Where do Helpers Look? Gaze Targets During Collaborative Physical Tasks

Susan R. Fussell, Leslie D. Setlock, Elizabeth M. Parker
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
+1 412-268-4003
susan.fussell@cmu.edu

ABSTRACT

This study used eye-tracking technology to assess where helpers look as they are providing assistance to a worker during collaborative physical tasks. Gaze direction was coded into one of six categories: partner's head, partner's hands, task parts and tools, the completed task, and instruction manual. Results indicated that helpers rarely gazed at their partners' faces, but distributed gaze fairly evenly across the other targets. The results have implications for the design of video systems to support collaborative physical tasks.

Keywords

Computer-supported collaborative work, video conferencing, eye gaze, empirical studies

INTRODUCTION

In this paper we examine people's use of visual information as they assist their partners during collaborative physical tasks—tasks in which two or more individuals work together to perform actions on concrete objects in the three-dimensional world. For example, an expert might guide a worker's performance of repairs or a doctor might direct a medical team. Because expertise is increasingly distributed, there is growing demand for technologies to support remote collaboration on physical tasks.

Collaboration on physical tasks requires extensive coordination: Helpers must determine when assistance is needed, how to phase their messages, and whether their partner has understood the instructions. When they are physically co-present—located at the same place at the same time—collaborators share a rich visual space that can help them achieve coordination. For example, a helper can identify when to provide the next instruction by observing that the worker has completed the previous step. Figure 1 lists six visual resources for physical tasks, along with some of their possible functions for collaboration.

CHI2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA.

When working at a distance, partners must rely on technologies such as video conferencing that limit visual resources. Bandwidth limitations and other factors make it impractical to support all sources of visual information; hence, most video systems provide only one or two. For example, traditional video conferencing systems show views of others' faces, head-mounted systems show views of workers' hands [e.g., 1, 2, 4], and scene-oriented systems show wider views of the work area [e.g., 2].

Visual Sources	*Sample Functions*
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

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

To date, decisions about what features to include in video systems have typically been theoretically- or technologically-motivated [e.g., 3, 5]. In the current study, we use eye-tracking technology to determine empirically the relative importance of each visual resource. Although we hypothesized that helpers would look least at their partner's faces, we had no specific hypotheses about their gaze toward other targets. Our goal was to better understand which visual resources need to be supported in video systems for collaborative physical tasks.

METHOD

Equipment and Materials

The Robotix Vox Centurion robot kit was used as the basis for the task. Pairs collaborated to complete the head of the robot (Figure 2). An instruction manual outlining the steps to be completed was created in PowerPoint.

Figure 2. Robot head built by partners.

An ISCAN head-mounted camera with eye-tracking was used to record helpers' gaze. The output from the eye-tracker was recorded on a Panasonic DV-VCR. Wireless microphones were used to record pairs' conversations.

Participants and Procedure

Nineteen participants (12 male, 7 female) served as "helpers" in a robot construction task. Participants first built the robot head alone, using the instruction manual. Then, the confederate worker was brought into the room. Participants were told that their task in this phase of the study was to provide instructions to a novice worker as he or she built the robot. The helper put on the head-mounted camera and the experimenter calibrated the eye-tracker. The pairs then performed the robot task, with the confederates following a preset script, which indicated how they should respond to each instruction. Sessions took approximately 30 minutes and were taped and transcribed

Gaze direction 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 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

Figure 3 shows the mean number of glances directed at each target. Substantially less gaze was oriented toward the worker's face than toward the other targets. Overall, there was a highly significant effect of target ($F [6, 102] = 13.97$, $p < .0001$). Post-hoc tests indicated that the robot, pieces, and worker's hands were glanced at significantly more often than all other targets ($p < .005$ or better) but did not themselves differ significantly. The manual was glanced at more often than the worker's face ($p < .02$).

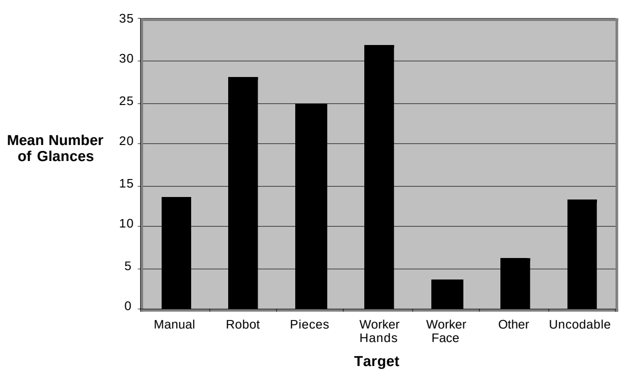

Figure 3. Mean number of glances by target.

Figure 4 shows the mean duration of each glance by target. Again there was an overall significant effect of target ($F [6, 102] = 5.88$, $p < .0001$). Post-hoc tests indicated that glances to the face lasted for significantly less time than other gaze orientations, but no other significant differences.

DISCUSSION AND FUTURE RESEARCH

Our findings suggest that remote collaborators on physical tasks would benefit most from video systems that include views of the object being constructed, task pieces and tools, and the worker's hands. Views of the worker's face and other aspects of the work environment appear to be much less important for this type of task.

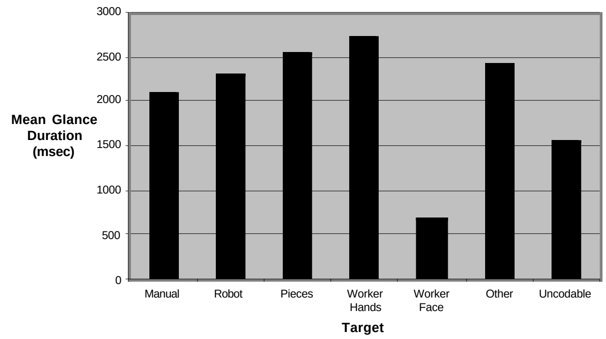

Figure 4. Mean glance duration as a function of target.

How video systems should be designed to provide all three primary sources of visual information requires further analysis. As noted earlier, it is difficult to provide multiple visual fields within one system due to bandwidth and other limitations. One possibility is to create a system that dynamically reorients its view depending upon task phase. We are currently examining relationships between helpers' speech and gaze to see if glances at particular targets can be matched with specific phases of work. A second possibility is to devise a video system that encompasses task objects, pieces/tools, and worker's hands simultaneously.

ACKNOWLEDGEMENTS

This study was conducted with support from the National Science Foundation grants 9980013 and 0208903. We thank Salma Ting for running experiments, Asa Sherrill for writing the coding software, Katherine Bankson for coding gaze, and Jie Yang for his technical and theoretical input.

REFERENCES

1. 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.
2. Fussell, S. R., Setlock, L. D., & Kraut, R. E. (in press). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of CHI 2003.*
3. 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.
4. 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.
5. Kuzuoka, H. (1992). Spatial workspace collaboration: A Sharedview video support system for remote collaboration capability. *Proceedings of CHI'92* (533-540). NY: ACM Press.