Assessing the Value of a Cursor Pointing Device for Remote Collaboration on Physical Tasks

Susan R. Fussell, Leslie D. Setlock, Elizabeth M. Parker, Jie Yang
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 assessed the value of a cursor pointer that allows remote collaborators to point to locations in a partner's workspace via a shared video feed. We compared performance with the cursor pointer with that in video-only and side-by-side conditions. Results indicated that participants found the cursor pointer of value for referring to objects and locations in the work environment, but that the pointer did not improve performance time over video-only. We conclude that cursor pointing is valuable for collaboration on physical tasks, but that additional gestural support will be required to make performance using video systems as good as performance working side-by-side.

Keywords
Computer-supported collaborative work, video conferencing, gesture, empirical studies

INTRODUCTION
In this paper, we consider a simple cursor tool as a device to allow remote pointing during collaboration on physical tasks—tasks in which two or more individuals work together to perform actions on concrete objects in the three-dimensional world. In the task we consider—a robot construction task—one party (the "worker") manipulates the parts and tools under the guidance of a remote expert (the "helper"). This scenario is similar to telemedical, distance learning, remote repair, and similar applications.

Conversations during collaborative physical tasks typically focus on the identification of target objects and locations, descriptions of actions to be performed on those targets, and confirmation that the actions have been performed successfully (Figure 1) [1, 2, 3]. When partners interact side-by-side, they can identify objects and locations easily using pointing and deictic expressions (e.g., "put that piece over there"). However, in most video systems, remote helpers are unable to point. Helpers often express frustration with this situation, saying, for example, "If I could just point to it, its right there," when struggling to identify objects for their partners.

Helper utterance	Task phase
Take the long purple piece with the six holes in it,	Object identification
Insert it at a 45 degree angle, clockwise	Procedural instructions
Do you see what I mean?	Comprehension monitoring

Figure 1. Sample task dialogue, from the helper's side.

Kuzuoko and colleagues have developed a series of elegant laser pointing devices [e.g., 4, 5]. In their GestureMan system, for example, a remote partner manipulates a robot with attached laser pointer that can be directed at any spot in the work space. However, such devices suffer from several limitations, including expense and availability, that make their widespread use unfeasible at present.

In the current study we take a simpler approach, based on a shared view of the workspace. A remote helper can point to objects and locations in this workspace and the output of the pointing is displayed on a monitor in front of the worker. This system has the benefits of being easy to use and easily implemented in most any video conferencing system. We hypothesized that this cursor pointer would improve performance over a video-only system.

METHOD
Participants
Pairs (N = 38) completed three robot assembly tasks, one in each of three media: side-by-side, video only, and video plus cursor pointer. One partner was randomly assigned to the "helper" role and the other to the "worker" role. Tasks, trials, and media conditions were counterbalanced.

Materials
The Robotix Vox Centurion robot kit (Figure 2) was used to create three tasks of equivalent difficulty (e.g., the robot's head). Instruction manuals outlining the steps to be completed were created for each task.

Figure 2. Robot used in the experiments

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

Three sets of online surveys were created: (a) a pre-test survey collected basic demographic information; (b) post-task surveys elicited participants' reactions to each media condition, and (c) a post experimental survey asked participants to assess the value of different video features.

Equipment

A video camera was positioned 5 feet behind and to the right of the worker and showed a view of the worker's hands, robot pieces, and part of the robot being completed. Helpers saw this view in the upper right of their computer screen; workers saw the view on a large monitor in front of their workspace. In the pointing condition, helpers could press a mouse button to create a pink circle on the video feed. The pink circle appeared in the same location on the workers' monitor (Figure 3).

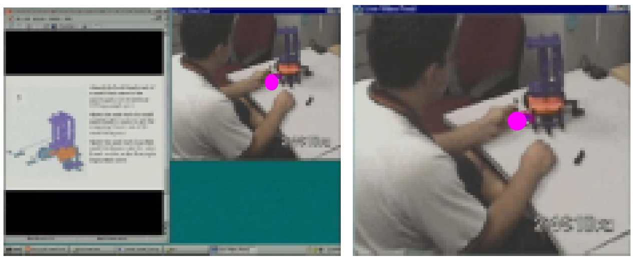

Figure 3. Cursor pointer from the helper's view (left) and the worker's view (right).

Procedure

Participants completed consent forms and pretests. They were then shown the robot and the communications technologies. The Helper was further instructed on use of the cursor pointing device. Pairs then began their series of 3 trials. Participants were told what technology would be available to the helper prior to each trial. Upon completion of the task, or after 10 minutes, the work was halted and participants completed post-task questionnaires. After all tasks were done, they completed the final questionnaire.

RESULTS

Survey and performance results were analyzed in 3 (task) by 3 (trial) by 3 (media condition) repeated measures analyses of variance (ANOVAs). We focus first on participants' ratings of how easily they could identify referents in each condition. For Helpers, there were significant main effects for task ($F [2, 87] = 4.78$, $p =.01$) and media condition ($F [2, 87] = 33.13$, $p < .0001$). Similarly, for Workers there were main effects for task ($F [2, 66] = 10.58$, $p <.0001$) and for media condition ($F [2, 66] = 16.57$, $p < .0001$). Both helpers and workers found it easiest to identify referents in the side-by-side condition, and easier in the video + cursor condition than in the video only condition (Figure 4). Post-hoc tests showed that all differences between conditions were significant at $p <.005$.

Analysis of performance times indicated main effects of trial ($F [2, 70] = 11.73$, $p < .0001$), task ($F [2, 70] = 27.73$, $p < .0001$) and media ($F [2, 70] = 22.92$, $p < .0001$). Performance was fastest in the side-by-side condition ($M = 6.47$ minutes), but there was no difference between the two video conditions ($M = 8.44$ and 8.49 minutes, with and without cursor respectively).

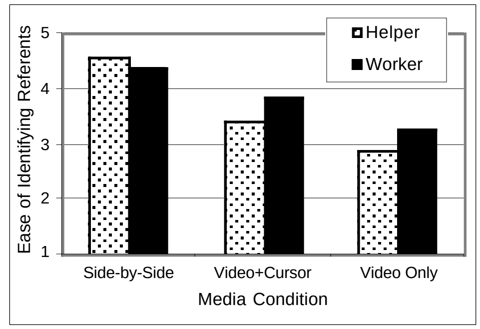

Figure 4. Ease of referent identification by participant role and media condition

DISCUSSION AND FUTURE RESEARCH

Participants found value in our cursor pointing device—it enabled them to refer more efficiently to task objects and locations than they could with video alone. However, the pointer was not sufficient to improve performance times over the video-only condition. In Figure 1 we showed that helping activity typically occurs in three phrases: identifying objects/locations, explaining procedures, and monitoring comprehension. The cursor pointer primarily benefits the identification phase, while procedural explanations take up the bulk of performance time. We are currently designing and testing systems that allow greater use of gesture during the procedural phase (e.g., to demonstrate actions), which we hypothesize will provide further benefits for remote collaboration on physical tasks.

ACKNOWLEDGEMENTS

This study was funded by National Science Foundation Grants 9980013 and 0208903. We thank William Kunz for developing the pointing software, Katherine Bankson, Marina Kobayashi, and Kathy Lau for running participants and preparing data, and Susan E. Brennan, Darren Gergle, and Robert E. Kraut for their valuable comments.

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., Miller, M. D., & Siegel, J. (1996) Collaboration in performance of physical tasks: Effects on outcomes and communication, *Proceedings of CSCW'96* (pp. 57-66). NY: ACM Press.
4. Kuzuoka, H., Kosuge, T., & Tanaka, K.. (1994) GestureCam: A video communication system for sympathetic remote collaboration, *Proceedings of CSCW 94* (pp. 35-43). NY: ACM Press.
5. Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., Mitsuishi, M. (2000). GestureMan: A mobile robot that embodies a remote instructor's actions. *Proceedings of CSCW 2000* (pp. 155-162). NY: ACM Press