Embodiments for Mixed Presence Groupware

Anthony Tang, Carman Neustaedter and Saul Greenberg
Department of Computer Science, University of Calgary
Calgary, Alberta CANADA T2N 1N4
+1 403 220 6087

{tonyt, carman, saul}@cpsc.ucalgary.ca

Abstract

Large surfaces such as tabletop and whiteboard displays naturally afford collocated collaboration, where multiple people work together over the shared workspace. As large digital displays become more ubiquitous, it becomes increasingly important to examine their role in supporting groups of distributed collaborators working over the digital work surface. In particular, *Mixed Presence Groupware* (MPG) is software that connects both collocated and distributed collaborators and their disparate displays via a common shared virtual workspace. We have built several MPG systems by connecting several distributed displays, each with multiple input devices, thereby connecting both collocated and distributed collaborators. By observing how these systems are used, we found that MPG presents a unique problem called *presenc1e disparity*: collaborators focus their energies on collocated collaborators at the expense of their distributed counterparts. Presence disparity arises because the physical presence of collaborators varies across the MPG workgroup: physically collocated collaborators are seen in full fidelity, while remote participants are represented by only virtual embodiments. Consequently, we propose four design principles for MPG systems that we believe will help mitigate the problem of presence disparity in MPG. We then introduce how these principles are realized in VideoArms, an embodiment technique that digitally captures people's arms as they work over large work surfaces, and redisplays them as digital overlays on remote displays. Our evaluation of VideoArms validates its use in principle as an effective embodiment technique for MPG systems.

Keywords
Mixed presence groupware, single display groupware, distributed groupware, consequential communication, embodiments, gestures.

1. INTRODUCTION

Large displays such as tabletop and whiteboard displays naturally afford collocated collaboration allowing multiple people to work together over the shared display. However, as large displays become more ubiquitous, it becomes increasingly important to examine their role in supporting *groups* of distributed collaborators.

> *Imagine you are part of a team of designers based in Seattle working on a new product in its early stages. Management has told you to hold a joint brainstorming session with another group in your company that is familiar with the type of product you are designing. However, the difficulty is that the other team works in the New York office. Fortunately, your company has a special meeting room setup in each city. Both meeting rooms are connected with an audio link and contain electronic whiteboards with special software. This software allows your team to draw your ideas on the wall using styli, and also lets your colleagues in New York see those drawings in real time. Just the same, you can see the drawing activities of the New York team on the same workspace.*

This scenario may not be difficult to imagine since the hardware that is needed to support this type of activity already exists. While the hardware is important, our research focus is on understanding and designing the type of software described in this scenario, *mixed presence groupware* (MPG): software that connects *both* collocated and distributed collaborators together in a shared visual workspace, often utilizing the collaborative opportunities presented by large screen displays. In practice, we have built MPG systems by connecting several distributed displays, each with multiple input devices, thereby connecting both collocated and distributed collaborators. Figure 1 shows an example MPG system where three groups of collocated collaborators work on large wall and table displays (top). Even though the three groups are separated by distance, MPG software creates a virtual shared space for the groups by connecting the large displays (Figure 1, bottom).

MPG presents a unique problem called *presenc1e disparity*, where collaborators focus their energies on collocated collaborators at the expense of their distributed counterparts [8]. Presence disparity arises because physically collocated collaborators are seen in full fidelity (they are in the same location working at the same display), while remote participants are represented by only embodiments— virtual presentations of their bodies. Most groupware systems reduce this virtual presentation to telepointers—usually a custom mouse cursor—which clearly cannot compete against the physical body of a collocated collaborator. Thus, presence disparity unbalances the collaborator's subjective experience because even dyadic collaborative dynamics will vary in terms of how one senses presence, engagement and involvement of collocated *vs.* remote partners.
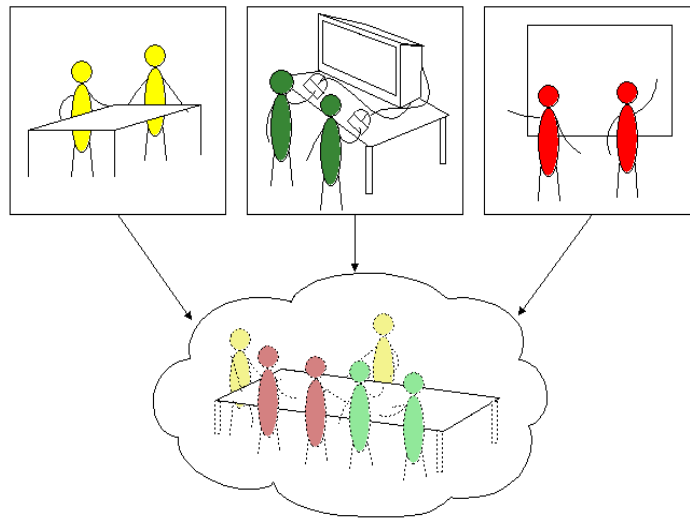
Figure 1. Three teams working in MPG over three connected displays (top), stylized as a virtual table (bottom).

The core problem of presence disparity arises from the physical distribution of participants in the virtual workspace—the physical *presence* of collaborators varies across an MPG workgroup. This has negative impacts on conversational dynamics because MPG collaborators cannot communicate (verbally and non-verbally) as effectively with remote collaborators as they can with those who are collocated and thus will tend to focus their communicative efforts toward their collocated partners [3]. Remote collaborators are less likely to be invited into informal discussions of work objects, and are therefore less likely to perform the task as effectively as collocated counterparts.

In this article we discuss the design and evaluation of VideoArms, an embodiment technique that aims to mitigate the problem of presence disparity in MPG. VideoArms captures people's arms as they work over large work surfaces, and redisplays them as digital overlays on remote displays. First, we describe VideoArms and the design principles behind it in greater detail. Second, we present an evaluation of VideoArms that validates its use as an effective embodiment for MPG systems.

2. VIDEOARMS: A VIDEO-BASED MPG EMBODIMENT

VideoArms is a video-based embodiment technique for MPG systems that captures collaborators' arms as they work over the workspace using a video camera, and redraws the arms at the remote location. Figure 2 illustrates a sample session of VideoArms. The top images show two connected groups of collaborators. Each group works over a large touch-sensitive surface—the left is a front-projected touch-sensitive horizontal DViT, while the right is a rear-projected vertical SmartBoard. Each surface displays the same custom MPG application that lets people sketch and manipulate images, while displaying video embodiments.

Figure 2 (bottom) also illustrates what users can *see* when using the VideoArms embodiment in this MPG application. First, collocated collaborators can see their own arms as local feedback, rendered semi-transparently, providing feedback of what others
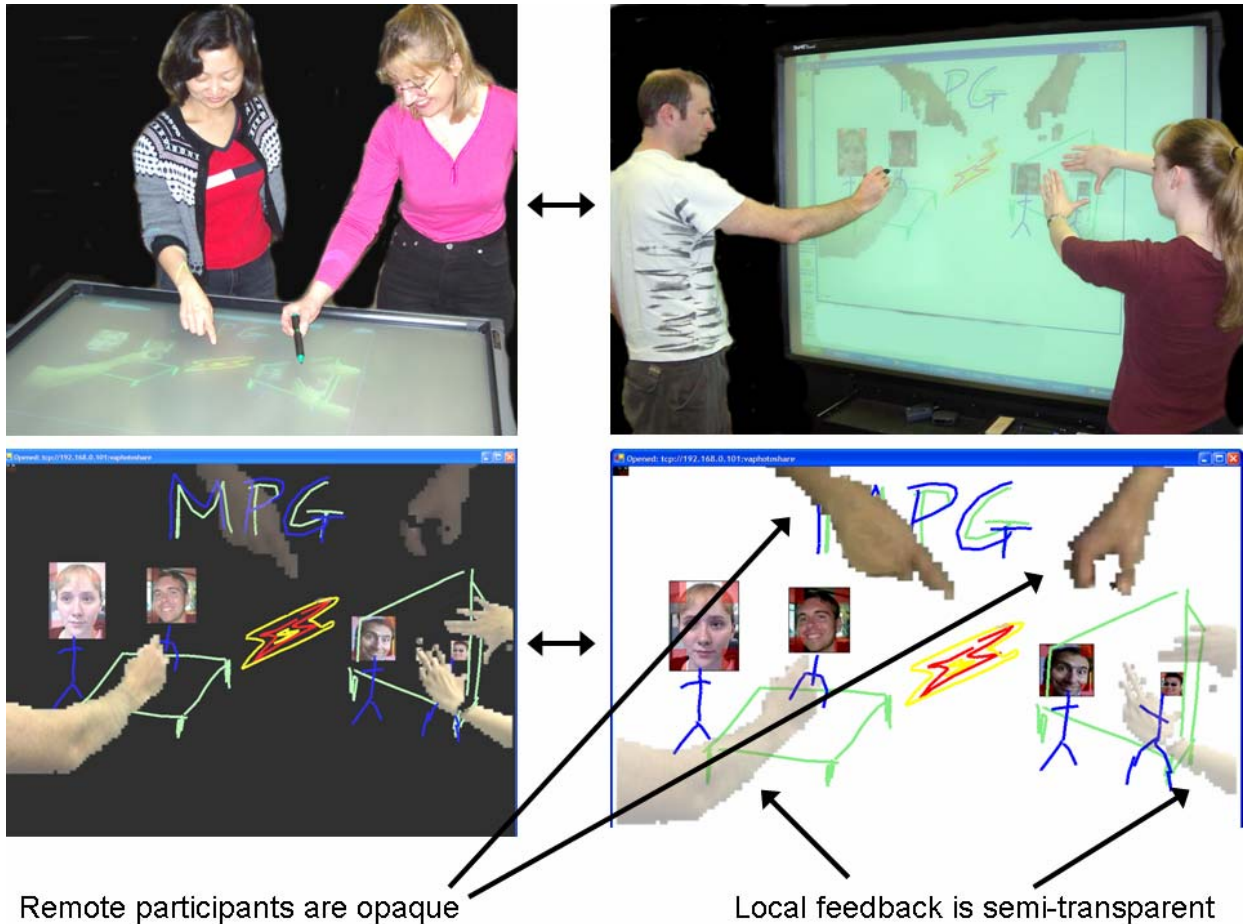
Figure 2. VideoArms in action showing two groups of two people working over two connected MPG displays (top) and a screenshot of what each side sees (bottom). Local and remote VideoArms are in all scenes, but local feedback is more transparent.

can see while minimizing interference. For example, the bottom right image of Figure 2 shows three semi-transparent arms as local feedback for the two collaborators working on the wall display (Figure 2, top-right).

Second, each group sees the solid arms of the remote participants in reasonable 2½-dimensional fidelity (while the images are not truly 3-dimensional, the system captures and reproduces colour-based depth-cues). For example, the bottom right image of Figure 2 shows two opaque hands which present the arms of the remote participants working on the table display (Figure 2, top-left) to the two people working on the wall display (Figure 2, top-right).

Third, the remote drawings of arms preserve the physical body positioning relative to the workspace. Both physical and video arms are synchronized to work with the underlying groupware application, where gestures and actions all appear in the correct location[1]. For

---

[1] VideoArms simply reproduces a video-captured image of the workspace. In principle, it can therefore support an infinite number of non-overlapping arms. While our goal was to develop a true MPG application with VideoArms, technical limitations imposed by the input devices (the actual SMARTBoards) meant that our final system only supported two simultaneous touches on one display; the other display could only support a single touch.
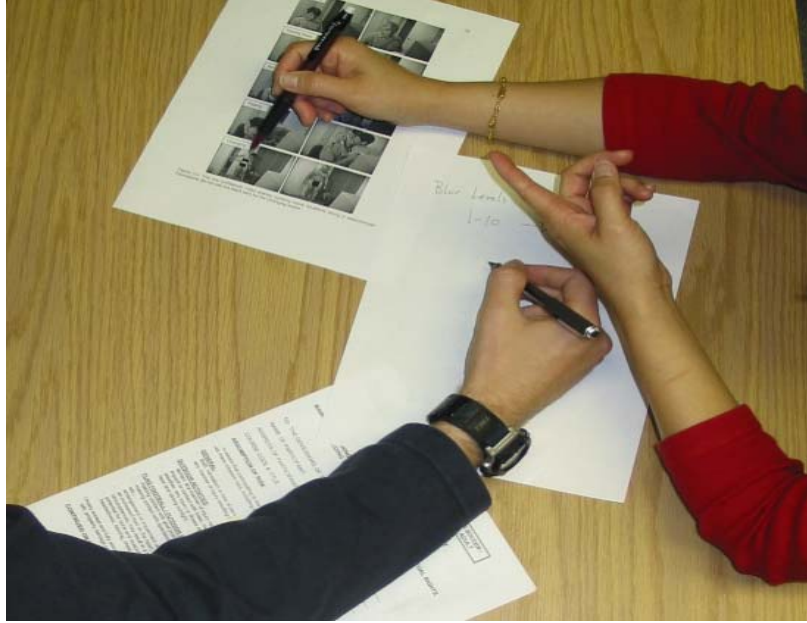
Figure 3. A bird's eye view of a physical workspace.

example, because the people at the table display (Figure 2, top-left) are positioned at the rear of the table, their arms appear on the vertical display as coming from the top (Figure 2, right).

Figure 2 also reveals communicative aspects of the embodiment. In this MPG setting, all participants can simultaneously gesture to the full, expressive extent of arms and hands. The system neither dictates nor implies any sort of turn-taking mechanism, and captures workspace and conversational gestures extremely richly. Furthermore, users are not tethered to any particular place in the workspace: using touch and pens to interact with the groupware application, users are free to physically move around the workspace as they see fit. For example, we can see the use of rich gestures in the top right image of Figure 2 when the woman uses her hands to indicate the intended size of an object. At the same time, the woman on the left at the table display (Figure 2, top-left) points to a particular object.

2.1 Design Principles
The VideoArms metaphor captures and presents the workspace from a bird's eye view of the workspace (c.f. "through the glass" metaphor from [4][10]). From this perspective, the arms are the primary indicators of a collocated collaborator's presence (Figure 3). To mitigate presence disparity for remote collaborators, VideoArms was designed to support four principles.

1. To provide *feedback* of what others can see, a person's embodiment should be visible not only to one's distant collaborators, but also to oneself and one's collocated collaborators.

2. To support *consequential communication* for both collocated and distributed participants, people should interact through direct input mechanisms, where the remote embodiment is presented at sufficient fidelity to allow collaborators to easily interpret all current actions as well as the actions leading up to them.

3. To support bodily *gestures*, remote embodiments should capture and display the fine-grained movements and postures of collaborators. Being able to see these gestures means people can disambiguate and interpret speech and actions.

4. To support bodily actions as they relate to the workspace *context*, remote embodiments should be positioned within the workspace to minimize information loss that would otherwise occur.

We perceive our own actions and the consequences of our actions on objects as *feedback*, and we constantly readjust and modify our actions as our perceptions inform us of changes to the environment, or changes about our bodily position [6]. Threading a needle when blindfolded is difficult because without our ability to perceive our own bodies as physical objects in the world, we cannot smoothly interact with it. Thus, the first design principle suggests that a person's embodiment should be visible not only to one's distant collaborators, but also to oneself and one's collocated collaborators.

Our bodies are the key source of information comprising *consequential communication*: the information unintentionally generated as a consequence of an individual's activities in the workspace, and how it is perceived and interpreted by an observer [7]. A person's activity in the workspace naturally generates rich and timely information that is often relevant to collaboration. For instance, how a worker is positioned in the workspace and the kinds of tools or artefacts being held or used tells others about that individual's current and immediate future work activities. Therefore, the second design principle addresses the need to support consequential communication by using direct input mechanisms and through high fidelity MPG embodiments.

While consequential communications comprises unintentional body actions, *gestures* are intentional bodily movements and postures used for communicative purpose [2]. Gestures play an important role in facilitating collaboration by providing participants with a means to express their thoughts and ideas both spatially and kinetically, reinforcing what is being done in the workspace and what is being said. For this reason, the third design principle speaks about the necessity for embodiments to capture and display the body gestures of collaborators.

Because consequential communication and gestures occur in the workspace, removing such actions from their context also removes much of their interpretation. For instance, the statement, "Put this object here," is meaningful in the context of Figure 2, but is unintelligible outside of the context of the workspace. This leads to our fourth design implication which stresses that embodiments should be placed within the context of the workspace.

From a collaborative standpoint, the VideoArms prototype theoretically satisfies our MPG embodiment principles.

1. Local participants know what remote people see because their own embodiments are shown as semi-transparent feedback.

2. Because the body is used as an input device on the touch sensitive surface, VideoArms supports consequential communication: other collaborators can easily predict, understand and interpret another's actions in the workspace as one reaches towards artefacts and begins actions.

Figure 4. The image on left is colour-segmented to find the skin-colour pixels (middle image). The two images are then combined to produce the VideoArms image on the right.

3. Rich gestures (coupled with conversation and artifact manipulation) are also supported well because the remote arms are displayed in rich 2½ dimensional fidelity and a reasonable framerate (~12 fps). While clearly not ideal, practical experience with the prototype showed that 12 fps was reasonable enough to interpret gestures. Finally, task-related gestures are easily interpreted because they are placed in the context of the workspace.

4. Collocated participants can use and interpret natural body language of their physical bodies as they collaborate. Because collaborators are not tethered to input devices, their actions are direct and in the workspace context; thus, an individual's physical body *is* the embodiment.

Next we describe the key implementation details of VideoArms.

2.2 Implementation Details

VideoArms uses inexpensive web cameras positioned approximately two meters in front of the display to capture video images of collaborators. The software extracts the arms (and other bare-skinned body parts) of collaborators as they work directly over the displayed groupware application. It then transmits these images to the remote workstation, where they are further processed to appear as an overlay atop the digital workspace. To provide local feedback, VideoArms overlays a local person's video on the work surface.

Frames captured by the camera are processed, transmitted and displayed in a four step process.

1. *Find frames that match skin colour.* This step uses a statistical quantity known as the Mahalanobis distance to determine the likelihood of a given pixel as being skin. During calibration, 10 pixel samples of skin are taken whose colour (R, G, B) values are read. These values seed a mean vector and covariance matrix representing skin tones. For each frame (Figure 4, left), VideoArms calculates the Mahalanobis distance for each pixel against skin tones. If it is typical of skin (an arbitrary cut-off), then the pixel is judged to be skin; otherwise, it is judged to not be skin. Morphological opening, a standard computer vision technique, is then applied to the skin mask to remove image noise while preserving the shape and size of larger objects. This process produces a silhouette mask of the collaborator's arms (Figure 4, middle), much like the shadow-like embodiments found in [1] and [10].

2. *Combine mask with original image*. A full-colour image of the arms is produced by overlaying the silhouette mask with the original image: black pixels of the silhouette are copied onto the original image (Figure 4, right).

3. *Transmit arm to listening clients*. The arm images are transmitted to listening clients via IP multicasting (clients include both the remote and local display). IP multicasting is used to reduce traffic on the network, and its use of UDP packets ensures quick delivery.

4. *Overlay images on the workspace*. Using standard GUI techniques, all received images are drawn on the groupware work surface, which creates a composite of local and remote arms.

The fully digital nature of VideoArms provides many benefits over its analog predecessors (e.g. [4] and [10]). For instance, VideoArms can be rendered in many ways: semi-transparent, outline, vector, and stylized arm representations. Furthermore, analog video systems suffer from the drawback of degraded image quality when multiple video signals are composited. The digital nature of VideoArms does not suffer from this drawback since image noise can be digitally removed, thereby making it more scaleable.

VideoArms is built using Python, the .NET Framework, PyIPP (a set of Python wrappers for the Intel Performance Primitives library), the Python Imaging Library, and the Python numarray open source libraries. To maximize performance, we use one workstation to process and transmit the video from the Logitech QuickCam Pro 4000 camera, and another to display the VideoArms and run the groupware application. On a Celeron 2.4GHz, video frames are processed at 320×240 resolution at 25 frames per second, and overlaid across a 640×480 groupware workspace. This resolution is sufficient for interpreting consequential communication and gestures, and improves upon [1], which works over a 640×480 workspace and a 176×144 video image on a 1GHz machine. While further optimizations are possible, our primary intention was to develop a system suitable to test our ideas and rather than to produce a production-level implementation.

3. EVALUATION

We ran an observational study to evaluate VideoArms' support of the four design principles, and to understand whether these collectively mitigate presence disparity. We first articulate a set of questions we were interested in addressing in our evaluation, followed by a description of the participants, the materials and tasks used in the evaluation. We then discuss the experimental procedure and justify the tasks in the evaluation. The results and discussion are presented in subsequent sections.

3.1 Questions of interest

VideoArms' approach to mitigating presence disparity is to present remote collaborators through the richness and fidelity of video. Theoretically, this approach allows gestures, workspace activities and consequential communication to be natural, easily conveyed and interpreted. Thus, in this study, we are interested primarily in uncovering the incidence and variety of gestures, the occurrence of consequential communication, and the level of engagement between remote participants.

- Do participants use gestures when there is a voice link? Are these gestures placed within the context of the workspace? For whom are these gestures intended? Remote, or collocated collaborators, or both?

- How natural are these gestures?

- Does consequential communication occur and is it used across the link in spite of the presence of a voice link and a collocated participant?

- Do participants make use of local feedback?

- How does correction (a common occurrence related to consequential communication) occur?

3.2 Participants

A total of 22 paid participants (12 female, 10 male) were recruited via a notice on the main UNIX server for the Department of Computer Science, and by an email sent to that department's graduate students. Participants, who ranged from 18-29 years of age, were all daily computer users, and 18 of 22 participants were computer science majors or graduate students (the remaining participants were students from other faculties).

Participants were recruited as groups, so each participant already knew his or her group members well. Six of these participants were pairs, while the remaining 16 participants were in groups of four (4 groups of 4). Thus, a total of 22 participants or 7 groups were observed using VideoArms.

3.3 Materials

Four Celeron 2.4Ghz machines, each with 256 MB RAM and connected on a 10Mbps hub on a private network, were used in the evaluation. Two of these had Logitech QuickCam Pro 4000 cameras, and were responsible for capturing and sending out images of collaborators. The remaining two machines drove the SMARTBoard and horizontal DVIT displays and the main application.

We used a rear-projected, touch sensitive SMARTBoard, which has a 167.6cm screen (diagonal). The horizontal DVIT display, located in a separate room, is similarly sized, but was front-projected. Although the DVIT could technically support two simultaneous touches, the SMARTBoard could not. To prevent this technical difference from affecting the results of the study, the software was written to allow only one touch per board for the study.

Participants were given yellow dishwashing gloves to use with VideoArms as their bright, uniform color provided better extracted arm images. While VideoArms was designed to pick up skin tones, configuring the system for each group would have been too time-consuming. Also, since our primary interest was not in the computer vision algorithm but rather in the collaborative aspects of the system, we felt this was a reasonable substitution.
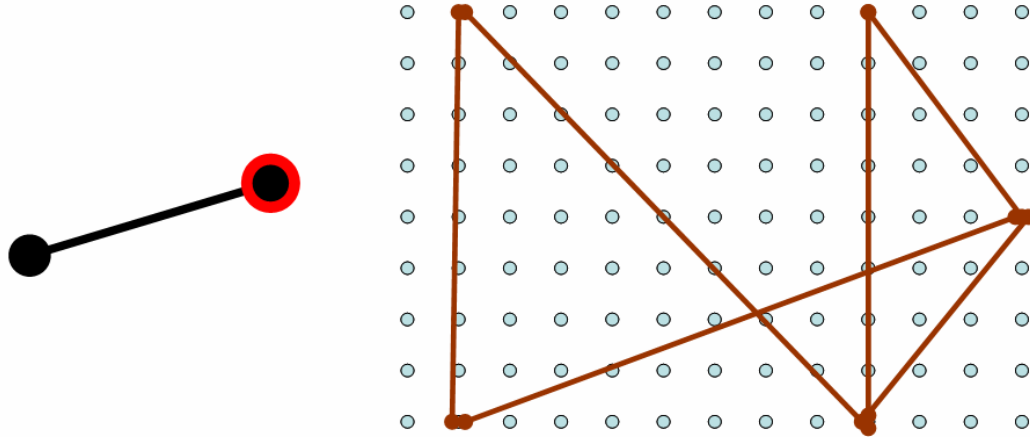
Figure 5. An elastic with nails on each end. The right nail has been selected, and is therefore highlighted with a halo (left). An elastics and nails target image (right).

3.4 Tasks

Pairs completed a task that we called "elastics and nails." Groups of four completed this task along with an open-ended design task.

The elastics and nails task was designed specifically to evaluate the effectiveness of VideoArms in supporting gesturing and consequential communication. The task is a drawing activity where participants construct images (e.g. Figure 5, right) with several randomly placed lines (elastics). Each elastic is a rubber band with two grab points (nails) that highlight when grabbed (Figure 5, left). To move an elastic, participants must grab opposite nails of the same elastic, whereupon it can be moved by simply dragging the nails around. If either party releases his or her nail, the elastic sets itself down in place. Simultaneous interaction with the elastic is required to move it. By requiring simultaneous interaction to move the elastics, the task precludes a divide-and-conquer strategy, thereby requiring closely coupled interaction.

The second task was an open-ended drawing task with a whiteboard-like application where participants could simultaneously sketch and erase. Participants were asked to design and sketch the user interface of a print dialog for digital photographs.

The design task was included to determine whether the behaviours we would see in the elastics and nails task would occur in a less constrained scenario. Since the design task is, by nature, a free form activity without low level constraints, behaviours in this task could more generally be attributed to the embodiment technique itself rather than the task.

3.5 Procedure

The groups of two and four followed slightly different protocols. The main difference was the setup: pairs did not have a voice connection, while groups of four completed all of their tasks with an audio link.

After being introduced to the displays and the elastics and nails task, each group was split, and one person (or pair) led into a room with a SMARTBoard display while the

other was led to the horizontal DVIT display. Participants then cooperatively controlled an elastic until they felt comfortable with both the display and the task.

Pairs completed eight trials of the elastics and nails task, each trial containing six elastics. To further encourage interaction, only one participant was given the target image (e.g. Figure 5a). This participant, called the *director,* ensured that the resulting image, cooperatively constructed with the *follower,* matched the target image. Four of these trials were with VideoArms, and four were with a telepointer-based embodiment. Each participant directed four trials, and followed with four.

Groups of four completed two trials of the elastics and nails task, each with 20 elastics (since we were interested in whether gestures would still be used given the audio channel). Only one participant on each side of the link was given a copy of the target image. Again, these participants were tasked with directing the action for the followers with the stipulation that directors were *not* to touch the elastics themselves. Participants followed with either VideoArms or the telepointer-based embodiment, and then directed with the other. Next, participants moved onto the open-ended design task where, with VideoArms, they designed a print dialogue for digital photographs.

Participants then completed a questionnaire and a semi-structured interview before being debriefed and paid.

3.6 Design Justification

At this early design stage, we are interested in validating VideoArms' potential as an effective embodiment, or understanding problems that hinder it from achieving its promise. Thus, our study is not intended to be a controlled experiment facilitating statistical decision making; instead, it is a fairly broad-brush observational study where we are looking for occurrences of large effects. That said, the tasks need some justification, particularly in terms of their external validity.

The elastics and nails is a contrived task; thus, its face-value external validity is suspect. However, the communicative acts and collaborative processes generated by this task are frequently found in real life workspace tasks: information transfer and the director/follower paradigm [4]; complex gesturing [9]; maintenance of awareness via consequential communication [4], and simultaneous joint activity [4]

The use of the director/follower paradigm appears suspect; however, consider that collaborators rarely go into meetings with equal sets of information. Often, one party has some information that is to be disseminated. In this case, the director plays that role, and disseminates that information to the followers using a combination of gestures and voice instructions. In particular, followers will *follow* the director through the workspace (physically); similarly, it is likely that the directors will (verbally) *direct* followers through the workspace.

In the design task, the specific task of creating a print dialog for digital photographs is perhaps not common in the real world. Yet, it typifies generative design tasks, which are common in any brainstorming activity or task where ideas emerge or are refined pictorially over time. Examples of such tasks include the design of the information flow through an application, or the structuring of a diagram showing different parts of a business' supply chain. All of these utilize the basic processes that this task requires:

communication, collaboration, and content creation. This is why they have been used as the basic task in a variety of other studies (e.g. [2], [4] and [9]).

In generative design activities, collaborators must collectively generate ideas, suggest and draw design elements, and decide how different elements fit together. They do this primarily by communicating both verbally and non-verbally, through drawing marks that comprise image components, and by creating textual lists of ideas [9]. As they draw, collaborators monitor and coordinate each others' actions to ensure the final product accords to their expectations [4].

4. RESULTS

In this section, we report and discuss our study results. We caution that our claims are tentative due to the modest number of participants; however, we stress that the behaviours observed across the 22 participants in seven groups were fairly consistent, and thus suggestive of generalizable behavioural patterns. In general, VideoArms was excellent in its capacity to engage remote participants by demonstrating superior support for complex gestures and consequential communication. We now describe our observations of the use of gestures, consequential communication, and local feedback, and then discuss VideoArms' ability to mitigate presence disparity.

4.1 Gestures

Participants used a wide variety of natural and easily interpreted static and motion-based gestures with VideoArms. With pairs, gestures often acted as audio substitutes. For example: waving to say hello, or "push it that way", or "bring it this way", an a-okay, a hold gesture (open hand with fingers apart), an open-handed wave as an error signal, or a thumbs-up to signal that something was correct.

Across all groups, the variety of VideoArms gestures observed was fairly extensive. Beyond kinetic, spatial and pointing gestures [2], we observed deixis (referential gestures relating to speech), as well as illustrations (gestures clarifying speech). Comments with respect to gesturing with VideoArms were positive, e.g.:

> C: I liked VideoArms because it feels more natural. I can signal her the way I normally would with this sort of sign language.

However, because the fidelity of VideoArms was low, participants generally exaggerated the nature of these gestures both in speed and in size. One possibility for why motion-based gestures were expressed slowly is that the frame rate for local feedback was fairly low (between 8-12 fps): two participants reported that they consciously slowed their actions so that their gestures could be interpreted with the reduced frame rate. To test this theory, we ran one group of four at a higher frame rate (20-25 fps), achieved by removing local feedback. In this condition, all four participants gestured more rapidly (in general) compared to participants in other stages of the evaluation. While this result is promising, we caution that it is preliminary since only one group was observed.
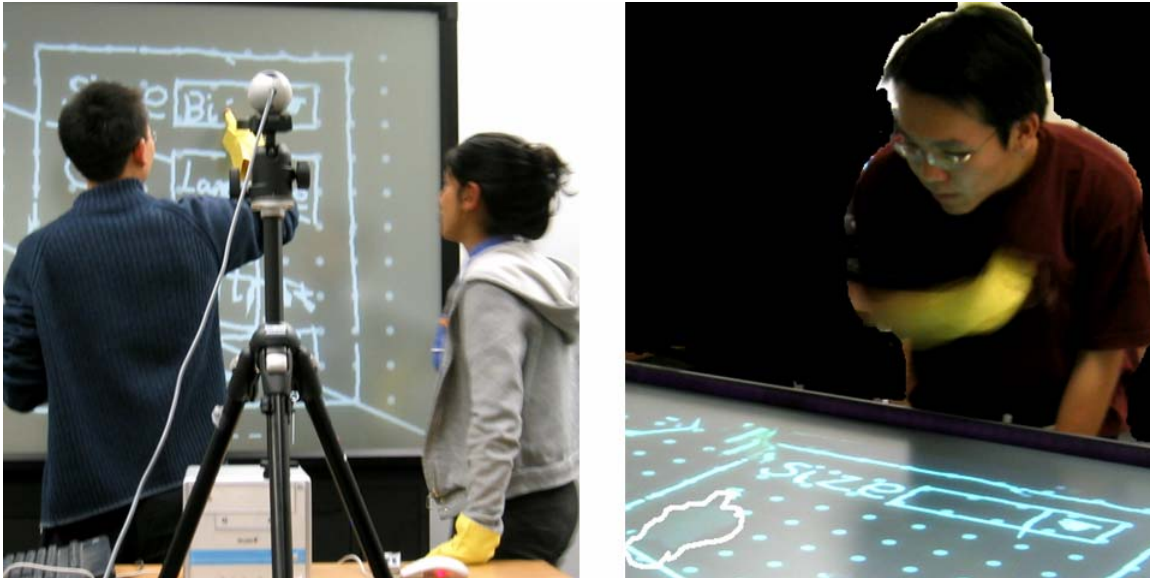
Figure 6. Participants spent a lot of time watching each other. On the left, H watches her collocated partner, W's activities. On the right, D also watches W's activities carefully via VideoArms (W's hand is outlined in white for clarity, bottom-left of table).

VideoArms provided a remarkably useful medium for participants. Participants were able to fluidly gesture and integrate those gestures into their interactions with collocated and remote participants.

4.2 Consequential communication

Participants spent a considerable amount of time observing their partners to understand the state of the activity (Figure 6). In elastics and nails tasks, directors would watch to ensure their partners had grabbed the correct nail, or had positioned the nail in the correct location. When directors detected an error (e.g. if the follower grabbed the wrong elastic or moved a nail to the wrong location), directors would redirect followers to the correct elastic or location. Followers would reciprocally watch directors' actions to determine which nail to pick up. As the task progressed, directors gestured less frequently at which nail to pick up as the consequential communication sufficed.

If an embodiment supports consequential communication, we should also expect to see users correcting the actions of others in the workspace. In the elastics and nails task, it may be to place an elastic in the correct location. In the open-ended design task, it may be to put a picture element in a different place. Corrections are predicated on understanding the state of the workspace and the activities of other participants— knowledge that is gained primarily through consequential communication.

We saw many instances of correction occurring across the link. For instance, one participant interfered with a remote participant by waving aggressively:

> R: I was mostly watching F, but I could also see what M was doing. When it seemed like M wasn't doing what we'd agreed on, I asked him what the heck he was doing.

When probed about what he had been watching, R responded:

> *R: At first, it wasn't so much what he was drawing. I could see that he was completely in the wrong place. When he started drawing, I knew he had the wrong idea.*

Yet the consequential communication provided by the embodiments was not perfect. For example, participants remarked that the imprecision in the video quality of VideoArms often made it difficult to understand their remote partners' activities at times. The problem of image noise meant that the specific location of remote participants was not clear. Furthermore, because remote participants were all represented as yellow latex gloves, it was difficult to identify who was doing what at the remote display. Thus, although VideoArms gave users the freedom to move around the workspace, this freedom actually caused problems in identifying remote collaborators arms. One user remarked, "It's hard to figure out who is who when they're moving around all the time." Situations like these made it difficult for users to determine the activities of specific individuals.

In spite of VideoArms' poor video quality, participants made constant use of the embodiment as a source of consequential communication. VideoArms helped to increase engagement as evidenced by the incidents of corrective acts across the link.

4.3 Local feedback
Participants' responses about the local feedback provided by VideoArms were mixed, reflecting the mixed usage of local feedback during the study. Some participants readily acknowledged the utility of the feedback while others questioned its value; however, an implementation issue may have been the cause of some participants' rejection of local feedback.

Some participants made use of local feedback to ensure their gestures were interpretable. For instance, some participants working in front of the upright display would deliberately stand on either side of the display (as opposed to directly in front of it) to give the camera an unobstructed view of his/her arms.

In contrast, other participants made no use of the local feedback, some because the camera placement limited the utility of feedback (Figure 7), and others because they failed to see the utility of the feedback.

> *A: I can see myself just fine. It doesn't make sense to draw it on the screen. [Local feedback] seems like a waste.*

Finally, as detailed earlier, participants generally gestured slowly with VideoArms because, based on the local feedback, they believed (correctly) that their actions would be sent to remote participants at a very low frame rate. In the single trial, we removed the local feedback altogether without any ill effects: participants did not note the lack of feedback.

The utility of the local feedback provided by VideoArms was not effectively evaluated in this study due to camera placement requirements of VideoArms. However, the single trial where feedback was removed suggests that local feedback may play an implicit role in participants' workspace behaviour.

4.4 Addressing presence disparity
Based on the results from our study, we believe that indeed, VideoArms mitigates presence disparity by supporting our four design principles.
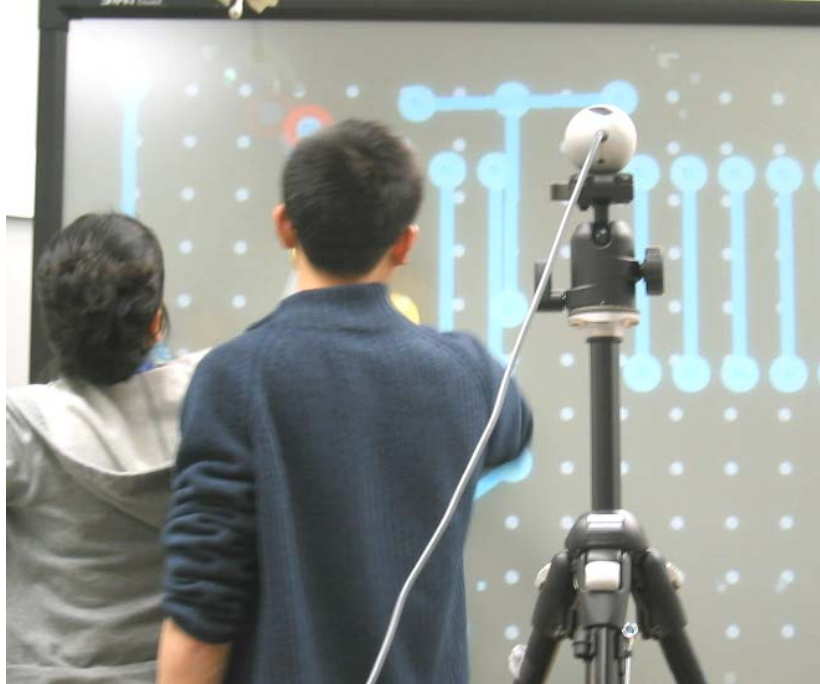
Figure 7. Some groups made no use of local feedback. In this case, H (left) is the director. H's pointing gestures (made with her right hand) cannot be seen remotely because the camera's view is blocked by her partner, W (right).

Participants clearly made use of VideoArms to gesture within the context of the workspace. We observed deixis, which generally have no interpretation out of context [4]. We also observed a wide variety of natural gestures with VideoArms, which persisted in the presence of a voice channel and a collocated collaborator. Importantly, gestures *were not replicated* for remote participants: a single gesture was generally sufficient to communicate to both collocated and remote participants

Participants also made use of VideoArms by carefully watching the arms of others in the workspace, lending support to the importance of consequential communication. Furthermore, we also observed instances of error-correction across the link, facilitated by consequential communication.

Finally, some participants made use of local feedback, in some instances contorting their bodies so that the camera could capture their arms. For one group, we removed local feedback with no clear ill effects.

It seems that participants were able to discover and make use of the features provided by VideoArms. Yet, did VideoArms, by providing some of the features of a physical embodiment, reduce presence disparity? Recall that we defined presence disparity as a marked difference in the level of engagement between collocated and remote participants. We suggested that by increasing the fidelity and richness of the embodiment of remote participants, we could increase the level of engagement with remote participants. In this evaluation, we observed that the increased fidelity of VideoArms enabled richer channels of communication, both by explicit and varied gestures, and with consequential communication. Thus, it is clear that VideoArms indeed reduced presence disparity.

5. DISCUSSION

The results of this study confirm primarily three statements. First, people use complex gestures in the workspace, and like doing so. Second, the current VideoArms implementation is an imperfect system that requires retooling. Finally, although VideoArms was a crude implementation, it increased both communication between remote participants and the overall level of engagement across the link, thereby reducing presence disparity.

As [9] observed, a large portion of the workspace activity involving hands are intentional gestures intended to attract attention or to convey an idea. These gestures are natural and fluid, occur in everyday conversation, and have accepted meanings. Given questionnaire and interview comments, VideoArms appears to support natural gestures in a manner much like what users desired and expected. Gestures in the study were also often combined or used in unpredictable ways (for instance, we observed clapping behaviour to signal "good job"). Furthermore, many participants made extensive use of the two-handed interaction provided by VideoArms.

In spite of this success, VideoArms had a key technical failure.

> J: *If the image quality were better, [the system] would be better. Sometimes, it wouldn't even show me properly.*

VideoArms' images were not clear and crisp enough for participants. First, the colour segmentation technique used was not perfect, producing on-screen artifacts or holes and sometimes confusing users. Second, a participant's body could easily obscure the camera's view of the participant's arms (Figure 7). Finally, the camera's 320×240 image was blown up to 640×480, with the resulting interpolation dulling the image.

VideoArms clearly provided a very rich medium for participants' interactions with remote parties. Participants enjoyed using the medium to work with remote parties (as indicated by questionnaire responses). It also provided a means for users to detect and correct errors by collaborators, indicating that it provided a rich medium for consequential communication. Finally, in comparison to telepointer-based embodiments, VideoArms provides a richer non-verbal communications medium, thereby increasing the overall level of interaction and engagement in the workspace.

The design principles behind VideoArms were formulated to target presence disparity, and the study presented here demonstrates that VideoArms, by supporting these four principles, mitigates presence disparity. In principle then, VideoArms is an effective embodiment technique for mixed presence groupware.

In spite of the successes of VideoArms, it is not a suitable technique in its current form beyond the laboratory setting. VideoArms' main failing is the camera placement with respect to the displays (Figure 8): while this distance allows the physical work surface to fill the camera frame, it neglects the fact that people rarely work with their arms completely and uncomfortably outstretched in the workspace.
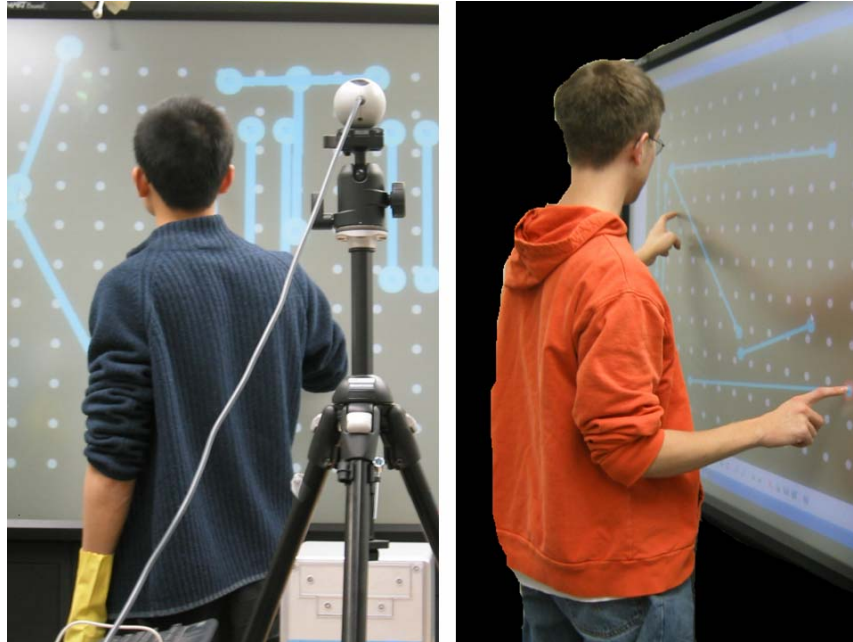
Figure 8. People generally like to work with their hands comfortably in front of themselves as opposed to outstretched.

Instead, people endeavor to maintain comfort. For instance, when working with an upright display, people are more comfortable with their arms positioned in front of them (Figure 8). Similarly, when working over tabletop displays, people will rarely work by extending their arms far from their bodies; instead, they will lean forward with their bodies. The consequence of these comfort-preserving behaviours is that the cameras no longer have an unobstructed view of participants' arms.

Thus, VideoArms fails as a practical embodiment system because, while it maintains the naturalness of gestures and hand/arm-based communication in the work space, it forces collaborators to position themselves awkwardly.

6. CONCLUSION

Our research makes two primary contributions. First, we offer four implications for the design of Mixed Presence Groupware (MPG) embodiments. We present an understanding of social issues in MPG systems, and in particular explain why embodiments should incorporate feedback, consequential communication and gestures to mitigate the presence disparity problem. Our recommendations give guidance to those designing MPG embodiments and technologies.

Second, we contribute VideoArms, a video-based embodiment technique for supporting collocated and distributed collaboration around large displays. Through an evaluation, we showed how VideoArms naturally supports feedback, intentional and unintentional gestures, and consequential communication over MPG groupware surfaces, thereby reducing the presence disparity problem.

VideoArms is not a total solution. For example, eye contact and body positioning, which have been found to be important to collaboration [4], are not supported at all. Yet VideoArms is a reasonable first step for a workspace-focused group because it presents the parts of the body that appear within the workspace context.

VideoArms is a working proof of concept, and as such there is still room to improve its interface as well as the underlying groupware system. These need to be fixed, at which point we will undertake a more thorough empirical evaluation to validate VideoArm's effectiveness as an MPG embodiment. At this point, however, we believe that we have forwarded MPG research into a space where we can begin to understand embodiment design and the tradeoffs between different types of embodiment types within MPG collaboration.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Apperley, M., McLeod, L., Masoodian, M., Paine, L., Philips, M., Rogers, B., and Thomson, K. Use of video shadow for small group interaction: Awareness on a large interactive display surface. *Proc 4th Australasian User Interface Conference (AUIC'03)*, 81-90, 2003.

[2] Bekker, M. M., Olson, J. S., and Olson, G. M. Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. *Proc ACM Designing Interactive systems (DIS'95)*, 157-166, 1995.

[3] Finn, K., Sellen, A. and Wilbur, S. *Video-Mediated Communication.* Lawerence Erlbaum Associates, Inc, 1997.

[4] Gutwin, C. (1997). Workspace awareness in real-time distributed groupware. Ph.D. Thesis, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada. December.

[5] Ishii, H. and Kobyashi, M. Integration of interpersonal space and shared workspace: Clearboard design and experiments. *ACM Transactions on Information Systems,* 11 (4), 349-375, 1993.

[6] Robertson, T. Cooperative work and lived cognition: a taxonomy of embodied actions. *Proc 5th European Conference on Computer Supported Cooperative Work (ECSCW '97)*, Kluwer, 205-220,1997.

[7] Segal, L. D. Designing team workstations: the choreography of teamwork. *Local applications of the Ecological Approach to Human-Machine Systems,* vol 2, P. Hancock, J. Flach, J. Caird and K. Vicente (eds.), LEA Press, 1995.

[8] Tang, A., Boyle, M. and Greenberg, S. Display and presence disparity in mixed presence groupware. *Proc 5th Australasian User Interface Conference (AUIC'04)*, 73-82, 2004.

[9] Tang, J. (1991). Findings from observational studies of collaborative work. *Intl J Man-Machine Studies*, 34(2), 143-160.

[10] Tang, J. and Minneman, S. VideoWhiteboard: Video shadows to support remote collaboration. *Proc ACM CHI'91,* 315-322, 1991.