How Coherent Environments Support Remote Gestures

Naomi Yamashita
NTT Communication Science Labs.
naomi@cslab.kecl.ntt.co.jp

Keiji Hirata
NTT Communication Science Labs.
hirata@brl.ntt.co.jp

Toshihiro Takada
NTT Communication Science Labs.
takada@brl.ntt.co.jp

Yasunori Harada
NTT Communication Science Labs.
hara@brl.ntt.co.jp

ABSTRACT
Previous studies have demonstrated the importance of providing users with a coherent environment across distant sites. To date, it remains unclear how such an environment affects people's gestures and their comprehension. In this study, we investigate how a coherent environment across distant sites affects people's hand gestures when collaborating on physical tasks. We present video-mediated technology that provides distant users with a coherent environment in which they can freely gesture toward remote objects by the unmediated representations of hands. Using this system, we examine the values of a coherent environment by comparing remote collaboration on physical tasks in a fractured setting versus a coherent setting. The results indicate that a coherent environment facilitates gesturing toward remote objects and their use improves task performance. The results further suggest that a coherent environment improves the sense of co-presence across distant sites and enables quick recovery from misunderstandings.

Categories and Subject Descriptors
H.4.3 Information systems applications: Communications applications – Computer conferencing, teleconferencing, and videoconferencing

Keywords
Computer-supported collaborative work, collaborative physical task, video-mediated communication, coherent environment, remote gesture

1. INTRODUCTION
Recent research on distance work has significantly demonstrated the importance of providing people with a coherent environment [3, 5, 2]. When the positional relationships between distant sites become fractured, as is often the case with conventional video systems, people tend to have difficulties in making sense of others' speech and gestures with the surrounding environment [2].

The problem becomes particularly serious in distance work where gestures play a significant role. Collaborative physical tasks [1] fall into such works, in which one or more individuals (workers) work with a concrete object under the guidance of a remote individual (helper).

AVI'08, May 28–30, 2008, Napoli, Italy.

Given that gesturing is so crucial to collaborative physical tasks, a variety of video systems are being developed to facilitate remote gesturing (e.g., DOVE [1], Agora [7]). They typically facilitate remote gesturing by introducing a coherent space (i.e., shared visual space) in which the relationships between helper's gestures and the remote objects are maintained.

While previous studies have indicated that the introduction of a coherent space improves task performance [5], researchers have so far focused exclusively on how the introduction of a coherent space affects the worker's understanding of the helper's gestures [5, 1].

Yet no one has investigated the influence of the introduction of coherent space on helper's gestures. In other words, we still lack an understanding of how coherence affects gesture usage in collaborative physical tasks. For example, does a coherent environment equally facilitate all types of gestures or only certain types? If the latter case is true, what types of gestures are facilitated, and are those gestures understood efficiently in relation to the surrounding environment? Furthermore, does a coherent environment enhance the collaborators' sense of co-presence? Answering such questions will provide guidance for designers of video-mediated technologies.

2. CURRENT STUDY

2.1 Re-classification of Gestures
Previous studies suggest that people use several types of gestures during collaborative physical tasks [1]. The classification of such gestures differs between systems [8], but all differentiate between pointing and representational gestures.

Pointing gestures are used to refer to objects and locations. Representational gestures are used to represent the shapes of objects and the nature of the actions to be done with the objects [8]. Representational gestures are further classified into three types that play a critical role in collaborative physical tasks [1]: iconic, spatial, and kinetic. Iconic representations form hand shapes to show what a particular object looks like; spatial gestures describe the distance between two objects by typically placing two fingers or hands a certain distance apart; kinetic gestures describe how actions should be performed on an object.

While researchers have mainly focused on the role of each gesture, we are more interested in the mediation of gesture across distant sites. To this end, we re-classify representational gestures into two types based on whether the gesture involves interaction with objects at a remote site: *remote-oriented representational gestures*, which involve interaction with remote sites, and *locally-closed representational gestures*, which do not involve interaction with remote sites.

We expect that gestures involving interaction with remote objects (i.e., both remote-oriented representational and pointing gestures) are only understood effectively in a coherent environment where the relationship between gesture and object is maintained. Those gestures will not be understood correctly in a fractured setting, in which positional relationships between distant sites are not preserved.

2.2 Hypotheses

If distant collaborators are provided with a coherent environment, we expect that collaborators will be able to understand each others' gestures better in relationship with the surrounding objects. This leads to several hypotheses regarding the performance of helper-worker pairs in the mentoring physical tasks explored in this study.

H1 (Gesture usage): Collaborators in a coherent environment will make greater use of remote-oriented representational and pointing gestures than a fractured environment. Conversely, collaborators in a fractured environment will make greater use of locally-closed representational gestures than a coherent environment.

H2 (Effects of Gestures): Higher use of pointing and remotely-oriented representational gestures is correlated with faster performance in a coherent environment, but not in a fractured setting.

When collaborators frequently gesture toward a remote site, we expect that collaborators will feel co-present with their distant collaborators and objects and tend to often use local deixis.

H3 (Sense of co-presence): Collaborators in a coherent environment will frequently use local deixis and achieve a greater sense of co-presence than in a fractured environment.

2.3 t-Room System

In this study, we investigate the value of the coherent environment by comparing mentoring collaborative physical work using the t-Room system [4].

Figure 1 shows the hardware design of the system. A single t-Room consists of six modules called Monoliths arranged octagonally and a worktable at the center embedded with LCD displays.

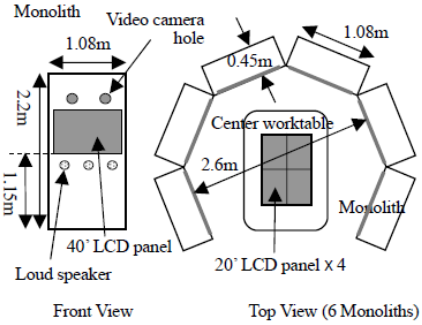

Figure 1 Hardware Design of t-Room

Users in the t-Room are surrounded by six 40-inch LCD panels (resolution of 1280 by 768), six HDV cameras, and 18 loudspeakers. An HDV camera is mounted inside each Monolith to capture the views inside the room, especially the heads and upper bodies of users. A polarized film is placed over each camera to eliminate infinite video feedback. LCD panels are positioned at the height of user heads and upper bodies, showing both local users' self-reflection images and remote users' images (Figure 3). The self-reflection images are intended so that the users can check how their own figures are projected at the distant site. An HDV camera is also hung from the ceiling to capture the scene at the worktable. In this way, collaborators can share the same views projected on the wall and table screens; collaborators are aware of exactly what the others can see of the work space.

2.4 Experimental Design

We installed two identical t-Rooms in the cities of Atsugi and Kyoto, which are approximately 400 km apart. A commercially available 100 Mbps optical fiber line connects the two rooms. The network delay for video and audio data transmission between Atsugi and Kyoto is around 0.7−0.8 and 0.4-0.5 seconds, respectively.

In the experiment, a helper and a worker performed a repair task on a personal computer (DELL OptiPlex 170L) in each of two media conditions: (a) *fractured setting*: a video system that fractures the relationships between gesture and the target object in a distant space; a handy camera that captures a partial view of the worker's task space, and a scene camera capturing the helper's upper body (see Figure 2). (b) *coherent setting*: a video system that provides collaborators with a coherent environment. Cameras and displays are setup so that the relationship between action and environment is maintained across distant sites.

The study included ten participants. The workers consisted of nine part-time employees who had never deconstructed a PC or used a video-mediated communication system before the experiment. We recruited a male helper who is a PC repair expert and had worked as an instructor at a PC technical college to provide guidance from the Atsugi t-Room to all nine workers in the Kyoto t-Room. Prior to the experiment, the helper practiced giving instructions with two extra participants, so that he could offer steady instructions throughout the experiment.

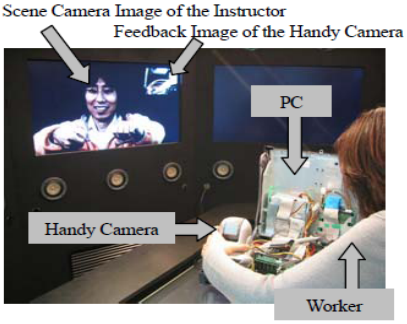

Figure 2 Media Condition (a): Fractured Setting

Figure 3 Media Condition (b): Coherent Setting

Table 1 Characteristics of Each Media Condition

	Condition (a)	Condition (b)
View	Narrow	Wide
Detailed Image	Yes	No
Camera Movement	Yes	No
Coherence	Fractured	Coherent

Table 1 summarizes the differences between the two media conditions focused on in our study. In condition (a), we setup a general video setting suitable for mentoring collaborative physical tasks; a worker can move the camera and control what the helper sees; he can zoom or/and focus on parts of the object to which he wants to draw the helper's attention. However, the camera view in the condition is relatively narrow and fractures the environment across sites. In condition (b), collaborators are provided with a wide view of each other's spaces, although they cannot control the camera views. The collaborators are also provided with a coherent environment, although the helper's gestures (particularly pointing gestures toward a remote object) are sometimes occluded by the actual object.

2.5 Procedure

The following was the experiment's procedure:

Procedure (1): Workers were given explanations how the system worked. The helper and a worker also engaged in a short-term pre-study task to become familiar with the t-Room environment and to grasp how to deal with a real object.

Procedure (2): Workers were given an overview of their roles in the experiment: to replace a broken PC. Then, the helper and a worker engaged in three tasks: exchanging a power supply unit, a hard disk drive, and a DVD unit, each in different system settings: fractured setting, coherent setting, and another setting, which is over the scope of this paper. Trials, tasks, and media conditions were counterbalanced. The pairs were instructed to complete the task as quickly as possible. They were allowed to freely communicate, but the helper was instructed to avoid giving workers information unrelated to their current task.

Procedure (3): Following the tasks, workers and the helper were interviewed about the ease of understanding each other's utterances, the usefulness of specific technological features, and their preference of technology.

3. RESULTS

Since the experiment was initially designed to compare three media conditions [9] (i.e. fractured vs. coherent vs. coherent with partially fractured space), results were analyzed in a trial by task by media condition repeated measures ANOVA.

Pairs completed the tasks in an average of 12.4 and 11.9 minutes under fractured and coherent settings, respectively. The differences in task completion times were not significant. Furthermore, all workers correctly exchanged the PC units in both conditions. However, two of nine workers misunderstood the helper's instruction and attached the PC cord to a different place during the fractured setting.

3.1 Effects of Gestures

3.1.1 Gesture Usage

The helper frequently gestured when instructing the workers; he gestured once every 11.8 seconds in the fractured setting and once every 9.4 seconds in the coherent setting. Analysis on the frequency of gesture indicated a significant main effect for media condition ($F[2,18]=6.29$, $p=<.01$). Post-hoc tests indicated that the helper gestured more frequently in the coherent than the fractured setting ($p<.05$).

To investigate how the helper's gestures differed between conditions, we classified them into three categories: Pointing, Remotely-oriented representational, and Locally-closed representational. Two independent coders classified gesture samples until they reached 90% agreement. They then each coded half of the videos. Table 2 shows the proportion of gestures in each of the three categories across each media condition.

Table 2 Proportion of Helper's Gestures in Each Category

Environment	*Pointing*	*Remotely -oriented*	*Locally -closed*
(a) Fractured	13%	19%	68%
(b) Regular t-Room	44%	31%	25%

Analysis on the proportion of each gesture usage indicated that the usage of gestures differed significantly across media conditions (pointing gestures: $F[2,18]=111.41$, $p<.001$; remotely-oriented gestures: $F[2,18]=23.18$, $p<.001$; locally-closed representational gestures: $F[2,18]=255.37$, $p<.001$). As predicted by *H1*, post-hoc tests indicated that the helper made greater use of pointing and remotely-oriented gestures in the coherent than in the fractured setting ($p<.001$).

3.1.2 Effects of Gestures on Completion Time

Although the helper frequently gestured toward the PC unit in the worker's site, not all his gestures could be seen by the worker; some were out of camera site. Approximately half of the helper's gestures were unable to see from the worker's site in both fractured and coherent settings (gestures were not deemed "viewable" when part of the view was missing). Regardless of many cameras used in the coherent setting, many of the helper's gestures were off the camera site, since the helper frequently gestured toward the object on the central table, which was slightly lower than the shooting area (i.e., side wall screens).

To examine *H2*, we first calculated the rate of viewable gestures per second and then examined the relation between the viewable gestures and task performance (Table 3).

As shown in Table 3, the rate of viewable remotely-oriented gestures were significantly correlated with faster performance times in the coherent setting, but not in the fractured setting. The rate of viewable pointing gestures slightly correlated with the task performance in the coherent setting. A higher rate of viewable locally-closed gestures was slightly correlated with faster performance in the fractured setting, but not in the coherent setting.

Table 3 Correlation between Viewable Gestures and Completion Time

Environment	*Pointing*	*Remote-oriented*	*Locally-closed*
(a) Fractured	r= -.12 (p= .77)	r= .41 (p= .28)	r= -.59 $^{+}$ (p=.09)
(b) Coherent	r= -.58 $^{+}$ (p= .09)	r= -.80 ** (p< .01)	r= -.29 (p=.45)

+ significant at 10% level; * significant at 5% level; ** significant at 1% level

3.2 Sense of Co-presence

Previous studies have shown that people feel co-present when they gesture a lot. We have seen in Section 3.1 that the helper gestured significantly more in the coherent setting than in the fractured setting.

To examine *H3*, we further calculated the number of local deixis in each utterance and compared the values across media conditions (Figure 4). Typically, people use local deixis (e.g., *here*, *this*, *these*) more often when they feel present in a remote environment and co-located with a set of distant objects [4].

Analysis on the numbers of local deixis per utterance indicated significant main effects for media condition ($F[2,18]=18.37$, $p<.001$), but no main task effect. Post-hoc tests indicated that the use of local deixis was significantly higher in the coherent setting than the fractured setting ($p<.001$).

Consistent with the quantitative results, several participants remarked in the post-experimental interviews that they felt more co-present with their remote collaborator in the coherent setting than the fractured setting.

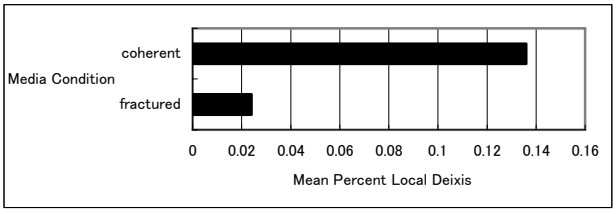

Figure 4 Proportion of use of Local Deixis per Utterance

3.2.1 Overcoming Misunderstandings

In the coherent setting, we found interesting cases where the helper instructed the worker as if they were in the same room, relying on the practices and resources of co-located collaboration; when the workers had trouble identifying a PC component, the helper sometimes walked around the table and directed the workers to look at the PC from his standing position as shown in the following excerpt.

Helper: Can you pull the loop like this? [Gestures how to pull the loop].
Worker: Yes. [Tries to pull out a different component].
Helper: Umm. Excuse me.
Worker: Yes?
Helper: Can you come over here? [Walks around the table] …stand over here?
Worker: Ok? [Walks around the table, and stands very close to the helper].
Helper: This orange cable. . . See it? Bend it down a little bit.
Worker: [Bends it down as told].
Helper: See the orange thing… looks like a wire? Something round.
Worker: Oh, I got it. This?
Helper: Yes. Pull it up.

Such a scene only makes sense when the participants in the rooms can move freely inside the rooms, while maintaining the spatial relationships between the two sites.

4. Conclusions

Our results demonstrate the value of providing distant collaborators with a coherent environment for collaborative physical tasks. First, a coherent environment improved the collaborators' sense of co-presence and enabled them to rely on the practices and the resources of co-located collaboration. For example, collaborators walked around the table to view an object from the same angle and quickly resolved misunderstandings.

Second, the environment facilitated collaborators' use of remotely-oriented gestures (i.e., representational gestures involving interaction with remote sites). Using such gestures in the environment facilitated grounding in the task procedure and was highly correlated with faster performance.

Regardless of such benefits of the coherent environment, the workers did not complete the tasks significantly faster in the coherent setting than in the fractured setting. Perhaps the visibility of the helper's remotely-oriented gestures was low (13% of the total gestures) so no effects on overall task performance times were visible.

5. ACKNOWLEDGEMENTS

We thank Yoshinari Shirai, Shigemi Aoyagi, and Junji Yamato for their assistance. We also thank Hideaki Kuzuoka and several anonymous reviewers for their valuable comments.

6. REFERENCES

1. Fussell, S. R., Setlock, L., Yang, J., Ou, J., Mauer, E., and Kramer, A. D. I. Gestures over video streams to support remote collaboration on physical tasks. *Journal of Human-Computer Interaction*, 19, (2004), 273-309.
2. Heath, C. and Luff, P. Disembodied Conduct: Communication Through Video in a Multi-media Office Environment. *Proceedings of CHI'91*, ACM Press, (1991), 99-103.
3. Heath, C., Luff, P., Kuzuoka, H., and Yamazaki, K. Creating Coherent Environments for Collaboration. *Proceedings of ECSCW'01*, Kluwer Academic Publishers (2001), 119-128.
4. Hirata, K., Harada, Y., Takada, T., Aoyagi, S., Shirai, Y., Yamashita, N., and Yamato, J. The t-Room: Toward the Future Phone, *NTT Technical Review*, 4, 12, (2006), 26-33.
5. Kirk, D., Crabtree, A. & Rodden, T. Ways of the Hands. *Proceedings of ECSCW'05*, Kluwer (2005).
6. Kramer, A., Oh, L., and Fussell, S. Using Linguistic Features to Measure Presence in Computer-Mediated Communication. *Proceedings of CHI'06*, ACM Press (2006), 913-916.
7. Kuzuoka, H., Yamashita, J., Yamazaki, K., Yamazaki, A. Agora: A Remote Collaboration System that Enables Mutual Monitoring. In CHI'99 Extended Abstracts, (1999), 190-191.
8. McNeill, D. *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press (1992).
9. Yamashita, N., Hirata, K., Takada, T., Harada, Y., Shirai, Y., and Aoyagi, S. Effects of Room-sized Sharing on Remote Collaboration on Physical Tasks. *IPSJ Journal*, Digital Courier, Vol. 3, (2007).