VideoArms: Embodiments for Mixed Presence Groupware

Anthony Tang[†], **Carman Neustaedter**[‡] **&**
Saul Greenberg[‡]

[†] *Human Communication Technologies Laboratory, University of British Columbia, Vancouver, B.C., Canada*

Email: *tonyt@ece.ubc.ca*

[‡] *Interactions Laboratory, University of Calgary, Calgary, Alberta, Canada*

Email: *{carman,saul}@cpsc.ucalgary.ca*

Mixed presence groupware (MPG) allows collocated and distributed teams to work together on a shared visual workspace. Presence disparity arises in MPG because it is harder to maintain awareness of remote collaborators compared to collocated collaborators. We examine the role of one's body in collaborative work and how it affects presence disparity, articulating four design implications for embodiments in mixed presence groupware to mitigate the effects of presence disparity: embodiments should provide local feedback; they should visually portray people's interaction with the work surface using direct input mechanisms; they should display fine-grain movement and postures of hand gestures, and they should be positioned within the workspace. We realize and evaluate these implications with VideoArms, an embodiment technique that captures and reproduces people's arms as they work over large displays.

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

Tang, A., Neustaedter, C. and Greenberg, S. (2006)
VideoArms: Embodiments for Mixed Presence Groupware. In N. Bryan-Kinns, A. Blanford, P. Curzon and L. Nigay (Eds), People and Computers XX - Engage (Proceedings of HCI 2006), September, Springer.

1 Introduction

Large surfaces such as tabletop and whiteboards naturally afford collocated collaboration, allowing multiple people to work together over the shared display. As large digital displays become more ubiquitous, we anticipate they will offer a shared workspace for not only collocated people, but distant collaborators as well.

> *Imagine you are a member of a design team located in Calgary. You schedule a brainstorming session with your Vancouver-based counterparts on a new product idea. Your company has special meeting rooms in each city, connected by audio links and containing large digital stylus-based whiteboard displays. Groupware allows members of your Calgary team and the Vancouver team to concurrently draw ideas on the display wall using styli, which everyone sees in real time.*

This scenario describes *mixed presence groupware* (MPG), software that connects both collocated and distributed collaborators together in a shared space. Although hardware support for this MPG scenario already exists, we do not yet know how to design software to support this kind of activity in a fluid, seamless way. MPG systems are still in their infancy: to date, only a few research systems have investigated this arrangement of collaborators [Apperley et al. 2003; Everitt et al. 2003; Tang et al. 2005]. Yet simply providing technological support for MPG ignores a core problem called presence disparity: in MPG workspaces, some collaborators are physically present, while others are not. The result of this discrepancy is that collaborators tend to focus their energy on collocated collaborators at the expense of their distributed counterparts [Tang et al. 2005].

One reason for this asymmetric interaction is that collocated collaborators are seen in full fidelity, while remote participants are represented by only embodiments – virtual presentations of their bodies. Most commercial groupware systems reduce this virtual presentation to a telepointer (remote mouse cursor), which clearly cannot compete against the communicative power of a physical body. Presence disparity unbalances a collaborator's experience of the group: maintaining awareness, sensing engagement and involvement and communicating is much easier with collocated collaborators compared to remote collaborators.

In this paper, we explore the problem of designing embodiments for MPG. First, we develop an understanding of the role collaborators' bodies play in collaborative work by exploring three concepts – feedback and feedthrough, consequential communication, and gestures. From these, we articulate four design implications for MPG embodiments to mitigate presence disparity:

1. embodiments should be visible to both collocated and remote collaborators;
2. embodiments should be driven by direct input mechanisms and presented in high fidelity;
3. embodiments should capture and display fine-grained movements and postures; and
4. embodiments should be positioned in the context of the workspace.

Second, we apply these implications to design a prototype system called VideoArms. As we will see, VideoArms provides a rich embodiment by digitally capturing people's arms as they work over large work surfaces, where it overlays these arms on the remote displays. Finally, we present the results of a pilot study that support our current VideoArms design directions for embodiments in MPG.

2 Background: Bodies in Collaborative Work

The physical body plays a large role in collocated collaboration, helping to explicitly convey information, and providing a means for others to maintain an awareness of our workspace activities [Gutwin 1997]. For embodiments in mixed presence groupware to reduce presence disparity, we need to understand the particular communicative affordances bodies bring the collaborative process so that we can recreate them for remote collaborators.

This section reviews three concepts that give some insight to how bodies contribute to collaborative work [Pinelle et al. 2003]: feedback and feedthrough, consequential communication, and gestures. Although these concepts are well known in the computer-supported cooperative work (CSCW) community, they manifest themselves differently in mixed presence groupware. By reviewing these concepts and reflecting on their consequences in naïve MPG implementations, we derive four design principles for MPG embodiments.

2.1 Feedback and Feedthrough: Perceiving Ourselves and Others

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 [Robertson 1997]. Our ability to perceive ourselves is important: without our ability to perceive our own bodies as physical objects in the world, threading a needle when blindfolded might otherwise be a painful experience.

In distributed groupware, feedback is echoed to other participants as *feedthrough*, the reflection of one person's actions on other users' screens [Dix et al. 1998]. In collaborative work it is important to be able to understand remote collaborators' actions and the effect they are having on the workspace. Within a distributed system, feedback and feedthrough play a dual role: feedback not only informs us of our own actions, but gives us insight to how our actions are being interpreted on the other side (the feedthrough).

In mixed presence groupware, one only needs to look at collocated collaborators to acquire full feedthrough. Because feedback and feedthrough are the same, the person doing the action also knows what the other person can see [Rodden 1996]. In contrast, one may see only partial feedthrough of a remote collaborator's actions. Because feedback and feedthrough may not be identical (e.g. due to network latency or other deficiencies in the system), the person performing the action (e.g. a gesture) can only intuit what remote collaborators might see. This dissimilarity between feedback and feedthrough for remote vs. local collaborators can introduce imbalance, confusion, and uncertainty in how people experience the interaction.

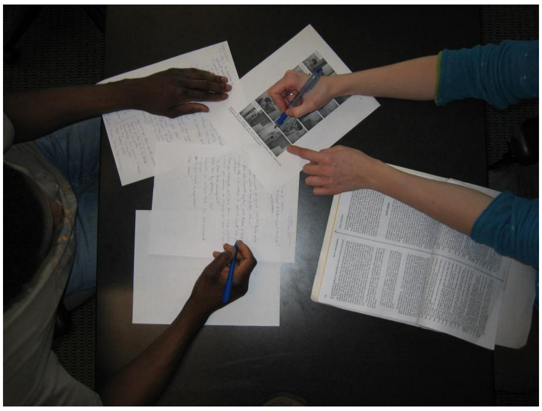

Figure 1: A bird's eye view of a physical workspace.

This imbalance between feedthrough and feedback suggests our first design principle for mixed presence groupware embodiments. *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.*

2.2 Consequential Communication: Watching Others Work

Our bodies are the source of *consequential communication*: information generated as a consequence of our activities in the workspace [Segal 1995]. A person's activity in the workspace naturally generates rich and timely information often relevant to collaboration. For instance, the way a worker is positioned, and the types of tools or artifacts being held and used tells others about that individual's current and immediate future work activities (e.g. Figure 1).

The graceful choreography of teamwork arises from the subtle role played by consequential communication. Segal [1995] found that pilots spend 60% of their time simply observing co-pilots' consoles while they were being manipulated. Further, he reports that pilots would often react smoothly to one another's actions without explicit verbal cuing. Similarly, Gutwin [1997] observed that 'participants would regularly turn their heads to watch their partners work' in small group interaction. Tang's [1991] reports of choreographed hand movements during group work over physical surfaces can also be understood in terms of consequential communication: by observing others' actions and activities in a shared workspace, one can fairly accurately predict others' future acts or intentions, thereby easily working with or around them. Consequential communication is an important conduit for maintaining an awareness of others, allowing us to monitor, understand and predict others' actions in the workspace without explicit action on their part [Gutwin 1997].

In mixed presence groupware, consequential communication between collocated vs. remote participants is out of balance, as people have different views of their collocated and remote participants. Collocated actions over the physical workspace allow others to observe individual atomic-level interactions with the workspace (e.g. reaching towards a stylus, fingers grasping the stylus, lifting the stylus, moving the stylus towards the display, touching its tip to the display, etc.), allowing them to predict future activities well. Indirect input devices (e.g. mice) can restrict consequential communication between collocated participants, since they can no longer see how bodies are attached to actions, or how actions are generated [Gutwin & Greenberg 1998]. For remote collaborators, one's ability to observe others depends directly on the embodiment's abstraction and fidelity. Yet virtual environments typically tend away from atomic-level interactions, often representing activities at a coarser level (e.g. a mouse pointer changes into a pen representing a mode change from pointing to drawing, or a pen suddenly appearing in an avatar's empty hand). This abruptness makes remote participants' actions less predictable.

MPG embodiments need to have a comparable range of expressiveness and fidelity compared to their corporeal counterparts if they are to provide parity in the consequential communication that is conveyed. The embodiment must capture appropriate information, and present it in an interpretable way: the closer an embodiment's presentation relates to the activities of the participant, the easier those activities are to interpret. This brings us to our second implication for the design of MPG embodiments. *To support consequential communication for both collocated and distributed participants, people should interact through direct input mechanisms, where the remote embodiment of how the input device is manipulated is presented at sufficient fidelity to allow collaborators to easily interpret all current actions as well as actions leading up to them.*

2.3 Gestures: Facilitating Intentional Communication

Gestures are intentional bodily movements and postures used for communicative purpose [Bekker et al. 1995; Kirk et al. 2005]. Gestures provide participants with a spatial and kinetic means to express their thoughts, reinforcing what is being done and said in the workspace. Gestures are a frequent consequence of how bodies are used in collaborative activity: Tang [1991] observed that 35% of hand activities in a physical workspace were gestures intended to engage attention and express ideas. Because intentional gesturing is so frequent, hindering the process – by not giving participants the ability to view or to produce gestures effectively – may negatively impact collaborative activities in mixed presence groupware.

Two classes of gestures facilitate the communication of ideas and coordination in group work: pure communicative acts, and those that relate to the workspace and its artifacts. Pure communicative gestures, which arise from a person's natural communicative effort, are used by both the speaker and listener for fluid interaction. People use such gestures to facilitate both speech production [Krauss et al. 1995], and interpretation [Riseborough 1981]. Gestures can also convey semantic information above and beyond speech alone (e.g. deictic gestures), and some replace speech entirely (e.g. yes or no via thumbs-up or thumbs-down). Similar gestures are also used to help coordinate conversational turn-taking (e.g. putting up one's hand to express a desire to speak, or gesturing at the next speaker).

The communicative value of these pure communicative gestures relies on our ability to produce gestures by animating our bodies, and upon others being able to see these gestures in detail. In mixed presence groupware, while collocated collaborators see these gestures in detail, remote participants do not. This leads to our third implication for the design of MPG embodiments: *To support bodily gestures, remote embodiments should capture and display the fine-grained movement and postures of collaborators. Being able to see these gestures means people can disambiguate and interpret speech and actions.*

Workspace-oriented gestures relate directly to the collaborative workspace and the artifacts contained within. These gestures typically refer to objects or locations in the workspace, or clarify verbal communication by illustration over the workspace [Harrison & Minneman 1994]. Bekker et al. [1995] identify three workspace-oriented gestures: kinetic (movement that illustrates an action sequence), spatial (movement that indicates distance, location or size), and point (pointing at a person, object or place, where targets may be concrete, abstract, denoting an attitude, attribute, effect, direction or location) – often referred to as a deictic reference. Bekker et al. [1995] also observed that gestures were often combined into *sequences*. For example, one common sequence in design activities is a *walkthrough*: a succession of kinetic gestures illustrating how something might be used. Since collaborators will often combine atomic-level gestures in novel sequences to express ideas, attempting to support remote gesturing by providing 'canned' gestures would be cumbersome.

Further, Bekker et al. [1995] highlighted the importance of the design role of gestures: those that relate to design activity, such as referring to objects, persons or places, showing distances, enacting the interaction between user and product, etc. This role shows that a gesture's semantic information is often heavily related to the context in which it is produced. For instance, gestures in the workspace often refer to objects or locations in the workspace (e.g. 'I think this should be this big').

In mixed presence groupware, collocated collaborators see exactly how these gestures are enacted over the workspace. Yet workspace-oriented gestures of remote participants are often shown via a telepointer: a crude surrogate where information fidelity is lost. Alternatively, gestures of remote collaborators are often seen in a video stream outside the workspace, which removes much of the meaning conveyed by the gestures. Thus, our fourth implication for the design of MPG embodiments is that: *To support bodily gestures as they relate to the workspace context, remote embodiments should be positioned within the workspace to minimize information loss that would otherwise occur.*

This discussion of gestures reinforces our second implication recommending direct input mechanisms. Since the ability to freely use gestures is important for fluent speech production, smooth interaction in MPG is necessarily best facilitated by un-tethered input devices (pens, touch surfaces) that interact directly with the display surface. This leaves people free to both gesture and work directly over the work surface. Tethering users to input devices such as keyboards or mice inhibits users from gesturing as a part of their communicative effort.

In closing, we should mention that our review does not consider the role of eye contact for interpersonal communication, and eye gaze for knowing where others are focusing their attention [e.g. Ishii & Kobayashi 1993]. Instead, we have reviewed

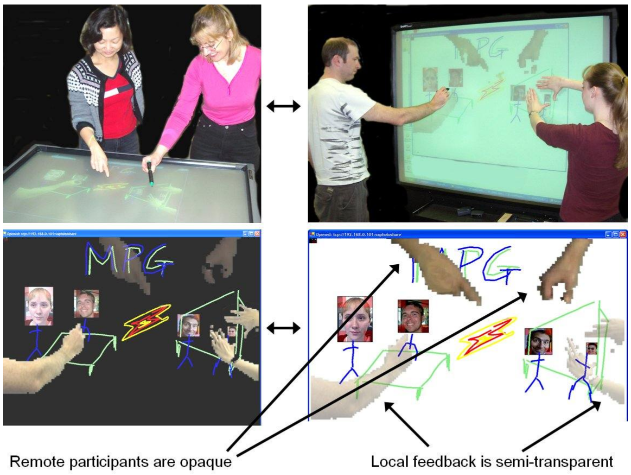
MPG
Remote participants are opaque
Local feedback is semi-transparent

Figure 2: A sample MPG session using VideoArms.

three concepts revealing how bodies – particularly the visible aspects of the body from a top-down view (Figure 1) – facilitate the collaborative process. Furthermore, we have suggested why these lead to presence disparity problems in mixed presence groupware, and recommend how this disparity can be mitigated through careful embodiment design. In the next section, we put these design principles to practice in building VideoArms, an MPG embodiment.

3 VideoArms: A Video-based MPG Embodiment

VideoArms is a prototype video embodiment mixed presence groupware system that visually recreates the part of the body normally seen over the workspace: people's arms. In this section, we give an overview of our VideoArms system, briefly explain its relationship with other similar systems and how it addresses each of our design principles. We then briefly describe its implementation.

VideoArms digitally 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. Two connected groups of collaborators (Figure 2, top) each work over different touch-sensitive surfaces. Each surface runs the same custom MPG application, allowing all participants to simultaneously see, sketch and manipulate artifacts within a common workspace. Figure 2 (bottom) gives a close up of what these participants can see when using the VideoArms embodiment in this MPG application:

1. collaborators see their own arms as local feedback, rendered semi-transparently;

2. each group sees the solid arms of remote participants in 2.5-dimensional fidelity (the system captures and reproduces colour-based depth-cues); and

	Local feedback of embodiment	Direct input mechanism	Rendering of fine-grain movement	Workspace-embedded embodiments
Agora [Kuzuoka et al. 1999]		✓	✓	✓
ClearBoard [Ishii & Kobayashi 1993]		✓	✓	✓
Designer's Outpost [Everitt et al. 2003]		✓		✓
Facetop [Stotts et al. 2004]	✓		✓	✓
LIDS [Apperley et al. 2003]	✓		✓	✓
Roussel [2001]			✓	✓
TeamWorkstation [Ishii & Kobayashi 1993]	✓		✓	
VideoDraw [Tang & Minneman 1991a]		✓	✓	✓
VideoWhiteboard [Tang & Minneman 1991b]		✓	✓	✓
WSCS-II [Miwa & Ishibiki 2004]		✓	✓	✓
VideoArms	✓	✓	✓	✓

Table 1: How various video-based embodiment techniques address the four design implications of MPG embodiments.

3. remote arms are painted to 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.

Figure 2 also reveals communicative aspects of the embodiment. In this MPG setting, 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. Finally, users are not tethered to particular locations 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.

3.1 Related Systems

The VideoArms metaphor captures and presents the workspace from a bird's eye view of the workspace, cf. 'through the glass' metaphor from earlier work [Ishii & Kobayashi 1993; Tang & Minneman 1991b]. From this perspective, the arms are the primary indicators of a collocated collaborator's presence (as in Figure 1). While VideoArms builds upon concepts of other non-MPG systems that integrate video feeds of remote collaborators within the workspace, it differs in several respects:

1. VideoArms' design is an attempt to solve the problem of presence disparity unique to MPG using the design implications described earlier;

2. VideoArms facilitates distortion-free composition of multiple video feeds and the evaluation of more abstract presentation techniques;

3. VideoArms is intended to support multiple collaborators at a site, allowing collaborators to see and interpret fine-grained activities of remote collaborators: most other systems assume only a single person per site.

Table 1 summarizes how embodiment techniques offered in other systems only partially address our four MPG design implications.

VideoDraw [Tang & Minneman 1991a], VideoWhiteboard [Tang & Minneman 1991b], TeamWorkstation and ClearBoard [Ishii & Kobayashi 1993] were all intended to connect a pair of distance-separated collaborators, each of whom could draw in a shared workspace. These systems used analog cameras to transmit both the images of the collaborators (their arms and bodies in VideoDraw and VideoWhiteboard, and their faces in TeamWorkstation and ClearBoard) and the contents of the workspace. While effective for their purposes, these systems suffered from two major limitations:

1. people were not able to manipulate each other's physical drawing marks (although later versions of ClearBoard addressed this problem using transparent digital displays); and

2. the analog video mixing technology limited the number of sites that could be composited without significant image degradation.

Facetop is a digital video-based system intended to support two remotely located extreme programmers that uses a ClearBoard-like metaphor [Stotts et al. 2004]. Roussel [2001] uses a chroma-key technique to address the image degradation issues. While both systems are excellent for two remote collaborators, the techniques do not adequately support collocated consequential communication due to the physical separation of the gesturing area and input area.

LIDS uses a fully digital system to recreate VideoWhiteboard for distributed PowerPoint presentations [Apperley et al. 2003]. LIDS captures the image of a person working in front of a shared display using consumer-grade cameras, and transforms this image via background subtraction and posturing techniques into a frame containing the digital shadow of the person. Three images are then overlaid to create the scene: the digital shadow, the PowerPoint slide, and another overlay that captures digital annotations. Similarly, the Distributed Designer's Outpost [Everitt et al. 2003] also captures digital shadows via rear-projection; however, the low fidelity of the shadows is only useful for showing another person's presence and very coarse gestures. As with VideoWhiteBoard, both approaches use shadows, which provide considerably less detail than full fidelity images – a desired feature according to users of Distributed Designer's Outpost.

Most of the preceding examples were designed to support collaboration between distributed individuals (instead of groups). With MPG, we explicitly design for collaboration between distributed sites with multiple individuals [Tang et al. 2005]. Only three of the systems were designed to support MPG explicitly: Agora [Kuzuoka et al. 1999], Distributed Designer's Outpost [Everitt et al. 2003] and WSCS-II [Miwa & Ishibiki 2004]. Agora builds on the analog approaches of ClearBoard and VideoWhiteboard to support two dyads, sharing the same limitation that physical artefacts cannot be manipulated in remote locations. WSCS-II's approach produces a shared virtual space, thereby allowing participants who are not actively engaged in the task to be embodied. In contrast, our focus is primarily in a shared work surface, and the active participants on the surface.

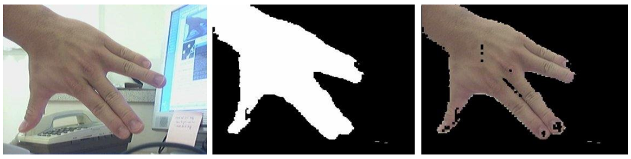

Figure 3: The image on the left is colour-segmented to find the skin-colour pixels (middle). The two images are then combined to produce the VideoArms image on the right.

While VideoArms builds on these prior approaches, it explicitly addresses the problem of presence disparity in MPG by supporting our four design implications:

- Local participants know what remote people see because their own embodiments are shown as semi-transparent feedback.
- Because the body is used as an input device that works directly on the touch sensitive surface, VideoArms supports consequential communication. Other collaborators (whether collocated or remote) can easily predict, understand and interpret another's actions in the workspace as one reaches towards artefacts and begins actions. Because collaborators are not tethered to input devices, their actions are direct and in the workspace context.
- Rich gestures (coupled with conversation and artifact manipulation) are well supported because the remote arms are displayed in rich 2.5-dimensional fidelity and a reasonable (although not ideal) framerate (~12 fps) that proved acceptable for interpreting gestural meanings.
- Task-related gestures are easily interpreted because they are placed in the context of the workspace.

3.2 Implementation Details

In this section, we show how all of the above design implications are realized by describing the key implementation details of VideoArms.

VideoArms uses inexpensive web cameras hand-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 (see Friedland et al. [2005] for a more robust implementation). Transmitted images are processed at the remote workstation to appear as an overlay atop the digital workspace. To provide local feedback, VideoArms overlays the local person's video on the work surface. To avoid image degradation (and thus facilitate scaling to multiple sites), VideoArms extracts and composites onto the workspace image only a person's body parts (such as one's arms): all other background visuals are removed.

Frames captured by the camera are processed, transmitted and displayed in a four step process (Figure 3). First, pixels matching skin colour (based on a Mahalanobis distance calculated against a sample of 10 or more skin sample pixels) are identified. Morphological opening is applied to this skin mask to produce a silhouette mask (Figure 3, middle). Second, this mask is combined with the original image (Figure 3, right). Third, the image is transmitted to all clients using UDP packets for quick delivery. Finally, standard raster graphics compositing techniques are used to paint the image on the groupware work surface.

VideoArms uses Python, the .NET Framework, the Intel Performance Primitives library, the Python Imaging Library, and the Python numarray open source libraries. 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. While further optimizations are possible, our primary intention was to develop a system suitable to test our ideas rather than to produce a production-level implementation [see Friedland et al. 2005].

4 Initial Experiences from an Exploratory Study

We conducted an exploratory study with pairs and groups of four to understand whether our approach to embodiment design had merit in terms of mitigating presence disparity. At this early design stage, we were interested in an initial validation of our design implications for mixed presence groupware embodiments. This exploratory study was aimed to be observational and fairly broad-brush, designed so that we could look for large effects and critical incidents:

- What problems would participants have with VideoArms?
- Would participants make use of the ability to gesture freely? Would they continue to gesture even if there was a voice link, and were these gestures intended for remote collaborators, collocated collaborators or both?
- Would consequential communication occur across the link?

In essence, our larger goal was to see if a richer, video-based embodiment of remote collaborators could mitigate the effects of presence disparity on the collaborative process as they worked on their natural activities. We also recognized that VideoArms might be an imperfect instantiation of our design implications, so our lesser goal was to look for specific design flaws and to iterate over our design.

4.1 The Study

Pairs and groups of four completed a series of collaborative workspace tasks (directed puzzle completion and a design task) using a custom mixed presence groupware application on two large displays (one table, one upright whiteboard) running across a remote link. The puzzle completion task was designed so that participants had asymmetric knowledge about how the finished puzzle should look (and therefore had to cooperate with one another to complete the task). With groups of four, one participant on each side of the link had knowledge of the finished puzzle, but these participants were restricted to directing the other participants in completing the puzzle (they were not allowed to directly work on the puzzle themselves). The design task allowed participants to freely sketch their ideas on the workspace (similar to a standard whiteboard), and asked them to design a photograph print dialogue.

These tasks are modified forms of the follower+director task from [Gutwin 1997] and the design task from [Tang 1991].

Participants worked over a custom-built MPG application on two different large displays. To simulate remote collaboration, displays were located in separate rooms. The first was a rear-projected, touch sensitive SMARTBoard, which has a 167.6cm screen (diagonal). The second was a similarly sized but horizontally mounted and front-projected DVIT display. The DVIT display could support two simultaneous touches, but the SMARTBoard could not. To prevent this technical difference from affecting the results of the study, the study software interpreted only one touch per board. Each group of participants was split in two: for groups of two, one participant worked in front his or her own display; similarly, groups of four were split into two pairs, and each pair worked in front of a shared display.

Using a partial within-subjects design, participants completed the puzzle completion tasks alternately with VideoArms, and then with telepointers only. Some groups had a voice link, some did not (to understand how voice affected gesture interpretation). Finally, groups of four completed the design task with only VideoArms. We videotaped the sessions, and collected field notes detailing the kinds of gestures that were used with the different embodiment techniques, and the kinds of interaction patterns that were evident.

We recruited 22 paid participants from the university computer science student population. We chose users familiar and comfortable with computers, and asked that they come in pairs (and in four cases, groups of four).

Finally, to expedite the calibration process, participants wore yellow dishwashing gloves to use with VideoArms (their bright, uniform colour facilitated easy extraction of arm images). While VideoArms was designed to pick up skin tones, we took this shortcut for two reasons:

1. we could calibrate the system for glove colour ahead of time (instead of recalibrating for each group); and
2. our primary interest was not the computer vision algorithm used to extract skin features, but on the collaborative aspects of the system – we did not expect the use of gloves to affect the outcome.

Indeed, if VideoArms proves worthwhile, we anticipate that computer vision specialists could rework our implementation to generate far more efficient implementations and faster calibration methods [Friedland et al. 2005].

4.2 Major Findings

We saw a consistent, constant mix of natural gesturing behaviour and consequential communication regardless of the embodiment (VideoArms vs. telepointers). However, the nature of the gestures was far more varied and natural with the VideoArms embodiment. Consequently, VideoArms was able to engage participants across the link in a far richer way regardless of the group size. This section reports on these observations of participant behaviour with illustrative vignettes from the sessions. We caution again that this is an exploratory study. Our claims are somewhat tentative due to the modest number of participants; however, we stress that the behaviours observed across our participant groups were fairly consistent, and thus suggestive of generalizable behavioural patterns.

Consistent use of 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 [Bekker et al. 1995], we observed deixis (referential gestures relating to speech), as well as illustrations (gestures clarifying speech). The following session transcript illustrates how participants appropriated VideoArms for two-handed gestures – something that was impossible in the telepointer condition:

(L and M are on opposite sides of the link.)

L: *With her left hand, L points to an artefact that M should grab. Once M has touched the artefact, L points to where M's artefact should go with her right hand. L then grabs her own artefact with her left hand and moves it in place (still pointing with her right hand), checking to see if M has moved hers to the right place.*

L: *Satisfied that M has moved it to the right place, L retracts her right hand, and makes a full-arm clapping motion.*

Because the fidelity of VideoArms was low (compared to real life), participants generally exaggerated the nature of these gestures both in speed and in size – a direct response to the local feedback of the embodiment (i.e. the feedback was not 'keeping up' to the speed of the gesture, or the gesture was too subtle to be seen).

Rich gestures used as part of the collaborative process. VideoArms provided a remarkably useful communications medium for participants. Participants were able to fluidly gesture and integrate those gestures into their interactions with collocated and remote participants. Further, these gestures were more varied and natural (accompanying speech) than those expressed with the telepointers:

(J & K are collocated, and separate from B & C.)

J: *'Okay, K, move yours over to here.' J points at a location.*

B: *In the meantime, B on the other side has directed C to move her artefact to a certain spot.*

J: *J sees that C has not moved it exactly to the right position. 'C, could you guys move it closer to right over here', J makes a jabbing motion with her finger, as if she could push C's hand to the right position.*

With the telepointer-based embodiments, many of the gestures were motion-based, including waving (to indicate presence or to garner attention), directed thrusting

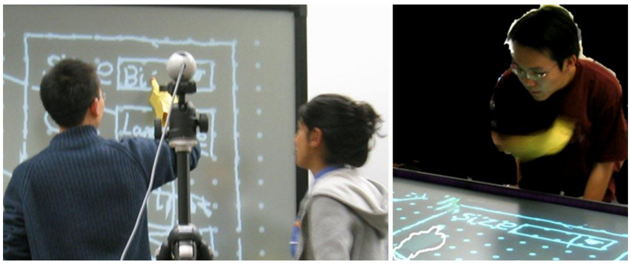

Figure 4: Participants spent a lot of time watching each other. On the left, H watches her colocated partner W's activities. On the right, D also watches W carefully via VideoArms.

to indicate a location, and so forth: artificially impoverished versions of real-life gestures. Most interestingly, we occasionally observed collaborators 'incorrectly' pointing with their hands instead of using the telepointer embodiment. This meant that those gestures would not be seen by remote collaborators. It also suggests that gestures are most naturally performed using the physical body – something that VideoArms supports by design.

Watching is an integral part of the collaborative process. Participants spent a considerable amount of time observing their partners (whether collocated or remote) to understand the state of the activity, regardless of the type of embodiment (Figure 4). In the puzzle task, directors would watch to ensure their partners had grabbed the correct artefact, or had positioned the artefact in the correct location. When directors detected an error (e.g. if the follower grabbed the wrong artefact or had moved it to the wrong location), directors would redirect followers to the correct artefact or location. Followers would reciprocally watch directors' actions to determine which artefact to pick up.

If an embodiment supports consequential communication, we should also expect to see users correcting the actions of others in the workspace. Of note, we saw many instances of correction occurring across the link in the groups of four conditions. This means that participants were sufficiently engaged with remote participants to suggest corrections instead of waiting for the mistake to be noticed. The previous vignette illustrates an instance of this occurrence.

5 Discussion and Conclusions

Based on the results from our observational study, we believe that our design principles are appropriate starting points for embodiments in mixed presence groupware. We saw evidence that VideoArms helped to mitigate presence disparity by promoting more varied yet natural communication across the link.

Participants used VideoArms to gesture in the workspace. We observed deixis, and 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. By increasing the level and style of engagement across the link, VideoArms helped to mitigate presence disparity.

Further iteration on VideoArms is required to make it a practical embodiment system. As a prototype system, VideoArms had two limitations:

1. poor image quality; and
2. impractical camera placement.

VideoArms' colour segmentation technique produced on-screen artifacts, leaving images not clear and crisp enough for participants. More robust implementations are available [e.g. Friedland et al. 2005; Wilson 2005]. Second, the placement and use of cameras poses practical problems: with a vertical display, a collaborator's body sometimes occluded the camera's view of his or her arms. As a consequence, participants sometimes worked with their arms uncomfortably outstretched so that remote collaborators could see. In spite of these shortcomings, we saw very convincing evidence of VideoArms' utility as a communication medium. We predict that collaborators would likely make even further use of a better implementation.

The first generation of groupware systems succeeded by making the impossible possible: by letting people share views of their computer display, they gained the ability to work in real time over computer artifacts. As groupware moved on to successive generations, attention was increasingly moved to the fine-grained nuances of communicating through technologies [Pinelle et al. 2003]: subtleties in how people maintained awareness of one another's actions in the workspace [e.g. Gutwin 1997; Gutwin & Greenberg 1998], the role of gestures [e.g. Bekker et al. 1995; Krauss et al. 1995; Tang 1991], eyegaze [Ishii & Kobayashi 1993], feedthrough [Dix et al. 1998], consequential communication [Segal 1995], etc.

Our research continues the quest to programmatically capture, transmit and display much of the rich information that makes up the collaborative process. In doing so, we make three primary contributions:

First, we suggest that careful embodiment design can mitigate the presence disparity problem in mixed presence groupware, and offer four implications for their design grounded in a theoretical understanding of how people socially interact over a workspace. We explain why embodiments should incorporate feedback, consequential communication and gestures to mitigate the presence disparity problem, hoping to guide those designing MPG embodiments and technologies.

Second, we contribute VideoArms as a method: a video-based embodiment technique for supporting collocated and distributed collaboration around large displays. We recognized the intellectual roots of VideoArms in its predecessor systems, showing VideoArms' method extends previously presented concepts to the MPG setting, while recognizing the varied design choices of these earlier systems.

Third, we present early observations and a critique of VideoArms, for we expect future researchers not only to build on our successes but to try to overcome our failures. We believe that VideoArms is a reasonable first step for a workspace-focused MPG group because it presents the parts of the body that appear within the workspace context. Yet we recognize that eye contact and body positioning, which have been found to be important to collaboration [Ishii & Kobayashi 1993] are not supported at all. Similarly, we point out technical limitations of VideoArms: it is currently a working proof of concept, and as such there is still room for better performance. Issues such as frame rate, image extraction, camera positioning, skin colour calibration, latency, and so forth need to be fixed and improved.

VideoArms is best considered as a first serious solution to solving the presence disparity problem in MPG. We believe we have forwarded MPG research into a space where we can begin to understand embodiment design, and the tradeoffs between different embodiment types within MPG collaboration.

6 Acknowledgements

We thank members of the Interactions Lab for their intellectual support, and NSERC for financial support.

References

Apperley, M., McLeod, L., Masoodian, M., Paine, L., Philips, M., Rogers, B. & Thomson, K. [2003], Use of Video Shadow for Small Group Interaction: Awareness on a Large Interactive Display Surface, *in* R. Biddle & B. Thomas (eds.), *Proceedings of the Fourth Australasian User Interface Conference (AUIC 2003)*, Australian Computer Society, pp.81–90.

Bekker, M. M., Olson, J. S. & Olson, G. M. [1995], Analysis of Gestures in Face-to-face Design Teams Provides Guidance for How to Use Groupware in Design, *in* G. Olson & S. Schuon (eds.), *Proceedings of the Symposium on Designing Interactive Systems: Processes, Practices, Methods and Techniques (DIS'95)*, ACM Press, pp.157–66.

Dix, A., Finlay, J., Abowd, G. & Beale, R. [1998], *Human–Computer Interaction*, second edition, Prentice–Hall.

Everitt, K. M., Klemmer, S. R., Lee, R. & Landay, J. A. [2003], Two Worlds Apart: Bridging the Gap Between Physical and Virtual Media for Distributed Design Collaboration, *in* V. Bellotti, T. Erickson, G. Cockton & P. Korhonen (eds.), *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI'03), CHI Letters* **5**(1), ACM Press, pp.553–60.

Friedland, G., Jantz, K. & Rojas., R. [2005], SIOX: Simple Interactive Object Extraction in Still Images, *in* S. Kawanda (ed.), *Proceedings of the Seventh IEEE International Symposium on Multimedia (ISM'05)*, IEEE Computer Society Press, pp.253–60.

Gutwin, C. [1997], Workspace Awareness in Real-time Distributed Groupware, PhD thesis, Department of Computer Science, University of Calgary.

Gutwin, C. & Greenberg, S. [1998], Effects of Awareness Support on Groupware Usability, *in* C.-M. Karat, A. Lund, J. Coutaz & J. Karat (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'98)*, ACM Press.

Harrison, S. & Minneman, S. [1994], A Bike in Hand: A Study of 3-D Objects in Design, *in* K. Dorst, H. Christiaans & N. Cross (eds.), *The Delf Protocols Workshop: Analyzing Design Activity*, John Wiley & Sons, pp.205–18.

Ishii, H. & Kobayashi, M. [1993], Integration of Interpersonal Space and Shared Workspace: Clearboard Design and Experiments, *ACM Transactions on Office Information Systems* **11**(4), 349–75.

Kirk, D., Crabtree, A. & Rodden, T. [2005], Ways of the Hands, *in* H. W. Gellersen, K. Schmidt, M. Beaudouin-Lafon & W. Mackay (eds.), *Proceedings of ECSCW'05, the 9th European Conference on Computer-supported Cooperative Work*, KluwerAcad, pp.1–21.

Krauss, R., Dushay, R., Chen, Y. & Rauscher, F. [1995], The Communicative Value of Conversational Hand Gestures, *Journal of Experimental Social Psychology* **31**(6), 533–52.

Kuzuoka, H., Yamashita, J., Yamazaki, K. & Yamazaki, A. [1999], Agora: A Remote Collaboration System that Enables Mutual Monitoring, *in* M. E. Atwood (ed.), *CHI'99 Extended Abstracts of the Conference on Human Factors in Computing Systems*, ACM Press, pp.190–1.

Miwa, Y. & Ishibiki, C. [2004], Shadow Communication: System for Embodied Interaction with Remote Partners, *in* J. Herbsleb & G. Olson (eds.), *Proceedings of 2004 ACM Conference on Computer Supported Cooperative Work (CSCW'04)*, ACM Press, pp.467–76.

Pinelle, D., Gutwin, C. & Greenberg, S. [2003], Task Analysis for Groupware Usability Evaluation: Modelling Shared-workspace Tasks with the Mechanics of Collaboration, *ACM Transactions on Computer–Human Interaction* **10**(4), 281–311.

Riseborough, M. G. [1981], Physiographic Gestures as Decoding Facilitators: Three Experiments Exploring a Neglected Facet of Communication, *Journal of Nonverbal Behavior* **5**(3), 172–83.

Robertson, T. [1997], Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions, *in* J. Hughes, W. Prinz, T. Rodden & K. Schmidt (eds.), *Proceedings of ECSCW'97, the 5th European Conference on Computer-supported Cooperative Work*, Kluwer Academic Publishers, pp.205–20.

Rodden, T. [1996], Populating the Application: A Model of Awareness for Cooperative Applications, *in* M. J. Tauber, B. Nardi & G. C. van der Veer (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground (CHI'96)*, ACM Press, pp.88–96.

Roussel, N. [2001], Exploring New Uses of Video with VideoSpace, *in* M. R. Little & L. Nigay (eds.), *Engineering for Human–Computer Interaction: Proceedings of the 8th IFIP International Conference (EHCI 2001)*, Vol. 2254 of *Lecture Notes in Computer Science*, Springer-Verlag, pp.73–90.

Segal, L. D. [1995], Designing Team Workstations: The Choreography of Teamwork, *in* P. Hancock, J. Flach, J. Caird & K. Vicente (eds.), *Local Applications of the Ecological Approach to Human–Machine Systems*, Vol. 2, Lawrence Erlbaum Associates.

Stotts, D., Smith, J. & Gyllstrom, K. [2004], Support for Distributed Pair Programming in the Transparent Video Facetop, *in* C. Zannier, H. Erdogmus & L. Lindstrom (eds.), *Proceedings of XP Agile Universe 2004*, Vol. 3134 of *Lecture Notes in Computer Science*, Springer, pp.92–104.

Tang, A., Boyle, M. & Greenberg, S. [2005], Understanding and Mitigating Display and Presence Disparity in Mixed Presence Groupware, *Journal of Research and Practice in Information Technology* **37**(2), 71–88.

Tang, J. C. [1991], Findings from Observational Studies of Collaborative Work, *International Journal of Man–Machine Studies* **34**(2), 143–60.

Tang, J. & Minneman, S. [1991a], Videodraw: A Video Interface for Collaborative Drawing, *ACM Transactions on Office Information Systems* **9**(2), 170–84.

Tang, J. & Minneman, S. [1991b], VideoWhiteboard: Video Shadows to Support Remote Collaboration, *in* S. P. Robertson, G. M. Olson & J. S. Olson (eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching through Technology (CHI'91)*, ACM Press, pp.315–22.

Wilson, A. [2005], PlayAnywhere: A Compact Tabletop Computer Vision System, *in* P. Baudisch, M. Czerwinski & D. Olsen (eds.), *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST'05)*, ACM Press, pp.83–92.