The Impact of Delayed Visual Feedback
on Collaborative Performance

Darren Gergle, Robert E. Kraut, Susan R. Fussell

Human Computer Interaction Institute

Carnegie Mellon University

Pittsburgh, PA 15232

dgergle+@cs.cmu.edu, robert.kraut@cmu.edu, susan.fussell@cmu.edu

ABSTRACT

When pairs work together on a physical task, seeing a common workspace benefits their performance and transforms their use of language. Previous results have demonstrated that visual information helps collaborative pairs to understand the current state of their task, ground their conversations, and communicate efficiently. However, collaborative technologies often impinge on the visual information needed to support successful collaboration. One example of this is the introduction of delayed visual feedback in a collaborative environment. We present results from two studies that detail the form of the function that describes the relationship between visual delay and collaborative task performance. The first study precisely demonstrates how a range of visual delays differentially impact performance and illustrates the collaborative strategies employed. The second study describes how parameters of the task, such as the dynamics of the visual environment, reduce the amount of delay that can be tolerated.

Author Keywords

Shared visual space, visual delay, empirical studies, language and communication.

ACM Classification Keywords

H.5.3. Information interfaces and presentation: Group and Organization Interfaces – collaborative computing, computer-supported cooperative work.

INTRODUCTION

Recently there has been a growing interest in the design of technologies to support a host of remote collaboration activities such as architectural planning, telesurgery, and remote repair tasks. These activities, when performed in a collocated environment, rely on a number of intricate dependencies between verbal communication and physical actions. When designing tools and technologies to support such tasks remotely, we need to understand how the introduction of technological mediation impacts the coordination mechanisms typically relied upon in collocated physical environments.

A number of studies have demonstrated that visual information is a key element to successful collaboration. Speakers and addressees take into account what one another can see [23]. They notice where one another's attention is focused [1, 2, 13], and they use cues available in the local visual context to help them coordinate both the language they use as well as the actions they engage in. When mediating an environment, we need to understand how technologies impinge upon these processes in ways that can disrupt the mechanisms that support successful collaborative activity. For example, how does the often unavoidable delay found in distributed applications, such as the type involved in network congestion, impact an individual's ability to maintain an awareness of their partner's actions? How do visual delays disrupt the critical language processes required for successful communication? And how do these delays impact the task strategies pairs use to successfully collaborate?

This paper aims to describe a basic function that governs the impact that delayed visual feedback has on collaborative task performance. It provides detailed insight into the amount of visual delay that can be tolerated before impacting collaborative performance, and how this range of tolerance is dependent upon features of the task. Finally, it presents both quantitative and qualitative descriptions of the strategic differences that occur across a range of visual delays.

The Impact of Delay on Collaborative Task Performance

A number of studies have examined the impact *audio delay* has on communication and collaborative task performance. As demonstrated by Krauss & Bricker [18], small audio delays of 300ms can have detrimental effects on communication processes, and delays as large as 900ms can severely impact a pair's ability to communicate. O'Connaill and Whittaker [22] found that audio delays between 410ms and 720ms led to reduced use of back-channels, fewer

interruptions, and less overlapping speech. Cohen [8] described how a simultaneous 705ms delay of audio and video resulted in longer conversational turns and decreased overlap between utterances, and Tang & Isaacs [24] found that a one-way delay of 570ms severely disrupted turn-taking behaviors. In summary, the work on audio delay and its impact on collaborative performance tends to find that delays below 300ms pose little problem. Delays between 450ms and 700ms can severely impact communication and coordination processes. And delays greater than 700ms drastically impact communication, coordination, and overall task performance.

While these studies examined audio on its own or combined with video, they do not provide insight into the communicative impact of the visual information itself. Do delays in updating *visual information*, of the sort introduced by video compression or network lags, undercut its value in a similar fashion?

Gutwin and colleagues [16, 17] examined delay[1] in the context of a tightly-coupled motor coordination task in a shared workspace. They reported that delays greater than 200ms led to a larger number of coordination errors in comparison to conditions with no delay. However, they found no impact of visual delay on overall task completion time. In addition, they do not report the differences between various levels of delay, nor do they describe the functional relationship between delay levels and task performance.

Our earlier work [20] examined the impact of delay on a collaborative task that had both language and motor components. We demonstrated that delayed visual feedback impacts performance since it is not synchronized with the state of the task nor the language that it needs to support. However, this work compared a rather long delay of 3000ms to no delay. We found that the delay impacted task performance times as well as the communication processes and language patterns of the collaborative pairs.

In addition to our work, Vaghi and colleagues [25] performed a naturalistic study of a collaborative virtual ball game where they provide qualitative evidence that strategic adaptations occur in delays ranging from 150ms to 1000ms, and they suggest that delays of 500ms or more may cause disruptions to collaborative performance.

While these studies provide evidence to suggest that a phenomenon is present, they provide a wide variety of suggestions regarding tolerable ranges of delay. We do not yet have a firm grasp on how visual feedback impacts collaborative task performance and strategies over a wide range of delays. Several questions remain. Is collaborative performance differentially impacted across a range of delays? If so, what is the corresponding functional form and how is it impacted by various task characteristics? Finally, what are the different coordination mechanisms impacted by visual delay?

To address these questions, we apply a methodology that allows us to derive the form of the function that governs the impact of visual delay. This provides a much more detailed account of the impact of visual delay on collaborative task performance and surrounding communication processes. Before proceeding with the studies, we begin with a brief theoretical discussion of the ways pairs use shared visual information in order to ground our investigation of these questions.

THEORETICAL BACKGROUND
Our theoretical understanding of the way pairs use visual evidence for collaborative purposes relies heavily on two psychological theories: Situation Awareness Theory and Grounding Theory (for further discussions see [19, 26, 27]).

Situation Awareness Theory holds that visual information helps pairs assess the current state of the task and plan future actions [10, 11]. It primarily focuses on how visual information influences the ability of groups to formulate a common representation of the state of the task, which in turn allows them to plan and act accordingly. This awareness supports low-level coordination for tightly-coupled interactions.

At a somewhat higher level, and having to do more with the language and communication surrounding a collaborative activity, Grounding Theory suggests that visual information can serve as an unambiguous source of evidence that allows conversational partners to generate efficient speech and more easily assess a level of understanding [4, 6, 7]. Visual information provides a means for coordinating language and generating efficient and understandable discourse surrounding a collaborative activity.

Together these theories predict that when groups have access to visual information they will coordinate their work better because, at a lower level, they can monitor the state of the task and plan and act appropriately, while at a higher level, they can deliver instructions and clarifications in an efficient and timely manner. Delay in providing the necessary visual information should impact these coordination mechanisms and ultimately have a negative impact on both communication processes and task performance. We now discuss the theories in more detail, highlighting the impact that delayed visual information has in each.

Visual Information in Support of Situation Awareness
According to Situation Awareness Theory, visual information is primarily valuable for coordinating the task itself. In order for collaboration to be successful, group members need to maintain an ongoing awareness of one another's activities, the status of relevant task objects, and the overall state of the collaborative task [10, 11]. This

---

[1] While Gutwin and colleagues distinguish between two forms of delay, latency and jitter, throughout this paper we use delay to refer to what they call latency.

awareness allows accurate planning of future activities and can serve as a mechanism to coordinate tightly-coupled interactions at a lower-level.

For example, Nardi and colleagues [21] describe how a scrub nurse on a surgical team uses visual information provided on an overhead monitor to assess the task state and anticipate the instruments a surgeon will need. If the nurse notices the surgeon has nicked an artery during a surgery, she can immediately begin preparation of the cauterization and suture materials and have them ready to present before the surgeon asks for them. However, if the visual information were delayed for some reason, such tight-coordination would not be possible and precious seconds could be lost.

In a similar fashion, but at an even finer temporal level, Gutwin and colleagues [17] describe how task coordination is supported by the availability of visual information during a tightly-coupled collaborative task in which pairs need to quickly move computational objects within a shared 2D workspace. When the view of the shared workspace is delayed, the pairs have difficulty assessing the state of their partner and the state of the task, and there is an increase in the number of errors they make by grasping the same piece.

At a micro-level, situation awareness of what is currently happening likely influences the next move or action engaged in. When pairs are performing tightly-coupled interactions in a distributed environment, a delay in the availability of the visual information may disrupt the formation and maintenance of such awareness, ultimately yielding coordination difficulties.

Visual Information in Support of Grounding

At a higher level, Grounding Theory suggests that visual information can improve collaborative task performance by supporting the verbal communication surrounding a collaborative activity. It states that successful communication relies on a foundation of mutual knowledge or common ground [4, 6, 7]. Speakers form utterances based on an expectation of what a listener is likely to know and then monitor whether the utterance was understood. In return, listeners have a responsibility to demonstrate their level of understanding. Shared visual information serves to support both the initial generation of utterances as well as to provide evidence of comprehension [3, 19].

Throughout a conversation, participants are continually assessing what one another knows and using this knowledge to generate subsequent contributions [3, 6, 7]. Clark & Marshall [6] propose that physical co-presence (i.e., visual access to a shared environment and the actions of a partner) allows speakers to anticipate what a partner knows. Hence, a person can point to an object in a shared physical environment and refer to it using the deictic pronoun "that" if she believes her partner can also see the object and her gesture. However, in distributed environments with delayed visual feedback such

communicative efficiencies may no longer be available and performance deficits can occur [15, 20].

Visual information also impacts comprehension monitoring in a number of ways. In a typical spoken interaction, partners can use explicit verbal statements (e.g., "I got it" or "do you mean the red one?") or back-channel responses (e.g., "uh-huh") to indicate their level of comprehension. In environments where visual information is available, the visual feedback itself can provide a critical resource for comprehension monitoring [3, 19]. Evidence can be deliberate (e.g., as in a pointing gesture) or as a side effect of proper performance of a desired action (e.g., by moving the correct object in a workspace), provided both parties are aware of what one another can see [14].

Recently, Clark and Krych [5] demonstrated that collaborative pairs use visual information to facilitate the precision timing required when discussants are introducing new entities to a discourse or changing their speech mid-sentence in response to their partner's actions. However, delays of the sort introduced by video compression are likely to undermine the value of this visual feedback. Our recent work demonstrated that large delays reduce the communication benefits of shared visual information [15, 20].

Although immediately available shared visual information generally improves collaborative task performance by supporting situation awareness and conversational grounding, the benefits it provides in any given situation will likely depend on both the accuracy of the visual information (e.g., if it is up-to-date or stale) along with the requirements for coordination imposed by the task structure. Any delays in the availability of the visual information are likely to impact these coordination mechanisms in different ways.

Hypotheses

We can summarize this discussion about the influence of delayed visual feedback on collaborative task performance in terms of a set of hypotheses that describe our expected findings. Study 1 aims to address the question of the functional form of the impact delayed visual feedback has on collaboration due to its impact on both lower level coordination tasks as well as higher level communication processes. In particular, we should expect that:

H1: A collaborative pair will perform their task more quickly when they share an immediately available shared view of the work space.

H2: A collaborative pair will perform their task more quickly as the linguistic complexity of the task objects decreases.

H3: An immediately available shared view of the work space will have additional performance benefits as the linguistic complexity of the task objects increase.

In addition to these general hypotheses about the overall impact of visual information, we should expect that the delays will differentially impact the coordination mechanisms on a different timescale,

H4: The immediacy of the shared visual information will have differential impact across a range of delays.

As the collaborative task becomes more dynamic or tightly-coupled, we should expect the tolerance a collaborative pair has for delays to become reduced. As the task becomes more tightly-coupled and dynamic, the pairs will experience performance deficits with shorter delays in comparison to less dynamic environments. We examine this in Study 2 where we manipulate the dynamics of the task environment and expect that,

H5: A collaborative pair will perform their task more quickly when the objects in the environment are less dynamic.

H6: A dynamically changing environment will reduce the tolerance a collaborative pair will have for delay in the visual feedback.

THE COLLABORATIVE PUZZLE TASK

We use the collaborative puzzle task paradigm originally presented in [20] to test our hypotheses. In this task, pairs of participants are randomly assigned to play the role of "Helper" or "Worker". The Helper needs to successfully describe a configuration of pieces to the Worker so she can properly arrange the pieces in their work space. The goal is to have the Worker correctly place the four solution pieces so that they match the target solution the Helper is viewing in the shortest amount of time.

In this task, the Helper and Worker are seated in separate rooms in front of a computer with 21-inch displays. The pairs communicate over a high-quality, full-duplex audio link with no delay. The experimental displays for the Helper and Worker are presented in Figure 1.

The Worker's screen (left) consists of a staging area on the right hand side where the puzzle pieces are held, and a work area on the left hand side where the puzzle is constructed. The Helper's screen (right) shows the target solution on the right, and a view of the Worker's work area in the left hand panel. Constructing the displays in a computational
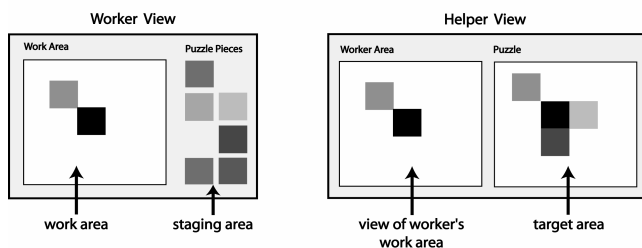

Figure 1. The Worker's view (left) and the Helper's view (right).

environment facilitates the manipulation of the Helper's view of the Worker's work area and allows us to introduce variable delay rates as well as explore different task and object features.

STUDY 1: THE IMPACT OF VISUAL DELAY ON COLLABORATIVE PERFORMANCE

In this study we were primarily interested in assessing the pair's performance over a wide range of visual delays. In addition, we were interested in examining the conversational and communication processes adopted by the pairs.

Method

We manipulated the amount of visual delay present in the Helper's view of the workspace (Visual Delay) along with the amount of conversational grounding that was required to describe the pieces in the environment (Linguistic Complexity).

Independent Variables

Visual Delay [60-3300ms]: We chose delay times from a distribution that provided a finer level of granularity at the shorter delays since prior literature suggested that task performance might be more sensitive to times in that range. The times were generated according to the following distribution:

$$f(n) = \begin{cases} T_1 = 60 \\ T_n = T_{n-1} \cdot e^{.05} \end{cases}$$

We then temporarily slotted these times into three sub-ranges for the sake of balancing participant assignment. *Low* delay was the range of [60-230ms], *Medium* delay was [230-850ms], and *High* delay was [850-3300ms]. Participants were selected to receive two levels from each bin and these times were crossed with the levels of linguistic complexity[2].

Linguistic Complexity (Primary vs. Plaid): We manipulated the linguistic complexity of the task by providing the pairs with two types of pieces. The pieces were either lexically simple, easy to describe primary colors (e.g., red, yellow, orange, etc.), or they were more complex tartan plaids that required negotiating a common naming convention. While the primary colors were likely to be part of a shared lexicon and therefore required very little grounding to name the objects, the plaid pieces were not and required the pairs to negotiate on the terms used to represent the various pieces. Figure 2 presents examples of the task objects.

---

[2] It is important to keep in mind that while we discuss bins here, the variable represents an essentially *continuous* range. The bins were only temporarily used in order to assign each pair to a number of delays that fell somewhere in the low, middle and high ends of the distribution. This was done to balance the pairs across the times and not to conflate any given pairs with the range of times they received.
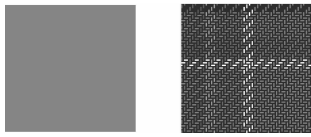
Figure 2. Primary pieces (left) and Plaid pieces (right).

Participants and Procedure
Participants consisted of 27 pairs recruited from the Pittsburgh area. They were randomly assigned to play the role of Helper or Worker and the pairs were balanced by gender. *Visual Delay* [60-3300ms] and *Linguistic Complexity* (Primary, Plaid) were manipulated within each pair. Each pair participated in a total of six experimental conditions, three different Visual Delay times for each level of Linguistic Complexity, counter-balanced. Pairs solved four puzzles within each experimental condition. This resulted in a total of 24 puzzles that were completed in approximately one hour.

Measures
Task Performance Measures. The pairs were instructed to complete the puzzles as quickly and accurately as possible. We used time to complete the puzzle as the primary measure of task performance. Nearly all puzzles were solved correctly, so error rates were a less useful indicator of performance.

Conversational Excerpts. To detail how processes the pairs used at varying rates of visual delay, we transcribed their interactions and present representative examples to demonstrate qualitative evidence of the communication patterns witnessed.

Statistical Methods and Analyses
We used a statistical technique known as Multivariate Adaptive Regression Splines (MARS) [12] to model the impact that the visual delay had on collaborative performance. This technique describes the impact of the independent variables (e.g., Visual Delay or Lexical Complexity) on the dependent variable (task completion time) as an optimized sequence of piecewise linear regressions[3]. The algorithm finds optimal breakpoints by examining points within the range of the independent variable where slope changes are most likely to occur. The resulting piecewise segments can then be examined to explain, for example, how much delay can be tolerated before group performance begins to suffer or to describe the rate at which collaborative performance is impacted over a particular range of delays.

The MARS algorithm is a two-stage process. The first stage begins with a forward selection process that adds functions (i.e., variables capturing the breakpoints) to the model. As

---

[3] While this is not strictly true in all cases, particularly in higher-dimensional interactions, it holds in the analyses presented here. For details on the procedure see [12].

more functions are added, the model begins to account for non-linear trends in the data. This cycle continues until a pre-defined number of functions have been added. At this point, the algorithm enters the second stage where it prunes the functions until it achieves an optimized tradeoff between the number of functions and the goodness of fit. The Generalized Cross Validation (GCV) measure is used as the model measure of goodness of fit (originally described in [9]; modified by [12]). The GCV measure strikes a balance between model complexity and quality of fit in a fashion similar to that of the Aikake Information Criterion (AIC) commonly used in parametric regression models.

In the models presented in this paper, we allowed the algorithm to construct up to 100 functions for inclusion. Each of the models was evaluated using a 10-fold cross validation technique. That is, each model is created over 10 trials, with each trial using 90% of the data to train, and the remaining 10% to test the model's performance. Performance is optimized based on the best fit as assessed by the GCV error measure.

Results
For the first stage of the analysis we used the MARS method to discover the optimal partitioning of the continuous Visual Delay factor. We found two major breakpoints at delays 939ms and 1798ms (the circles seen in Figure 3). These results were then used to construct an appropriate random effects piecewise linear regression model where [Visual Delay], [Visual Delay – 939], [Visual Delay – 1798], Linguistic Complexity (*Primary, Plaid*), Block (*1-6*), and Trial (*1-4*) were repeated factors. We also included all 2- and 3-way interactions, for each Visual Delay segment, in the analysis. Because each pair participated in 24 trials, observations within a pair were not independent of one another and were modeled as a random effect. The final model achieved a good fit to the data (*Adj-R^2* = .532; *GCV-R^2* = .497; *p* < .001).

Task Performance
Linguistic Complexity. Consistent with *H2*, the manipulation of linguistic complexity had a large impact on the speed with which the pairs could solve the puzzles. Overall, the pairs were consistently faster in the trials in which the puzzle pieces were Solids than when they were Plaids (38.0sec vs. 61.7sec; $F_{1,610}$ = 270.6, *p* < .001).

Visual Delay. Consistent with *H1*, the quicker the visual feedback provided, the faster the pairs were at completing the puzzles. However, this result was not consistent across the entire range of delays. Similarly, the results addressing *H3* were only found for delays greater than 1798ms.

For delays between 60ms and 939ms, we found no evidence to indicate any impact of delayed visual feedback on task performance (β = 0.48, *SE* = (2.87), $F_{1,610}$ = .028, *p* = .87). As can be seen in Figure 3, the slope for this segment is relatively flat. In this range of delay there was no impact for
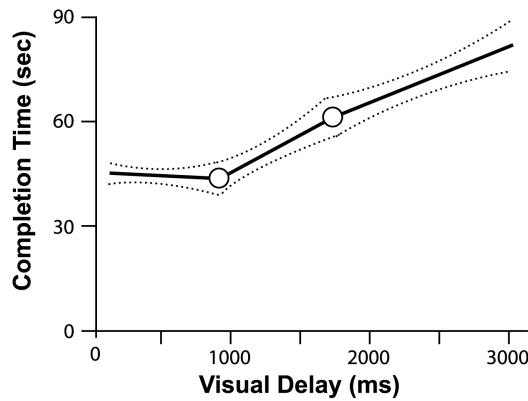
Figure 3. Effect of Visual Delay on Task Completion Time. Main effect graph of piecewise linear regression fit line (solid) with learned breakpoints (circles) and corresponding 95% confidence intervals (dashed).

either the *Primary* or *Plaid* pieces. In other words, we found no evidence of a *[Visual Delay] × Linguistic Complexity* interaction ($F_{1,610} = .71$, $p = .40$).

However, for delay rates between 939ms and 1798ms there is a significant impact on task performance ($F_{1,610} = 13.57$, $p < .001$). This can be seen in Figure 3, where the slope for this segment is rather steep. In this range, every 100ms increase in visual delay slowed the pair's completion time by an additional 2.3 seconds (holding constant the mean level of all other variables in the model). The impact of delay was equally important for both the *Primary* and *Plaid* pieces, as evidenced by the fact that we found no *[Visual Delay – 939] × Linguistic Complexity* interaction ($F_{1,610} = 1.74$, $p = .19$).

Delay rates greater than 1798ms also demonstrated a significant impact on task performance ($F_{1,610} = 15.28$, $p < .001$). While Figure 3 illustrates the mean increase across the two levels of linguistic complexity, there was a significant *[Visual Delay – 1798] × Linguistic Complexity* interaction ($F_{1,610} = 10.46$, $p = .001$). In support of *H4*, decomposition of this interaction reveals that the slope for the Plaid pieces remains strong and positive, while for the Primary pieces it is flat to slightly negative. In the higher range of delays, the impact of the delay appeared to additionally affect the Plaid puzzles. This suggests that when the delays were greater than 1798ms, it appeared as though they impacted the conversational grounding processes required to talk about the plaid pieces, while having little additional impact on the primary colored pieces which were already a part of the pairs' shared lexicon. We describe these differences in more detail with the following qualitative descriptions of the pairs' performance.

Conversational Excerpts

Figure 4 presents an example of the types of problems that arose for both Primary and Plaid puzzles when the delay was in the range of [939-1798ms]. In this range, the pairs

demonstrated a number of coordination errors that signified they had misaligned awareness of one another's task state. In this example, the Helper describes a piece and where to put it (line 1). However, the delayed visual feedback causes him to reiterate his directive (line 3) since he assumes his partner did not hear or did not understand. However, when the Worker hears this, her puzzle state already indicates the correct move (line 3), and therefore she interprets his reiteration as a clarification and incorrectly adjusts the piece to the lower left of the workspace (line 4). The Helper then sees the delayed view and believes everything is fine and confirms the placement (line 5). Unfortunately, the Worker believes this confirmation to be about her new placement. Shortly thereafter, the Helper sees the incorrect move and they begin a repair sequence. This example demonstrates how the delay led to misaligned views of the task state, ultimately resulting in coordination problems that impacted task performance.

In the highest range of delays [1798ms – 3300ms], the differences between views becomes readily apparent, and the pairs demonstrate a strategic shift whereby they exhibit less behaviors that rely on tight integration between speech and the visual information. At these higher levels of delay, the pairs tended to attempt to complete the puzzles simply using spoken language. This was evidenced by their relative lack of use of deictic pronouns such as "that one," and "this." Instead, the pairs relied on lengthier verbal descriptions to describe the objects and their arrangement. However, this posed a greater problem for Plaids than it did for the Primary pieces. When the pieces were part of their shared lexicon, pairs could quickly describe the colors and

Worker View	Helper View	Speaker	Speech / [Actions]
		Helper	Now take that one right to the left of it and put it in the bottom left hand corner
		Worker	[Correctly moves piece to bottom left of previous piece]
		Helper	Move it to the bottom left corner
		Worker	[Incorrectly re-positions piece to the bottom left of the workspace]
		Helper	OK, now…
		Helper	…no, no, no

Figure 4. Excerpt demonstrating a coordination error resulting from a lack of shared situation awareness (at a delay of approximately 1100ms).

Primary Pieces	Plaid Pieces
H: then there's like a-a red	**H:** um horizontal white stripe...
W: okay	**W:** any blue?
H: make it touch the corner	**H:** and two and two...and two like hor- vertical gray stripes
W: okay	**W:** horizontal white stripe with two vertical gray stripes?
H: then there's another red but it's more rosey...	**H:** yeah
W: rosey, okay	**W:** this one?
H: make it touch the red one on the left hand side...	[pause 2.5 sec]
	H: uh...no...
	W: oh...
	H: it's horizontal white stripe and two vertical stripes...
	H: yeah that one

Figure 5. Excerpt demonstrating grounding difficulties in Plaids (right) versus Primary colors (left) at a delay of approximately 2700ms.

where to place them while using verbal acknowledgements to keep on track, reserving use of the visual information for delayed confirmation that the task had been performed correctly (see Figure 5, left excerpt). However, when negotiation about the names of the pieces was required, as was the case with the Plaids, the inefficiencies of the linguistic medium became much more evident (for similar findings see [20]). As can be seen in Figure 5 (right), the Helper and Worker both became much more active in negotiating their descriptions. However, when the visual evidence was needed for disambiguation or confirmation, they had to wait to receive the information. This shift in strategy likely led to the additional impact of delay on task performance for the Plaids over the Primary colors in the high range of delays.

STUDY 2: THE IMPACT OF TASK DYNAMICS AND VISUAL DELAY ON COLLABORATIVE PERFORMANCE

While the results in Study 1 are consistent with the findings in Kraut et al., [20] they contrast with the timings presented by Gutwin and colleagues [16, 17]. In this study we aim to clarify this inconsistency by examining how the dynamics of the task objects interact with the visual delay to impact the lower-level coordination mechanisms required for successful collaborative performance. In this study we demonstrate how the dynamics of the task interact with the amount of visual delay that can be tolerated before impacting task performance.

Method
We manipulated the amount of visual delay present in the Helper's view of the workspace (Visual Delay), and the dynamics of the task objects by providing puzzle pieces that changed colors at various rates (Object Dynamics).

Independent Variables
Visual Delay [100-3119ms]: We chose a similar distribution as Study 1. However, the initial delay was set at 100ms. The times were generated according to the following distribution:

$$f(n) = \begin{cases} T_1 = 100 \\ T_n = T_{n-1} \cdot e^{.05} \end{cases}$$

We temporarily slotted these times into three sub-ranges for assignment. *Low* delay was the range of [100-306ms], *Medium* delay was [319-977ms], and *High* delay was [1020-3119ms]. Participants were selected to receive two levels from each bin and these were crossed with similarly binned levels of the Object Dynamics.

Object Dynamics (Moderate, Fast, and Very Fast): We manipulated the dynamic complexity of the task objects by allowing the colors of the blocks to cycle. Each piece changed its color, smoothly moving through the color palette. At the *Moderate* cycle rate, the pieces experienced a major perceivable color change (e.g., from "red" to "orange", or "blue" to "purple") approximately every 6-8 seconds. It took roughly one second of continuous observation to notice whether any given piece was changing. In the *Fast* cycle rate, the pieces achieved a major perceivable color change approximately every 2-3 seconds. While at the *Very Fast* cycle rate, the pieces were rapidly changing colors at a rate of approximately one perceivable color change every second or less. It should be noted that these values fluctuate somewhat due to the fact that people do not perceive change equally across the color spectrum.

Participants and Procedure
Participants consisted of 27 pairs recruited from the Pittsburgh area. They were randomly assigned to play the role of Helper or Worker and the pairs were balanced by gender. *Visual Delay* [60-3300ms] and *Object Dynamics* (Moderate, Fast, Very Fast) were manipulated within each pair. Each pair participated in a total of nine experimental conditions that varied across a range of delays crossed with a range of object cycle rates. Pairs solved four puzzles within each experimental condition. This resulted in a total of 36 puzzles that were completed in approximately an hour and a half.

Measures
We used the same measures of task performance as in Study 1, once again choosing task completion time over errors since the number of final errors was very low.

Statistical Methods and Analyses
The analyses in this study are the same as those described in Study 1, with one exception. We ran separate models for each level of the Object Dynamics (Moderate, Fast, and Very Fast) in order to discover the optimal breakpoints for each object cycle rate.

Results
For ease of exposition we focus our description of the results on the overall model fits, their breakpoints, and the slopes of their initial two segments. This allows us to

**The Relationship Between Object Change Rate
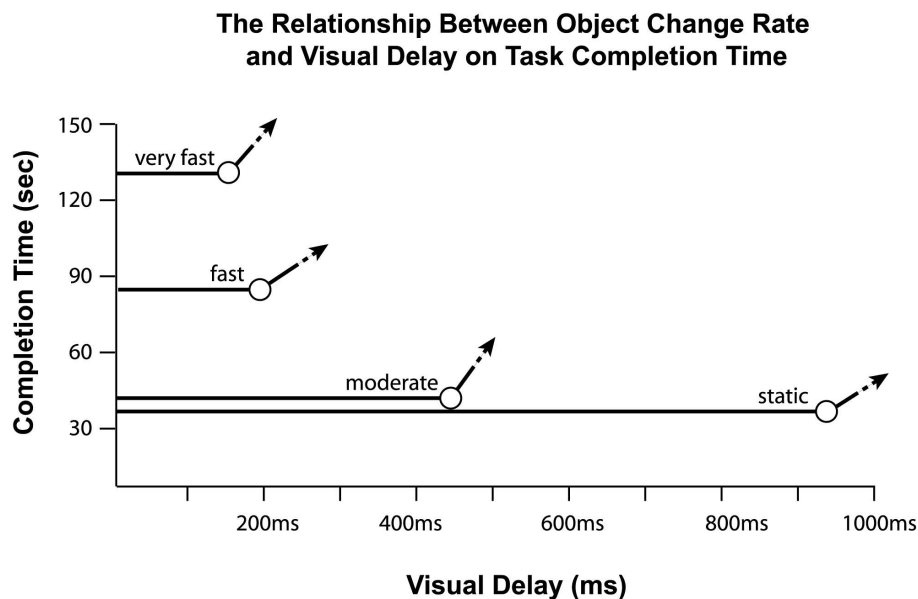and Visual Delay on Task Completion Time**

Figure 6. This illustration presents a stylized view of the data. It represents the initial breakpoints (circles) across a range of object dynamics. The lines up to the breakpoints are slopes that are not significantly different from zero, and the subsequent trajectories represent the slope changes. From top-to-bottom the lines represent the various levels of change rate: Very Fast, Fast, Moderate (Study 2), and Static (Study 1).

clearly demonstrate how the dynamics of the environment shift the range of tolerable delays when a more dynamic environment is in play (as overviewed in Figure 6).

Model of the Moderate Change Rate

We discovered two optimal breakpoints in delay at 431ms and 558ms, for the objects that changed at a *Moderate* rate. These results were then used in a random effects piecewise linear regression model where [Visual Delay], [Visual Delay – 431], [Visual Delay – 558], Block (*1-9*), and Trial (*1-4*) were repeated factors. We included all 2- and 3-way interactions in the analysis. Because each pair participated in 36 trials, observations within a pair were not independent of one another and were modeled as a random effect. The model achieved a good fit to the data (Adj-R^2 = .522; GCV-R^2 = .498; $p < .001$).

When examining the impact of visual delay in the *Moderate* condition, we found that there was no impact on task performance when the delay was below 431ms (β = -2.16(13.8), $F_{1,265}$ = .024, p = .88). In other words, the slope for the range of delays between 100ms and 431ms was essentially flat. Note however, that the dynamics of task objects did substantially shorten the range of tolerable delays in comparison to those found in Study 1 when the pieces were non-changing solid colors (see Figure 6).

However, once the delay reached 431ms there was a significant impact on task performance ($F_{1,265}$ = 8.362, p = .004). At this point there was evidence of a drastic impact of the visual delay on task performance, with every 100ms increase in visual delay increasing the pairs' completion time by approximately 14.5 seconds.

Model of the Fast Change Rate

For the *Fast* paced changing objects the optimal breakpoints were found to be at 191ms and 1783ms. We used the same model as above with the following changes, [Visual Delay], [Visual Delay – 191], [Visual Delay – 1783]. The model achieved a good fit to the data (Adj-R^2 = .528; GCV-R^2 = .442; $p < .001$).

When examining the impact of visual delay at the *Fast* level, we found the range of tolerable delays to be greatly reduced in comparison to the *Moderate* change rate (see Figure 6). For delays under 191ms, there was no evidence of the impact of delay on task performance (β = -68.4(60.3), $F_{1,278}$ = 1.288, p = .26). Once again, the slope for this initial segment was essentially flat.

Once the delay reached 191ms, the trend towards an impact on task performance appeared in the appropriate direction ($F_{1,278}$ = 2.24, p = .13). However, while the plots appear to indicate an upward swing consistent with the other models, indicating an impact of delay on task performance, the slope of this shift was not significant as it was in all the other models. This may be due in part to an increase in the amount of noise in the data given the increasing complexity of the task.

Model of the Very Fast Change Rate

For the *Very Fast* paced changing objects the optimal breakpoints were found to be at 154ms and 450ms. We used the same model as above with the following changes, [Visual Delay], [Visual Delay – 154], [Visual Delay – 450]. The model achieved a reasonable fit to the data (Adj-R^2 = .443; GCV-R^2 = .367; $p < .001$).

When examining the impact of visual delay when the objects were rapidly changing, we found the range of tolerable delays to be smaller than for any other condition (see Figure 6). For delays under 154ms, the slope was again flat and there was no evidence of the impact of delay on task performance (β = -341.2(214.9), $F_{1,254}$ = 2.52, p = .12). This suggests that there was no impact of the delay on performance in the range between 100 and 154ms.

However, once the delay reached 154ms, there appeared to be a marginal impact on task performance ($F_{1,254}$ = 3.23, p = .07). At this point there was evidence of a drastic impact of the visual delay on task performance.

DISCUSSION

In these studies we demonstrated the application of a statistical method that allows us to examine collaborative task performance over a continuous range of visual delays. This method provides detailed insight into the range of delays within which collaborative task performance is not affected, as well as uncovers the points at which performance begins to break down. In addition, examination of the corresponding slope coefficients provides an indication of the relative impact additional delays have on performance. This method allows us to extend the findings of earlier work that examined discrete levels of delay but could not pinpoint the precise time at which collaborative performance breaks down in the presence of delayed visual information [16, 17, 20].

In Study 1 we found that the visual delay had no impact on task performance when it was less than 939ms. However, for the range between 939 and 1798ms the delay impacted both the Primary and Plaid puzzles equally. The conversational transcripts suggest that these deficits in performance may be due in part to the fact that the lower-level coordination processes supported by shared situation awareness are disrupted. The delayed visual information can lead to misalignments in a pair's model of the current state of the shared task (as demonstrated in Figure 4). Such misalignments, or inaccurate mental representations of task state, can severely impact low-level coordination on the part of the pairs.

At delays greater than 1798ms, the impact of the delay seemed to shift to conversational grounding processes. This was evidenced by the fact that the *[Visual Delay – 1798] × Linguistic Complexity* interaction was significant, and that there remained an increasing slope for the Plaid pieces while the slope for the Primary pieces leveled off. The transcripts revealed that this may be due to the fact that for the Primary pieces the pairs simply resorted to using linguistic terms to describe the objects and their placement and only used the visual information for delayed confirmation (for additional evidence of this see [20]). However, when attempting to use this strategy to describe the Plaid pieces, the pairs suffered a much greater penalty for not being able to use the efficiencies of the visual space for supporting grounding on the object terms. Instead, they had to use rather inefficient linguistic descriptions in an attempt to ground on terms to represent the Plaid pieces. In this case, the pairs were much more reliant on the visual information to play a role in disambiguation and comprehension monitoring.

In Study 2, as is captured in Figure 6, and in support of *H5* and *H6*, we found that when the dynamics of the task objects increased, visual delay began to have an impact at much shorter time intervals. This was demonstrated by the tendency for the first breakpoints to move closer to the 100ms lower bound. Together, these results provide evidence that the dynamics of the task objects and environment have a major impact on the range of delay that can be tolerated before collaborative task performance begins to suffer. The range of tolerable delays found in this study are much more in line with those described in Gutwin's work [16, 17]. In our moderately dynamic environment we found that pairs could accommodate up to a 431ms delay. However, as the dynamics of the task increased to our fast rate, we found that the pairs appeared to suffer performance deficits once the delay reached 191ms (although this result was only a trend), and for our very fast dynamic rate the pairs could only tolerate delays up to 154ms before their performance began to degrade.

Together, these results suggest it is not as simple as picking a single number to serve as a hard threshold for dictating whether or not a given delay is tolerable for collaborative task performance. Instead, a detailed task-analysis needs to be performed in order to establish the collaborative requirements of the task.

CONCLUSIONS
We examined the effects that delayed visual feedback has on collaborative task performance. Our results demonstrate that a number of factors come into play when assessing the tolerance pairs have for visual delay. A solid understanding of the degree to which the collaborative pairs rely on situation awareness to perform their task, knowing the amount of shared domain they share, and understanding the complexity and dynamics of the task environment are all keys to understanding how well a given technology may serve a particular group.

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

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

3. Brennan, S.E. (2005). How conversation is shaped by visual and spoken evidence. In Trueswell, J.C., and Tanenhaus, M.K. Eds. *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, 95-129. MIT Press, Cambridge, MA.

4. Clark, H.H. (1996). *Using language*. Cambridge University Press, New York, NY, US.

5. Clark, H.H., and Krych, M.A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory & Language*, *50* (1), 62-81.

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

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

8. Cohen, K. (1982). Speaker interaction: Video teleconferences versus face-to-face meetings. In *Proceedings of Teleconferencing and Electronic Communications*, 189-199. University of Wisconsin Press.

9. Craven, P., and Wahba, G. (1979). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, *31*, 317-403.

10. Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors Special Issue: Situation Awareness*, *37* (1), 32-64.

11. Endsley, M.R., and Garland, D.J. (2000). *Situation awareness analysis and measurement*. LEA, Mahwah, NJ.

12. Friedman, J.H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, *19*, 1-141.

13. Fussell, S.R., Setlock, L.D., and Parker, E.M. (2003). Where do helpers look? Gaze targets during collaborative physical tasks. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2003) (Extended Abstracts)*, 768-769. NY: ACM Press.

14. Gergle, D., Kraut, R.E., and Fussell, S.R. (2004). Action as language in a shared visual space. In *Proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW 04)*, 487-496. NY: ACM Press.

15. Gergle, D., Kraut, R.E., and Fussell, S.R. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language & Social Psychology*, *23* (4), 491-517.

16. Gutwin, C. (2001). The effects of network delays on group work in real-time groupware. In *Proceedings of European Conference on Computer-Supported Cooperative Work*, 299-318.

17. Gutwin, C., Benford, S., Dyck, J., Fraser, M., Vaghi, I., and Greenhalgh, C. (2004). Revealing delay in collaborative environments. In *Proceedings of ACM Conference on Computer Human Interaction (CHI 2004)*, 503-510. NY: ACM Press.

18. Krauss, R.M., and Bricker, P.D. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America*, *41* (2), 286-292.

19. Kraut, R.E., Fussell, S.R., and Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human Computer Interaction*, *18* (1), 13-49.

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

21. Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S., and Sclabassi, R.T. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. In *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems*, 327-334.

22. O'Conaill, B., and Whittaker, S. (1997). Characterizing, predicting, and measuring video-mediated communication: A conversational approach. In Finn, K.E., Sellen, A.J., and Wilbur, S.B. Eds. *Video-Mediated Communication*, 107-131. LEA.

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

24. Tang, J.C., and Isaacs, E.A. (1993). Why do users like video? Studies of multimedia-supported collaboration. *CSCW: An International Journal*, *1*, 163-196.

25. Vaghi, I., Greenhalgh, C., and Benford, S. (1999). Coping with inconsistency due to network delays in collaborative virtual environments. In *Proceedings of VRST '99*, 42-49.

26. Whittaker, S. (2003). Things to talk about when talking about things. *Human-Computer Interaction*, *18* (1-2), 149-170.

27. Whittaker, S., and O'Conaill, B. (1997). The role of vision in face-to-face and mediated communication. In Finn, K.E., Sellen, A.J., and Wilbur, S.B. Eds. *Video-Mediated Communication*, 23-49. LEA.