Information and Media Technologies 3 (1) : 91–102 (2008)
reprinted from: IPSJ Digital Courier 3: 788–799 (2007)

Regular Paper

Effects of Room-sized Sharing on Remote Collaboration on Physical Tasks

Naomi Yamashita,[†] Keiji Hirata,[†] Toshihiro Takada,[†]
Yasunori Harada,[†] Yoshinari Shirai[†] and Shigemi Aoyagi[†]

This paper outlines some of the benefits of providing remote users with consistent spatial referencing across sites when collaborating on physical tasks. Two video-mediated technologies are introduced: room-sized sharing that enables remote users to see similar things to what they would actually see if in the same room and a snapshot function that enables users to gesture at remote objects. We examine the impact of these technologies by comparing remote collaboration on physical tasks in a regular video conferencing system with a handy camera versus a room duplication system versus a room duplication system with a snapshot function. Results indicate that room-sized sharing facilitates remote collaborators' sense of co-presence and supports remote gesturing, which is closely aligned to normal co-present gesturing. Although such benefits did not contribute directly to the overall decrease of task performance, room-sized sharing and the snapshot function helped remote collaborators construct appropriate messages, efficiently establish joint focus, and monitor each others' comprehension when conducting complicated physical tasks.

1. Introduction

Collaborative physical tasks are defined as "tasks in which two or more individuals work together with concrete objects in the three-dimensional world" [7]. Because expertise is increasingly distributed across space, demand is growing for technologies that support remote collaboration on physical tasks.

Research on collaborative physical tasks has emphasized the importance of providing collaborators with richer spatial context across sites [6],[16],[17]. Insight has been basically derived from comparisons between face-to-face and video-mediated collaboration. However, it remains unclear how well current video-mediated technology can provide spatial context and whether providing collaborators with such an environment using video-mediated technology actually improves collaborative work.

The aim of this paper is to investigate how and in what situations, rich spatial context across sites, particularly room-sized sharing, helps collaborators perform collaborative physical tasks. Of particular interest: does room-sized sharing solve the problem of maintaining meaningful spatial references across sites? If not, what are the remaining issues to solve the problem? Answering such questions will help provide a foundation for designing video-mediated collaboration for physical tasks.

To examine our research interests, we developed a video-mediated communication system called "t-Room" that aims to provide consistent spatial referencing across sites. t-Room is a distributed video system that supports room-sized sharing and collaboration with multiple screens and cameras on the walls and tables. The room layout is replicated at all sites, and everything inside the rooms is mutually projected on another room's screens; collaborators using the system can see similar things to what they would actually see if in the same room. Therefore, we expect that t-Room features more advances in sharing gestures, gaze, and objects across sites than previous video-mediated communication systems.

In the remainder of this paper, we first describe the theoretical framework guiding our work. Next, we present a study that aims to test empirically the value of providing collaborators with richer spatial context in a complex collaborative task: replacing a computer component. We conclude with a discussion of the implications of our findings for the design of video systems to support remote collaboration on collaborative physical tasks.

1.1 Collaborative Physical Tasks

In this paper, we focus on "*mentoring collaborative physical tasks* in which one person directly manipulates objects under the guidance of one or more experts" [7].

[†] NTT Communication Science Laboratories

Remote mentoring of a collaborative physical task requires extensive coordination between helper and worker. According to Fussell, et al., there are three main conversational subtasks in mentoring collaborative physical tasks[6]: (1) helper and worker must identify what their partners are attending to and establish joint focus, (2) the helper must assess the worker's level of comprehension by monitoring his/her actions and/or task status, (3) they (especially the helper) must create efficient communicative messages by ascribing to the principle of "least collaborative effort[4]."

1.2 Conversational Grounding

Communication becomes more efficient when people share a greater amount of common ground, i.e., mutual knowledge, beliefs, attitudes, expectations, goals, etc[3],[4]. Studies have demonstrated that conversational grounding[1],[2], an interactive process by which communication partners work together to accrue common ground, is enhanced when collaborators share visual access to each others' work spaces[9],[16]. For example, when collaborators are co-present and have smooth access to views of shared work spaces, they can easily identify each others' focus of attention, monitor facial expressions, body orientations and actions, and assess whether their utterances have been adequately comprehended.

Initial research demonstrated that views of facial expressions ("talking heads") provide almost no support for conversational grounding[21]. However, more recent research has shown that facial expressions do indeed provide support for conversational grounding, especially when collaborators have very different backgrounds, i.e., people who need to negotiate common ground[25]. Also, a series of studies have shown that visual access to "task space" extensively improves the performance of collaborative physical tasks[6],[16].

1.3 Integrating Multiple Video Feeds

While studies have shown that sharing various visual information of remote sites facilitates mutual understanding, such visual information must be presented in a way that preserves the relationships between space, speech, and gestures[11],[12],[14],[18],[23],[24]; when relationships become fragmented (as in most current video-mediated communication systems), a user may not be able to comprehend the gestures of remote collaborators[19].

Gaver, et al.[8] provided collaborators with the ability to switch between multiple video feeds so that they could see each others' work spaces from various positions. In the study, they discovered that the switching ability made it difficult for collaborators to identify which part of the visual information was shared. Fussell, et al.[7] also did a study that provided a helper with simultaneous views of two kinds of worker's task spaces: a view from a camera mounted on the worker's head and another view from a camera showing a wider view of the worker's task space. In their study, they found that collaborators faced the same problem as in Gaver's experiment (workers had difficulty identifying which part of the visual information was shared) and helpers had difficulty dividing their attention between the two cameras as well.

Thus, when providing participants with shared visual information using multiple video feeds, the visual information must not be presented independently, but should be integrated so that people can collaborate in a coherent environment.

2. Current study

From the above discussion, we developed the idea of duplicating an entire room in distant places. Since the room allows remote collaborators access to wide views of shared work spaces as well as coherent environments in which to accomplish action and interaction, we expect that such a communication system will incur less cost on conversational grounding between remote collaborators working on physical tasks, resulting in improved task performance.

While it may be impractical to have identical rooms at more than one location, it is worth investigating how much room duplication can improve spatial awareness in collaboration and how such an environment impacts remote collaboration of physical tasks.

2.1 Hypotheses

As described above, the ability of remote collaborators with access to wide views of shared work spaces, as well as coherent environments in which to accomplish action and interaction, influence both the creation and understanding of interactions along with facilitating conversational grounding. This leads to several hypotheses regarding the performance of helper-worker pairs in the mentoring physical task explored in this study.

Since room duplication allows distant users to see similar things to what they would actually see if in the same room, we expect users to experience a greater sense of co-presence than regular video conferencing systems (VCS).

- *H1* (*Sense of co-presence*): Collaborators will achieve a greater sense of co-presence when using t-Room than regular VCS settings.

When remote users feel co-located, we expect them to easily understand the intention of the gestures of remote users in combination with space, speech and facial expressions, etc. Therefore, users will be able to better infer remote user's levels of comprehension and create appropriate messages or assistance accordingly. This leads us to the following hypothesis:

- *H2* (*Creating appropriate instructions*): Workers will ask fewer questions and/or confirm understanding less often when using t-Room compared to regular VCS settings.

We also expect room duplication to contribute to the efficiency of remote users establishing joint focus and monitoring others' sites and/or levels of comprehension, since these are particularly difficult when collaborators are at different sites. Thus, the following benefits are expected:

- *H3* (*Establishing joint focus*): Collaborators will more efficiently assess each others' focus of attention and establish joint focus when using t-Room rather than regular VCS settings.
- *H4* (*Monitoring worker's level of comprehension*): The helper will more efficiently monitor the worker's site and infer his/her level of comprehension when using t-Room rather than regular VCS settings.

As a result of the above benefits, we expect room duplication to improve communicational efficiency and task performance:

- *H5* (*Communicational efficiency*): Collaborators will perform their task with fewer utterances when using t-Room rather than regular VCS settings.
- *H6* (*Task performance*): Collaborators will complete their tasks faster when using t-Room than regular VCS settings.

3. Method

3.1 Room Duplication System: t-Room

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

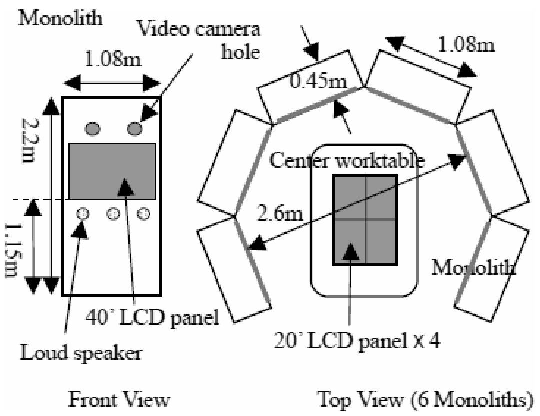

Fig. 1 Hardware design of t-Room.

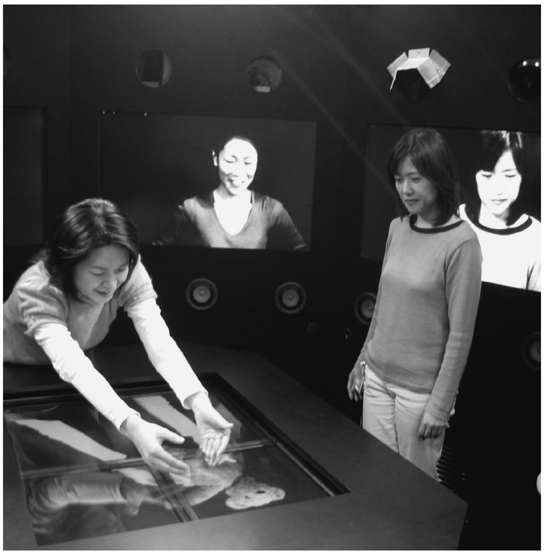

Fig. 2 People collaborating through t-Room.

Users in the t-Room are surrounded by six 40 inch LCD panels (resolution of 1280 by 768, i.e., WXGA), 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 polarizing film is placed over each camera to only capture views in front of the opposite display; the film eliminates infinite video feedback. LCD panels are positioned at the height of user heads and upper bodies, showing both user self-reflection images and remote users' images, as in **Fig. 2**. 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[10].

3.1.1 Remote Pointing Function: Snapshot

Although remote collaborators in t-Room can

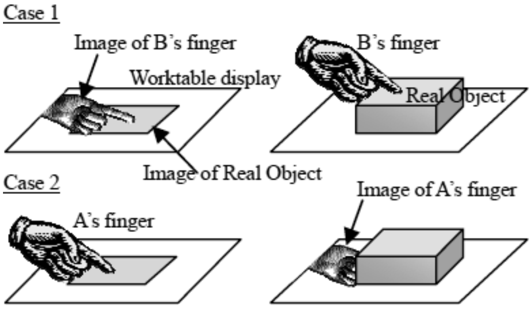

Fig. 3 Real object hiding shared image.

share identical visual images projected on the screens, it is still impossible for a user to point to a remote object. Consider the case where users A and B in distant locations are collaborating on a physical task. When user B points to an object located in his site, user A will be able to see what user B is pointing at, as in Case 1 shown in **Fig. 3**. However, when user A points to a displayed image of the remote object, user B will not be able to see exactly where user A is pointing because the scene on the remote worktable is displayed life-size on the local worktable, and the real object on the local worktable hides the displayed image, as in Case 2 in Fig. 3.

Since pointing is a critical function that facilitates grounding in collaborative physical tasks[7], we prepared a snapshot function that enables a user (the helper, in our experiment) to point at remote objects. A user first takes a snapshot image of any screen image he/she wishes to point to and then displays the snapshot image on one of the screens. Once the snapshot image is displayed on a screen, collaborators can freely share and point to the image (see Fig. 5).

Note that the snapshot function is aligned with the "mixed ecology" approach[13),14)], which advocates the use of unmediated representation of hand gestures. Also, the function fits people's natural behavior, as reported in Kuzuoka's experiment[18)] where helpers often pointed at objects displayed on the screen with their fingers, even if they knew that these actions were not reflected back to the workers.

3.2 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 (i.e., FTTH) 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. Workers performed three repair tasks on a personal computer (DELL OptiPlex 170L) with the assistance of a helper. A helper and a worker performed one task in each of three media conditions: (a) *regular VCS setting*: a video-mediated communication system with fewer spatial cues; a scene camera captures the helper's upper body, and a handy camera captures a partial view of the worker's task space. The former image is projected on the worker's wall screen, and the latter image is projected on the helper's wall screen. Note that the latter image is also fed back to the worker so that he/she gets an idea of what the handy camera is capturing (see **Fig. 4**); (b) *room duplication setting*: a video-mediated communication system with rich spatial cues; regular t-Room setting; a distributed video system that supports room-sized sharing and collaboration with multiple screens and cameras on walls and tables; and (c) *room duplication with snapshot*: a video-mediated communication system with rich spatial cues and a function that enables collaborators to point to objects at each others' sites; t-Room with a snapshot function (See **Fig. 5**).

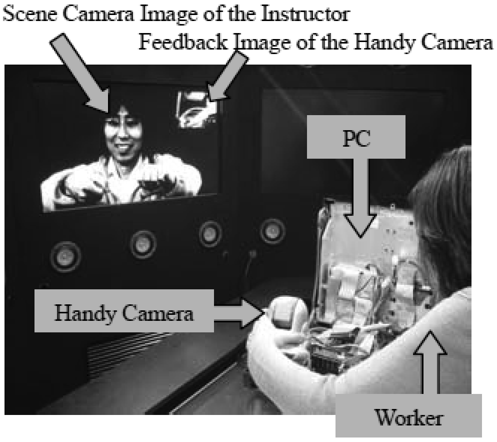
Scene Camera Image of the Instructor
Feedback Image of the Handy Camera
PC
Handy Camera
Worker

Fig. 4 Regular VCS setting.

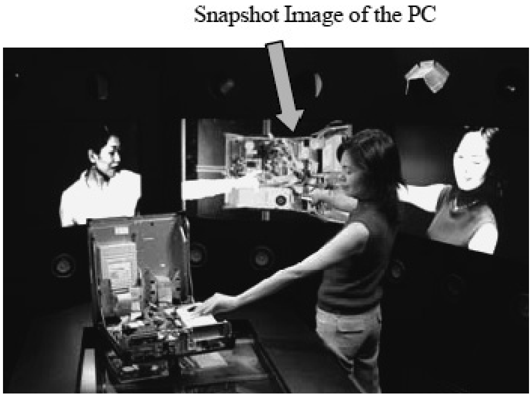

Fig. 5 Room duplication with snapshot.

To reduce factors other than spatial cues between conditions (a) and (b), we set up a regular VCS setting using the t-Room system, as in Fig. 4; one screen was used, and the other screens were blanked out during regular VCS settings.

Exposure to experimental trials was counterbalanced to control the effects of order, and each pair exchanged three different PC units to avoid the effects of practice between trials. Tasks and media conditions were also counterbalanced over participants.

3.3 Participants

The study included ten participants who were paid for taking part in the experiment. Workers consisted of nine part-time jobbers and undergraduate university students (seven females and two males) who had never opened a PC before the experiment. They also had never used a video-mediated communication system before the experiment. Their ages ranged from 20–40. We recruited a male helper who is a PC repair expert and had worked as an instructor at a PC technical college for two years. He provided advice and guidance from the Atsugi t-Room to all nine workers in the Kyoto t-Room.

3.4 Procedure

The helper was told to instruct the workers how to replace the broken PC units (a DVD unit, a hard disk drive, and a power supply unit). He practiced giving instructions with two extra participants prior to the experiment, so that he could give steady instructions throughout the experiment.

The experiment's procedure was as follows:

- Procedure (1): Workers were given explanations how the system worked.
- Procedure (2): The helper and a worker engaged in a short term pre-study task; the worker received a map with a path drawn on it that she was asked to memorize. Then, she was told to explain the path to the helper, who had the same map without a path. The worker could use any function available. The pre-study was intended to allow workers to become familiar with the t-Room environment and grasp how to share a real object.
- Procedure (3): Workers were given an overview of their roles in the experiment: to replace a broken PC.
- Procedure (4): The helper and a worker engaged in three tasks: exchanging a power supply unit, exchanging a hard disk drive, and exchanging a DVD unit, each in different system settings: regular VCS, regular t-Room, and t-Room with snapshot. They were instructed to complete the task as quickly as possible. Also, they were allowed to freely communicate, but the helper was instructed to avoid giving workers unnecessary information, such as names and roles of PC units which were not related to their current task.
- Procedure (5): Following the three tasks, workers were interviewed, as described below.

Table 1 Utterance types.

Category	*Definition*
Joint focus	Utterances relevant to establishing joint focus. (e.g., H: "There's an orange cord." H: "Do you know which one I'm talking about?" H: "Look over here." W: "You mean this one?")
Monitor	Utterances monitoring others' site and/or level of comprehension. (e.g., H: "Got it?" W: "Yes." H: "How is it going?" W: "It's hard. . . ")
Procedural	Instructions and advices furthering task completion. (e.g., H: "Next, push the white lever." W: "Ok..")
Others	Non-task communication and utterances irrelevant to any of the above subtasks. (e.g., H: "Oops, I think the system just froze." H: "Let's wait for a while." W: "Ok." H: "Ok, let's continue.")

3.5 Interviews

At the end of the three tasks, we interviewed the workers about task difficulty and the ease of conducting each task, the ease of understanding the helper's instructions, the appropriateness of the helper's assistance, the usefulness of specific technological features, and their preference of technology. After the whole experiment, we also asked the helper similar questions, including the ease of instructing the workers and the usefulness of specific technological features.

3.6 Conversational Coding

Video and audio recordings of the sessions were the basis for verbatim transcripts and more detailed, post-experimental coding of communication. To examine the influence of room duplication and the snapshot function on the three subtasks in the collaborative physical task, we classified each utterance into one of the following four categories listed in **Table 1**.

Two independent coders classified utterance samples until they reached 90% agreement. Disagreements were resolved through discussion. They then each coded different transcripts, periodically coding a common transcript to ensure that the categories did not shift during coding.

4. Result

All workers and the helper answered in post-experimental interviews that exchanging the power supply unit was the most difficult task, much more difficult than the other two tasks (exchanging the DVD unit and HD drive).

4.1 Sense of Co-presence

To investigate *H1*, how media condition influenced participants' sense of co-presence, we counted the number of local deixis used in their conversations. 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 [15].

We performed a repeated measures analysis of variance (ANOVA) on the numbers of local deixis, using media conditions and tasks as repeated factors. Results indicated significant main effects for both media condition ($F[2, 18] = 13.80$, $p < .001$) and task ($F[2, 18] = 11.32$, $p = .001$) but no interactions (**Fig. 6**). Post hoc tests indicated that the use of local deixis was significantly higher when using t-Room or t-Room with snapshot function than regular VCS setting ($p = .001$ and $p < .001$, respectively).

Consistent with the quantitative results, it appears that participants felt more co-present with their remote collaborator and objects when using t-Room; in the post-experimental interview, a worker said,

> When using the handy camera, I had to be aware of what my instructor was seeing... In the [t-Room setting], I felt relieved because I could concentrate on fixing and because I felt that my instructor was watching over me all the time.

Another worker said:

> Although it took me a little while to understand the setting [t-Room setting], I gradually realized that I was sharing the same space with my instructor.

Indeed, the helper and workers sometimes used the room as a co-located (single) space; the helper sometimes asked workers to move around the table and look at the PC from his angle when workers had trouble understanding what he was saying. The following excerpt shows one such scene using the regular t-Room setting without snapshot.

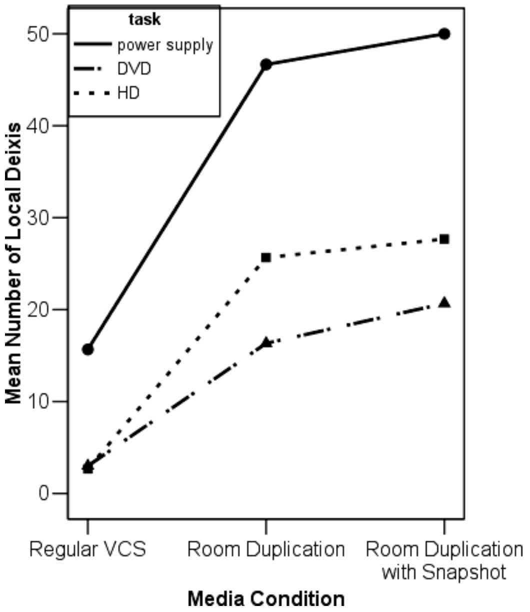

Fig. 6 Mean number of local deixis by media condition.

Helper: Now, let's remove this flat cable over here.
Worker: Ok.
Helper: There's an orange loop attached to it...
Worker: Ok.
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? ...stand over here?
Worker: Ok? [Moves around the table].
Helper: This orange cable... See it? Bend it down a little bit.
Worker: [bends down as told].
Helper: See the orange thing... like a wire? Something round.
Worker: Oh, I got it. This one?
Helper: Yes, yes. Pull it up.

4.2 Creating Appropriate Instructions

In our task, helpers must determine what assistance is needed, when and how to phrase it, and whether the message has been correctly understood. That is, assistance must be coordinated with the worker's actions and the current task status.

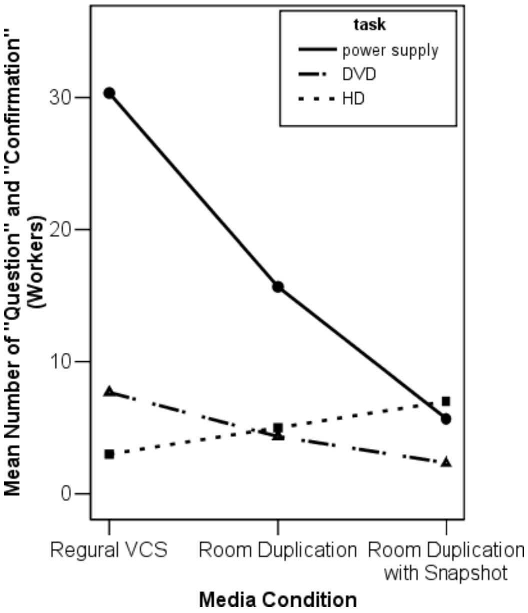

Fig. 7 Mean number of "Questions" and/or "Confirmations" asked by workers per task by media condition.

To examine *H2* whether room duplication (and snapshot function) benefited helpers to better assess the worker's level of understanding and to give appropriate instructions, we performed a task by media condition repeated measures ANOVA on the number of "questions" and "confirmations" workers asked the helper.

Results indicated a borderline main effect for media condition ($F[2, 18] = 2.74$, $p = .091$) and a significant main effect for task ($F[2, 18] = 7.27$, $p < .01$). Post hoc tests indicated that workers tended to ask fewer questions or confirmed fewer understandings using t-Room with snapshot function than regular VCS setting ($p = .078$).

Figure 7 shows the mean number of questions and/or confirmations asked by workers per task by media condition. Although we could not find support that room duplication contributed to the helper's assistance (resulting in fewer questions or confirmation) with easier tasks (exchanges of the DVD and HD), the figure suggests that room duplication contributed to better assistance with a complicated task (exchange of the power supply unit).

4.3 Establishing Joint Focus of Attention

To examine *H3*, how the media condition influenced the efficiency of establishing joint focus between remote users, we compared the number of messages classified into the "Joint focus" category between media conditions.

The number of "Joint focus" utterances was analyzed in a task by media condition repeated measures ANOVA. Results indicated significant main effects for media condition ($F[2, 18] = 8.65$, $p < .01$) and task ($F[2, 18] = 42.40$, $p < .001$). Post hoc tests indicated that distant users using t-Room with snapshot function established joint focus of attention with significantly fewer utterances than regular t-Room and regular VCS settings ($p < .05$ and $p < .01$, respectively).

Although we could not find support for *H3*, collaborators will more efficiently establish joint focus when using t-Room rather than regular VCS settings, we found that snapshot functions helped them identify which components their partners were concentrating on. It seems that the helper's pointing ability assisted them in identifying which component to focus on. In the interview, the helper said:

> The room duplication with snapshot setting was the best because we could both point to a PC component... In the regular VCS setting, I had to explain everything in words because I couldn't point to a PC unit. In the room duplication setting, I sometimes felt frustrated because I had to supplement with words where I was pointing at, even though I was able to point at a PC image.

A worker said,

> When conducting my second task [exchange of the power supply unit], I sometimes got frustrated using the handy camera. The task was complicated, and the handy camera sometimes showed images beyond my intention... In the room duplication setting, I could roughly tell where to look, but not exactly. I had to confirm with my instructor which component he was talking about... I thought using the snapshot image was an easy and reassuring way to identify which component I should focus on.

Figure 8 shows the mean number of utterances of remote users establishing joint focus per task by media condition. Similar to the previous section, the result suggests that although there is almost no difference in the efficiency of establishing joint focus with easier tasks (exchanges of the DVD and HD), room duplication and snapshot function influenced the efficiency of establishing joint focus with a complicated task (exchange of the power supply unit).

4.4 Monitoring Worker's Level of Comprehension

We found similar utterance tendencies in

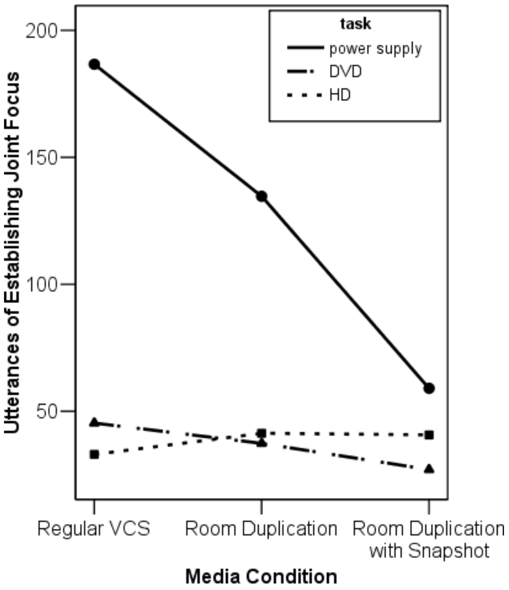

Fig. 8 Mean number of utterances of establishing joint focus per task by media condition.

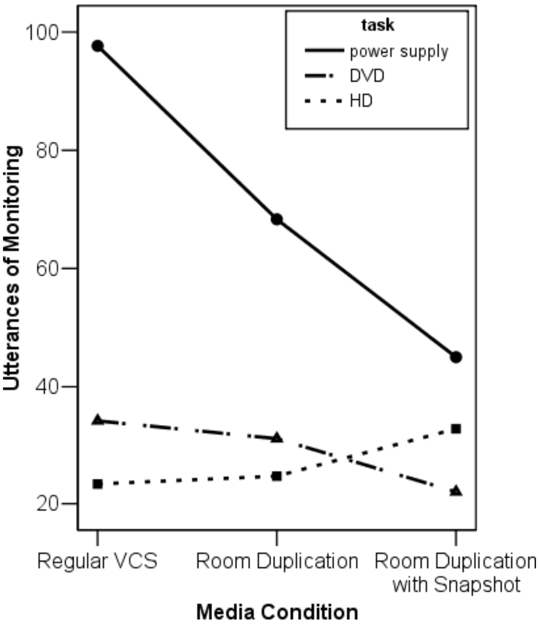

Fig. 9 Mean number of utterances of monitoring remote site and partner's comprehension per task by media condition.

monitoring others' sites and/or levels of comprehension. Repeated measures ANOVA on "Monitoring" utterances indicated a slight main effect for media condition ($F[2, 18] = 3.32$, $p = .059$) and a significant main effect for task ($F[2, 18] = 23.29$, $p < .001$). Post hoc tests indicated that distant users using t-Room with snapshot function efficiently monitored each others' space more than regular VCS settings ($p < .05$).

Similar to "Joint focus" utterances, **Fig. 9** indicates that room duplication and snapshot function help remote users efficiently monitor each others' spaces when conducting a complicated task (exchange of the power supply unit), but they have almost no influence when conducting easier tasks (exchange of the DVD and HD)☆.

4.5 Communicational Efficiency

Identical to the previous two sections, we performed a repeated measures ANOVA for "Procedural" utterances. Although we found a significant main effect for task ($F[2, 18] = 46.02$, $p < .001$), we did not find a significant main effect for media condition.

To examine the impact of the media condition on the overall performance of communication efficiency, we performed a repeated measures ANOVA on the number of utterances. Results indicated a slight main effect for media condition ($F[2, 18] = 3.29$, $p = .061$) and a significant main effect for task ($F[2, 18] = 69.12$, $p < .001$).

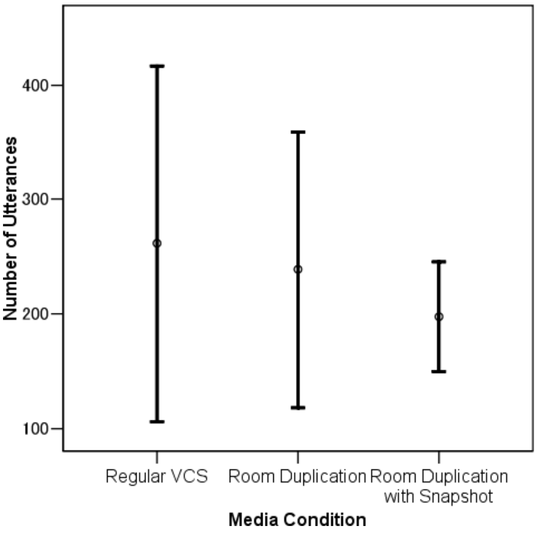

Fig. 10 Error bars of number of utterances in each media condition.

Figure 10 shows an error bar☆☆ plot for the confidence interval of the mean number of utterances for each media condition. A Levene

☆ When exchanging an HD unit, the helper took multiple snapshot images (PC images from different angles) and projected them on the t-Room sidewalls. He often used multiple images when instructing the workers.; this made the helper inevitably stand longer beside the side walls, rather than peering down to the central table. We expect that the multiple use of snapshot images and the distance between the side table and the central table made it hard for the helper to monitor the worker's task space in detail.

☆☆ The circles in the middle of the error bar represent the mean score. The "whiskers" represent the 95% confidence interval.

test for equality of variances on the number of utterances indicated that variance significantly differed across media conditions ($p < .01$). The differences in variance between media conditions can be attributed to the results presented in the previous two sections. While workers have difficulty conducting a complicated physical task, resulting in a wide dispersion of the number of utterances, in regular VCS settings, room duplication and snapshot function help them efficiently conduct the task, which results in narrower dispersion.

4.6 Task Performance

All workers correctly exchanged the PC units at the end of the experiment, although some workers made mistakes during repair. Thus, we defined task performance as the time required to complete the task.

Task completion time was analyzed in a task by media condition repeated measures ANOVA. Results indicated a significant main effect for task ($F[2, 18] = 48.38$, $p < .001$), but no significant main effect for media condition.

5. Discussion

Previous studies have suggested providing remote collaborators with a wide angle, static view of each others' workspaces[7),14)]. In this study, we developed "t-Room," which meets these conditions. t-Room supports room-sized sharing as well as pointing; a remote collaborator can take a snapshot image of a remote object and point at it. We examined the impact of room-sized sharing and snapshot function by comparing remote collaboration on physical tasks in a regular VCS setting versus t-Room (without snapshot function) versus t-Room with snapshot function.

The results provide insights into the effects of room-sized sharing and snapshot function on the collaboration of physical tasks. From our experiment, we found that (1) room-sized sharing facilitates remote collaborators' sense of co-presence and supports remote gesturing, which is closely aligned to normal co-present gesturing, such as walking around a table and viewing the object from the same position; (2) room-sized sharing and snapshot function help remote collaborators construct appropriate messages, efficiently establish joint focus, and monitor each others' comprehension when conducting complicated physical tasks.

5.1 Impact of Task Difficulty

In our study, t-Room outperformed regular VCS setting in establishing joint focus and monitoring each others' space only when exchanging the power supply unit. However, it remains unclear what factors of the task were more difficult in the regular VCS setting compared to t-Room settings.

To answer the question, we focused on task differences and investigated how the differences might influence the collaborators' communication. From the investigation, we realized that the size of the power supply unit was big, and the wires from the unit were connected to various parts of the PC. Since the portable camera could only capture a partial view of the PC, the units (of focus) were sometimes outside the view of the camera, and the helper often asked the workers to move the camera. The proportion of such utterances in the exchange of the power supply unit was far more (approximately twice as much) compared to other tasks (power supply unit: 9.4%; DVD: 5.4%; HD: 4.7%).

We infer that the wide use of the working space (i.e., the lack of a wide view in regular VCS setting) made it hard for the collaborators to have a joint focus and monitor their working space. The result reminds us of Fussell, et al's study; she compared conversations between a helper and a worker attempting to repair a bicycle using a head mounted camera and a static scene camera, and also pointed out the need to provide helpers with a wider field of view[7)].

5.2 Limitations of Room-sized Sharing

We found evidence that t-Room (or room-sized sharing) significantly facilitates the remote collaborators' sense of co-presence and even witnessed remote collaborators using t-Room as if they were in the same room (that is, the helper asked four workers to walk around the table and look at the PC from his angle when workers had trouble understanding what he was saying). Yet we could not find evidence that such an environment significantly helped remote collaboration on physical tasks; the environment only contributed to collaboration on a complicated physical task. The result evokes Kraut's study where a shared visual space was more useful in visually complex tasks[16)].

In our experiment, the helper and most workers answered in the interviews that room duplication with snapshot function was the best to collaborate on physical tasks. However, two workers believed that the regular VCS setting was superior. In the interview, one of the workers said,

> By using the handy camera, I could show my instructor a detailed image of the PC. I could control what my instructor was seeing by moving the handy camera... When I was replacing the HD unit [using t-Room setting], there was a component hidden behind a big one... that I couldn't share with my instructor. If I could have used the handy camera, I could have moved the camera inside the PC and shown it to my instructor.

It seems that users adept with handy cameras found regular VCS settings more useful than static room-sized sharing. Users unable to fully exploit the handy camera found room-sized sharing useful.

5.3 Usage of Snapshot Function

In our study we found that the snapshot function provided support for helper's pointing gestures. We found two ways to use the snapshot function: (a) taking sequential snapshot images of the PC and lining them up on the t-Room wall screens to help workers remember the previous states of the PC; (b) taking snapshot images of the PC from several different angles to use pointing gestures[☆].

Although the snapshot function was useful in remote collaboration on physical tasks, projecting snapshot images on the wall screens apparently created frustration when the helper gave instructions to the workers. In the interview, the helper said,

> When I gave instructions using the snapshot image, I had to switch between the terms up-down and right-left, which was quite confusing.

5.4 Implications for Video System Design

Our findings and the above discussion suggest the following recommendations for the design of future video-based systems to support remote collaboration on physical tasks: When providing remote users with shared visual information using multiple video feeds, such visual information should not be presented independently, but instead should be integrated to allow collaboration in a coherent environment that creates a sense of being co-located in the same space. It is also preferable to provide remote users with a full range of each others' working space.

Room duplication supports remote gesturing, which is closely aligned to normal co-present gesturing. It particularly supports collaboration on physical tasks that require users to walk around during collaboration, such as when objects are big.

When projecting an image of a real object, maintaining identical physical relationships with the real object is better so that user gestures toward the image can be recognized in relationship with the real object with less cognitive load.

5.5 Future Work

We are currently conducting a study on how room-sized sharing might influence collaborators' use of gestures and the transmission of those gestures. The preservation of gestures is particularly important in accomplishing effective collaboration[24)].

Next, we will compare collaboration using t-Room with face-to-face collaboration. We are interested in investigating how well room duplication supports situational awareness[5),6)]. We are also interested in investigating how room size and room shape impact collaboration, and how room-sized sharing (t-Room) impacts collaboration that requires more than two users.

We also plan to augment the snapshot function to include a record and replay function to allow remote users to operate remote objects over time. We are interested in investigating such usage and effects.

Acknowledgments The authors are grateful to Dr. Takehiko Ohno for his valuable comments and the contributions to the t-Room project at the early stage. Dr. Junji Yamato gave us precise suggestions about the camera and display arrangement.

References

1) Clark, H.H.: Using language, Cambridge University Press (1996).
2) Clark, H.H. and Brennan, S.E.: Grounding in communication. Perspectives on socially shared cognition, pp.127–149 (1991).
3) Clark, H.H. and Marshall, C.E.: Definite reference and mutual knowledge, Elements of discourse understanding, Cambridge University Press, pp.10–63 (1981).
4) Clark, H.H. and Wilkes-Gibbs, D.: Referring as a collaborative process, *Cognition*, Vol.22, No.1, pp.1–39 (1986).
5) Endsley, M.: Toward a theory of situation awareness in dynamic systems, *Human Factors*, Vol.37, pp.32–64 (1995).
6) Fussell, S.R., Kraut, R.E. and Siegel, J.:

[☆] The helper asked the workers to move the PC around. He then took a snapshot image of the PC and used the images for pointing to PC components.

Coordination of Communication: Effects of Shared Visual Context on Collaborative Work, *Proc. CSCW'00*, ACM Press, pp.21–30 (2000).

7) Fussell, S.R., Setlock, L. and Kraut, R.E.: Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks, *Proc. CHI'03*, ACM Press, pp.513–520 (2003).

8) Gaver, W., Sellen, A., Heath, C. and Luff, P.: One is not enough: Multiple views in a media space, *Proc. Interchi'93*, ACM Press, pp.335–341 (1993).

9) Gergle, D., Kraut, R.E. and Fussell, S.R.: Action as language in a shared visual space, *Proc. CSCW'04*, ACM Press, pp.487–496 (2004).

10) Heath, C. and Luff, P.: Media Space and Communicative Asymmetries: Preliminary Observations of Video Mediated Interaction, *Human-Computer Interaction*, Vol.7, pp.315–346 (1992).

11) Heath, C., Luff, P., Kuzuoka, H. and Yamazaki, K.: Creating Coherent Environments for Collaboration, *Proc. ECSCW'01*, Kluwer Academic Publishers, pp.119–128 (2001).

12) Hutchins, E. and Leysia, P.: Constructing Meaning from Space, Gesture, and Speech. Discourse, Tools and Reasoning: Essays on Situated Cognition, Springer-Verlag, pp.23–40 (1997).

13) Kirk, D., Crabtree, A. and Rodden, T.: Ways of the Hands, *Proc. ECSCW'05*, Kluwer Academic Publishers, pp.1–21 (2005).

14) Kirk, D.S. and Fraser, D.: Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks, *Proc. CHI'06*, ACM Press, pp.1191–1200 (2006).

15) Kramer, A., Oh, L. and Fussell, S.: Using Linguistic Features to Measure Presence in Computer-Mediated Communication, *Proc. CHI'06*, ACM Press, pp.913–916 (2006).

16) Kraut, R.E., Gergle, D. and Fussell, S.R.: The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence, *Proc. CSCW'02*, ACM Press, pp.31–40 (2002).

17) Kraut, R.E., Miller, M.D. and Siegel, J.: Collaboration in performance of physical tasks: Effects on outcomes and communication, *Proc. CSCW'96*, ACM Press, pp.57–66 (1996).

18) Kuzuoka, H., Kosaka, J., Yamazaki, K., Suga, Y., Yamazaki, A., Luff, P. and Heath, C.: Mediating Dual Ecologies, *Proc. CSCW'04*, ACM Press, pp.477–486 (2000).

19) Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K. and Oyama, S.: Fractured Ecologies: Creating Environments for Collaboration, *Special Issues of the HCI Journal: "Talking About Things: Mediated Conversation about Objects,"* Vol.18, pp.51–84 (2003).

20) Morikawa, O. and Maesako, T.: HyperMirror: Toward Pleasant-to-use Video Mediated Communication System, *Proc. CSCW'98*, ACM Press, pp.149–158 (1998).

21) Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S. and Sclabassi, R.: Turning away from talking heads: The use of video-as-data in neurosurgery, *Proc. Interchi'93*, ACM Press, pp.327–334 (1993).

22) Tan, D., Gergle, D., Scupelli, P. and Pausch, R.: With Similar Visual Angles, Larger Displays Improve Spatial Performance, *Proc. CHI'03*, ACM Press, pp.217–224 (2003).

23) Tang, A., Neustaedter, C. and Greenberg, S.: VideoArms: Embodiments for Mixed Presence Groupware, *Proc. 20th British HCI Group Annual Conference*, ACM Press (2006).

24) Tang, J.C.: Findings from observational studies of collaborative work, *International Journal of Man-Machine Studies*, Vol.34, pp.143–160.

25) Veinott, E., Olson, J., Olson, G. and Fu, X.: Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other, *Proc. CHI'99*, ACM Press, pp.302–309 (1999).

(Received April 25, 2007)
(Accepted September 3, 2007)
(Released December 5, 2007)

Naomi Yamashita received her B.Eng. and M.Eng. degrees in applied mathematics and physics and a Ph.D. degree in Informatics from Kyoto University in 1999, 2001 and 2006, respectively. She is interested in Computer Supported Collaborative Work and Interaction Analysis.

Information and Media Technologies 3(1): 91–102 (2008)
reprinted from: IPSJ Digital Courier 3: 788–799 (2007)
© Information Processing Society of Japan

Keiji Hirata received his Doctor of Engineering degree in Information Engineering from the University of Tokyo, Japan, in 1987. He then joined NTT Basic Research Laboratories. He spent 1990 to 1993 at the Institute for New Generation Computer Technology (ICOT). Dr. Hirata received the IPSJ Best Paper Award in 2001 and the IPSJ Yamashita Memorial Award in 2003, and is a director of IPSJ at present. His research interests include musical knowledge programming and interaction.

Toshihiro Takada received B.E. and M.E. degrees in Information Science from Tokyo Institute of Technology, Japan in 1986 and 1988. In 1988, he joined Nippon Telegraph and Telephone Corporation (NTT). His research interests are in networked information systems, networked computation, real-space computing, and human-computer interaction. He is a member of ACM, IPSJ, JSSST and Human Interface Society in Japan.

Yasunori Harada received his Doctor of Engineering in Information Engineering from Hokkaido University in 1992. He then joined the NTT Basic Research Laboratories. He spent 1998 to 2001 at PREST, Japan Science and Technology (JST). His research interests include visual language and object-oriented programming.

Yoshinari Shirai received an ME from the Graduate School of Media and Governance at Keio University in 2000. He then joined the NTT Communication Science Laboratories. His research interests include ubiquitous computing and interaction design. He is a member of ACM, IPSJ and Human Interface Society in Japan.

Shigemi Aoyagi received B.E. and M.E. degrees in Information Science from Tokyo Institute of Technology, Japan in 1988 and 1990. In 1990, he joined Nippon Telegraph and Telephone Corporation (NTT). His current research interests include parallel processing, distributed algorithms, image understanding, object recognition, and distributed systems for multimedia content. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), IPSJ, and Japan Society for Software Science and Technology (JSSST).