Supporting Fluid Tabletop Collaboration across Distances

Naomi Yamashita[1], Hideaki Kuzuoka[2], Keiji Hirata[1], Shigemi Aoyagi[1], Yoshinari Shirai[1]

[1] NTT Communication Science Laboratories
Kyoto and Atsugi, Japan.
naomiy@acm.org
hirata, aoyagi, shirai.yoshinari@brl.ntt.co.jp

[2] University of Tsukuba
Tsukuba, Ibaraki, Japan.
kuzuoka@iit.tsukuba.ac.jp

ABSTRACT

In this study, we examine how remote collaborators' upper body view affects collaboration when people engage in multiparty fluid tabletop activities across distances. We experimentally investigated the effects of upper body view on four-person group tabletop collaboration, two-by-two at identical locations: shared tabletop vs. shared tabletop plus upper body view. Although previous research has often failed to illustrate the advantages of showing remote participants' upper body view, our study showed that task performance was significantly higher in conditions with upper body view. Furthermore, participants with upper body view tended to take a step away from their remote partners to effectively glance at them while taking a comparable perspective of the tabletop objects. Detailed analysis of the video recordings revealed that upper body view was effective for fluid tabletop collaboration because it helped achieve joint perspective and helped estimate the timing and rough location of subsequent tabletop activity.

Author Keywords

Video-mediated communication, remote gesture, tabletop, group-to-group collaboration, upper body view

ACM Classification Keywords

H.5.3 Information interfaces and presentation (e.g., HCI): Group and organization interfaces–Computer-supported cooperative work

General Terms

Human Factors, Experimentation

INTRODUCTION

Tangible tabletops play a significant role in many small-group co-located activities [21]. This results from their abilities to naturally afford collaboration on a shared tabletop workspace while allowing users to maintain awareness of others' actions and intentions [1, 27].

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

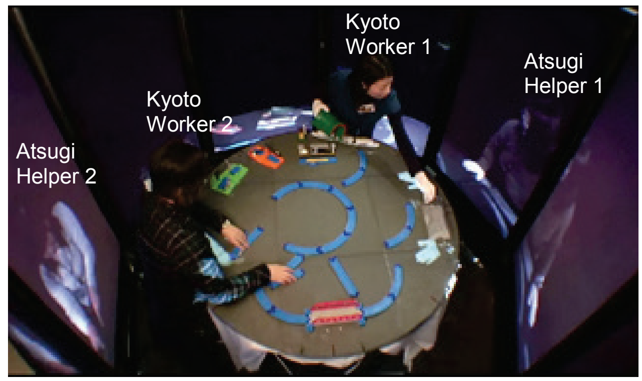


Figure 1. Multiparty fluid tabletop collaboration across distance using t-Room

To gain the same benefits of tabletop activities in distributed environments, a variety of systems have been developed and examined (e.g., [11, 16, 19, 26]). These various systems, however, are built along different lines, particularly whether they show the remote member's upper body images.

While visual information about workspaces (or work objects) are useful in improving communication [8, 14, 29], the advantages of showing visual information about work participants have been less persuasive. Although "upper body" view facilitates interpersonal relationships [23, 28], it seems to add little or no improvement when people are engaging in object-based collaboration [2, 3, 18, 20].

Those studies, which examined the effects of upper body view in object-based collaboration, however, placed restrictions on user positions during collaboration; most distributed tabletop systems do not allow users to change their positions or walk around the table during collaboration. Under such restrictions, it is quite obvious that the upper body view does not contribute to effective collaboration, because people know where others' hands will appear and can easily attribute each other's hands from their fixed positions [25].

Meanwhile, recent studies on tabletops point out the importance of supporting flexible user arrangement around the table [21]. For example, people tend to walk around the table when they experience difficulty reaching items located across from them or when they are attempting to interact with small components. Although flexible user arrangement around the table is common in many real-world tabletop activities (such as planning and design), virtually no one has developed a system that supports such flexible user arrangement across distance; nor has anybody investigated the effects of upper body view in such tabletop activities.

In this paper, we introduce a distributed multiparty tabletop system called t-Room that supports flexible arrangement around a table (Figure 1). By mutually displaying the entire view above and around distributed tabletops, t-Room provides remote collaborators with equal and direct access to the shared tabletop as well as a coherent environment in which to accomplish interaction.

Using this system, we investigate two major issues. First, by comparing the tabletop activities with and without upper body view, we examine whether upper body view improves collaboration when multiple users engage in distributed fluid tabletop collaboration. Although it may sound obvious that upper body view helps users identify the positions of remote users and anticipate their forthcoming actions, it remains unclear how significant that is for effective collaboration. It also remains unclear how people collaborate differently with and without remote user's upper body images.

Second, we examine how people position themselves around the table when they can take arbitrary arrangements across distance. While most video conferencing systems force face-to-face configuration by placing the monitors in front of the local users, studies on distributed tabletop systems suggest an overlapping configuration (i.e., collaborators look as if they are sitting in one another's laps) [5, 12, 25]. Yet we don't know how users position themselves when they can choose arbitrary arrangements during collaboration. It also remains unclear whether users prefer different arrangements between conditions with and without upper body view.

Answering such questions will provide a foundation for designing distributed tabletop systems that support flexible user arrangements around tables.

In the remainder of this paper, we first draw on prior research and predict how remote user's upper body view might influence tabletop collaboration. Next, we present a system called t-Room that enables multiparty tabletop collaboration with flexible arrangement around the table. Then, using t-Room, we describe a laboratory study that compared four-person groups divided across two locations engaging in remote mentoring tasks with and without remote user's upper body view (Figure 2 and 3). We conclude with a discussion of the implications of our findings and issues raised by our study.

PREVIOUS WORK

Shared Tabletops

A considerable body of research has explored how to support collaboration over artifacts in a shared workspace. Studies have significantly demonstrated the value of sharing a visual workspace across distance [3, 8, 12, 14]. For example, Kraut et al. studied pairs working on collaborative repair and suggested that shared visual workspaces facilitated grounding by providing common knowledge of the task artifacts and situational awareness of the workspace [14].

To achieve smoother coordination, research has further demonstrated the importance of embodying gestures into the workspace (e.g., [5, 9, 13]). While various technologies are used to represent remote gestures (e.g., pointers [9] and pen-based drawings [5]), gestures in distributed tabletop systems are typically represented through video-based embodiments [11, 12, 16, 24]. Such embodiments are particularly useful in shared tabletops because they enable all users to directly access artifacts using natural hand gestures [12, 17].

Adding Upper Body View to Shared Tabletop

Besides the utility of a shared tabletop, upper body view has proven useful in clarifying communication ambiguity [28] and supporting certain types of negotiation tasks [23].

To add such utilities of upper body view to distributed tabletop activities, researchers have frequently sought to merge the two utilities into a single system. During its evolution, researchers have suggested the importance of constructing a spatially coherent environment in which to accomplish interaction so that collaborators can exploit daily social skills (e.g., gaze awareness) [7, 10]. More recently, Luff et al. pointed out the importance of providing users with a complete view of a gesture so that they can anticipate the unfolding actions of distant members [17].

Regardless of such progress, most studies still question the value of adding upper body view to a shared workspace [4, 25]. For example, Fussell et al. reported that helpers providing assistance to a remote worker rarely looked up to see the remote worker during assistance in a collaborative physical task [4]. Similar results were also reported in Tang's study in which three people (one person per site) participated in a distributed tabletop activity [25]. They both argue that users can adequately communicate through activities over the shared tabletop, without exploiting upper body view.

These studies, however, are based on systems designed to support distant collaboration with a single participant per site in a fixed position. Tables are also designed relatively small so that everyone can reach all the artifacts without moving around the table. In such a situation, participants know where the hands of others will appear and can easily attribute them from their fixed positions. These features also allow them to effectively achieve "*joint perspective*" during collaboration [29]. Thus, it is understandable why upper body view does not contribute to effective tabletop collaboration.

Fluid Collaboration around the Table

Compared to such fixed tabletop collaboration, fluid tabletop collaboration obscures the remote participant positions, requiring far more effort to understand the ongoing activities and to anticipate the distant members' unfolding actions. Consequently, participants need to pay more attention to each user's position (i.e., who is where) and their locus of attention (i.e., doing what).

Since such awareness cues are likely to be gained through upper body view, we expect that it plays a more significant role in fluid collaboration than in fixed collaboration. To examine the impact of upper body view on distributed fluid tabletop collaboration, we conducted an experiment using a system called t-Room.

t-Room is a distributed tabletop system that supports multiparty fluid interaction around a table across distant sites. A single t-Room consists of a shared tabletop and vertical wall displays (i.e., upper body view) placed around the table. The vertical displays are capable of conveying the awareness cues of distant collaborators' positions and their locus of attention during collaboration.

So far, studies using the system have shown how different seating arrangements across sites affect communication in terms of their sense of unity and satisfaction with the group's decision [30]. Yet it remains unclear whether and how upper body view influences multiparty fluid tabletop collaboration across distant sites.

CURRENT STUDY

Using t-Room, we experimentally investigated the effects of upper body view in a multiparty fluid tabletop collaboration. We were particularly interested in answering the following questions:

- Whether/How upper body view improves task performance when multiple participants engage in distributed fluid tabletop collaboration?
- How do distributed participants position themselves when they can take arbitrary arrangements around the table? Does upper body view interfere with user positioning around the table?

DISTRIBUTED TABLETOP SYSTEM: T-ROOM

t-Room is a room-duplication system that supports distributed tabletop activities. It provides remote collaborators with equal and direct access to a shared tabletop workspace as well as a coherent environment in which to accomplish interaction.

t-Room consists of multiple cameras, screens, a table, and a projector (Figure 2); its walls consist of eight 65-inch LCD panels and High-Definition Video (HDV) cameras placed

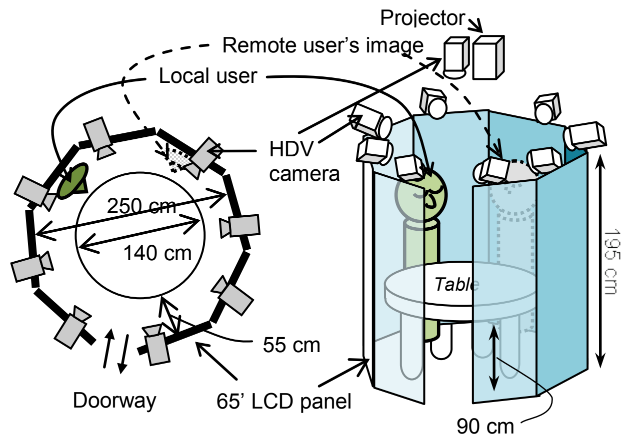


Figure 2. Hardware Design of t-Room: Top-view (left) and Bird's eye-view (right)

above the walls; a round table is located at the center of the room, with a HDV camera, and a projector.

Local activities above and around the table are captured through these cameras and displayed on the remote tabletop as well as on the LCD panels in a remote t-Room. The arrangements of video cameras and screens are configured to maintain spatial relationships between distant sites [22]. Local sounds are also captured with wireless microphones attached to each user and delivered to the remote participants' wireless ear phones. Audio is not spatialized in this study so that we can screen out audio factors and examine only the visual effects of the participants.

By introducing a space in between the table and the side wall panels, t-Room enables the participants to take arbitrary arrangements around the table. In return for such flexibility, however, this space introduces a discontinuity between the remote users' upper body and their hand gestures displayed on the tabletop. To mitigate such discontinuity between the side walls and the tabletop, a small area of the tabletop (rim of tabletop) is presented on the side wall panels in addition to the remote user's upper body images [17]. Such adjustment is expected to "*provide a resource to see hands moving just above a tabletop and from the projected surface, and hence for tying the display of conduct in one space with that in another*" [17].

METHOD

We based our study around four-person collaboration to go beyond existing 1-on-1 studies; although tables are inherently designed to support group work, the vast majority of studies on distributed tabletops have explored pairs (with a few exceptions like [25]). As mentioned in [25], "the number of collaborators working together exponentially increases the complexity of possible interactions, increasing the likelihood of misinterpretation and misunderstandings." Thus, we were interested in knowing whether/how upper body view would improve such collaboration.

Experimental Design
In this study, four-person groups participated in two collaborative tabletop tasks with different media conditions (Figs. 3 and 4): (I) a shared tabletop and (II) a shared tabletop plus upper body view. The order of the two media conditions and the two tasks (used in each media condition) were counterbalanced across groups.

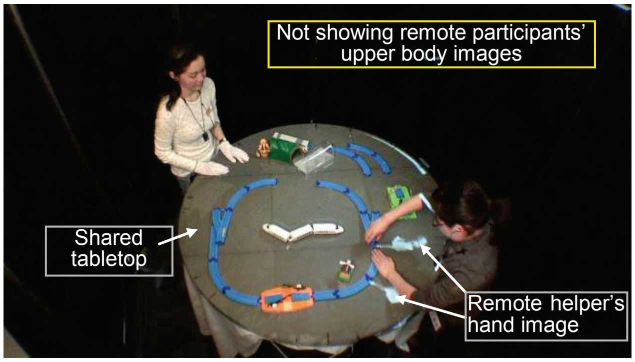


Figure 3. Media Condition I: Four collaborators working around a shared tabletop (*without* upper body view)

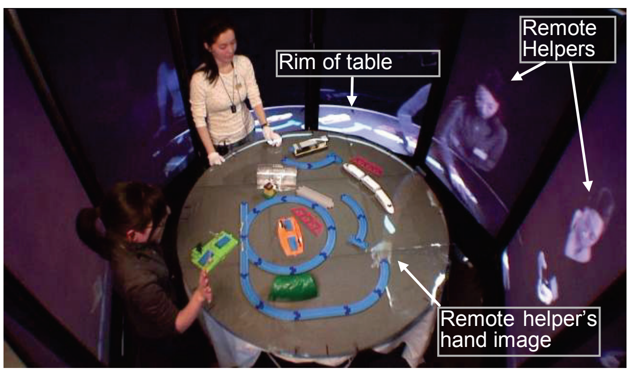


Figure 4. Media Condition II: Four collaborators working around a shared tabletop (*with* upper body view)

Apparatus
Two identical t-Rooms were installed in the cities of Atsugi and Kyoto, which are approximately 150 miles apart. The two rooms are connected by a gigabit network. The network delay for video and audio transmission between the two cities is around 0.3-0.4 and 0.2-0.3 seconds, respectively; audio and video are not synchronized.

Participants
Eight groups of four (32 participants) were recruited for this study and were paid for their participation. To screen out extraneous factors that may influence the results, we recruited only females who did not know each other; we chose females to avoid gender factors that might have interfered with participant positions around the table. Choosing strangers allowed us to avoid familiarity factors that might also have interfered with collaboration efficiency. Participants had never used a video-conferencing system or a shared tabletop system prior to the experiment.

Task
The experiment was run with two participants collocated in one location (Kyoto) and two in another location (Atsugi). Each collocated group was assigned a separate role: the Kyoto group was the worker group and the Atsugi group was the helper role. Those assigned the role of worker were asked to assemble a railway kit. Yet, the final completion maps (Figure 5) were disclosed only to the helpers. Although the workers had physical access to the actual components, the helpers could only access them virtually. One helper and one worker were then randomly paired to form helper-worker pairs who worked with different pieces: blue track or scenic pieces. This was a weak constraint (i.e. participants were allowed to give/receive help beyond pairs when required) aimed for reproducing an intensive working scene and preventing any one of them from being there doing nothing.

We prepared two tasks (i.e., final completion maps) for this study, and each used a different media condition. To even out the difficulty levels between the two tasks, we adjusted the maps prior to the experiment by equalizing their completion times; it took approximately eight minutes for four collocated users to complete each task.

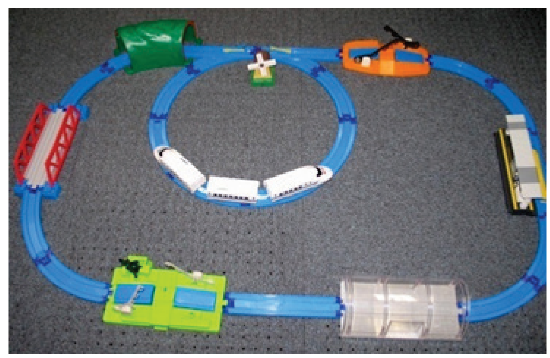


Figure 5. Railway kit used in our task

This task mimics a collaborative designing task (specifically, urban planning) in which users frequently change their positions during collaboration [21]. It also falls within a general class of "mentoring collaborative physical tasks [4, 13, 15]," in which helpers typically instruct workers to assemble artefacts into a particular arrangement. Since the task encompasses a variety of generic task elements like item selection, pattern matching, physical manipulations and error checking [12], it offers the opportunity to explore some of the demands that may be placed on real world applications in tabletop activities using real objects.

Procedure
Step (1): On arrival, participants in the same group were placed in separate rooms to avoid getting acquainted before the experiment. Next they completed experimental consent forms and moved to the t-Room. Participants were given a brief introduction to the study and the t-Room system and were introduced to each other.

Step (2): Participants were randomly divided into two helper-worker pairs: the "rail pair" and the "scenic pair." The scenic pair wore white gloves so that participants could immediately distinguish their remote partners by looking at the hand images displayed on the shared tabletop [25]. Each pair was given an opportunity to practice giving and receiving instructions using the railway kit pieces.

Step (3): The following procedure was repeated two times with different media conditions:

> Step (3-1): Helpers were asked to temporarily leave the t-Room. Outside the t-Room, the helpers were provided with a final completion map and given two minutes to memorize their parts so that they could instruct the workers without looking at the map. This process eliminated the possibility of workers viewing the final map by the t-Room.
>
> Step (3-2): Participants were given 20 minutes to complete each task and were told the following: a) their task was to assemble the rail kit so that it matches the final map; b) they should basically work within pairs but they can help others as required; c) they could use any strategy to accomplish the task; and d) they should complete the task as quickly as possible. They received a bonus if they finished it within 10 minutes.
>
> Step (3-3): Following the completion of the task, participants completed post-task questionnaires about the work they had just experienced.

Step (4): Next, participants completed a final questionnaire on the effects of the upper body view. Finally, they were interviewed about their different experiences while working between the two media conditions.

Measures

Video Recordings. Each session was video-taped, and we logged participant activities inside each t-Room. The video data were used for analyzing participant standing positions around the table and for detailed analysis on how upper body view influenced collaboration.

Task Performance. Since the participants were instructed and motivated to complete the task as quickly as possible, task performance was the time required to complete the task.

User Position. To see how each pair positioned themselves around the table, we developed a coding scheme before the experiment; we divided and marked the table rim into eight equal segments and used the marks as a scale to measure the distance between each helper-worker pair. Every ten seconds from the beginning of each session, we classified each pair's positional relationship into one of four categories based on their distance: overlap (*Distance.<1*), one-hop (*1≤Distance.<2*), orthogonal (*2≤Distance.<3*), and opposite (*3≤Distance.<4*). For instance, the distance between Helper 1 and Worker 1 in Figure 1 is categorized as "one-hop," and the distance between Helper 2 and Worker 2 is categorized as "overlap."

Questionnaires and Interviews. In the post-task questionnaires, participants were mainly asked about the ease of conducting the task, awareness of the remote participants, and the reasons for walking around the table. In the final questionnaire, participants were asked about their preference of media conditions and the differences in their standing positions across media conditions. Finally, in the post-experimental interviews, we asked the participants about any differences they noted between the two conditions and possible reasons for them. These interviews also helped explain some observed events during the sessions and to guide further research.

RESULTS

The data were first analyzed to examine the effects of upper body view on task performance. A second analysis was then carried out to determine whether upper body view made a difference in the participants' standing positions.

Task Performance

To see whether upper body view aided tabletop collaboration, we compared the completion times in a repeated measures analysis of variance (ANOVA), using task and media conditions as repeated factors.

As shown in Figure 6, completion times differed significantly across media conditions ($F(1,15)=9.91$, $p<.01$). Participants in the upper body view condition completed the tasks considerably faster than those without upper body view. There was no significant main effect for task and no significant interaction between the task and media conditions.

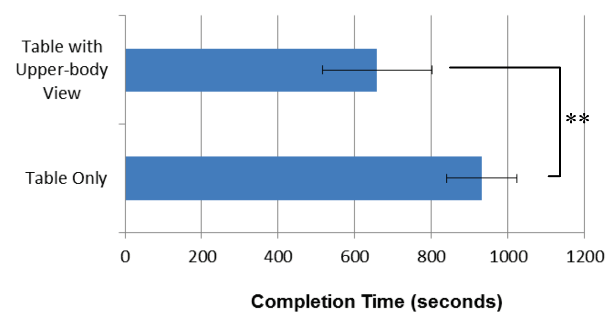


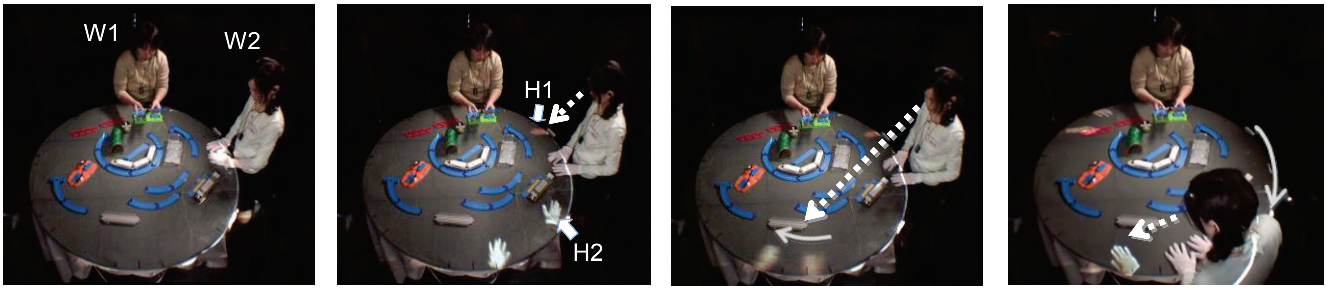
Figure 6. Completion times by media condition (sec.): "" denotes significant difference at p<.01 level**

Positioning between Helper and Worker

To investigate how helper-worker pairs positioned themselves, we classified each helper-worker's positional relationship of every ten seconds into one of four categories: overlap, one-hop, orthogonal, and opposite.

Table 1 Helper-Worker Positional Relationship in Each Media Condition

	Overlap	*One-hop*	*Orthogonal*	*Opposite*
Table Only	**33.4%**	20.9%	26.1%	21.7%
	∨	∧		
Table + Upper body View	22.9%	**30.8%**	25.1%	20.7%



a b c d

```
H1: sorewo eh-tto chotto (0.8) eh-ttodesune
     they   uum   a little      uum
                               *a
H2:                                         Hyakkei sa:n
                                            Hyakkei Miss  (3.1)
                                                 *b    *c     *d
```

Figure 7. Helper leading worker's perspective by hands. In this script, the first line is Japanese and the second is a word-by-word gloss. "*" indicates time when corresponding image is captured.

Table 1 shows the average proportion of the helper-worker positional relationship in each of the four categories for each media condition. The positional relationships between the helper-worker pairs were significantly different across media conditions (confirmed by Chi-square test: χ^2=24.28, p<.01).

When an upper body view was not available, as shown in Table 1, helper-worker pairs chose the overlapping position the most. Helpers approached the workers and gave them instructions in an overlapping position, particularly when the workers could not immediately understand their instructions. In contrast, helper-worker pairs chose the "one-hop position" the most in condition with upper body view. Workers tended to take a step away from the helpers and glanced at them when asking questions (e.g., Worker 1 in Figure 1) or when they finished the work at hand and were waiting for subsequent instructions.

Summary of Experiment Findings

Although previous studies questioned the value of adding upper body view to a shared tabletop, our results clearly show that upper body views exerted a significant influence on multiparty fluid tabletop collaboration across distance. From our experiment that compared four-person distributed tabletop collaboration between two media conditions (i.e., shared tabletop vs. shared tabletop plus upper body view), we found the following: (1) task performance significantly improved when upper body view was added to the shared tabletop; (2) helper-worker pairs selected overlapping positions the most in the shared tabletop condition and tended to take a step away from each other in the condition with upper body view.

Even though previous research reported that users rarely looked at remote user's upper body view during tabletop collaboration [4, 25], the participants in our study often looked up to see the remote participants. Indeed, participants with upper body view tended to take a step away from their remote partners so that they could effectively glance at them while taking a comparable perspective of the tabletop objects.

To better understand how upper body views improved fluid tabletop collaboration, we closely reviewed the interview notes and video recordings of our experiment.

HOW UPPER BODY VIEW HELPED COLLABORATION

First, the interview notes provided a hint to consider how participants exploited the upper body view. When they were asked about their impressions of the media use during the experiment, most responded that they experienced some confusion or anxiety during collaboration without upper body view. Additional video analysis, based upon such interview notes, revealed why and how upper body view helped collaboration. In this section, we focus on participants' large movements over and around the table, since they were limited/prohibited in previous research.

Anticipating the Subsequent Working Area

Based on post-experiment interviews, we found that upper body view played a significant role in anticipating the subsequent working area, particularly during a large shift in their working spaces. For example, a worker without upper body view mentioned the difficulties of anticipating the forthcoming working area when her partner gave instructions right after walking around the table. Furthermore, several participants mentioned that they intentionally put their hands on the table to compensate for the absent positional cues when upper body view was unavailable. The hand images seemed to help them reveal their own position and to attract the attention of the remote participants. A typical example of such cases is provided in Figure 7.

In this fragment, W1 is connecting two pieces of track, as instructed by H1, and W2 is waiting for the next instruction from H2. Prior to this fragment, H1 was standing across from W1 and H2 was standing beside W2 (in between W1 and W2), with their hands on the table. However, as H1 says "they, uum, a little, uum," H1 and H2 start to walk around the table and their hands disappear from the tabletop (Fig. 7a); H1 approaches W1 to tell her where to connect the tracks, and H2 walks away from W2 to move to the next instruction. As H2 walks around the table, she glances at W2's hand image and notices that W2 is standing still at the same position. H2 soon puts both of her hands back on to the table and calls W2 by name: "Miss Hyakkei" (Fig. 7b). She then drags her hands along the edge of the table (Fig. 7c) so that W2 can follow her hand movements. W2, on the other hand, first glances at H1's hands when her name was called (Fig. 7b), but soon realizes that they are not her partner's hands because they do not have gloves. Then she notices H2's hand images moving on her left (Fig. 7c) and starts following them (Fig. 7d).

The fragment shows how H1 eventually succeeded in leading W1's perspective to the next working area with her hands. However, the process is quite laborious and time consuming. In contrast, participants with upper body view could easily anticipate the subsequent working area by looking at their remote partners on the side wall displays. For example, if a worker sees the helper walk around the table, she could anticipate that the helper will soon give her instructions near the end point. This awareness also helps the worker anticipate subsequent instructions and prepare for them. The example below captures this tendency.

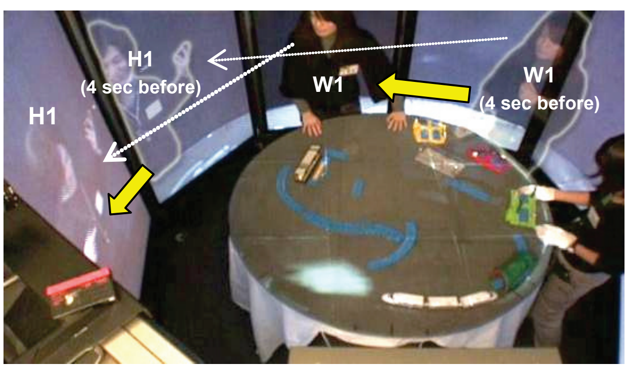


Figure 8. Worker anticipating and preparing for next instruction. Participants with fading colors and white outline shows their past figures and yellow arrows indicate their movements. White dotted arrows indicate W1's line of sight.

Figure 8 shows a scene where a worker (W1) follows her partner (H1) by looking at her upper body image. In this scene, W1 is monitoring H1 while waiting for the next instruction. Although H1 has not said anything yet, W1 senses from H1's upper body movement (i.e., walking around the table while gazing at a component on the table) that she will provide an instruction soon. W1 follows H1 while maintaining a certain distance from H1 so that she can keep monitoring H1 and quickly respond to forthcoming instructions.

Upper body view seemed to allow the participants to effortlessly lead their remote partners to the subsequent working area, achieve joint perspective, and smoothly move to the next activity. Participants without upper body view had to compensate their positional cues by continuously placing their hands on the table.

To examine whether such a behavioral difference (i.e., placing their hands on the table) was significant between conditions, we developed the following coding scheme: every ten seconds from the beginning of each session, we classified each participant's hand positions based on whether they were placed on the table. We used the remote tabletop as a criterion to classify each user's hand position; when the hands or even part of a hand was visible from the remote tabletop, we classified it as a "hand placed on the table."

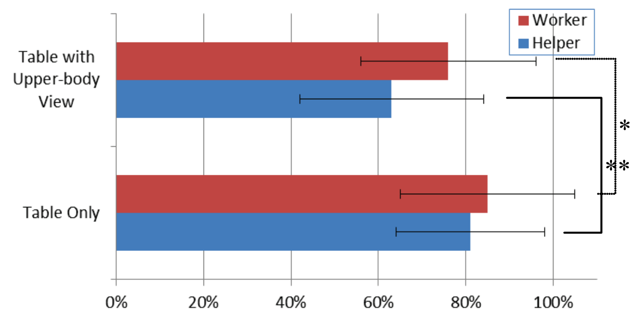
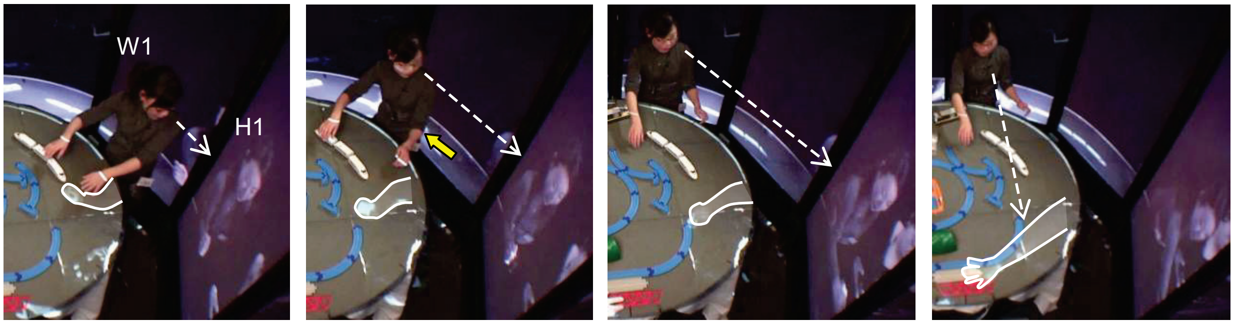


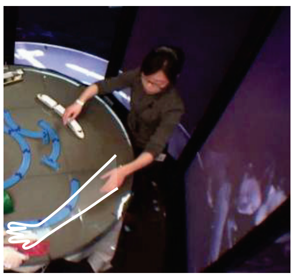
Figure 9. Mean proportion of participants placing their hands on table (conveying positional cues) by experimental role by media condition. "*" and "" denote significant differences at p<.05 and p<.01 levels.**

Figure 9 shows the mean proportion of hands placed on the table per task by conditions for each experimental role. Results indicate that both workers and helpers placed their hands more frequently in the condition without upper body view. We performed a repeated measures ANOVA on the proportion of hands placed on the table, using task and media condition as repeated factors. Results showed a significant main effect for media condition (Helper: $F(1,31)=22.61$, $p<.01$; Worker: $F(1,31)=6.76$, $p<.05$) but no main effect of task or interaction.

Regardless of such participant efforts in conveying their positional awareness in the condition without upper body view, their task performance was still significantly lower than those with the upper body view (Figure 6). An immediate suspicion that arises from the numerical result is in the sporadic drop-off of their positional cues during collaboration. From Figure 9, we see that participant appearances were completely invisible about 20% of the task time in the condition without upper body view. Although positional cues during migration are supposedly more important than those during fixed collaboration, participants tended to withdraw their hands while moving around the table (Figure 7a and H1 in Figure 8). We infer that such lack of positional awareness during migration is



a b c d



e

```
H1: kono(0.4)re:ru nandesukedo:
    This     track   regarding
W1:                               ha:i (1.8)
                                  yes
                                     *a   *b

H1: korewo ikkai totte kudasai kono (0.2) mijikai yatsudesu
    This   for now  detach      this       short    one
    *c      *d                               *e
```

Figure 10. Worker estimates timing and location of a helper's gesture. White dotted arrows indicate W1's line of sight and a yellow arrow indicates W1's movement. For legibility, we traced H1's gesture.

one reason why collaboration without upper body view was inefficient.

Finally, we point out the benefits of conveying gaze awareness, head directions, and body postures through upper body view, particularly during a large shift in their locus of attention over the table. We elaborate on this issue in the following section.

Anticipating the Remote Partner's Locus of Attention

Drawing our attention back to the interview comments, we found that participants without upper body view had difficulty following the remote partner's locus of attention, especially when there was a big leap in their attention over the tabletop. Indeed, participants mentioned that they had to concentrate on the tabletop surface more when upper body view was unavailable because they could not estimate the approximate location of the referent.

Although this problem was found in both conditions, it seemed less frequent and easily resolved when an upper body view was available. In particular, the helper's gaze, head directions, and posture on the side wall displays seemed to help the workers estimate the timing of when the helper's gestures would appear on the tabletop and also to help them infer the rough location of the forthcoming referent.

In Figure 10, the remote helper (H1) is asking the worker (W1) to detach a short piece of track located next to the red bridge. As H1 starts her instruction (saying "Regarding this track"), she leans forward and gazes the target track on her left. W1 who is standing close to H1 immediately provides a short acknowledgement to H1 and turns to her left (Fig. 10a). While reaching out her hands toward the unused tracks (which is not the one H1 intended), she takes a step away from H1 to take a better look at her upper body (Fig. 10b). When W1 is about to reach the unused track, H1 extends her arm towards the target piece saying "this one" (Fig. 10c). Since W1 was monitoring H1 on the side wall display, W1 is capable of anticipating when the H1's gesture appears on the tabletop. From H1's posture, W1 is also capable of anticipating the rough location of the target track. This enabled W1 to instantly shift her attention from H1's upper body to the appropriate area on the table (Fig. 10d) allowing her to effectively follow H1's instruction (Fig. 10e).

Design Implication

In this study, we used t-Room as a test bed to investigate the research questions stated earlier; we are not trying to argue that t-Room is necessary for effective collaboration. Consequently, our findings and the above analysis suggest some important implications that might be useful for designing future systems on fluid tabletop collaboration across distance.

- A coherent space must be constructed in which spatial relationships are maintained across distance. By doing so, people can observe the remote participants' standing position, gaze, and head direction in relation to the tabletop objects, allowing them to effectively achieve joint perspectives during multiparty fluid tabletop collaboration.

However, most existing distributed tabletop systems are not equipped with vertical displays like t-Room. Thus, we consider some implications that might be useful for those systems.

- *Conveying Positional Cues*: Participants in our study tried to convey their positional cues by placing their hands on the table. By doing so, participants were able to lead the perspectives of their remote partners and to achieve joint perspective with them. One way to convey such positional cues without using upper body view is to constantly show each remote participant's standing position on the tabletop rim by some kind of graphical representation.
- *Increasing the Visibility of Remote Gestures*: Since it is probably impractical to show various upper body cues (including gaze, head direction, and posture) through the tabletop, we suggest a different approach that might be useful for following the remote participant gestures: increasing the visibility of the remote gestures. One way to do this is to augment the remote participants' gesture images so that participants can find the referent even if they fail to follow the remote gestures. For example, trace pearl [25] could be useful, which tracks each point of contact for each user with a trail that fades after a few seconds.

CONCLUSIONS

Our study is significant in two ways. First, our study is the first to show a clear benefit (using objective measures) of upper body view in object-based collaboration across distance. While previous work has generally found that people do not use nor benefit from the upper body view, our study showed that people engaging in multiparty fluid tabletop collaboration did use and benefited from the upper body view: from our experiment that compared four-person distributed tabletop collaboration between two media conditions (c.f., shared tabletop vs. shared tabletop plus upper body view), we found out that (1) task performance was significantly higher in the condition with upper body view; (2) helper-worker pairs took overlapping positions the most in the shared tabletop condition, but they tended to take one step away from each other in the condition with upper body view.

Second, our study shows how upper body view helped collaboration and how it influenced collaboration; from our analysis, we identified two main reasons why upper body view effectively improved their task performance. First, participants with upper body view could effortlessly lead their remote partners to the subsequent working area, achieve joint perspective, and smoothly move to the next activity by monitoring their movement around the table. Second, by conveying gaze, head direction, and the posture of remote participants, upper body view helped the participants estimate the timing of when the helper's gestures would appear on the tabletop and also helped the workers infer the rough location of the forthcoming referent. These cues seemed particularly useful when there was a large shift in the remote participant's locus of attention. We speculate that participants took a step away from their remote partners so that they could effectively attain those cues from their upper body view while taking a comparable perspective of the tabletop objects.

Although we successfully showed the benefits of upper body view in multiparty fluid tabletop collaboration, further study is needed to understand its limitations. Task performance was still significantly lower than face-to-face tabletop collaboration; three groups of four collocated people (different from our study's participants) performed the same task in a single t-Room and completed it substantially faster (43% on average) compared to distant collaboration with upper body view. Many factors such as the flatness of the display [6], occlusion, transmission delay [9], the spatial gap between the side wall displays and tabletop displays [17] might account for the differences. Further investigation is required on how the combinations of these factors affect task performance. We also need to investigate how participants' fluidity around the table impact their collaboration (including task performance) by comparing conditions between fixed and flexible positions around the table.

ACKNOWLEDGMENTS

This work is partially supported by Grant for Scientific Research (C) 20500122. We gratefully thank Prof. Paul Luff for his insightful comments on this work. We would also like to thank the t-Room members for their support in setting up the experiment.

REFERENCES

1. Buxton, W. Integrating the Periphery and Context: A New Model of Telematics. *Proceedings of GI'95*, ACM Press (1995), 239-246.
2. Chapanis, A., Ochsman, R. B., Parrish, R. N., and Weeks, G. D. Studies in interactive communication: The effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors, 14*, (1972), 487-509.
3. Fussell, S.R., Kraut, R.E., & Siegel, J. Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of CSCW'00*, ACM Press (2000), 21-30.
4. Fussell, S. R., Setlock, L., and Parker, E. M. Where do helpers look? Gaze targets during collaborative physical tasks. *CHI'03 (Extended Abstracts)*, (2003), 768-769.
5. Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E. and Kramer, A. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction, 19*, (2004), 273-309.
6. Gaver, W. The affordances of media spaces for collaboration. *Proceedings of CSCW'92*, ACM Press (1992), 17-24.

7. Gaver, W., Sellen, A., Heath, C., and Luff, P. One is not enough: multiple views in a media space. *Proceedings of Interchi'93*, ACM Press (1993), 335-341.
8. Gergle, D., Kraut, R.E., and Fussell, S.R., Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. *Journal of Language and Social Psychology, 23(4)*, (2004), 491-517.
9. Gutwin, C. Improving interpretation of remote gestures with telepointer traces. *Proceedings of CSCW'02*, ACM Press (2002), 49-57.
10. 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.
11. Ishii, H. and Kobayashi, M. ClearBoard: a seamless medium for shared drawing and conversation with eye contact. *Proceedings of CHI'92,* ACM Press (1992), 525-532.
12. Kirk, D., Crabtree, A., and Rodden, T. Ways of the hands. *Proceedings of ECSCW'05*, Springer (2005), 1-21.
13. Kirk, D. S. and Fraser, D. Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. *Proceedings of CHI'06*, ACM Press (2006), 1191-1200.
14. 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. *Proceedings of CSCW'02*, ACM Press (2002), 31-40.
15. Kraut, R. E., Miller, M. D., and Siegel, J. Collaboration in performance of physical tasks: Effects on outcomes and communication. *Proceedings of CSCW'96*, ACM Press (1996), 57-66.
16. Kuzuoka, H., Yamashita, J., Yamazaki, K. and Yamazaki, A. Agora: A Remote Collaboration System that Enables Mutual Monitoring, *Extended Abstracts of CHI'99*, ACM Press (1999), 190-191.
17. Luff, P., Heath, C., Kuzuoka, H., Yamazaki, K., Yamashita, J. Handling documents and discriminating objects in hybrid spaces. *Proceedings of CHI'06*, ACM Press (2006), 561-570.
18. Matarazzo, G. & Sellen, A.J. The role of video in work at a distance: Addition or distraction? *Behaviour and Information Technology, 19 (5)*, (2000), 339-348.
19. Minatani, S., Kitahara, I, Kameda, Y. and Ohta, Y. Face-to-Face Tabletop Remote Collaboration in Mixed Reality. *Proceedings of ISMAR'07*, (2007), IEEE Computer Society, 43-46.
20. 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. *Proceedings of Interchi'93*, ACM Press (1993), 327-334.
21. Scott, S. D., Grant, K. D., and Mandryk, R. L., System guidelines for co-located, collaborative work on a tabletop display. Proceedings of ECSCW 2003, Kluwer Academic Publishers (2003), 159-178.
22. Sellen, A. Buxton, B. and Arnott, J. Using spatial cues to improve videoconferencing. *Proceedings of CHI'92*, ACM Press (1992), 651-652.
23. Short, J., Williams, E., & Christie, B. The social psychology of telecommunications. London: Wiley, (1976).
24. Tang, A., Neustaedter, C., and Greenberg, S. VideoArms: embodiments for mixed presence groupware. *Proceedings of British-HCI'06*, Springer (2006), 85-102.
25. Tang, A., Pahud, M., Inkpen, K., Benko, H., Tang, J. C., and Buxton, B., Three's Company: Understanding Communication Channels in Three-way Distributed Collaboration. *Proceedings of CSCW'10,* ACM Press, (2010), 271-280.
26. Tang, J. and Minneman, S. VideoDraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems 9,* 2 (1991), 453-469.
27. Tang, J. Findings from observational studies of collaborative work. *Man-Machine Studies 34*, 2 (1991), 143-160.
28. 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. *Proceedings of CHI'99*, ACM Press (1999), 302-309.
29. Whittaker, S. Things to talk about when talking about things. *Human-Computer Interaction, 18, 1*, (2003), 149-170.
30. Yamashita, N, Hirata, K, Aoyagi, S, Kuzuoka, H, Harada, Y. Impact of Seating Positions on Group Video Communication, Proceedings of *CSCW'08*, ACM Press (2008), 177-186.