Three's Company: Understanding Communication Channels in Three-way Distributed Collaboration

Anthony Tang[1], Michel Pahud[2], Kori Inkpen[2], Hrvoje Benko[2], John C. Tang[2], Bill Buxton[2]

[1]Human Communication Technologies Lab
University of British Columbia, BC, Canada

[2]Microsoft Research
Redmond, WA, USA

tonyt@ece.ubc.ca {mpahud | kori | benko | johntang | bibuxton}@microsoft.com

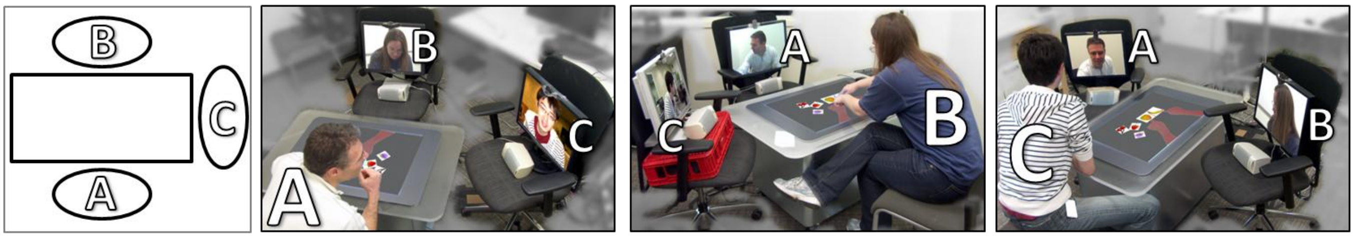

Figure 1. Three collaborators working around a shared tabletop (left: overhead schematic). In each physical space, remote participants are embodied as surrogates (with display, camera, microphone and speaker). Note that the spatial relationships are preserved in this setup. As B works in the space, her arm shadows are propagated to remote surfaces.

ABSTRACT

We explore the design of a system for three-way collaboration over a shared visual workspace, specifically in how to support three channels of communication: person, reference, and task-space. In two studies, we explore the implications of extending designs intended for dyadic collaboration to three-person groups, and the role of each communication channel. Our studies illustrate the utility of multiple configurations of users around a distributed workspace, and explore the subtleties of traditional notions of identity, awareness, spatial metaphor, and corporeal embodiments as they relate to three-way collaboration.

Author Keywords

Shared workspace, tabletop, media space, video-mediated communication.

ACM Classification Keywords

H5.3. Group and Organization Interfaces: *Computer-supported cooperative work*..

General Terms

Design, Experimentation, Human Factors

INTRODUCTION

In designing systems to support distributed collaborative activity, researchers have frequently sought to merge the utility of video conferencing systems with collaborative workspaces (e.g., [32,16]). Video conferencing systems support interpersonal interaction, allowing distributed users to discuss ideas while also supporting non-verbal interaction such as facial expression, body language and gesture (e.g., [3,7,25,29,22]). On the other hand, collaborative workspaces enable co-workers to share artifacts and data that constitute or support the purpose of meetings (e.g., [13,30]). These bodies of work enable rich remote collaboration, each focusing on a specific aspect or channel of interaction [5]: video-conferencing systems enable *person space* for relationship and trust development [3], while shared collaborative workspaces are a *task space*, where work is accomplished. In this work, we consider the intersection of these domains, articulating and exploring the communicative and coordinating role of *reference space* [5]—mechanisms allowing collaborators to reference, point, and relate with one another in task space.

Our interest is in supporting collaboration in distributed teams. Industry attention has primarily focused on video conferencing, from consumer-level webcams to dedicated video conferencing installations (e.g., Cisco TelePresence). Yet there is still a pressing need to better support collaboration over digital artifacts associated with the meeting (e.g., documents, diagrams) [35]. The focus of the present work is to explore collaborative activity over connected digital tabletops which support a shared sense of presence amongst people and artifacts associated with the task. In particular, we focus on supporting users' ability to employ a rich gestural vocabulary to directly refer to various aspects of the artifacts. Through the iterative design and study of such a system, we bring attention to issues of configuration, workspace awareness, and coordination. In particular, our studies point to two conclusions:

CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA.

- Person and reference space may be spatially disjoint, so task characteristics can be used to determine the optimal configuration;
- The design of mechanisms to support reference space should consider issues of fidelity, identity, and saliency.

We arrive at these conclusions based on studies of three-person, distributed collaboration. As illustrated in Figure 1, each workstation employs a multi-touch tabletop surface as a collaborative workspace, as well as two LCD monitors—physical surrogates for the remote collaborators (e.g., [29,20]). Each surrogate includes a camera, speaker and microphone which correspond to the remote person's eyes, mouth and ears. Additional cameras above the tabletops capture images of the users' arms as they move over the workspace, transmitting them to remote workstations. Varying the spatial configuration of the surrogates and where the shadow of remote hand and arm gestures appear on the tabletop workspace revealed participant behaviour that underscored the utility of different configurations, as well as aspects of the shared reference space.

The vast majority of explorations into distributed collaboration employing video conferencing and shared workspaces has studied pairs (e.g., [16,17,18,21,26,32], cf. [2,30,37]). It is unclear whether these designs scale beyond two users, and indeed, it seems like many will not. The number of collaborators working together exponentially increases the complexity of possible interactions, increasing the likelihood of misinterpretation and misunderstanding. We chose to base our studies around 3-way meetings as a means of extending beyond the dyad while still keeping things practically manageable. While these findings may not immediately scale beyond groups of three, they identify directions for further study regarding group size.

This paper makes three contributions: first, we articulate the role of reference space in distributed collaboration (as distinct from person and task space) in a way that builds on [5]; second, we report on studies of three-person collaboration, identifying subtle issues in supporting reference space that are less pertinent in dyadic scenarios; and finally, we show how spatial configurations of person and reference space affect behaviour in task space.

In the next section, we outline prior work that provides the theoretical foundation for this work, drawing on the same literature to articulate a vocabulary that describes the components of reference space. Next, we discuss the design of the system we employed to study distributed collaboration. We then describe our iterative studies of the system, drawing on observations from those studies to illustrate the points outlined earlier. Finally, we discuss the implications of this work and outline plans for future work.

BACKGROUND

We situate our work within a long lineage of research that explores the intersection of video media spaces and shared interactive workspaces [1,7,14,22].

Video Media Spaces

Originating primarily in the seminal work of Xerox PARC/EuroPARC, BellCore, US West and the University of Toronto, video media spaces (VMSs) were borne out of the need to connect distributed collaborators (e.g., [1,7,14,22]). VMSs provided collaborators with always-on audio and video connections, which were as much an augmentation of physical architecture as telephony. Having established a shared space, to speak with a collaborator, one only needed to glance over and begin talking. As exemplified in [16, 29], the spatial configuration of these spaces was important if collaborators were to be able to exploit everyday social skills, such as those based on gaze awareness. Because time and spatially-multiplexed monitors inhibit gaze awareness [6], the evolution of these spaces has begun to transition to more tangible surrogates of remote collaborators (e.g., [29,20]), which we mimic.

Shared Visual Workspaces

A considerable body of research has explored how to support collaboration over artifacts in a shared workspace. For instance, [13] explore how augmenting conventional desktop applications with shared workspace awareness tools (awareness of others' interactions in the workspace) facilitate smoother workspace coordination. The concept of workspace awareness encompasses notions of presence (is Janet present), communicative gesture (pointing and deictic reference), consequential communication (background awareness of others' interactions), and feedthrough (equivalent of feedback, but sent to remote parties).

These aspects of workspace awareness are somewhat implicit in the design of several video-based shared workspaces [32,16,37,21,36] where videos of remote and local spaces are fused into a single workspace. In many systems, video cameras also capture remote collaborators' arms and bodies, providing a rich sense of embodiment and presence, as well as communicative context.

Of interest here, the systems employed fairly diverse spatial metaphors with regard to placement of collaborators around the virtual workspace. ClearBoard [16] employed an elegant "separated by glass" metaphor, while Agora [37], and its successor [21] employed physical table metaphors, placing collaborators around or on opposite sides of a table. VideoDraw [32], VideoDesk [19] and C-Slate [17] depart from these physical metaphors, placing collaborators' video surrogate in a face-to-face configuration, but configuring the workspaces so that users shared the same perspective (i.e., they effectively sat in one another's laps). Our first study explores the consequences of these spatial metaphors.

These shared spaces, of course, are distributed analogs to our everyday collocated workspaces (i.e., tabletops). Researchers have also studied collocated collaboration, and for example, how the design of digital tabletop software can support or inhibit coordination [24,15,28]. For instance, [31,28] observed that in non-digital tabletop collaboration, users partition work artifacts both semantically and spatially to help coordinate activity. Our first study also

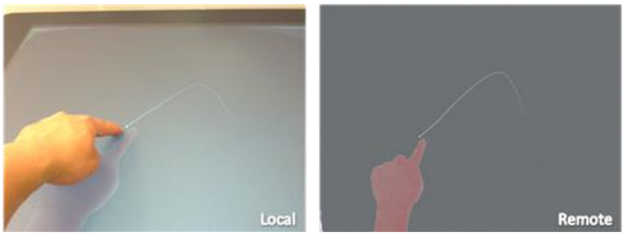

Figure 2. Arm shadows are displayed locally as feedback, as well as at remote sites. Finger contacts are transmitted to remote sites, and conveyed via trace pearls: the contact point is represented by a small disc, and the trail fades over time.

explores this partitioning practice in the context of a distributed tabletop system.

Embodiment in Distributed Tabletops

The video-based embodiments of collaborators' arms from works like [16,19,32] have received renewed interest (e.g., [18,26,30,33,36]), in part due to a growing literature on distributed tabletop workspaces. Motivated by Tang's original study of collaborators in collocated, paper-and-pencil design activities [31], these embodiments were postulated to aid communication and coordination in similar distributed workspaces. In principle, users' hand and arm gestures, learned over a lifetime of day-to-day interaction, would be supported by such embodiments.

As described by [18] and [30], these embodiments provide many such communicative functions in distributed workspaces: drawing attention, supporting mimicry, a means for deictic reference, and so forth. Such embodiments also provide an important coordinating resource when multiple users occupy a shared space (e.g., [30,33,37]), though their use to coordinate territorial behavior has been questioned [33]. We address this idea in our second study.

An equally pressing concern is whether the arm embodiments (being in the reference space) need to be "connected" somehow to the video-based embodiment of a remote collaborator in person space. Luff et al. for example, suggest that the interpretation of a gesture is not only of its final manifestation, but of the entire *production* of a gesture [21]. The immediate implication of this work raises the question of how connected the person space need to be with reference space. Several designs have taken painstaking efforts to mimic and reproduce real-life spatial metaphors, keeping reference and person space connected (e.g., [21,37]). Other designs have clearly placed priority on the document or workspace (e.g., [32,17]), breaking the spatial relationship between the workspace embodiment of a collaborator and his video-based embodiment. In continuing this line of inquiry [26], we explore the benefits and drawbacks of different configurations in our first study.

REFERENCE SPACE AND SYSTEM DESIGN

Buxton [5] outlines a technology-independent vocabulary to describe video conferencing and shared workspace systems. In particular, he suggests three distinct types of spaces for distributed collaboration:

- *person space*: where verbal and facial cues are used for expression, trust and gaze—typically realized as video and audio connections;
- *task space*: where the work appears—typically realized through a shared workspace application, and
- *reference space*: where remote parties can use body language to refer to the work—often realized as mouse pointers, though also as video embodiments of arms.

This conception separates distributed collaboration into three (relatively) independent components, allowing researchers to focus on one or more of these spaces. In this section, we outline the design of our own system, paying particular attention to our designs for supporting reference space. The discussion underscores several design requirements we derived from prior work.

Design Requirements for Reference Space

The following design requirements were derived from previous research (e.g., [30]) and our exploration with several prototype systems.

Support foreground "use" of reference space. The embodiment needs to enable users to perform deictic gestures in the workspace to support meaningful communication with other collaborators.

Support background "use" of reference space. At the same time, the embodiment should be easily ignored, allowing remote parties to maintain an awareness of others' activities in the workspace while performing their own activities.

Support both coarse and fine-grained activity. The embodiment should enable the interpretation of both coarse activity (e.g., presence or approach of remote collaborators) and fine-grained activities (e.g., artifact manipulation).

Support local feedback. The system should "reflect back" what is being transmitted to remote stations. This enables collaborators to modify their gestures and behaviours *in situ* so they will be "correctly" interpreted by remote parties.

Implementation

Reference Space. Figure 2 illustrates the two mechanisms that we employed to facilitate reference space: arm shadows and trace pearls. Eight-bit grayscale images of a collaborator's arms are captured at 320×240 at 15 fps using an infrared camera mounted above the tabletop (Point Gray Research FireFly MV). These images are alpha-blended atop the workspace at both the remote (60% opacity) and local (20% opacity) workstations. The trace pearls are akin to telepointer traces [12], tracking each point of contact for each user with a trail that fades after 2s.

System. Our system connects three workstations through the local network (Figure 1). The Surface tabletops ran custom-built WPF shared workspace applications, communicating via GT# [11]. Each surrogate employed a 17" display, a webcam/microphone placed at the top and centre of the

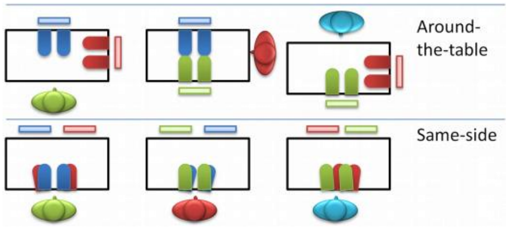

Figure 3. The two configurations investigated in Study 1. Around-the-table mimics real-life, while same-side allows users to share the same perspective on the task space.

display, and a speaker. These surrogates enabled spatially-correct gaze awareness and spatialized audio.

STUDY 1: CONFIGURATIONS FOR 3-PERSON COLLABORATION

There are many different ways to structure a distributed, three-way collaboration using a tabletop display. Of the configurations explored by prior work, only one has addressed the case of more than two distributed collaborators [37]. Our first study considers the two configurations illustrated in Figure 3. The first, *around-the-table*, mimics the real-life metaphor of people sitting around a table (from [37]), where each user has a unique position and perspective, and any hand and arm gestures are seen by others to emanate from that position. The second, *same-side*, has all three collaborators see the table contents from the same perspective. Since hand and arm gestures emanate from the viewing side, the perception in terms of the reference space was that all three participants sit in each others' laps. However, each participant saw the two remote people—by way of their video surrogates—sitting side-by-side across the table. Thus, *person* and *reference* spaces of the two remote participants were spatially disjoint.

Each configuration has potential strengths and weaknesses. The *around-the-table* configuration enables users to rely on their intuitions of space; however, each user's unique perspective may be problematic for oriented tasks (e.g., reading text). Nevertheless, this configuration does provide a spatial connection between *person space* and *reference space*. On the other hand, the *same-side* configuration overcomes the problems of oriented tasks since all collaborators share the same view of the workspace; however, the configuration does not map to a physical analog—thus, micro-coordination activities such as territoriality may be impacted negatively. Furthermore, this disjoint configuration may also introduce problems, such as gaze awareness, and identifying which remote person is making a particular gesture (cf. [21]).

Method

Configurations. Users completed activities using each condition: 1) *around-the-table* and 2) *same-side*.

Tasks. We used two different types of tasks: an orientation-free task, and a single-orientation task. Both tasks involved moving and arranging a set of tiles, initially piled in the centre of a shared workspace. In the orientation-free task

Figure 4. On the left, examples of tiles participants were given to construct approximations of the logos on the right.

("logo" task), participants were asked to recreate two of four possible logos (Figure 4, right) using a set of tiles containing various shapes (Figure 4, left). Participants were required to use at least eight tiles and the tiles could be rotated and translated. This task mimics the photograph sorting/organization of many tabletop studies (e.g., [15]), where the content of the tiles and logos are less strongly oriented than the text-based task..

In the single-orientation task ("text" task), tiles contained two sentences of text (17-37 words), and the tiles were all oriented in a single direction. Participants were asked to select and order a subset of these tiles according to a prescribed set of conditions. As an example, one version of this task was derived from Gottman's Desert Island Survival Task [9]. Users had to discuss and negotiate to come to a consensus about which tiles to select, and decide on an ordering in terms of importance. This task mimics the core activities in shared document editing such as gross and fine references to paragraphs and text, editing (moving of tiles), and discussion. The tiles in this task could be translated, but not rotated. In the *around-the-table* condition, tiles were oriented toward the short side of the table, while in the *same-side* condition tiles were oriented toward the participants (long side of the table).

Measures. We were primarily interested in qualitative measures of activity: examining users' interactions in each configuration, tracking their problems, and identifying interesting patterns of use. Each session was videotaped and we logged participants' interactions with the workspace. Finally, post-study questionnaires assessed preferences.

Participants. Four groups of three (12 participants – one female) were recruited for this study. Groups knew each other beforehand, but were not from the same familial unit. Participants' ages varied (18-42). Participants had little or no prior experience with touch-interfaces to computers, and were provided a gratuity for their participation.

Design. We employed a counter-balanced 2×2 within-subjects design, with two configurations (*around-the-table*, *same-side*) and two task types (*text*, *logo*). Each group completed both configurations with one task type before moving on to the second task type.

Procedure. Participants were given a brief introduction to the study, the equipment being used, and were given an opportunity to interact with tiles on the Surface. They were then led to their individual workstations (in separate rooms) and provided with paper-based instructions. Participants were given 15 minutes to complete each task, then completed a paper-based survey. After each task, depending

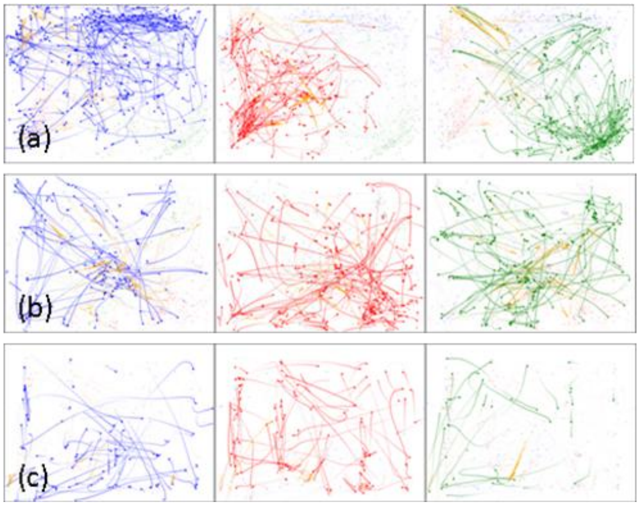

Figure 5. Contact traces for one group in three separate trials. Each cell represents a user, and each row is a trial. (a) *around* configuration, logo task; (b) *same* configuration, logo task; (c) *around* configuration, word task. Orange streaks represent instances of where two users attempted to move the same tile.

on the condition, the space was reconfigured. Once all four tasks were complete, they answered a final survey.

Analyses. We developed a custom visualization application to analyze the log data. During each trial, we also collected field notes, and corroborated these notes with observations from the video data. We were particularly interested in problems arising from three-way collaboration that would have been unique from dyadic scenarios.

Findings

Configuration Preference. With regard to overall preference, 9 out of 12 participants preferred the *same-side* configuration. When asked for preference on a per-task basis, for text tasks, 8 of 12 participants preferred the *same-side* configuration. For the logo tasks, 8 of 12 participants preferred the *same-side* configuration for the logo task.

Identity. Attributing workspace activity to a remote collaborator is difficult when there is more than one remote collaborator. Problems with identity attribution were not observed in the *around-the-table* condition, suggesting that the spatial distribution of users provided a sufficient spatial cue to know who was doing what. However, in the *same-side* configuration (when all of the hands came out from the same side of the table) participants were sometimes confused about whose hands were whose. The following excerpt provides an example:

***Jack** studies the workspace, and suddenly sees a tile being moved into the wrong place.*
***Jack:** No, wait. Larry, Larry…*
***Larry**: I'm not touching anything, dude!*
***Roper** retracts his hands quickly upon realizing that Jack was referring to him.*

Disembodiment. We were interested in whether the disembodiment of users' arm shadows from their video surrogates would be problematic in the *same-side* configuration. Users did not appear to have any trouble with the fact that the reference space was disjoint from person space, and none expressed any concern over this disembodiment. We return to this idea in Study 2.

Clutter. Although previous work shows that providing arm shadows for distributed colleagues benefits collaboration, it was unclear whether this approach would scale effectively given the additional clutter of shadows resulting from additional users. Despite the fact that we regularly saw three, and occasionally saw six arm shadows (two per person) in the task space, users did not seem to be distracted by this additional clutter. In tasks where colour is relevant, this design approach would need to be revisited.

Readability. Similar to Pauchet et al. [26], the single-orientation tasks (i.e., text tasks) caused problems for participants in the *around-the-table* configuration. When sitting with a compromised view of the text (i.e., not oriented with the text), participants frequently contorted their bodies and heads to read the tiles. As such, participants preferred the *same-side* configuration:

"Sitting together allows us to see everything from the same perspective, which was helpful."

Territoriality. We observed some spatial partitioning consistent with [28]. Figure 5 illustrates representative contact traces from three trials by the same group, showing that for the logo task, in the *around-the-table* configuration (Figure 5a), participants partitioned their activities in different regions of the workspace, whereas in the *same-side* configuration (Figure 5b), they did not. Figure 5c illustrates the text task in an *around-the-table* configuration. The spatial partitioning behaviour that we observed in the orientation-free task (Figure 5a) is noticeably absent here, showing that the semantics of the tiles in the text task changes the way in which users interact with the space.

Many participants asked for space where they could try out ideas without being observed by others. Consistent with Tang's observations [31], the workspace functions as a type of "stage" for behaviour, and the inability to rehearse, or try out ideas in a personal space was a concern for users. While personal territories may still form when using a larger workspace, the character of these may be somewhat different with a *same-side* configuration.

Conflict. Instances where participants attempted to touch and manipulate the same tile simultaneously were coded as "conflicts". A t-test of log data showed that for the orientation-free tasks (logo tasks), there were significantly fewer conflicts with the *around-the-table* configuration (mean=22, σ=11) compared to the *same-side* configuration (mean=36, σ=17), t_3=2.8, p<.05. For the single-orientation tasks (text tasks), this was not the case: groups in the *around-the-table* configuration experienced a similar number of conflicts (mean=8.3, σ=4.9) compared to the *same-side* configurations (mean=13, σ=8.3), t_3=1.4, p=0.31

(although statistical power is low given the small sample size). In general, there were fewer conflicts overall for the text tasks, which may suggest that the task structure reduced conflict.

In the around-the-table configuration, participants remarked that it was easier to see what everyone else was doing. This likely contributed to the fewer number of conflicts since the users could see remote collaborators' arms as they approached a tile. However, in the *same-side* configuration, the user's own hand often occluded the shadow of the remote collaborator's hand, making it difficult to be aware of concurrent activity.

Discussion of Study 1

Both configurations we explored had utility and our results indicate that the advantages of arm shadows (in terms of reference space) do scale to three-way collaborations. The *around-the-table* configuration demonstrated two chief benefits: enabling spatial partitioning, and identification (therefore awareness) of others' activities. The *same-side* configuration facilitated reading and shared perspective. The *same-side* configuration was also preferred by many participants. Separation of person space and reference space did not seem to cause any difficulties; however, occlusion of arms shadows may be a concern in *same-side* configurations, especially if the number of participants is increased even further.

Although multiple configurations are possible and likely useful—the appropriate configuration likely depends on task characteristics. Some relevant issues include: orientedness of the task artifacts (e.g., document editing vs. photo sorting) and parallelism of task (demand for simultaneous vs. serial interaction). For shared orientation tasks, such as text documents and spreadsheets, the benefits of the same-side configuration may well out-weigh the problems, whereas for orientation-free tasks, around-the-table may well be the better choice.

An interesting observation from this study was that participants did not make extensive visual use of the video surrogates. While they continually spoke to discuss and coordinate activities, participants rarely looked up from the task space. To better understand the role of arm shadows in the reference space, as well as the video surrogates of person space, we designed a second study.

STUDY 2: AFFORDANCES OF REFERENCE SPACE

Although previous research has demonstrated the efficacy of arm-based embodiments in the workspace [32,17,18,30,26,33], we were interested in further exploring the relationship between our proposed communicative "spaces". In particular, given that the participants in Study 1 did not seem to use the video surrogates a great deal, we were interested in the relationship between *person space* (i.e., our video and audio channels), and *reference space* (i.e., the video arm shadows and trace pearls). We wanted to better understand the importance of these channels and the degree to which a collaborative activity would be hindered if these channels were not present.

Method

Conditions. We varied the presence and absence of the communication channels, resulting in three conditions: (1) *audio + arm shadows* (A+S), (2) *audio + video* (A+V), and (3) *audio + video + arm shadows* (A+V+S). Given that individual and group variability can be high for collaborative activities, we also wanted to gather data from a baseline condition. As such, every group also completed the task in a co-present, face-to-face configuration.

Tasks. We used the same orientation-free "logo" task from Study 1, and created four equivalent versions of the task.

Measures. We collected log data from interactions with the workspace and videos of each session. We repositioned the video cameras so that we could see where participants were looking, though we were only able to collect this data from two of the three participants.

Apparatus. We used the *around-the-table* workstation configuration from Study 1 since the task for this study did not have an orientation. The hardware and software remained the same except that we slightly tinted the arm shadows (red, blue and green) to provide stronger identification cues.

Participants. We recruited six new groups of three (18 participants – four females) from the local population.

Design. We employed a three-way within subjects design with three main conditions: (*A+V*, *A+S*, *A+V+S*), fully counterbalanced. In addition, we ran a control, face-to-face condition for each group. The face-to-face condition was completed first for half of the groups, while the remaining groups completed the face-to-face condition last. Each group completed one task for each condition.

Procedure. We followed the same procedure as Study 1 (introduction to study and equipment, 15 minutes per task, paper-based surveys following each condition). For the face-to-face condition, participants were brought into the same room.

Analyses. We coded video from three of the six sessions, specifically looking for: (1) instances of interaction problems (between collaborators), and (2) any glances, looks or long stares from one collaborator to another (be it through the surrogates, or in face-to-face). Finally, we used field notes collected from each trial to group observations.

Findings

Use of Reference Space. Participants made consistent use of the arm shadows—both for explicit communication and for maintaining awareness of collaborators' activities. For example, we observed three types of deictic reference: (1) pointing at or waving their hands over a set of tiles while speaking ("over here", or "these ones"); (2) using a combination of task space feedthrough along with the shadows (e.g., wiggling a tile to draw attention to it), and

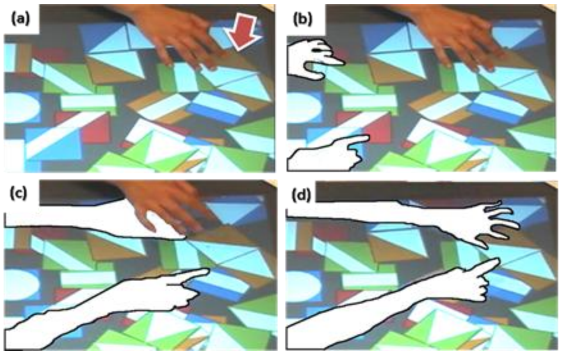

Figure 6. This sequence illustrates how the arm shadows helped users avoid conflict. Lasting 2s, we see (a) the local user manipulating a tile; (b) a remote user's arm shadow approaches from the left; (c) the remote user arrives, and the local user begins to withdraw his hand; (d) finally, only the remote user is manipulating the tile. For print legibility, we have traced the arm shadow and made it opaque.

(3) drawing transient outlines or circles on the workspace itself (where tiles were not) with the trace pearls.

Consequential communication provided by the reference space was also of significant utility to participants. Many participants expressed that the shadows could be used to maintain an awareness of the remote collaborators:

"I liked the hand shadows because you could tell who was doing what and what piece people were talking about"

Awareness: Confusion and Assists. The presence of the arm shadows reduced confusion, promoting awareness among collaborators of each others' activities. The questionnaire asked participants to rate their level of confusion over who was doing what on a 7-point scale. The average ratings for each condition were: A+V (mean=4.2, σ=1.9), A+S (mean=2.2, σ=1.3), and A+S+V (mean=2.2, σ=1.4) A Friedman test indicated that the conditions were significantly different from one another, $X^2(2)=14.45$, $p<0.001$. Post-hoc pairwise comparison using the Wilcoxon signed-ranks test, with a Bonferonni correction, revealed that participants were significantly more confused in the A+V condition (without shadows) than either of the conditions with shadows (A+S and A+S+V), $p<.01$. No significant difference was found between the two conditions with shadows (A+S and A+S+V).

Accordingly, the video data revealed clear instances where the arm shadows aided micro-coordination, preventing potential conflicts (a relevant indicator of coordination between the partners). Figure 6 illustrates such an example where the local collaborator withdraws his hands as he sees the approach of a remote collaborator, and understands the remote collaborator's intention of "fixing" the tile.

The arm shadows also enabled a fluidity of collaborative interaction that was characteristically absent from the A+V conditions: when arm shadows were present, groups were able to aid each other in retrieving tiles. These types of assists were comparatively absent in the A+V condition.

Table 1. Coded number of glances per group/condition.

Group ID	*A+V*	*A+V+S*	*F2F*
x	29	25	14
y	33	33	14
z	38	22	7

Glances in Person Space. Beyond the arm shadows, we were also interested in what role person space (i.e., the audio/visual surrogates) had in the task. Table 1 reports the number of times participants glanced or looked at one another. Interestingly, that total number of glances is relatively low, and most of the glances were extremely brief (i.e., less than one second). By way of comparison, we also recorded the number of glances that occurred during the F2F sessions, and again, these numbers are surprising low.

Given that the *task* the participants were completing was embedded in the task space (i.e., the Surface), we speculate that there may have been little to be gained from looking up at the video surrogates, whereas a great deal could be gained from studying the task space and the embedded reference space (i.e., watching other users' activity). Given participants' ability to communicate through the task space (e.g., wiggling tiles), reference space (e.g., pointing or waving), and through person space (i.e., speech), it is perhaps understandable why groups did not make extensive use of the video surrogates.

We performed a secondary analysis trying to understand *when* glances and use of the video space were occurring. Although not our initial focus, we observed that participants made consistent and extended use of the video space for conversation during setup *between* trials (e.g., when experimenters were setting up the next condition, or once participants had completed the task and were filling out the questionnaire) rather than *during* the trials. This "in-between" use was often social in nature, with participants joking and laughing with one another. During the trial itself, we noted that glances occurred primarily at the beginning (to discuss strategy or coordinate on ideas) and end (to confirm when all participants thought the logo designs were complete), or when someone was uncertain of what another had said or needed clarification (consistent with [34]). Participants' questionnaire feedback corroborated this analysis:

"Audio: Extremely necessary, w/o it we would be lost. Hand: helped see what others were doing. Video: not extremely useful, but it made the experience more personal"

"I never looked at the video screen because it was not helpful to look at another person. Audio was most important because of directions, and shadows helped to not grab the same objects"

"While audio and hand shadows were important the video made the exercise feel more group like and not sitting alone in a room."

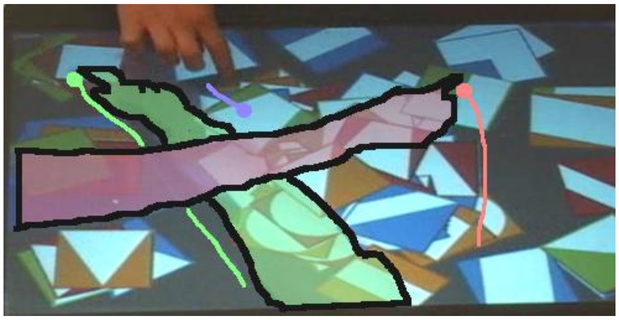

Figure 7. Participants freely overlapped their arms in the space in ways that would not have been physically possible in face-to-face. Note that arm shadows and pearl traces have been artificially recoloured for print legibility.

When asked to choose two of the three channels (audio, video, hand shadows), 13 participants out of 18 selected audio and arm shadows.

Corporeal Embodiments. We found interesting differences in the way groups worked when face-to-face compared to distributed, particularly in how corporeal embodiments mandated serial access to shared space, and how they enabled additional ways to communicate and gesture.

While the Surface itself supported multiple simultaneous contacts, when the group sat together at the same tabletop, each participant's physical access to the Surface appeared to be considerably reduced. For example, if a participant reached into the "shared" region of the Surface, she would not only claim a tile, her arm would also block both visual and physical access to the same space. The only space a participant would typically have uncontested access was the space immediately in front of herself. The physicality of participants' bodies thus afforded only serial access to large portions of the shared space. During distributed trials, participants appeared comparatively unfettered, and took advantage of the absence of others' bodies, freely crossing over one another's arm shadows without hesitation. Figure 7, from a participant session, illustrates how the absence of others' bodies allows simultaneous, parallel activity.

Thus, note that the mere *presence* of another's arm shadow did not necessarily inhibit interaction with a workspace artifact—instead, the arm shadow, combined with the trace pearls *mediated* a user's understanding of the state of the workspace. The *arm shadow*'s presence indicates that a remote user is present, but the absence/presence of *trace pearls* helps users distinguish whether an artifact is actively being used by the remote user. Because arm shadows frequently cover items that people are not actively using, the fine-grain awareness provided by the trace pearls is of benefit to this micro-coordination.

The saliency of users' physical bodies also played a role in how users formed and used gestures compared to their use of the arm shadows. Arm shadow gestures are projected onto and into the tabletop workspace, thereby implying a workspace location—even if it is not intended. Thus, users employing arm shadow gestures tended to be fairly precise when referring to tiles—often going so far as to tap the tile in question, since gestures made above Surface could be misinterpreted if the arm shadow appeared in the wrong place. In contrast, during the face-to-face condition, participants used more gestures that were comparatively vague—often away from the surface of the tabletop itself. It may be that the more sequential nature of activity in face-to-face trials meant that deictic references were as much temporal references as spatial ones (i.e., a reference to "that one" might as easily be a reference to the last tile that had been touched as the one being explicitly pointed to); in contrast, with more parallel activity of distributed conditions, such references would be more ambiguous.

Perhaps the most poignant example of this off-surface gesturing in face-to-face trials was how participants were able to communicate ideas. In particular, when participants were attempting to convey an idea during distributed trials, they would frequently resort to illustrating the ideas using the tile pieces directly (i.e., actually *performing* the idea). With face-to-face trials, although we still saw similar instances of such performances, users also *mimed* ideas: for instance, talking about and acting out (in mid-air) interactions on tiles (without actually using the tiles). Group members seemed capable of deciphering these mid-air gestures.

Discussion of Study 2

This second study allowed us to better understand the function of person space, reference space and task space independent and in concert with one another. Independently, the results confirm the theory postulated by Buxton [5], where person space is used to resolve ambiguity and to maintain social contact (even between tasks, as it turned out), task space for actual work activity, and reference space for deictic and workspace-relevant gestures. With regard to reference space itself, our implementation of arm shadows combined with the trace pearls provided both coarse-grained presence awareness as well as fine-grained manipulation information, both of which were used by participants in different capacities. Perhaps what is most interesting here is the tight coupling and synergistic relationship between task space feedthrough and reference space. In concert, they form a powerful source of layered information: the arm shadows provide awareness of presence, while the feedthrough and pearl traces provide more detailed information about remote users' activities. For example, during a pilot session, the participants were able to immediately detect synchronization errors in the software, even before it became apparent to the experimenters. The participants stared at the shared workspace, visibly puzzled, and then commented, "What are you doing? It looks like you're doing something [referring to the remote collaborator], but nothing's moving." This example of behaviour illustrates that users were attending to both the arm shadows, as well as interaction in the task space to understand others' activities.

As demonstrated in Study 1, the spatial relationship between reference space and person space is evidently not as critical:

the relative disembodiment between remote users' arm shadows and video surrogates did not ever appear to be a problem.

Our findings align with a long line of inquiry that questions the value of video in person-space [25]. One conclusion might be that for the particular task being investigated, it had negligible value, and that perhaps for other types of tasks (e.g., negotiation [3, 25], or where there is more ambiguity [27,34]), it would provide more value. Such a focus on task may sidestep the more subtle issue that perhaps we need to reconsider what is being measured with regard to the video. In this work, coding for number of glances was admittedly disappointing; however, trying to understand *when* the glances occurred provided considerable insight.

Finally, this study raises interesting questions about how users' corporeal arms are used differently when collocated compared to the arm shadows when distributed. The inherent nature of arm shadows causes gestures to be fairly explicit, whereas in face-to-face trials, we saw gestures that were more vague, yet still communicative. The projected, flattened arm shadows have some obvious, and perhaps some subtle property differences that users appropriate differently. Further, since the arm shadows could not physically impede on users' access to the tabletop, distributed setups enabled more parallel interaction when compared to collocated settings.

OPPORTUNITIES FOR DESIGN AND CONCLUSIONS

The work presented here builds on a distributed workspace systems that provide strong embodiment—both atop the workspace (reference space) and outside of it (person space). We investigated, in particular, these spaces as they relate to one another spatially, and our own reference space design in three-party collaboration. Our findings suggest several opportunities for design, as well as promising avenues for exploration.

Role of Social Coordination. This study reinforces the argument that with sufficient awareness of others' activities in a shared workspace, users can smoothly coordinate their activities, obviating the need for certain types of "locks" on shared entities at the application level. This was evident in both participants' avoidance of collisions, as well as the micro-coordination required for assisting one another. An interesting question is the extent to which this coordination can be achieved using more simple reference space mechanisms (e.g., telepointers).

Embodiment Presentation and Control. Our use of video for users' arm shadows resulted in characteristically anthropomorphic embodiments, although it is worth considering non-video and/or non-anthropomorphic approaches, which may have lower network bandwith demands (e.g., [30]). The projection of arms onto the planar tabletop also fails to capture the proximity information that is present in face-to-face scenarios—an awareness cue that might be of utility for coordination. Finally, some participants asked for more control over the saliency of the arm shadows—suggesting that the saliency of arm shadows can also be a coordinating mechanism (e.g., more opaque when performing explicit gesture vs. more transparent when working independently). A fruitful avenue may be semi-automated control, where the opacity of the arm shadows is linked to speech from each workstation (e.g., if you are speaking, your arm shadows become more opaque).

Configuration and Spatialization. This work provides further evidence that fluid collaboration can still occur when person and reference space are divorced. We demonstrated two configurations of these spaces, suggesting that the appropriate configuration will depend on task characteristics. Thus, there are likely a whole host of configurations that may be of differential utility—for example, tools that allow users to flexibly control their orientation to the workspace. Spatialization was important to us, and we fixed several parameters with regard to person space (e.g., we carefully aligned speaker, LCD and webcam for each surrogate). An interesting consequence is in how users respond to this spatialization: in one trial, where the video channel was removed, we observed a user lean in and chastise the speaker that was associated with the collaborator he was arguing with—in effect, responding to the audio source of that collaborator.

Finally, users' desire to identify the owners of embodiments and actions underscores both the utility of studying three-person collaboration, and the need for scalable designs. Spatializing users and tinting the video helped to provide this identity cue, but other approaches may also be viable.

In this work, we have sought to build on existing system designs, attempting to extend them for three-way collaboration. We have explored several design choices, and empirically studied the consequences of these choices. Our studies have raised issues of spatial configuration, identity, scalability of embodiment designs, and workspace coordination, which will continue to be of pertinence in this research space.

ACKNOWLEDGEMENTS

We thank Sara Bly, Carman Neustaedter and Saul Greenberg for discussions about the ideas in this paper. We also thank Libby Hanna, Koji Yatani, and Chris Brooks.

REFERENCES

1. Bly, S., Harrison, S., and Irwin, S. Media spaces: bringing people together in a video, audio, and computing environment. *CACM 36*, 1 (1993), 28-47.
2. Birnholtz, J. P., Grossman, T., Mak, C., and Balakrishnan, R. An exploratory study of input configuration and group process in a negotiation task using a large display. In *Proc CHI 2007*, (2007), 91-100.
3. Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z. Effects of four computer-mediated communications channels on trust development. In *Proc CHI 2002*, (2002), 135-140.

4. Buxton, W. Integrating the Periphery and Context: A New Model of Telematics In *Proc GI '95*, (1995), 239-246.
5. Buxton, W. Mediaspace – meaningspace – meetingspace. In Harrison, S. (Ed) *Media Space 20+ Years of Mediated Life*, Springer, 2009.
6. Buxton, W., Sellen, A., and Sheasby, M. Interfaces for multiparty videoconferencing. In K. Finn et al. (Eds.) *Video Mediated Communication*. Erlbaum, Hillsdale, N.J., 385-400.
7. Dourish, P. and Bellotti, V. Awareness and coordination shared workspaces. In *Proc CSCW 1992*, (1992), 107-114.
8. Finn, Sellen, A. & Wilber, S. (Eds.). *Video Mediated Communication.* Hillsdale, N.J.: Erlbaum (1997).
9. Gottman, J. *Marriage Clinic.* Norton, NY, 1999.
10. Greenberg, S., and Kuzuoka, H. Using digital but physical surrogates to mediate awareness, communication and privacy in media spaces. *Personal Technologies 4*, 1 (2002), 182-198.
11. GT#. http://www.hci.usask.ca/research/gt/
12. Gutwin, C. Improving interpretation of remote gestures with telepointer traces. In *Proc CSCW 2002*, (2002), 49-57.
13. Gutwin, C. and Greenberg, S. A descriptive framework for real-time groupware. *Computer Supported Cooperative Work 11*, 3-4 (2002), 411-446.
14. Harrison, S. (Ed) *Media Space 20+ Years of Mediated Life*, Springer (2009).
15. Hinrichs, U., Carpendale, S., and Scott, S. Evaluating the effects of fluid interface components on tabletop collaboration. In *Proc AVI 2006*, (2006), 27-34.
16. Ishii, H. and Kobayashi, M. ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In *Proc CHI 1992*, (1992), 525-532.
17. Izadi, S., Agarawal, A., Criminisi, A., Winn, J., Blake, A., and Fitzgibbon, A. C-Slate: exploring remote collaboration on horizontal multi-touch surfaces. In *IEEE Tabletop 2007*, 3-10.
18. Kirk, D., Crabtree, A., and Rodden, T. Ways of the hands. In *Proc ECSCW 2005*, Springer (2005), 1-21.
19. Krueger, Myron, W. *Artificial Reality II*. Reading, MA: Addison-Wesley (1991).
20. Kuzuoka, H. and Greenberg, S. Mediating awareness and communication through digital but physical surrogates. *Ext. Abstracts CHI 1999*, (1999), 11-12.
21. Luff, P., Heath, C., Kuzuoka, H., Yamazaki, K., and Yamashita, J. Handling documents and discrimination objects in hybrid spaces. In *Proc CHI 2006*, (2006), 561-70.
22. Mantei, M., Baecker, R., Sellen, A., Buxton, W., Milligan, T., and Wellman, B. Experiences in the use of a media space. In *Proc CHI 1991*, (1991), 203-209.
23. Minneman, S. and Bly, S. Managing a trois: a study of a multi-user drawing tool in distributed design work. In *Proc CHI 1991*, (1991), 217-224.
24. Morris, M. R., Ryall, K., Shen, C., Forlines, C., and Vernier, F. Beyond "social protocols": multi-user coordination policies for co-located groupware. In *Proc CSCW 2004*, (2004), 262-265.
25. Olson, G. M. and Olson, J. S. Distance matters. *HCI 15*, 2 (2000), 139-178.
26. Pauchet, A., Coldefy, F., Lefebvre, L., Picard, S., Bouguet, A., Perron, L., Guerin, J., Corvaisier, D., and Collobert, M. Mutual awareness in collocated and distant collaborative tasks using shared interfaces. In *Proc INTERACT 2007*, 59-73.
27. Ranjan, A., Birnholtz, J. P., and Balakrishnan, R. An exploratory analysis of partner action and camera control in a video-mediated collaborative task. In *Proc CSCW 2006*, (2006), 403-412.
28. Scott, S., Carpendale, S., and Inkpen, K. Territoriality in collaborative tabletop workspaces. In *Proc CSCW 2004*, (2004), 294-303.
29. Sellen, A., Buxton, B., and Arnott, J. Using spatial cues to improve videoconferencing. In *Proc CHI 1992*, (1992), 651-652.
30. Tang, A., Neustaedter, C., and Greenberg, S. VideoArms: embodiments for mixed presence groupware. In *Proc British-HCI 2006*, Springer (2006), 85-102.
31. Tang, J. Findings from observational studies of collaborative work. *Intl J. Man-Machine Studies 34*, 2 (1991), 143-160.
32. Tang, J. and Minneman, S. VideoDraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems 9*, 2 (1991), 453-469.
33. Tuddenham, P. and Robinson, P. Territorial coordination and workspace awareness in remote tabletop collaboration. In *Proc CHI 2009*, (2009), 2139-2148.
34. Veinott, E., Olson, J. S., Olson, G. M., and Fu. X. Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other. *Conf. Compan. CHI 1999*, (1999), 302-309.
35. Whittaker, S., Frohlich, D., and Daly-Jones, O. Informal workplace communication: what is it like and how might we support it?. In *Proc CHI 1994.* (1994), 131-137.
36. Wilson, A. PlayAnywhere: A Compact Tabletop Computer Vision System. In *Proc. UIST 2005*, (2005). 83-92.
37. Yamashita, J., Kuzuoka, H., Yamazaki, K., Miki, H., Yamazaki, A., Kato, H., and Suzuki, H. Agora: supporting multi-participant telecollaboration. In *Proc HCII 1999*, Erlbuam (1999), 543-54.