaw puzzle. The goal was for the Worker to arrange their pieces so that they matched the target that the Helper was viewing.

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 difficulty of the task itself (*Puzzle Difficulty*).

Fidelity of the Visual Space. We varied whether the Helper could see the state of the Worker's puzzle. In any trial, the Helper could either see 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. These are, respectively, the *Immediate, Delayed,* and *None* visual space conditions.

Color Drift. We varied the lexicality of the puzzle pieces by changing whether the colors of the blocks were static (e.g., red) or constantly cycling (e.g., red to orange to yellow to…). In the *Static* condition, pieces were chosen randomly for each experimental condition from a palette of easily distinguishable colors (see Figure 1). In the *Drift* condition, each piece slowly and continuously changed its color, cycling through the colors in the color palette.

Puzzle Difficulty. We varied the difficulty of the puzzles by having configurations where the pieces simply abutted edges (*Easy*) or overlapped one another (*Difficult*).



(a) Worker's Display **(b) Helper's Display**

Figure 1 Worker's display (a) and Helper's display (b).

Apparatus

The Helper and Worker were each seated in front of separate desktop computers with 21-inch monitors. They communicated over a high-quality, full-duplex audio link with no delay. 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 puzzle target on the right, holding the goal state. In the left, it showed one of the three views of the Worker's work area. This view was either an exact copy of the Worker's work area (Immediate), showed the work area with a three-second delay (Delayed) or remained black (None).

Participants and Procedure

Participants consisted of 12 pairs of Carnegie Mellon University undergraduate students. Participants received $10.00. The participants were randomly assigned to play the role of Helper or Worker. Color Drift was manipulated across pairs of participants, while both Visual Space and Puzzle Difficulty were manipulated within each pair. Six pairs participated in the Static condition and six in the Drift condition. Each pair participated in six experimental conditions, once in each Visual Space by Puzzle Difficulty combination, counter-balanced. Pairs solved four puzzles within each experimental condition.

Measures

Two types of dependent measures were analyzed: task performance and conversational coding.

Task Performance Measure

The pairs were instructed to complete the task as quickly as possible, so task performance was the time it took to properly complete the puzzle. Overall, the vast majority of the puzzles were solved correctly so differences in error rates would be a less useful indicator of task performance.

Conversational Coding

To investigate the relationship between the shared visual space and dialogue we developed a coding scheme to capture the primary purpose of each utterance. A subset of the categories used for analyses in this paper is presented in Figure 2.

Utterance Types

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

Figure 2. Types of utterances coded.

Another area of interest was the use of deictic pronouns and spatial deictic expressions. Both 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. Whereas in the phrase, "It's the one on top of the red block", "on top of" uses the relative position of objects to refer to them. Figure 3. presents the types of deixis coded for in this analysis.

Deictic Expressions

Deictic Pronouns	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".

Figure 3. Types of deictic expressions coded.

Two independent coders classified a sample of utterances until they reached 90% agreement. 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

Our analysis of performance uses time to complete a puzzle as the dependent variable. The analysis is a repeated measures analysis of variance in which Block (1-6), 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.

When we conducted analysis of conversational efficiency, we included time to complete the task as a covariate. When conversational content is the dependent variable, we included both time and number of words as covariates.

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

RESULTS

Task Performance Analysis

Manipulations checks. The performance measure is seconds (in the log scale) to complete a puzzle. The manipulation of Puzzle Difficulty had a significant impact on the speed with which the pairs solved the puzzles. The pairs were faster in the trials in which the puzzles were simple and pieces abutted than when they were difficult and pieces overlapped (LS Means[1]: 61.9 sec vs. 73.8 sec, $p = 0.002$).

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 than when they were drifting (LS Means: 54.3 sec vs. 84.1 sec, $p < 0.001$).

Shared visual space. This experiment was designed to examine the impact of the fidelity of shared visual spaces on performance for different types of tasks. 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 ($p < .0001$) or the No Shared Visual Space condition ($p < .0001$), (LS Means Immediate = 57.9 sec; Delayed = 79.04 sec; None = 81.4). Consistent with hypothesis one, the results show that a shared view of the work area benefited performance, but only when the view was kept up-to-date. Even a three-second delay eliminated its benefit.

[1] Because the independent variables were not completely orthogonal, 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.

The interaction between the Visual Space and Color Drift manipulations demonstrates that having a shared view of the work area had greatest benefit when the objects being discussed were lexically complex and difficult to describe (see Figure 4; for the interaction F (2,256) = 7.13; p < .001).

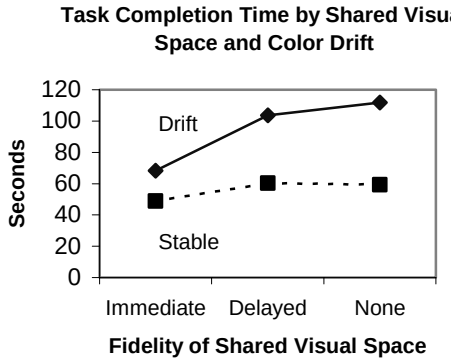


Figure 4. Effect of shared visual space and color drift on task completion times.

The Immediate Shared Visual Space condition was substantially faster than the No Shared Visual Space condition when colors were drifting than when they were stable (for the interaction, t = -3.74, p = .0002). 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 (for the interaction, t = -1.96, p = .05). 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 [5], temporal synchrony matters a great deal (see Figure 4.).

It is instructive that the Visual Space by Puzzle Difficulty interaction, while in the hypothesized direction, was not statistically significant (F (2,256) = 1.06, p > .35). Visual complexity itself did not raise the value of a shared view of the work area. It was primarily when the task was dynamic and the environment was changing that having the display was most beneficial.

These data are consistent with hypotheses one and three. A shared visual space was important for this collaborative communicative task, and there is greater benefit when the task environment is rapidly changing. In addition, the findings are consistent with hypothesis four. The value of the shared visual space is decreased when there is a delay in the visual information. However, we did not find support for hypothesis two.

The next stage of analysis explores the ways in which the communication between the Helper and the Worker varies when the shared visual space is perturbed.

Conversational Coding Analysis

The performance data demonstrated that the manipulations of the fidelity of the shared visual space, object stability and task difficulty all play a part in the ability of the pairs to quickly and efficiently solve the puzzles. However, this tells us little about the *process* by which the teams operated when faced with different configurations of shared visual space.

Rate of Word Production

We explored the rate at which the pairs produced words (in the log scale) in order to examine the efficiency with which they communicated. Pairs should be able to describe the puzzles with less effort (i.e., fewer words per unit time) when there is a shared visual space available. We examined word rate (the number of words, controlling for time) to test this prediction. The model used for the word rate analyses was similar to the analysis of variance model 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 by Visual Space by Speaker Role, they were removed from the model.

In support of hypothesis five, the pairs produced more efficient speech when they had higher fidelity shared visual space. That is, pairs used fewer words, controlling for time (LS Means Immediate = 19.4 words per puzzle; Delayed = 30.1; None = 45.0). The Immediate Shared Visual Space condition was more communicatively efficient than both the Delay condition (t = -2.55, p = 0.01) and the No Shared Space condition (t = -4.84, p < .0001). In turn, the Delay condition was more efficient than the No Shared Space condition (t = 5.775, p < .017).

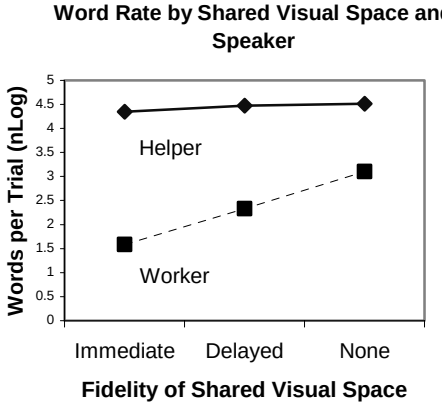


Figure 5. Effect of shared visual space and speaker role on word rate.

An examination of the interaction between speaker role and shared visual space reveals that the fidelity of the shared visual space influenced the Worker's efficiency rather than the Helper's (see Figure 5; for the interaction, F (2,110) =

10.80, $p < .0001$). Because the Worker could always see the work area, these changes in Worker behavior reflect their accommodation to differences in the Helper's view of the workspace. This provides further support for hypothesis five.

Content Analysis and Qualitative Descriptions

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 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 analysis of variance model for examining word rate with one exception. It 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 further investigate the changes in the dialog structure as suggested by hypothesis five.

Acknowledgements of Behavior

An initial look at the overall number of acknowledgements in response to behaviors revealed no overall differences across the different shared visual space conditions. However, the partner producing the acknowledgement changes depending on the fidelity. Figure 6 demonstrates a typical example of how the pairs acknowledge behaviors with and without a shared visual space. Workers took over the responsibility of assessing and communicating the state of the task when the Helpers did not have up-to-date visual information.

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

Figure 6. Shifts in responsibility in assessing and communicating correctness of performance.

In the Immediate Shared Visual Space, the Helper issued nearly as many behavioral acknowledgements as the Worker. However, when the shared visual space was limited, the Workers made up for the difference by increasing their production of acknowledgements (see Figure 7; for the interaction, $F(2,106) = 32.42$, $p < .0001$). This interaction between shared visual space and speaker is less severe between the Immediate and Delayed conditions ($t = 1.75$, $p = .084$) than it is between the Immediate and No Shared Visual Space ($t = 7.59$, $p < .0001$). In other words, as the fidelity of the shared visual space decreases, the Workers must take a much more active role in producing acknowledgements of behavior.

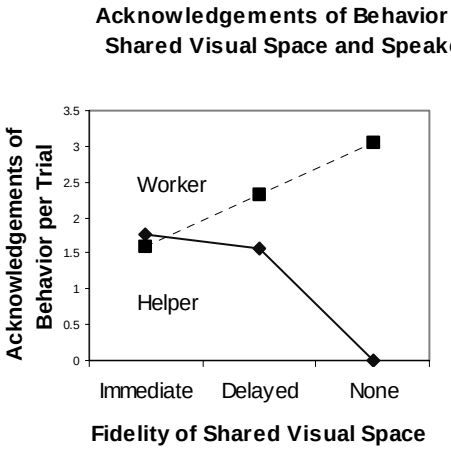


Figure 7. Effect of shared visual space and speaker role on proportional production of acknowledgements of behavior.

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 explicitly state their understanding of instructions, since the Helpers can infer this by observing whether they performed correctly. However, when the fidelity of the space decreases, the Workers must be more explicit in communicating their understanding.

The data are consistent with this reasoning. Overall, the pairs used acknowledgements of understanding less when they had an immediate shared visual display than when it was not available ($p < .0001$). However, there was little difference between having an immediate display and having a delayed one ($p = .59$) (LS Means Immediate = 1.30 (.27); Delayed = 1.51 (.28); None = 3.11 (.30)). Thus, the pairs were most explicit in stating their understanding when they had no shared visual space.

The interaction between the shared visual space and the speaker demonstrates that it was more important for the Workers to explicitly state their understanding when the shared visual space was of lower fidelity (see Figure 8; for the interaction $F(2,107) = 8.752$, $p = .0003$). Closer examination reveals that this is even more evident in the No

Shared Visual Space pairs than it is for the Immediate ($t = 4.05$, $p < .0001$), while there appeared to be less difference between the Immediate and Delayed Shared Visual Space ($t = 1.36$, $p = .1777$). The Helpers typically produce about the same portion of acknowledgements regardless of the degree of the shared visual space. However, the Workers significantly increase their rate of production.

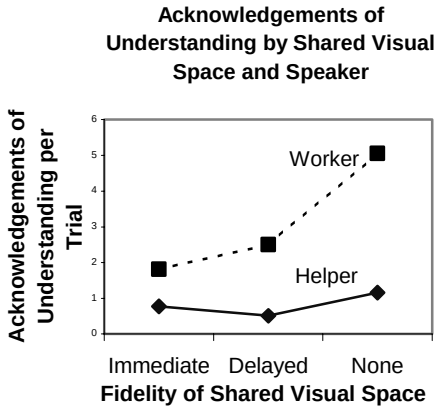


Figure 8. Effect of shared visual space and speaker role on proportional production of acknowledgements of understanding.

The pairs also increased their production of acknowledgements when the colors of the pieces were changing. This led to more acknowledgements of understanding in the No Shared Visual Space condition than in the Immediate Shared Visual Space condition. Workers increased their acknowledgements of understanding more than the Helpers when the colors were drifting (for the interaction, $F(2,107) = 5.129$, $p = .007$).

Deictic Expressions

Since 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. Deictic pronouns are words that are used to "point" objects out and typically increase the efficiency of speech. If the pairs were trying to maximize their communication efficiency, we should expect to see greater use of deictic terms in the trials where the visual space was of higher fidelity. Figure 9 shows excerpts typical of the use of deictic pronouns when the pairs have a shared view and when they do not.

Immediate Shared Space	***No Shared 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.

Figure 9. Use of deictic pronouns with and without shared visual space.

Overall we found higher use of deictic pronouns in the Immediate condition than in either the Delayed condition ($p = .08$) or the No Shared Visual Space Condition ($p = .001$) (LS Means Immediate = 1.30 (.27); Delayed = 1.51 (.28); None = 3.11 (.30)).

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". The use of spatial descriptions is expensive. They are less efficient than a simple noun phrase (e.g., the blue one) or a deictic pronoun (e.g., that one). We expected that due to the relative inefficiency of such referring expressions, they would be used less when the shared visual space was immediately available. However, when the shared visual space was delayed, it was not clear whether subjects would use spatial descriptions or incur a three-second delay to visually verify whether or not a piece was in the right position. We expected the use of spatial deixis to increase substantially in the absence of a shared visual space, since this is one of the primary ways in which the pairs could describe the layout.

We found a trend for the pairs to use a higher proportion of spatial deixis in the No Shared Visual Space trials than in the Immediate Shared Visual Space ($p = .11$). A similar result was found when comparing the Delayed and Immediate Shared Visual Space trials ($p = .02$) (LS Means Immediate = 2.82 (.29); Delayed = 3.64 (.30); None = 3.41 (.31)).

There was also a trend for the shared visual space to affect the Helpers' use of spatial deixis more than the Workers' (see Figure 10; for the interaction $F(2,107) = 2.187$, $p = .117$)). Further examination revealed that this interaction was greater for the comparison between the Immediate and Delay conditions ($t = -2.01$, $p < .05$) than it was for the Immediate vs. No Shared Visual Space ($t = -1.58$, $p < .12$). The general trend here is that the Helpers increase their production when the fidelity of the display is decreased, while the Workers tend to produce a consistent number of spatial deixis per puzzle regardless of the view.

In addition, we found that the lack of a shared visual space increased use of spatial deixis more when the colors were drifting (for the interaction, $t = 2.18$, $p = .043$). The lack of a shared visual space also increased use of spatial deixis more when the puzzles were more difficult (for the interaction, $t = 3.70$, $p = .03$). 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.

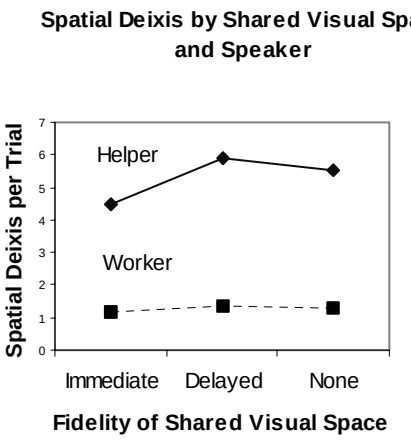


Figure 10. Effect of shared visual space and speaker role on proportional production of spatial deixis.

DISCUSSION

Communication media influence how well people collaborate. This research shows that collaborative pairs can perform more quickly and accurately when they have a shared view of a common work area. Overall, the shared visual space improved performance and conversational efficiency, and adding delay degraded the benefits of having a shared visual space. The value of a shared visual space depended to a degree on the task being performed. The shared visual space helped performance and conversational efficiency most when the tasks were complex. When the tasks required temporal accuracy or were more visually complex, having a shared visual space produced greater benefit.

The shared visual space also changed the cost of understanding the task state and establishing common ground differentially for the Helper and Worker. For example, when there was no shared visual space, the Helper no longer had an up-to-date view of the task state. In order to get this information, the Helper had to query the Worker to give an explicit description, and the Worker needed to respond with lengthier descriptions of the task state.

Implications for Theory

Overall, we found broad support for Clark's thesis that common ground is crucially important for conversation. We also found specific support for Clark & Brennan's [6] hypothesis that different communication features change the cost of achieving common ground. We extended this work by illustrating how features of the task interact with features of the communication setting to influence the grounding process.

The data demonstrate general support for the cooperative model of communication. This support was demonstrated by the way in which the Workers adapted their communication and behavior to compensate for what the Helper could or could not see. They elaborated their descriptions of objects much more when they knew their partners could not see them.

The shared visual space provided an important resource that allowed participants in this experiment to comprehend the degree to which their partners understood an utterance. That is, the shared visual space provided a resource for grounding. Consider, for example, the finding that the Helpers were more likely to use elaborated spatial descriptions when they had no view of the workspace or when the view was delayed. One might think that this spatial elaboration by the Helper was unnecessary because the Workers could always see the work area. The Helpers were using a costly method to help ensure correspondence between their spatial descriptions and the Worker's understanding of them. When the Helpers could see the work area with no delay, they knew if the Worker had understood the instruction by watching the Worker's behavioral response. The Worker's placement of a piece in the correct place was immediate, costless evidence that they understood. However, without this evidence, the Helper continued to elaborate the spatial description until they got explicit confirmation from the Worker about understanding.

Manipulating delay in the shared visual space had a strong impact on its value. We found that a three second delay made a large difference in the pair's performance and in many cases rendered the shared visual space useless. This was especially the case in a dynamic work environment, where objects were changing.

Implications for Practice

The interactions between the fidelity of shared visual space and the task manipulations demonstrate the importance of understanding the task 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 settings with many objects in a variety of spatial relationships to one another (e.g., medical setting, 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.

In this study, task objects changed rapidly in the drift condition; hence, temporal delays 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 enable distance collaboration.

Limitations of the Study

The stylized task used in this paper 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 three-second delay, and an immediate visual space. The three-second delay was unrealistically high for today's technologies. It might have been worthwhile to manipulate delay as a continuous variable, in order 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 messages and a method for monitoring comprehension. We have demonstrated a technique for experimentally manipulating features of a shared visual space and have observed their effects on performance and communication. The work we have presented here is a first step in understanding which features of a shared visual space are most important. By using these techniques and combining the results with findings from more realistic, but less precise studies of real world use, we hope to further our understanding of shared visual space.

ACKNOWLEDGEMENTS

This work was made possible through the support of National Science Foundation Grant No. IIS-9980013. In addition, the authors would like to thank Susan Brennan for her comments on early drafts of this paper. We would also like to thank James Hanson, who implemented the experimental apparatus, and John Lee and Darrin Filer, who refined the implementation.

REFERENCES

1. Brennan, S. E. (1990). Seeking and providing evidence for mutual understanding. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
2. 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.
3. Brennan, S. E., & Lockridge, C. B. (in preparation). How visual co-presence and joint attention shape speech planning.
4. 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(6), 487-509.
5. Clark, H. H. (Personal Communication).
6. 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*, 127-149. Washington, DC: APA.
7. Clark, H.H., & Marshall, C.R. (1978). Reference diaries. In D.L. Waltz (Eds.) *Theoretical Issues in Natural Language Processing – 2*, 57-63. NY: ACM Press.
8. Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1-39.
9. Fussell, S.R., Kraut, R.E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of CSCW 2000*, 21-30. NY: ACM Press.
10. Isaacs, E., & Clark, H.H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1), 26-37.
11. Karsenty, L. (1999). Cooperative work and shared visual context: An empirical study of comprehension problems and in side-by-side and remote help dialogues. *Human-Computer Interaction*, 14(3), 283-315.
12. Krauss, R.M., & Bricker, P.D. (1967). Effects of transmission delay on the efficiency of verbal communication. *Journal of Acoustical Society of America*, 41(2), 286-292.
13. Kraut, R. E., Fussell, S. R., Brennan, S., & Siegel, J. (2002). Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. P. Hinds & S. Kiesler (Eds.), *Technology and Distributed Work*, 137-162. Cambridge, MA: MIT Press.
14. Nardi, B., Schwartz, H., Kuchinsky, A., Leichner, R., Whittaker, S., & Sclabassi, R. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. *Proceedings of INTERCHI '93*, 327-334. NY: ACM Press.
15. Short, J., Williams, E., and Christie, B. (1976). *The Social Psychology of Telecommunications*. London, U.K.: Wiley.
16. Veinott, E., Olson, J., Olson, G., & Fu, X. (1999). Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. *Proceedings of CHI'99*, 302-309. NY: ACM Press.
17. Whittaker, S., & O'Conaill, B. (1997). The role of vision in face-to-face and mediated communication. In K. Finn, A. Sellen & S. Wilbur (Eds.) *Video-mediated communication*, 23-49. Mahwah, NJ: Erlbaum.
18. Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, 84(5), 963-976.