

Using Eye-Tracking Techniques to Study Collaboration on Physical Tasks: Implications for Medical Research

SUSAN R. FUSSELL AND LESLIE D. SETLOCK

Human Computer Interaction Institute

Carnegie Mellon University

5000 Forbes Avenue, Pittsburgh, PA USA

This paper discusses eye-tracking as a technique to study collaborative physical tasks—tasks in which two or more people work together to perform actions on concrete objects in the three-dimensional world. For example, a surgical team might collaborate to save treat a patient. We first consider the use of eye-tracking as a dependent measure—that is, the recording of gaze as people perform their tasks. We review studies applying eye-tracking to individual performance of physical tasks and interpersonal communication, then present a study on gaze in a collaborative construction task. Next, we consider eye-tracking as an independent measure—a factor that is manipulated in studies of remote collaboration on physical tasks. We discuss how the use of eye-tracking can be used to assess the importance of gaze awareness information for collaboration and present results of a study using this technique. We end by considering limitations and theoretical issues regarding eye-tracking as a research tool for collaborative physical tasks.

Video, eye-tracking, interpersonal communication, collaborative work

Introduction

There is growing interest in understanding collaborative physical tasks—tasks in which two or more individuals work together to act on concrete objects in the three-dimensional world. Examples can be found across a variety of domains, including technical assistance (e.g., an expert guiding a worker's performance of aircraft repairs), education (e.g., students collaborating to built a science project), and medicine (e.g., a surgical team working together to save a patient's life). A better understanding of how people execute collaborative physical tasks can benefit training and education, error

prevention, and the design of systems to enable remote collaboration on such tasks (e.g., telemedicine).

Observational studies of physical collaboration show that people's speech and actions are intricately related to the position and dynamics of objects, other people, and ongoing activities in the environment (e.g., Ford, 1999; Goodwin, 1996; Nardi et al., 1993; Tang, 1991). Conversations during these tasks include identifying target objects, describing actions to be performed on those targets, and confirming that actions have been performed successfully. During the course of the task, the objects may undergo changes in state as people act on them (e.g., a malfunctioning piece of surgical equipment may undergo repair) or as the result of outside forces (e.g., a patient might start hemorrhaging).

Because of the complex interactions among actions, speech and environment in collaborative physical tasks, there are numerous points at which errors and miscommunications may arise. For example, a nurse might misunderstand a doctor's request for a particular implement, have difficulty finding the implement amongst a set of alternative tools, or be otherwise engaged in behaviors that conflict with his/her delivering the tool in a timely fashion. Understanding the sources of errors and miscommunications during physical collaborations is essential to devising strategies to minimize them. The complexity of collaborative physical tasks also makes it difficult to devise suitable technologies to permit their remote accomplishment. If, for instance, a surgeon is guiding an operation at another location, what sorts of tools (video, audio, and so on) must we provide that surgeon in order for him/her to successfully interact with the other team members? To address these issues, we need a deeper understanding of the dynamics of face-to-face collaboration on physical tasks. We need to know what techniques collaborators use to coordinate their activities and where their coordination may break down.

Collaborative physical tasks are typically fast-paced and can involve multiple participants as well as tools, parts, and the like. Hence, it can be difficult for researchers to study them in real time—too much is happening at any given moment for a single observer or small set of observers to capture the full experience. For this reason, video

recording can be invaluable for understanding the precise dynamics that arise over the course of the interaction.

Standard video techniques provide coarse-grained information about people's focus of attention (e.g., whether a surgeon is looking at the patient, a monitor, or another member of the medical team). In some cases, however, it may be valuable to understand in finer detail where a person's attention is focused. The increasing availability of mobile eye-tracking units now allows researchers to combine study of participants' gaze patterns with other video-based analyses of interaction. Although the relationship between gaze and attention is not invariant (cf. Velichkovsky et al., 2000), gaze nonetheless provides an excellent cue for inferring attention. Eye-tracking technologies combined with video allow researchers to study in depth where people are looking over the course of a task, the patterns with which they scan the scene, and the relationships among these gaze targets and patterns and ongoing activities.

Eye-tracking has been fruitfully applied in numerous areas of cognitive and applied psychology (see Duchowski, 2002; Jacob & Karn, in press, Rayner, 1998; and papers in Hyona et al., in press; and ETRA 2002), including studies of complex cognitive tasks such as driver distraction (e.g., Land & Horwood, 1995; Sodhi, et al., 2002; Sodhi et al., in press; Velichkovsky et al., 2000), and pilot eye movements (e.g., Anders, 2001; Kasarskis et al., 2001). Within the medical domain, investigators have used eye-tracking to study the radiological image interpretation (e.g., Krupinski & Roehrig, 2002; Mello-Thoms et al., 2002) surgical eye control (Tchalenko et al., 2001) and anesthesiologists' monitoring behaviors (Seagull et al., 1999). To date, however, eye-tracking has been rarely used in studies of physical collaboration.

In the remainder of this paper we examine eye-tracking as a tool for understanding collaboration on physical tasks. We first consider the use of eye-tracking as a dependent measure—recording of gaze direction, fixations, and the like as people perform their tasks. We review studies applying eye-tracking to individual performance of physical tasks and to interpersonal communication, and then present a study we have done on gaze in a collaborative robot construction task. Next, we consider eye-tracking as an independent measure—as a way to study the importance of seeing others' gaze for collaboration on physical tasks. The role of gaze awareness is difficult to study in face-

to-face settings; instead, we use a paradigm in which remote partners collaborate with and without gaze awareness in order to assess the effects of this awareness on interaction. We end by considering limitations and theoretical issues regarding eye-tracking as a research tool for collaborative physical tasks.

Eye-tracking as a Dependent Measure

Traditionally, eye-tracking is used as a dependent measure. Individuals are presented with tasks such as target identification, web page evaluation, or driving a car in a simulator, and their eye movements are recorded as a series of coordinates using eye-tracking software. Many eye-trackers also include a small camera that records the scene as viewed by the participant at the same the eye-movements are recorded. In static settings, such as a single viewer looking at a single computer monitor, the eye tracking system software can compute the percentage of time a participant looks at predefined areas in the scene, and to display gaze patterns overlaid on the static scene view. However, in mobile settings, such as a hospital operating room or automobile, the scene is constantly changing and alternative methods of identifying gaze targets must be employed. In the ISCAN system we use in our lab (<http://www.iscaninc.com>), for example, eye gaze is recorded as an “X” overlaid on the output of the head-mounted camera. By using the camera and eye-tracker output together, we can identify gaze targets as a person moves around the environment.

The ability to track gaze targets (and thus estimate attention) over the course of a collaboration allows investigators to investigate a variety of research issues that would otherwise be difficult to study. For example, we can determine the percentage of time a surgeon looks at the patient as opposed to other objects and individuals in the operating room. The fine-grained time intervals of the gaze recordings also allow us to assess how quickly a person identifies a problem once it arises (e.g., if a patient starts hemorrhaging), or finds a tool or completes a task after it has been requested. Typical patterns of gaze during successful performances of a given task can be identified and used as a basis for distinguishing experts from novices or evaluating how well the task has been learned.

In the remainder of this section we provide examples of the application of eye-tracking in two areas related to collaborative physical tasks—the performance of solo physical

tasks, and interpersonal communication. We then briefly describe work in progress examining gaze during a collaborative robot construction task.

Eye-tracking research on (non-collaborative) physical tasks

Recently, a number of studies have applied eye-tracking to non-collaborative physical tasks, with the aim of understanding the relationships between gaze and actions. For example, Land and colleagues (Land et al., 1999; Land & Hayhoe, 2001) used eye tracking to study gaze and hand movements during the performance of well-learned physical tasks (e.g., making tea [Land et al., 1999], making peanut butter and jelly sandwiches [Land & Hayhoe, 2001], and handwashing [Pelz & Canosa, 2001].) Other studies have used eye-tracking to examine complex athletic behaviors (e.g., Fairchild et al., 2001; Oudejans et al., 1999; Vickers, 1999).

This line of research has made considerable progress identifying the typical patterns of eye movements people make while performing physical tasks. Land et al. (1999), for instance, were able to identify four basic categories of eye movements: locating an object, directing an object to a goal, guiding two objects together, and checking the status of an object. Several studies have shown that eye movements are directly related to task behaviors, and precede hand movements to the same targets by about 500 msec. (Johansson et al., 2001, Land et al., 1999, Land & Hayhoe, 2001; Pelz & Canosa, 2001). Ballard et al. (1995), using a task in which participants arranged colored building blocks to match a model, were able to identify predictable gaze patterns (e.g., from model, to block, to model, to construction area). Additional research has found predictable relationships among gaze, head position, and behaviors in physical tasks (e.g., Pelz, Hayhoe and Loeber, 2001; Smeets, et al., 1996).

As a whole, this research demonstrates the applicability of eye-tracking to the understanding of physical tasks. To date, however, few studies have studied gaze in tasks requiring coordination among multiple participants, such as would be typical in collaborative medical procedures (one exception is a study of table tennis by Land & Furneaux, [1997]). In addition, the solo performers in the studies reviewed above had no need to communicate with partners as they performed their tasks. Since communication presents its own demands for visual attention (e.g., Argyle & Dean, 1976), we anticipate

that the need to talk during a collaborative physical task would complicate the regular patterns of gaze found in these studies of solo tasks. As we discuss in the next section, eye-tracking research has also been fruitful in understanding processes of message production and comprehension.

Eye-tracking research on interpersonal communication

A second relevant line of research relevant to our interest in applying eye-tracking methodology to collaborative physical tasks focuses on using the technology as a tool to understand human communication. For example, studies investigating how quickly a named object is visually fixated have been used to test theories of language comprehension (e.g., Brown-Schmidt et al., 2002; Chambers et al., 2002; Eberhard et al., 1995; Hanna et al., under review; Keysar et al., 2000; Metzger & Brennan, under review). The majority of these studies have used a referential communication task in which one person (typically a confederate, in the eye-tracking studies) provides a series of descriptions of objects for another person, who must find the target in an array of alternatives. Investigators have manipulated such variables as the extent of common ground between speaker and listener to test theories of the role of common ground in message comprehension. Note that this task of object identification is common to many collaborative physical tasks, and this same research paradigm might be used to investigate the effects of, say, nurse experience on speed of identifying a requested surgical implement.

Other studies have used eye-tracking to determine people's focus of attention in conversation. Vertegaal et al. (2001) examined gaze at partners during a four-person conversation about current events and found that gaze strongly indicated participants' focus of attention. Stiefelhagen & Zhu (2002) also studied gaze during four-party conversations with a focus on how head and eye movements were associated as cues of attention. Gullberg (Gullberg & Holmqvist, 1999; Gullberg, 2003) studied fixations towards gestures in conversational settings and found that consistent with previous research, (e.g., Argyle & Dean, 1976) most attention was paid to partners' faces and only a small number of gestures were directly fixated. In a slightly different vein, Dabbish and Kraut (in progress) are using eye-tracking to investigate the effects of the degree of detail

presented in online awareness notifications about a partner's status on the timing of electronic communications.

With the exception of the Dabbish and Kraut study, none of the above research looked at conversations in which participants had to manipulate objects or perform other physical activities while they were conversing. In the next section, we consider how eye-tracking might be used to study tasks that combine both speech and action.

Eye-tracking research on collaborative physical tasks

The sections above suggest the usefulness of eye-tracking for understanding physical tasks and interpersonal communication individually, but none of the studies examined gaze in conversations *during* collaborative physical tasks. For example, when properly executed, a doctor may ask a nurse for an implement as he/she is working on a task, and the nurse will quickly visually identify and pass him/her that implement. To date, there is little research using eye-tracking in this context (but see Pomplun et al., 1998).

In a recent study (Fussell et al., under review), we used eye-tracking to examine people's use of visual information as they assist their partners during a collaborative robot construction task. The robot task falls within a general class of "mentoring" collaborative physical tasks, in which one person directly manipulates objects with the guidance of one or more experts. In our task, one person—the "worker"—builds a large toy robot. A second person—the "helper"—provides guidance to the worker during the task but does not actually manipulate objects, tools or parts. The relationship between helper and worker is thus similar to a teacher guiding a student's lab project or a head resident instructing new doctors on patient care.

In mentoring collaborative tasks, helpers must determine what assistance is needed and when, how to phrase their messages such that the worker understands them, and whether the message has been understood as intended. As we have discussed in more detail elsewhere (Kraut et al., 2002), successful assistance requires both *situation awareness*—ongoing awareness of what the worker is doing, the status of the task, and the environment (Endsley, 1995) and *conversational grounding*—working with the addressee to ensure messages are understood as intended (Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986).

When they are physically co-present—located at the same place at the same time—collaborators share a rich visual space. They can monitor one another’s facial expressions, watch each other’s actions, and jointly observe objects within the environment. This shared visual space facilitates both situation awareness and conversational grounding (e.g., Daly-Jones et al., 1998). For example, helpers can identify when to provide the next instruction by observing that a worker has completed the previous step (e.g., Fussell et al., 2000), a nurse can identify the right time to provide a next medical instrument by observing the doctor’s task progress (Nardi et al., 1993), or an instructor can identify when to intervene by observing a student making an error.

In Table 1, we consider six sources of visual information in our collaborative robot construction task—participants' heads and faces, participants' bodies and actions, the robot under construction, task objects and tools, the instruction manual and work environment—along with some of their possible functions for collaboration. Our goal was to use eye-tracking to better understand the extent of helpers’ reliance on each of these visual resources as they instruct workers in building a toy robot. Although we hypothesized that helpers would look least at their partner’s faces, we had no specific hypotheses about their gaze toward other targets.

| <i>Visual Sources</i> | <i>Sample Functions</i> |
|-------------------------|---------------------------------------------------|
| Partner’s head/face | Monitor comprehension |
| Partner’s hands/actions | Observe if partner is ready for next step |
| Task parts and tools | Identify parts for next step |
| Task object | Monitor task status |
| Instruction manual | Compare task status and actions with instructions |
| Work area | Monitor events that might affect task progress |

Table 1. Visual resources in a collaborative robot construction task.

Method

Nineteen participants served as “helpers” in a robot construction task. They collaborated with a partner to build the head of a large toy robot (see Figure 1). Participants first built the head alone, using an instruction manual. Then, the worker (a

confederate of the experimenter) was brought into the room. Participants were told that their task was to instruct a novice worker on how to building the robot.

Figure 1 [Robot]

Participants' gaze was recorded using an ISCAN head-mounted camera with eye-tracking. The system included an IBM-compatible computer with a 16 inch monitor and three 9 inch monitors used to calibrate the eye tracker (Figures 2 and 3). The output from the eye-tracker was recorded on a Panasonic DV-VCR. Wireless microphones recorded pairs' conversations. The system can record data with frame rate of 60 Hz, and precision is greater than one degree.

Figures 2 and 3 [ISCAN & Helper Set up]

Sessions took approximately 30 minutes and were taped and transcribed. Gaze was coded using an in-house system. Coders pressed a key for the onset of gaze towards each of 7 targets: the instruction manual, the robot head under construction, robot pieces, worker's hands, worker's head, other targets and uncodable. The software generated onset and offset times and total gaze duration for each glance.

Results and Discussion

The results suggest the utility of eye tracking for studying collaborative physical tasks. (see Fussell et al., 2003, for statistics and further details). First, we found significant differences between targets in terms of the number of glances directed toward those targets as helpers provided their instructions (Figure 4). Significantly fewer gazes were directed at the worker's face than at the other possible targets, and the most gaze was directed at the robot being constructed, the robot parts, and the worker's hands. Second, glance duration varied by target (Figure 5). Again there was an overall significant effect of target, such that glances to the face lasted for significantly less time than other gaze orientations, but there were no other significant differences.

Figures 4 & 5 [no. of glances, glance dur.]

Our findings suggest that when providing instructions for a collaborative construction task, helpers look predominantly at targets relevant to gathering information about the steps to be completed (manual) and task status (robot, robot parts, and worker's hands), rather than at the worker's face. Note that these results agree with those from the studies of gaze during solo physical tasks reviewed above, but are quite different from those typically found in studies of gaze during narrative conversations, in which the preponderance of gaze is usually directed towards others' faces (e.g., Argyle & Dean, 1976). We are currently coding helper and worker utterances during the task, in order to understand better how gaze is temporally related to messages of assistance. We are also identifying patterns of gaze, such as those identified by Ballard et al. (1995) for their block construction task. Once identified, we can use these patterns as baselines for comparing collaboration under different conditions (e.g., novice vs. expert helper; different media conditions).

Although the current study is only a preliminary step in understanding gaze during collaborative physical tasks, it demonstrates the potential of this technique for understanding medical teamwork. The same strategy can be used to examine, for instance, how nurses monitor doctors' behaviors and patient status in order to anticipate doctor's requests (Nardi et al., 1993). One could also use eye-tracking to assess how experts monitor equipment (Seagull et al., 1999), or to assess how differences in physician eye movements are related to task success (as was done by Kasarskis et al. [2001] for pilot expertise and success of landing).

Eye-tracking as Independent Measure

A second, less common, way of using eye-tracking in studies of collaborative physical tasks is as an independent variable—to assess the value of gaze information for partners in a collaboration. When people interact face-to-face, they can see where their partners are looking, and this visual information facilitates communication and enhances task awareness. For example, a nurse might anticipate what implement a surgeon needs next by observing that the surgeon is looking at a specific point on the patient (Nardi et al., 1993). It is difficult to quantify the effects of seeing partners' gaze in a face-to-face

setting, however, because there is not way to tease apart the contributions of gaze awareness from the other visual information in the environment. That is, we cannot manipulate the presence or absence of gaze awareness when partners are co-located.

Eye-tracking in combination with video conferencing provides a technique with which we can assess the value of gaze awareness. Instead of studying co-located teams, we can ask physically distributed teams to perform a task using one of several video systems. By manipulating the presence or absence of gaze awareness in the video system, we can begin to understand the role that gaze awareness plays in face-to-face collaboration on physical tasks. For example, Brennan and Lockridge (in preparation) compared performance on a cooperative task in which remote helpers had either coarse information about their partner's visual attention (the output of a head-mounted camera), precise information about the partner's visual attention (the eye gaze cursor superimposed over the output of the head-mounted camera), or no information about the partner's visual attention. Their results showed how helpers use gaze awareness information provided by the eye gaze cursor to better time their assistance.

An Example: Value of gaze information for collaboration on physical tasks

In a recent study (Fussell et al., 2003), we investigated the value of providing remote helpers with gaze awareness information in the context of a robot construction study. As discussed earlier, co-located collaborators can focus their gaze on a range of targets including fellow participants' faces, others' bodies and actions, task objects, and other entities in the work environment. Our goal was to assess the extent to which being able to see others' gaze benefited collaboration.

To address this question, we compared collaboration using a head-worn video camera with built-in eye-tracking, which provided gaze awareness, with that using a stationary scene camera that showed a wider but less detailed view of task objects and work environment. We hypothesized that both video systems would increase communication efficiency, improve helpers' ability to provide assistance, and improve performance over an audio-only link, and that combining the two cameras would improve communication and performance over either camera alone, by providing a complementary set of visual

cues. However, because all three video configurations provide less visual information than actual being in the same place, we hypothesized that pairs working side-by-side would perform best.

Method

Thirty-eight pairs of participants performed five robot construction tasks (e.g., building the left ankle or head of the robot; see Figure 1 above). One participant, the “worker”, performed the tasks with the assistance of his/her partner, the “helper”. Pairs performed one task in each of five media conditions: side-by-side, head-mounted camera with eye-tracking, scene camera, head-mounted and scene cameras, and audio-only. Tasks and media conditions were counterbalanced over participants.

As they built the robot, workers wore the ISCAN head-mounted camera and eye-tracker described previously. In addition, a video camera, positioned 5 feet behind and to the right of the worker, was used as the scene camera. It showed a view of the worker’s hands, the robot parts, and the part of the robot being completed. In the remote conditions, helpers were provided with a computer monitor. The manual was displayed on the left side of the monitor, and small (3 X 4 inch) windows to display output from the head- and scene cameras appeared on the right (Figure 6).

Figure 6 [Helper Monitor]

Participants completed a post task survey after performing each of the five tasks, in which they provided their subjective reactions about the success of each collaboration. In addition, their conversations were taped and transcribed.

Results And Discussion

Because our focus is on demonstrating the value of using eye-tracking as a way to study the role of gaze awareness in interaction, we focus here on only a subset of our findings (but see Fussell et al. [2003] for details and statistical analyses).

Consistent with previous studies (Fussell et al., 2000; Kraut et al., 1996; Kraut et al., in press) performance was significantly faster in the side-by-side condition (see Figure 7). Performance with the scene camera was faster than with audio-only, but performance

with the head-camera that provided gaze awareness information was not. Surprisingly, performance with both cameras together was not as good as performance with the scene camera alone, and did not differ significantly from performance in the audio-only condition.

Figure 7 [Completion Times]

Analysis of pairs' messages suggested that participants found the visual cues from the scene camera to be more valuable than those from the head-mounted camera plus eye tracking (see Figure 8). Both helpers and workers exchanged significantly fewer messages in the side-by-side condition than in all other conditions. Workers used significantly fewer words with the scene and scene+head cameras than with the head-mounted camera. Consistent with this, on the post-task surveys remote partners rated their ability to help workers significantly higher in the side-by-side condition than in all other conditions, and higher in the scene and scene+head-mounted camera conditions than in the audio-only and head-mounted camera conditions.

Figure 8 [Mean Words per Trial]

On the whole, the results show clear value for the scene camera, with its wider but less detailed view of the work area, but no benefit from the head-mounted camera with eye-tracking capability. These results could be interpreted to mean that awareness of partners' gaze is less useful to partners than we hypothesized. Alternatively, problems with the instantiation of gaze information in this study may have reduced the information's utility. For example, as workers move around, the eye-tracker may slip such that it is no longer properly calibrated. We discuss these and other limitations of eye-tracking technology in further detail in the next section.

We believe that the results of this study demonstrate the usefulness of eye-tracking methodology for understanding how collaborators on physical tasks use gaze awareness to coordinate their behavior. The same technique could be applied in a variety of medical settings. For example, one could compare the value of head-worn cameras with eye

tracking versus stationary scene cameras for telesurgery in order to understand how expert surgeons' awareness of others' gaze helps them coordinate complex procedures.

Eye-Tracking as a Research Tool: Issues and Limitations

The studies presented above and those we have reviewed suggest the value of eye-tracking as a method for investigating collaborative physical tasks in medical and other domains. Eye-tracking can be used to evaluate participants' attention as they collaborate on a task, or as a way to study the importance of gaze awareness during team collaboration. Despite the potential value of eye-tracking methodology, however, there are a number of issues and limitations incorporating the data into studies of collaborative physical tasks. In this section, we discuss several of these issues and limitations.

Limitations

The eye tracker typically cannot be calibrated correctly for a sizeable proportion of participants (up to 20%, in some studies, cf. Jacob & Kam, in press). Furthermore, the head-mounted device may slip over the course of a task, requiring recalibration to avoid data loss. This creates problems in collecting high-quality data. This is especially acute for longer studies and those requiring considerable movement among participants—as is true for most studies of collaborative physical tasks. Improvements in the design of head-worn eye tracking systems may minimize or eliminate these problems in the fairly near future.

Gaze data also requires considerable effort to process. Eye-tracking software has the ability to record the X and Y coordinates of the gaze points, enabling fairly automatic processing of stationary scenes; unfortunately, automatic processing is not possible when the scene is constantly changing. For our studies, we devised a software coding system that makes it easy to code which of a small set of targets is being fixated upon. However, manual coding could quickly become unwieldy in a setting with many, many possible targets (e.g., an OR with all the associated personnel, equipment, instruments, and the

like). New tools to help investigators code gaze in such settings would strongly benefit this area of research.

Theoretical Issues

Perhaps more problematic than current technological limitations to eye-tracking data collection are theoretical issues concerning the use and interpretation of data.

One challenging issue is how best to aggregate data from the eye tracking sources. In the first study presented above, we averaged fixations across all participants in the helper role, without regard to the way in which they are providing instructions and without regard to the workers' own gaze. While this is a good starting point for understanding gaze in collaborative physical tasks, a closer look at the relations among gaze targets, speech and actions is needed to fully understand the processes of collaboration. The integration of eye-tracking data with speech, gestures, behavioral data, and other participants' gaze is a quite complex task for which no adequate software currently exists. Yet, it is necessary if we are to uncover patterns of gaze such as those reported by Ballard et al. (1995) for solo physical tasks.

Even if we are able to coordinate the various sources of data (gaze, speech, gestures, actions), it may still be difficult to come up with patterns characterizing a specific collaborative physical task, due to the large number of variables at play (e.g., different numbers of surgeons or nurses in the OR, different patient characteristics). One possibility is to follow the strategy of Sodhi et al. (in press) in their studies of driver distraction. Sodhi and colleagues start by investigating gaze patterns in absence of any distractions to create a basic model of user gaze within the driving context; then, they can systematically examine the effect of distractions on gaze patterns. Another possibility might be to link eye tracking to cognitive processes via cognitive modeling, such as has been done by Salvucci (1999, 2000), although to date these cognitive models have not incorporated the social-interactional factors at play in collaborative physical tasks.

A final theoretical issue of particular importance in mentoring collaborative physical tasks, and in medical applications of eye-tracking, is the relationship between gaze and performance. In a few cases, such as the studies of listeners' identification of target objects based on their description described earlier, time to fixate and pick up a target

object provides a clear cut performance measure. Some studies comparing expert vs. novice gaze patterns likewise use tasks with clear-cut performance measures (e.g., detection of known tumors in radiology tasks). But in the complex tasks we focus on in our own research, performance measures are too gross (e.g., total task time, number of errors) to allow us to ascertain how specific gaze patterns benefit or hurt interaction. The problem is compounded by the inconsistent linkage between eye fixation and focus of attention: a person can be fixated on a target while thinking of something entirely different (for example, eyes fixated on the road, but mind focused on a cell phone conversation). Substantial theoretical development is needed before eye-tracking measure can fruitfully predict task performance in collaborative physical tasks.

Conclusion

Eye-tracking, in combination with video recordings from head-worn cameras, has tremendous potential as a tool for studying collaborative physical tasks such as those found in education and medicine. Eye-tracking can function as both a dependent measure, to help us understand people's gaze and focus of attention as they coordinate their activities with task partners, and an independent measure, to help us assess the value of gaze information for remote collaboration on physical tasks. Although there are currently technical and theoretical limitations to using eye-tracking as a research tool, we anticipate that as the number of investigators using these techniques increases, we will be able to make substantial progress on these limitations. In addition, by studying gaze across a variety of collaborative physical tasks, we may identify basic patterns of visual behavior that are fundamental to coordinating speech and actions among multiple individuals. Studying basic gaze patterns may prove useful in understanding the errors that arise during task performance, predicting the effects of adding in new tools or personnel to the workplace, and/or devising better task training strategies.

Acknowledgements

This work was conducted with support from the National Science Foundation grants #9980013 and #0208903. We thank Katherine Bankson, Terry Chan, Marina Kobayashi,

Kathy Lau, Sheel Mohnot, Phillip Odenz, Elizabeth Parker, Salma Ting and Kristin Weinziert for running participants and preparing data, Asa Sherrill for writing the gaze coding software, Nasri Haijj for implementing the online surveys, and Tom Pope for technical assistance. We also thank Susan E. Brennan, Darren Gergle, Robert Kraut, Jane Siegel, Yan Xiao, Jie Yang and several anonymous reviewers for their valuable comments on various stages of this research.

References

- Anders, G. (2001). Pilot's attention allocation during approach and landing—eye- and head-tracking research in an A330 Full Flight Simulator. In *International Symposium on Aviation Psychology (ISAP)*.
- Argyle, M., & Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.
- Ballard, D. H., Hayhoe, M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7, 66-80.
- Brennan, S. E., & Lockridge, C. (in preparation). *Monitoring an addressee's visual attention: Effects on referring in conversation*. Manuscript in preparation, SUNY Stony Brook.
- Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2002). Reference resolution in the wild: On-line circumscription of referential domains in a natural, interactive problem-solving task. *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*.
- Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, 47, 30-49.
- Clark, H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39
- Clark, H. H. & Marshall, C. E. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.
- Dabbish, L., & Kraut, R. (in progress). *Managing attention: Awareness displays and interruption over electronic communication*. Human-Computer Interaction Institute, Carnegie Mellon University.
- Daly-Jones, O., Monk, A. & Watts, L. (1998). Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies*, 49, 21-58.
- Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, and Computers*,
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language processing in natural contexts. *Journal of Psycholinguistic Research*, 24, 409-436.
- Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- Fairchild, M. D., Johnson, G. M., Babcock, J., & Pelz, J. B. (2001). *Is your eye on the ball?: Eye tracking golfers while putting*. Unpublished manuscript, Rochester Institute of Technology.
- Ford, C. E. (1999). Collaborative construction of task activity: Coordinating multiple resources in a high school physics lab. *Research on Language and Social Interaction*, 32, , 369-408.

- Fussell, S. R., Setlock, L. S., & Parker, E. M. (2003). Where do helpers look? Gaze targets during collaborative physical tasks. *CHI 03 Extended Abstracts*.
- Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of CSCW 2000* (pp. 21-30). NY: ACM Press.
- Fussell, S. R., Setlock, L. D., & Kraut, R. E. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of CHI 2003*.
- Goodwin, C. (1996). Professional vision. *American Anthropologist*, 96, 606-633.
- Gullberg, M. (2003). Eye movements and gestures in human face-to-face interaction. In J. Hyona, R. Radach, & H. Deubel, (Eds.) *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements* (pp. 685-703). Oxford: Elsevier Science.
- Gullberg, M., & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics and Cognition*, 7, 35-63.
- Hanna, J. E., Tanenhaus, M. K., Trueswell, J. C., & Novick, J. M. (under review). *Linguistic form, common ground, and referential interpretation: The effects of linguistic form and common ground on domains of referential interpretation*. Unpublished Manuscript, University of Rochester.
- Hyona, J., Radach, R. & Deubel, H. (Eds.) (in press). *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Oxford: Elsevier Science.
- Jacob, J. K., & Karn, K. S. (in press). Commentary on Section 4. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R Radach, & H. Deubel (Eds.) *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Oxford: Elsevier Science.
- Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *The Journal of Neuroscience*, 21, 6917-6932.
- Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A. & Wickens, C. (2001). Comparison of expert and novice scan behaviors during VFR flight. Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH, 2001.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32-38.
- Kraut, R. E., Fussell, S. R., Brennan, S., & Siegel, J. (2002). A framework for understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. P. Hinds & S. Kiesler (Eds.), *Technology and Distributed Work*. Cambridge, MA: MIT Press.
- Kraut, R. E., Miller, M. D., & Siegel, J. (1996) Collaboration in performance of physical tasks: Effects on outcomes and communication, *Proceedings of CSCW'96* (57-66). NY: ACM Press.
- Krupinski, E. A., & Roehrig, H. (2002). Pulmonary nodule detection and visual search: P45 and P104 monochrome versus color monitor displays. *Academic Radiology*, 9, 638-645.
- Land, M., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research*, 41, 3559-3565.
- Land, M., & Horwood, J. (1995). Which parts of the road guide steering. *Nature*, 377, 339-340.
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28, 1307-1432.
- Mello-Thoms, C., Dunn, S., Nodine, C. F., Kundel, H. L., & Weinstein, S. P. (2002). The perception of breast cancer: What differentiates missed from reported cancers in mammography? *Academic Radiology*, 9, 1004-1012.

- Metzing, C., & Brennan, S. E. (under review). *When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions*. Unpublished manuscript, SUNY Stony Brook.
- Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S. & Sciabassi, R. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. *Proceedings of Interchi '93* (327-334). NY: ACM Press.
- Oudejans, R. R. D., Michaels, C. F., Bakker, F. C., & Davids, K. (1999). Shedding some light on catching in the dark: Perceptual mechanisms for catching fly balls. *Journal of Experimental Psychology: human Perception and Performance*, 25, 531-542.
- Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, 41, 3587-3596.
- Pelz, J. B., Canosa, R., & Babcock, J. (2000). Extended tasks elicit complex eye movement patterns. *Proceedings of Eye Tracking Research Applications Symposium (ETRA) 2000*. NY: ACM Press.
- Pelz, J. B., Canosa, R., Kucharczyk, D., Babcock, J., Silver, A., & Konno, D. (2000) Portable eyetracking: A study of natural eye movements. *Proceedings of SPIE, Human Vision & Electronic Imaging*, San Jose, CA, 2000.
- Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139, 266-277.
- Pomplun, M., Prestin, E., & Rieser, H. (1998). Eye-movement research and dialogue structure. Report 98/12. Situierte Künstliche Kommunikatoren SFB 360. Universität Bielefeld.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.
- Salvucci, D. (1999). Mapping eye movements to cognitive processes. Doctoral dissertation, Carnegie Mellon University 1999). UMI Dissertation Express, 9950036.
- Salvucci, D. (2000). A model of eye movements and visual attention. In *Proceedings of the International Conference on Cognitive Modeling* (pp. 252-259). Veenendaal, The Netherlands: University Press.
- Seagull, F. J., Xiao, Y., MacKenzie, C. F., Jaber, M., & Dutton, R. P. (1999). Monitoring behavior: A pilot study using an ambulatory eye-tracker in surgical operating rooms. *Proceedings of 1999 Human Factors and Ergonomics Society Annual Meeting*.
- Smeets, J. B. J., Hayhoe, M., & Ballard, D. H. (1996). Goal-directed arm movements change eye-head coordination. *Experimental Brain Research*, 109, 434-440.
- Sodhi, J., Reimer, B., & Llamazares, I. (in press). Glance analysis of driver eye movements to evaluate distraction. *Behavior Research Methods, Instruments, & Computers*.
- Sodhi, M., Reimer, B., Cohen, J. L., Vastenburg, E., Kaars, R., & Kirschenbaum, S. (2002) On-road driver eye movement tracking using head-mounted devices. *Proceedings of ETRA 2002* (pp. 61-68). NY: ACM Press.
- Stiefelhagen, R., & Zhu, J. (2002). Head orientation and gaze direction in meetings. In *CHI 2002 Extended Abstracts* (pp. 858-859). NY: ACM Press.
- Tang, J. C. (1991). Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, 34, 143-160.
- Tchalenko, J., Dempere-Marco, L., Hu, X.P., & Yang, G. Z. (2001). Quantitative analysis of eye control in surgical skill assessment. 11th European Conference on Eye Movements, August 22-25, 2001. Turku, Finland.
- Velichkovsky, B. M., Dornhoefer, S. M., Pannasch, S., & Unema, P. J. A. (2000). Visual fixations and level of attentional processing. *Proceedings of ETRA 2000* (p. 79-85). NY: ACM Press.
- Vertegaal, R., Slagter, R., van der Veer, G., & Nijholt, A. (2001). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. *Proceedings of CHI 2001* (pp. 301-308). NY: ACM Press.

Vickers, J. N. (1996). Visual control when aiming at a far target. *Journal of Experimental Psychology: human Perception and Performance*, 22, 342-354.



Figure 1. The large toy robot used in our experiments.



Figure 2. ISCAN eye tracker set up, with computer running tracking software (right) and three small monitors (left). The left small monitor shows the scene; the right small monitor shows the participants' eye with tracking information overlaid.



Figure 3. Set up for Helper Gaze Study. Left: Helper in back, wearing eye tracker; worker in foreground, working on the robot. Right: View from the Helper's head-worn camera, with + overlaid to show focus of attention.

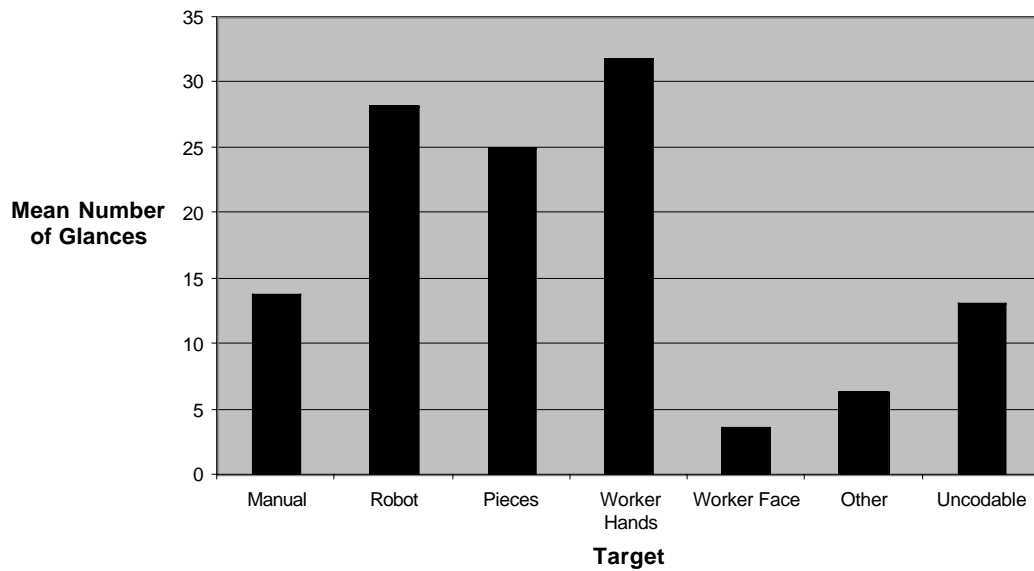


Figure 4. Mean number of glances by target.

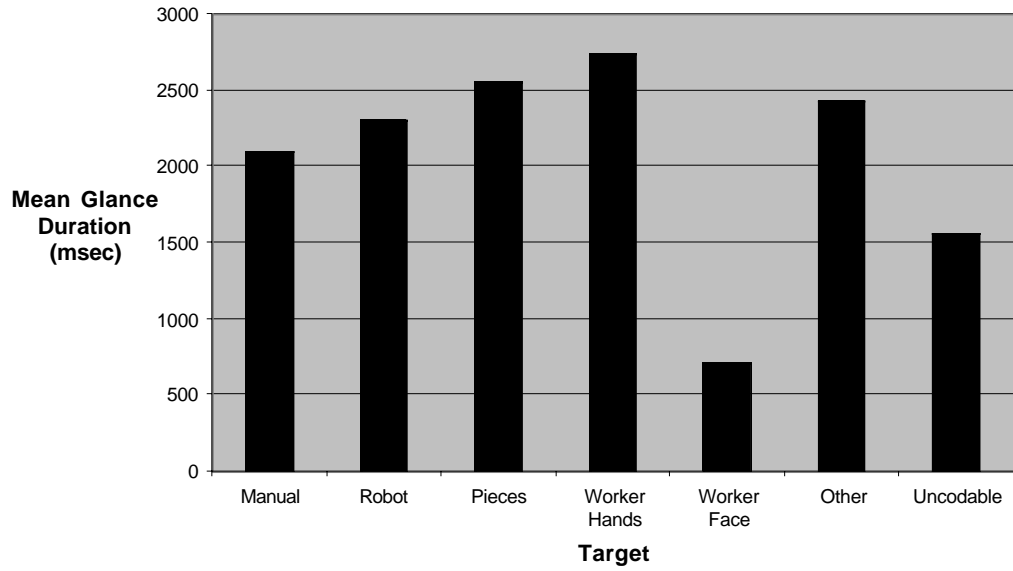


Figure 5. Mean glance duration as a function of target.

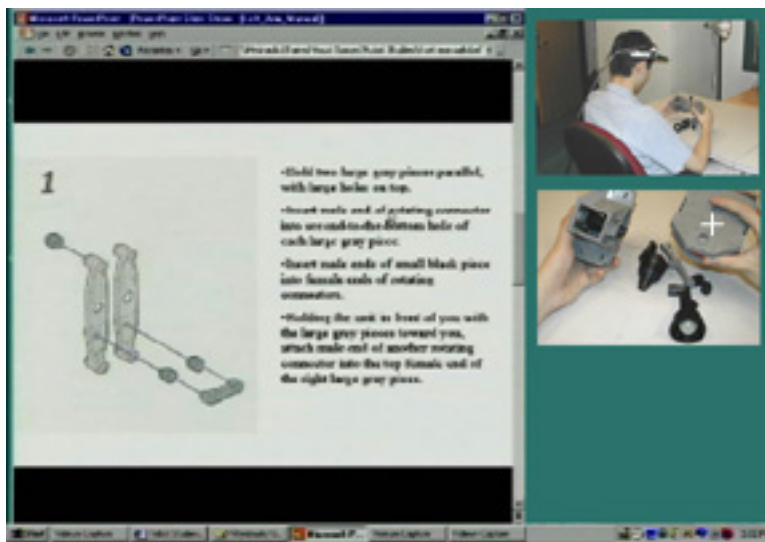


Figure 6. Helper monitor with online manual (left) and views from scene camera (top right) and head-mounted camera (center right, “+” shows focus of attention within the view).

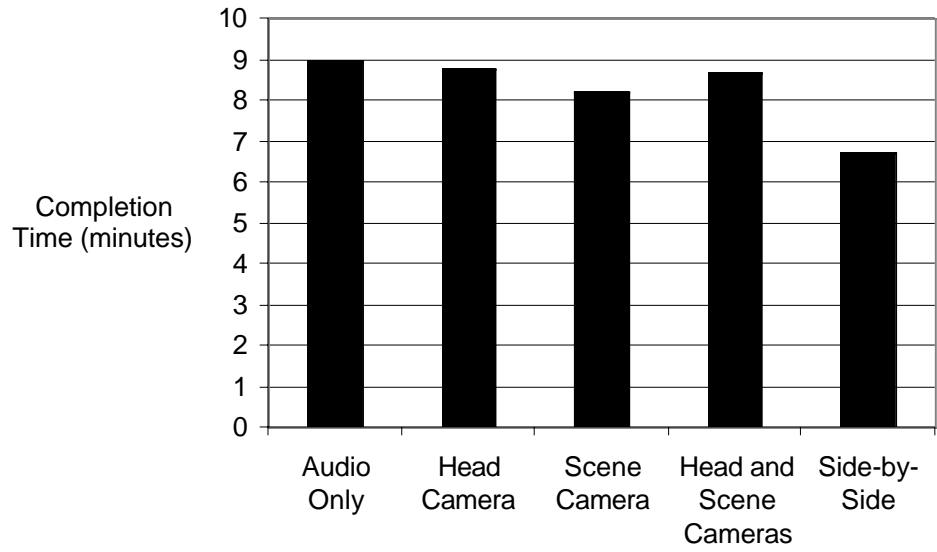


Figure 7. Completion times by media condition (mins.).

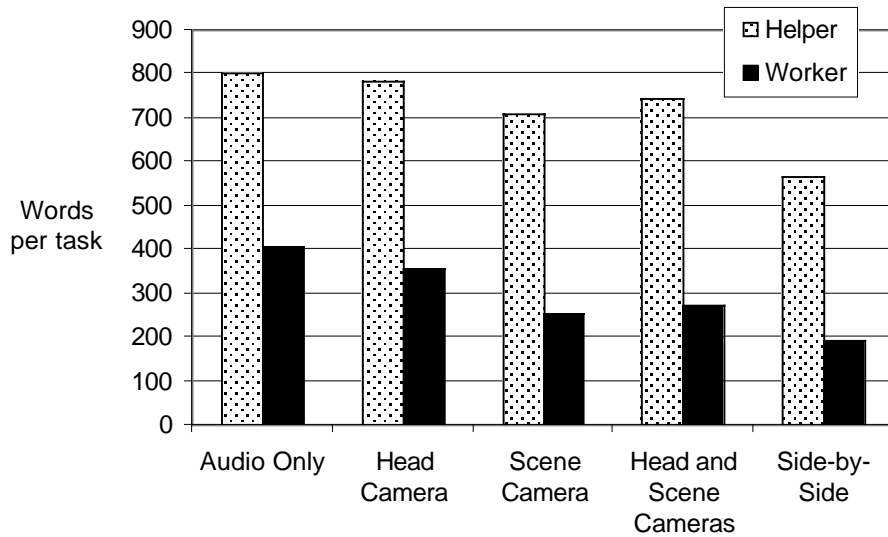


Figure 8. Mean number of words per trial by experimental role and media condition.