Research Article

A Study of Gestures in a Video-Mediated Collaborative Assembly Task

Leila Alem and Jane Li

CSIRO ICT Centre, P.O. Box 76, Epping, NSW 1710, Australia

Correspondence should be addressed to Leila Alem, leila.alem@csiro.au

This paper presents the results of an experimental investigation of two gesture representations (overlaying hands and cursor pointer) in a video-mediated scenario—remote collaboration on physical task. Our study assessed the relative value of the two gesture representations with respect to their effectiveness in task performance, user's satisfaction, and user's perceived quality of collaboration in terms of the coordination and interaction with the remote partner. Our results show no clear difference between these two gesture representations in the effectiveness and user satisfaction. However, when considering the perceived quality of collaboration, the overlaying hands condition was statistically significantly higher than the pointer cursor condition. Our results seem to suggest that the value of a more expressive gesture representation is not so much a gain in performance but rather a gain in user's experience, more specifically in user's perceived quality of collaborative effort.

1. Introduction

Collaboration with remotely located participants represents today's working situations, where necessary resources (materials or expertise) may not always be on site for solving the task at hand. Examples include maintenance operations, when an expert remotely guides a worker repairing a machine, and telemedicine when a specialist remotely leads a team to manage the care of a patient [1].

Numerous studies have explored remote collaboration on physical tasks in which an individual guides another over distance to perform a certain physical task [2–5] or collaborations involving small groups [6]. These studies focus on different aspects of remote collaboration, including investigating perceptual factors such as the person's gaze [3, 4] or comparing different representations of gesture such as overlaying hands and cursor pointer [5, 7, 8]. Additionally, these studies have discussed and demonstrated the importance of gesture in collaborative physical tasks.

Studies in this area have recognized the complexity of coordination processes and the need for representing nonverbal dialogue when the interaction is distributed. It has been acknowledged that in copresent communication a physically shared workspace allows gesture production from any individual and this leads to an effective interaction. The sharing of a task space, remotely or otherwise, is critical to activities like determining the partner's need for assistance, instruction efficiency and feedback [4, 9].

In our study we intended to investigate the relative effect of the two most popular gesture representations, overlaying hands and cursor pointer, on the collaboration in physical tasks. The study is performed within a video-mediated condition by using video-conferencing technology.

2. Related Work

In the past ten years, there has been an interest in HCI and CSCW research in studying collaboration on physical task using representations of gestures such as pointing, sketching and hand [2]. A variety of systems have been developed to facilitate remote gesturing (DOVE [2, 5], GestureMan [10, 11], "MixedEcology" [7, 8, 12]. Most of these systems were built with the intention of enabling a helper (or expert) to guide the actions of a worker located at some remote worksite. Results have suggested that using gesture representation could increase performance speed in these situations. Several research groups have explored this issue from different perspectives.

The DOVE system used the tool of "drawing over video" which overlayed a computer pointer or sketches on a video representation of the worker's task space and displayed the mixed image onto the worker's monitor [13]. Evaluation experiments of DOVE have shown the superiority of the digital sketches over cursor pointer in improving performance [2].

Another approach from a "mixed ecologies" perspective was to project the representation of gesture into the worker's task spaces [7, 8]. The GestureMan system [10] and the Wearable Active Camera/Laser (WACL) system [14] combined a laser pointer and a camera on a mobile basis with a helper controlled laser pointer pointing directly on the helper's task space. User studies have proved the feasibility of this approach [11]. Kirk and Fraser [7] demonstrated that gesturing with an unmediated representation of the hands led to an improved performance over the mediated representation (such as sketch). The authors also made the claim that the utility of remote gesturing systems was beyond pointing devices and gesturing systems were used at early stages of interaction to affect patterns of dialogue and prevent interruptions [12].

Although task performance was the major measurement in the studies of remote guidance, other interaction issues have also been explored in this scenario. Coordination, efficiency of communication [2] and learning aspects [8] have been investigated using questionnaire. Kirk et al. in [8] reported a poorer perception of involvement of a helper. They found that the helper was less involved in determining the manipulations being undertaken and less rapport emerged between the helper and worker during instructions.

Fussell et al. [2] reported a series of studies in this space. One of the studies compared video only with video and cursor and reported no difference in task performance. Another study compared video only and video with an overlay gesture tool and reported an improved performance in the gesture overlay condition. Kirk and Fraser compared a gesture drawing to a hand overlay system and reported that the pairs using the hand overlay completed more of the task in the allotted time period, but did not find any improvement in task accuracy [7].

3. Research Question

Drawing on Kirk et al. and Fussell et al.'s findings, our aim was to investigate how overlay hands and a cursor pointer affect aspects of collaboration. The following research question formed the core of the study:

> what is the relative effect of different gesture representations on task performance, satisfaction and perceived quality in a collaborative effort? More specifically, what are the factors that influence the process as well as the product of the collaboration.

With this study we intended to extend Kirk et al. and Fussell et al.'s work by investigating the relative effectiveness and other assessments such as user satisfaction and perception

of the collaboration quality. We believed that there was room for additional evidence for the effectiveness of gesture representations, particular in comparing two-gesture devices within one single media condition.

We included process components in terms of perceptual factors in the analyses. When study participants qualified their experience in video-mediated interaction, they self-reported on the process of interaction/collaboration. Attempting to characterize the process of interaction is aligned with the research view in the area of computer supported collaborative learning (CSCL), that is, that the process of the collaboration, in addition to traditional outcome measures, should receive closer attention [15]. This study allowed us to have a more comprehensive understanding of the components influencing the outcome and the process of collaborative activities in physical tasks.

4. Experimental Design

We conducted a within-subject design to compare differences between the following two-gesture representation conditions with respect to effectiveness and user assessments. The gesture representation was systematically varied in two ways.

(i) Hands condition: the video of the helper's hands was transmitted into the shared workspace, enabling him/her to use his/her hands to guide the worker through the task.

(ii) Cursor pointer condition: the mouse activity, shown in the form of an arrow cursor, communicated gestures appearing in the shared workspace view in both the helper and worker interface.

A between-subject design assessed interpersonal factors for the participants involved in one condition. In each condition, the participants assembled a part of a LEGO toy. Each trial was timed and recorded on video.

4.1. Task Description. Assembly of a Lego kit is a common task in the literature [4, 5, 8]. According to Kirk and Fraser [7], this task incorporates generic elements such as selection, pattern matching, rotating, inserting, and attaching: allowing investigation of the demands placed on real-world applications.

Assembly in this study was completed on a Bionic Lego toy (Lego Bionicle Piraka Avak, see Figure 1). In each condition, participants were asked to collaboratively build two different body parts of the Lego toy (i.e., leg and body) by assembling 12 pieces of Lego toys in 11 steps.

An instruction manual was provided to the helper to guide the worker. The manual contained a description of the steps required to successfully build the Lego toy.

4.2. Technical Setup

Environment. Much of the technical setup was based on the work by Kirk et al. [8]. The physical environment in terms of hardware and setup was similar on both ends (see Figure 2). Each participant faced a standard desktop monitor

FIGURE 1: Lego tasks and final object (far right).

and had a mat (30.5 cm × 41.0 cm) on their desk which acted as the shared workspace. A camera was positioned directly above the mat with a field of view encompassing the entire mat. A video feed was distributed via a local network to the helper's computer, broadcasting a shared workspace onto both the helper's and worker's screens. The Virtual Tea Room technology [16] functioned as the technical platform for this experiment. This application uses digital video over IP to provide an extensible and flexible telecollaboration environment for simultaneous multisite conferencing.

While helpers had the manual visible within their interface, workers had the physical pieces in front of them. The physical workspace of the worker contained task pieces laid out on a mat divided into two sections: a pieces bay and an assembly area in which they manipulated the pieces with their hands.

Interface. In each condition, a shared view of the workspace was available whereby the pair could see the pieces of Lego toy and the hands of worker in a single top-down view (workspace). Additionally, a "talking head" view of the remote partner was available (Figure 3).

4.3. Measures

Task Performance. Task performance was captured by measuring the time taken (assembly time) as well as number of mistakes (assembly accuracy). The assembly accuracy was defined by comparing the final task object with the "master solution" or correct figure.

Perception of the Quality of Collaborative Effort. In order to assess the process while engaged in a collaborative effort, we assessed several components of the subject's perception of their collaborative effort. Dimensions of their perception of collaboration included: perception of the collaborative assembly task itself, perception of the participants engaged in collaboration (self and partner), and perception of the environment in which the collaboration took place (technical setup). These self-assessment items were captured after each condition and measured on a 5-point scale (range 0 to 4).

Satisfaction. Satisfaction is an index used to qualify the user's feeling of adequacy with a given situation. Typically, if the technology or tool is adequate, the user will be "satisfied". The overall satisfaction was captured after each condition for each participant by self-assessment on a 5-point scale (range 0 to 4).

Preference. Preferences are defined as the perceived desired choices of users based on their assessments of the gesture representations. Participants' preference for one-gesture condition was captured at the end of the trial.

Other Variables. Various other variables were collected: basic demographic information (i.e., age, gender), participants' familiarity with the task, the computer environment, and familiarity with their partner.

4.4. Procedure. The participants were grouped into randomly assigned pairs with one worker and one helper. In separate rooms, each participant was given an overview of the study, required to sign a consent form and complete an entry questionnaire. Helpers first constructed each part of the robot toy in order to familiarize themselves with task materials. Pairs then performed their tasks by building two objects in each condition via video-conferencing. The conditions were randomized and were not assigned to a specific assembly task. Participants completed a postcondition questionnaire at the end of each condition, and completed an exit questionnaire at the end of the session. Following this, the participants were debriefed and compensated. Sessions took approximately 60 to 75 minutes.

Altogether 21 trials were conducted. Each trial consisted of one helper and one worker (a total of 42 participants). Five trials were excluded from the analyses due to technical difficulties experienced during the assembly task. As a result, 16 trials were used for the analysis. The participants were on average 23 or 24 years old and the majority were male. The percentage of native and nonnative English speakers was balanced. Most subjects were university students, and a few worked full-time.

5. Results

Both descriptive and inferential statistics were used to analyze the data. *T*-tests for paired samples are used (two-tailed with 5% confidence level). The identification of relationships was conducted by using correlative analysis (Pearson's correlation coefficients, 5% confidence level).

Assembly Time. First, the two subtasks for each object were compared statistically within and across the conditions in order to identify differences in the time on tasks. The analyses indicated no differences between the subtasks within each condition and no differences across the conditions. As a result, the subtasks were merged and each task was compared across the two conditions.

The assembly time was compared for the conditions based on 16 trials. For the hands condition the average time on task was 8:06 minutes (standard deviation (SD) = 2:23 minutes). For the cursor pointer condition, the participants needed an average assembly time of 9:42 minutes (SD = 3:22 minutes), the tests statistics indicate no significant difference between the two although there was a tendency towards a faster task performance in the hands condition ($t(1, 15) = -1.824$, $P = .088$).
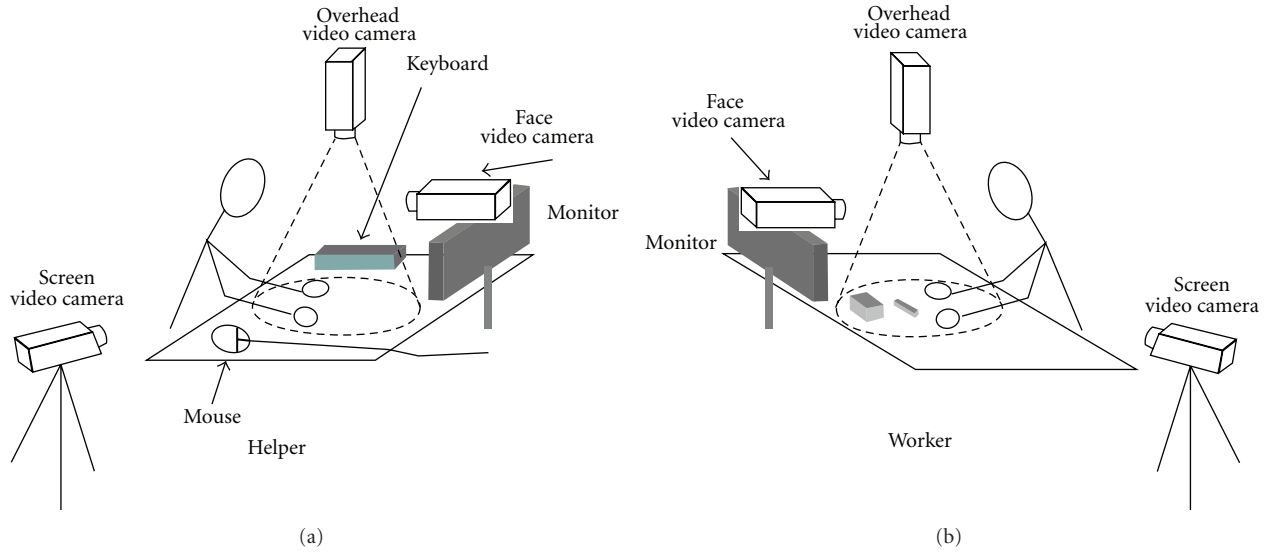
(a)

(b)

FIGURE 2: Technical setup (adapted from Kirk et al. [8]).
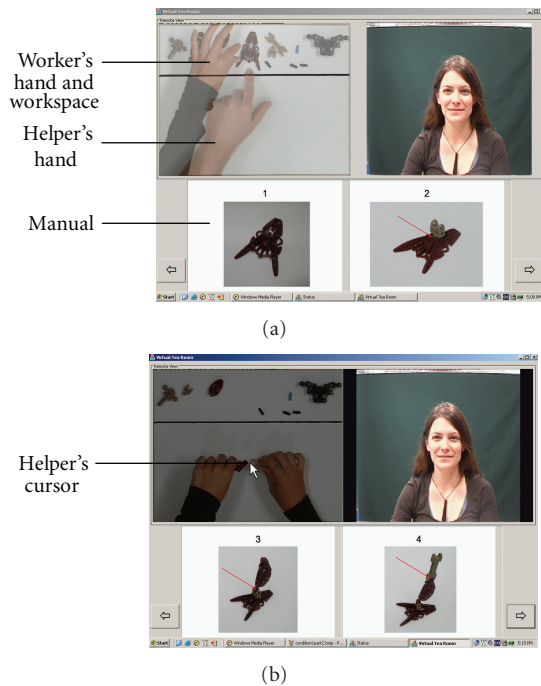

(a)

(b)

FIGURE 3: Helper's view, in hand's (a) and cursor pointer condition (b). (Note: the worker's view was similar in each condition, with the exception that the manual on the bottom of the screen was not visible).

Assembly Accuracy. Similar to the assembly time analysis, the accuracy scores for the two conditions were compared statistically. On average, there was a single mistake out of eleven steps per trial, the assembly accuracy for both conditions was fairly high (hand's condition: mean (M) = 11.00, SD = 1.26; cursor pointer condition: M = 10.69, SD = 1.66). No statistical differences are reported between the conditions ($t(1, 15) = .718, P = .484$).

TABLE 1: Helper and worker's satisfaction. (M: mean; SD: standard deviation; n.s.: not significant).

Condition	Role	M	SD	Significance
Hands	Helper	3.19	0.54	$t(1, 15) = 2.076, P = .055$
Cursor	Helper	2.88	0.72	
Hands	Worker	3.26	0.45	n.s.
Cursor	Worker	3.21	0.63	

TABLE 2: Perception of quality of collaborative effort scores.

Condition	Role	Mean	SD
Hands	Helper	3.30	.51
Pointer	Helper	3.03	.47
Hands	Worker	3.21	.41
Pointer	Worker	3.01	.66

Satisfaction. Both helper and worker were satisfied in every condition (see Table 3, means are all above 2 in the assessment on a 5-point scale ranging from 0 to 4). No significant statistical differences are reported in the satisfaction rating. For the helpers, there is a slightly higher satisfaction for the hands condition (see Table 1).

Perception of the Quality of the Collaborative Effort. Both the helper and worker reported a high score in the perception of the quality of their collaborative effort (see Table 2). Collaboration scores were reported significantly higher ($P = .032$) in the hands condition than the pointer condition for the helper. In the worker group collaboration scores were also higher in the hands condition however there as no significant difference.

In general, most of the participants experienced the collaboration in both conditions as symmetrical (see Table 3)

Table 3: Perceived contribution to task solving in percentage.

During the interaction	Hand's condition		Pointer Condition	
	Helper	Worker	Helper	Worker
I felt my contribution was less important than my partner's contribution	0.0	15.8	6.3	26.3
I felt my contribution was as important as my partner's contribution	87.5	63.2	68.8	68.4
I felt my contribution was more important than my partner's contribution	12.5	21.1	25	5.3

Table 4: Helper perception of the interaction in the two conditions (Reported are means and (standard deviation), the range was 0–6).

Dimension	Hands condition	Pointers condition	Significance
Understanding of partner's explanation	3.19 (0.66)	2.06 (0.93)	$t(1, 15) = 3.435$, $P = .004$
Helpfulness of partner's gestures	3.31 (0.48)	2.81 (0.66)	$t(1, 15) = 2.739$, $P = .015$
Transparency of partner's actions in workspace	5.06 (0.93)	4.00 (1.93)	$t(1, 15) = 2.455$, $P = .027$

and would judge their contribution to the task solving of equal importance to that of their partner.

Both helpers and workers assessed the transparency of the partner's actions in the workspace in terms of whether they could see clearly what the partner was doing in the workspace. This was significantly higher for the hands condition (see Table 4; for worker: hands condition: M = 4.94 (SD = .97), cursor pointer condition: M = 4.11 (SD = 1.59), $t(1, 15) = 2.39$, $P = .028$).

Ease to Explain. 64% of the helpers pointed out that it was easiest for them to explain the assembly steps when they used their hands, compared to 21% saying that the cursor pointer was the device easiest to explain with. 14% of the helpers saw no differences between the devices. For the workers, both hands and pointer were considered similarly easy for understanding the partner's explanation (each 42%). 16% of the workers stated that they did not experience a difference between the two-gesture representation with respect to the difficulty to understand explanations.

Ease of Indicating Objects. With respect to the perceived ease of indicating objects by helper or worker in relation to the devices, the helpers had no preference but would chose both devices equally as often (each 43%) and 14 percent would see no difference between hands and pointer with respect to their difficulty to indicate specific objects. On the other hand, 53% of the workers would see the pointer as easier to indicate specific objects and 26% of worker would see the hands as easier to indicate objects. 21% of them would see no differences.

Choice of Gesture Device. Concerning the explicit preference for one-gesture representation, 50 percent of the helpers and 53 percent of the workers preferred a pointer when solving comparable tasks. Compared to that, 43 percent of the helpers and 38 percent of the workers would choose hands as the preferred gesture representation.

6. Discussion

We have presented the results of the investigation of two-gesture representations in support remote collaboration on physical tasks, focusing on the effectiveness, user's satisfaction and perception of collaboration quality.

With respect to the comparative effectiveness of hands versus pointer as gesture representations, both conditions were similarly effective judged by assembly time and accuracy for the task, although the performance was slightly higher in the hands condition. In fact, the subjects did very well when using either device, making few (if any) mistakes.

These results are somewhat surprising as they contradict related findings. Kirk and Fraser [7] identify significant performance benefits in the hands-only condition compared to hands and sketch and digital sketches. Furthermore, Fussell and colleagues found sketch devices to be superior to laser pointers in remote collaboration tasks. In terms of effectiveness and visibility of the cursor a laser pointer would seem similar to a mouse with pointer. However, there were no significant differences found in our study. Previous studies may have observed differences because they compared gestures across media conditions and not within one media condition.

Our results showed that both helpers and workers reported an overall preference of using pointer functionality than that of hands. Hand representation is richer than a cursor pointer in terms of representations of rotation and orientation in the assembly task, but is not a commonly understood computer mediation communication tool like cursor pointer. The impermanence of the pointer as well as its small size makes them appropriate for pointing/direction gestures and could be sufficient enough for solving certain physical tasks.

However in terms of participant's perception of the quality of their collaborative effort, our study shows that the hands condition was ranked higher than the cursor condition. Hands represent the most intuitive way to gesture and could have impact on user's perception of interaction. We found that both helpers and workers perceived their interaction as "more transparent" when seeing their partner's actions in the hands condition over seeing a pointer in the pointer condition. The helpers assessed the hands condition superior with respect to the understanding of their partner's verbal explanations and the helpfulness of their gestures. A significant difference between the perception of interaction in hands condition and that in pointer condition was found

in the helper group. Also the overlaying hands condition leads to higher reports of both perception of the partner and perception of the quality of their collaboration effort with their partner.

This last finding suggests strongly that the level of "bandwidth" of a gesture has a significant impact on a participant's collaborative effort. The ability to perform more complex gestures (using hands) is more likely to influence collaboration than the ability to point and perform deictic actions (using a pointer).

Our results based on measuring performance and quality of collaborative effort suggest that the value of a more expansive gesture representation is not so much a gain in performance but more a gain in the user's experience and specifically in the user's perceived quality of collaborative effort.

7. Limitations of the Study

We have observed a number of spatial orientation issues during the sessions. Participants reported not understanding the point of view of their partner, as their view of the workspace could not be totally aligned with their partner's spatial view of the workspace. Shared viewpoint is important, as Alibali [17] states, "Gestures not only coincide with spatial information; they also reveal speakers' viewpoint on that information". This misalignment could have affected participant experience. Aligning the helper and worker spatial view may be achieved by providing a view as if the helper is looking over the shoulder of the worker, thereby addressing some of the orientation issues.

In addition, the helpers' manual for the Lego assembly task is provided on screen, requiring mouse clicks or keyboard presses to move through. This may have created overhead (and a negative bias) when switching to manual gestures. It needs to be considered though that Kirk's setup, using a physical manual, might have produced a similar bias towards the hands-only condition.

8. Conclusion

This paper presents an experimental investigation of gesture representations in remote collaboration on physical tasks. This research field has been studied by Fussell et al., Kirk et al. and other researchers. Their work has provided a valuable foundation for our study. In light of their investigations, we have compared two remote gesturing representations: the overlaying hands, and the pointer, with specific attention to the perceptual factors in addition to the task performance and user satisfaction. With this study we were able to further the work of Fussell et al. and Kirk et al. by shedding some light into the value of gesture representation from a user experience perspective, more specifically from the users' perceptions of the quality of their collaborative effort.

Acknowledgments

References

[1] L. Alem, S. Hansen, and J. Li, "Exploration of clinician's sense of presence in the virtual critical care unit," in *Proceedings of Presence*, pp. 103–104, 2006.

[2] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. I. Kramer, "Gestures over video streams to support remote collaboration on physical tasks," *Human-Computer Interaction*, vol. 19, no. 3, pp. 273–309, 2004.

[3] J. Ou, L. M. Oh, S. R. Fussell, T. Blum, and J. Yang, "Analyzing and predicting focus of attention in remote collaborative tasks," in *Proceedings of the 7th International Conference on Multimodal Interfaces (ICMI '05)*, pp. 116–123, October 2005.

[4] S. R. Fussell, L. D. Setlock, and E. M. Parker, "Gaze targets during collaborative physical tasks," in *Proceedings of the ACM Human Factors in Computing Systems Conference (CHI '03)*, pp. 768–769, 2003.

[5] S. Fussell, L. D. Setlock, E. M. Parker, and J. Yang, "Assessing the value of a cursor pointing device for remote collaboration on physical tasks," in *Proceedings of the ACM Human Factors in Computing Systems Conference (CHI '03)*, pp. 788–789, 2003.

[6] J. Ou, Y. Shi, J. Wong, S. R. Fussell, and J. Yang, "Combining audio and video to predict helpers' focus of attention in multiparty remote collaboration on physical tasks," in *Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI '06)*, pp. 217–224, November 2006.

[7] D. S. Kirk and D. S. Fraser, "Comparing remote gesture technologies for supporting collaborative physical tasks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, pp. 1191–1200, April 2006.

[8] D. S. Kirk, D. S. Stanton-Fraser, and T. Rodden, "The effects of remote gesturing on distance instruction," in *Proceedings of the Computer Support for Collaborative Learning Conference (CSCL '05)*, pp. 301–310, 2005.

[9] J. Ou, L. M. Oh, J. Yang, and S. R. Fussell, "Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, pp. 231–240, April 2005.

[10] H. Kuzuoka, J. Kosaka, K. Yamazaki et al., "Mediating dual ecologies," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '04)*, pp. 477–486, Chicago, Ill, USA, November 2004.

[11] S. Ohta, H. Kuzuoka, M. Noda et al., "Remote support for emergency medicine using a remote-control laser pointer," *Journal of Telemedicine and Telecare*, vol. 12, no. 1, pp. 44–48, 2006.

[12] D. Kirk, T. Rodden, and D. S. Fraser, "Turn it this way: grounding collaborative action with remote gestures," in *Proceedings of the 25th SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*, pp. 1039–1048, San Jose, Calif, USA, April-May 2007.

[13] J. Ou, S. R. Fussell, X. Chen, L. D. Setlock, and J. Yang, "Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks," in *Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI '03)*, pp. 242–249, Vancouver, Canada, November 2003.

[14] N. Sakata, T. Kurata, T. Kato, M. Kourogi, and H. Kuzuoka, "WACL: supporting telecommunications using wearable active camera with laser pointer," in *Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC '03)*, October 2003.

[15] K. Nurmela, T. Palonene, E. Lehtinene, and K. Hakkarinen, "Developing tools for analyzing CSCL process," in *Designing for Change*, B. Wasson, S. Ludvigsen, and U. Hoppe, Eds., pp. 333–342, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[16] M. Hogan and A. Krumm-Heller, "Design of a multicast interactive video application," in *Proceedings of CSIRO ICT Centre Conference*, 2004.

[17] M. W. Alibali, "Gesture in spatial cognition: expressing, communicating, and thinking about spatial information," *Spatial Cognition and Computation*, vol. 5, no. 4, pp. 307–331, 2005.