H. Gellersen et al. (eds.), ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work, 18-22 September 2005, Paris, France, 1–22.

Ways of the Hands

David Kirk, Andy Crabtree and Tom Rodden
School of Computer Science & IT, University of Nottingham, Nottingham, UK.

Abstract. This paper presents an ethnographic analysis of the nature and role of gestural action in the performance of a remote collaborative physical task. The analysis focuses on the use of a low-tech prototype gesturing system, which projects unmediated gestures to create a mixed reality ecology that promotes awareness in cooperative activity. CSCW researchers have drawn attention to the core problem of the distortion effect along with the subsequent fracturing of interaction between remote ecologies and have emphasized the need to support the 'projectability' of action to resolve this. The mixed ecology resolves the distortion effect by enabling a remote helper to project complex object-focused gestures into the workspace of a local worker. These gestures promote awareness and thus enable helper and worker to coordinate their object-focused actions and interactions. Analysis of the socially organized use of the system derives key questions concerning the construction of mixed ecologies more generally, questions which may in turn be exploited to drive the design of future systems.

Introduction

The theme of 'awareness' has come to play a central role in CSCW research and a range of technical solutions have been developed to explore ways in which computer-based systems and applications might enable people engaged in cooperative activity to 'take heed' of one another and *align* and *integrate* their individual actions. Understanding and promoting awareness through design is of fundamental concern to CSCW then, as awareness is a primary means by which *coordination* 'gets done' (Hughes et al. 1994). Early research in the field was devoted to exploring the potential benefits of audio-video 'media space' technologies to promote awareness amongst remote parties. However, as Schmidt (2002) points out, the expected benefits from these technologies never materialized. The root of the problem was that media spaces 'distorted' participants' natural awareness practices (Heath and Luff 1991), subsequently 'fracturing' interaction. Of particular concern was the effect of media space technologies on the range of human gestures that are central to the face-to-face coordination of collaborative tasks.

> The emergence of a gesture, its progressive movement and graduated appearance within the local environment, its complex operation on the periphery of the visual field, is distorted by the technology. It renders the recipient insensitive to the interactional tasks of the movement … It is interesting to note that in video mediated communication, individuals assume and attempt to preserve the presupposition of a common frame of reference and the interchangeability of standpoints. Indeed their visual conduct is systematically designed with respect to a common mutual environment. Speakers for example, shape their gestures as if the potential recipient will view it in the way in which it is designed. Yet by providing limited access to the other and 'transforming' the appearance of a gesture, the technology introduces an incongruous mutual environment and a curious subsidence into the foundations of socially organized interaction. (Heath and Luff 1991)

Research subsequently diversified from the communication paradigm that underpinned the development of media spaces, and interest in the socially organized properties of space burgeoned in particular. The spatial paradigm sought to promote the continuity of awareness through the development of 'shared spaces' or digital ecologies, situating audio and video alongside other media in collaborative virtual environments (CVEs) where participants were 'immersed'. The spatial paradigm has been explored by a great many researchers and in many different ways, though it has notably being extended through the development of *mixed reality boundaries* (e.g., Koleva et al. 1999). Mixed reality boundaries situate the shared ecology in the physical environment beyond the desktop and make it possible for remote participants to engage in 'collaborative physical tasks' – i.e., object-oriented tasks that take place in the real world but which are, at the same time, mediated through a digital environment. Actually accomplishing collaborative physical tasks via CVEs and in mixed reality settings inevitably brings the communication paradigm back into consideration however (Hindmarsh et al. 2000), which in turn leads us back to the distortion effect and the fracturing of interaction between ecologies that emerges from it (Luff et al. 2003).

Despite well-founded criticism (Dourish et al. 1996), researchers did not abandon the communication paradigm (or the face-to-face organizational aspects of at least) and recent attempts to address the problem of distortion have led to the emergence of *dual ecologies* (Kuzuoka et al. 2004). Dual ecologies are at the present time remote audio-video environments that seek to support collaboration between distributed parties through the further development of communication devices that resonate with participants' natural interactional practices (and therefore, it is hoped, prevent the fracturing of interaction between ecologies). Of key concern is the development of support for the 'projectability of action' between remote ecologies.

> "By 'projectability' we mean the capacity of participants to predict, anticipate, or prefigure the unfolding of action … The discussion of projectability is inseparable from the issue of ecology. A participant's actions are afforded by the properties of the environment. One participant can project the forthcoming actions of another participant by making sense of his/her actions in relation to the surrounding objects and environment … " (Kuzouka et al. 2004)

In this paper we seek to address the distortion effect and extend the field of research by marrying the communication and spatial paradigms together through the development of a *mixed reality ecology* to support the effective projectability of gesture and thereby resolve the distortion effect. Whereas the nature of dual ecologies is to have two separate ecologies (one for each participant, remote and local), with communication between them mediated by some interaction device, a mixed reality ecology consists of a mixed reality surface that is overlaid onto the real world ecology. This allows a remote participant to talk to, monitor and intervene in the task actions of another participant (Kirk and Stanton Fraser 2005). Bodily actions are projected into the physical task space, thereby conjoining both the communicative and spatial elements of two workspaces into one hybrid workspace. Specifically, the mixed reality ecology enables one participant to project her hands into the physical space of the other participant and to coordinate his physical task actions. The mixed reality ecology provides the 'common frame of reference' that is presupposed in and essential to naturally occurring interaction and retains the relational orientation and unmediated nature of naturally occurring gesture. The technology employed to create the mixed reality ecology (derived from low-tech prototyping methods) is well established (consisting of video cameras, a projector, and a TV in this case). What is novel and of importance here is the mixed reality ecology created by the technology, its ability to redress the distortion effect, and its consequences for 3D object-focused interactions. In order to understand the potential of mixed reality ecologies to promote awareness in cooperative physical activities, we conducted a series of lab-based experiments, which were subject to qualitative analysis. Below we describe the technological arrangement of the mixed reality ecology, the nature of the experiments, how they were studied, what we found, and the implications that these findings have for CSCW systems design.

Supporting the Projectability of Gesture

There are currently two main classes of systems support for projectability: *linked* gesture systems and *mediated* gesture systems. Linked systems support collaboration around the construction of shared (often digital) 2D artefacts. Linked systems largely emerged from the efforts of designers to support remote collaboration amongst themselves using video connections. Research into the potential of linked systems has led to the development of a range of video-based technologies. These include VideoDraw, which provides a 'virtual sketchbook' that allows participants to see each other's drawings and projects video of their accompanying hand gestures directly onto one another's drawing space (Tang and Minneman 1990); VideoWhiteboard, which extends VideoDraw by projecting digital shadows of hand and upper body gestures (Tang and Minneman 1991); and ClearBoard, another shared drawing tool that projects video feed of head and gaze movements through the shared surface (Ishii and Kobayashi 1992). More recent developments have extended linked systems to support distributed multi-party interactions with a wider variety of applications. These include the Agora system (Kuzuoka et al. 1999), which projects video of the documents on a local desk and the gestures, gaze, and bodily orientations implicated in using them onto a remote desk and vice versa; and VideoArms (Tang et al. 2004), which captures the arm movements of participants in particular locations and digitally recreates them as 'virtual embodiments' that are available across the shared workspace.

Mediated systems are more diverse in character and seek to exploit computational devices to articulate gesture. Early systems such as Commune (Bly and Minneman 1990) explored the use of digital tablets and styli as a means to articulate gesture around shared drawings. More recent research has exploited 'telepointer traces' – i.e., interaction histories for virtual embodiments - that visualize a participant's recent movements in a shared workspace (Gutwin and Penner 2002). Current research seeks to move beyond articulating gestures in 2D environments to understand how they might be manifest in the real world to support the physical manipulation of 3D objects. This class of mediated system is largely concerned to support remote help giving or instruction. The DOVE system (Ou et al. 2003, Fussell et al. 2004) supports remote interaction between local workers and remote helpers by allowing helpers to overlay pen-based gestures onto a video stream of the worker's task space, the results of which are displayed to the worker on a linked video window adjacent to their task space. Whilst local workers are made implicitly aware of the remote helper's view in the DOVE system, the worker still needs to extrapolate from the illustrated gestures presented on a video feed to their own local ecology. In other words, the local worker needs to 'decode' and realign the gesture in relation to her perspective on the task space; needs to *embed* the gesture in her local ecology in order to articulate its meaning and align and integrate her actions accordingly. Efforts to embed gesture in the worker's local environment are currently being pursued through robotics. The development of GestureMan (Kuzuoka et al. 2004) seeks to move gesture beyond the interface and *situate it in the remote worker's local workspace*, though distortion and fracturing of the two ecologies has proved to be a continuing problem (Luff et al. 2003).

Our own research seeks to explore the use of gesture in collaborative physical tasks. We exploit the direct and unmediated representation of gestures within a linked system to promote mutual awareness between participants located in asymmetric ecologies. We support the accomplishment of collaborative physical tasks through a mixed reality surface that aligns and integrates the ecologies of the local worker and the remote helper. To effectively embed remote gestures in the local ecology we exploit direct video projection (Figure 1a). Specifically, a video camera is used to capture images of the remote helper's hands and the gestures she makes are then projected onto the desk of the worker. The projection overlays the helper's hands on top of the worker's hands (rather than being face-to-face or side-by-side). This provides her with the same orientation to the worker's local ecology and creates a mixed reality surface at the level of the task space. The actions carried out at the mixed reality surface are captured by a second video camera and passed back to a TV monitor situated on the remote helper's desk. This arrangement establishes a common frame of reference. It allows the worker to see the helper's hands and hand movements in his local ecology. It allows the helper to see objects in the worker's local ecology, to see the worker's actions on the objects in the task space, and to see her own gestures towards objects in the task space. In turn, this 'reciprocity of perspectives', where the helper can see what the worker is doing and the worker can see what the helper is doing, provides for mutual awareness and enables the helper and worker to align and integrate their actions through the effective combination of gesture and talk (Figure 1b).

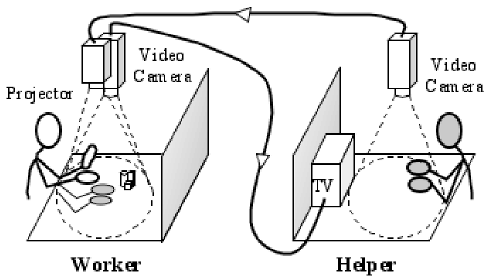


Figure 1a. Projecting gesture into the worker's local ecology.

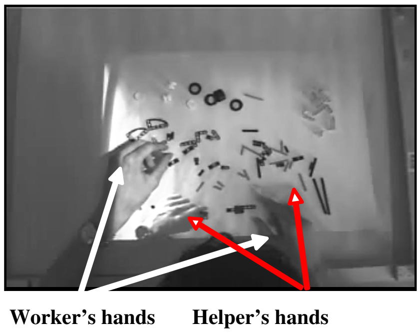


Figure 1b. The mixed reality surface: a reciprocity of perspectives.

Experiments in the Mixed Reality Ecology

In order to establish the *prima facie* efficacy of the mixed reality ecology we carried out a series of lab-based experiments involving 24 pairs of participants (student volunteers from various backgrounds) in the performance of a collaborative physical task, namely, assembling a Lego® kit. Participants were randomly assigned roles, one becoming a remote helper and the other a local worker. The remote helper had the kit's instructions and a video-gesture link to the local worker's task space. The local worker had physical contact with the items to be assembled, but no specific knowledge of how to assemble the pieces in front of him, and relied on the assistance of the remote helper to assemble the Lego model correctly. Both parties were located in the same room but were not able to see each other. The participants could talk to each other however, so no technical configuration was required to transmit audio. Pre-experiment demonstration introduced the participants to the system and they engaged in a simple assembly task to familiarize themselves with the technology. Once participants understood how the system operated they were instructed that they had 10 minutes to construct as much of the model as they could.

Assembling a Lego kit encompasses a variety of *generic task elements*, such as item selection, pattern matching, physical manipulations (rotate, insert, attach) and error checking. In this respect the assembly task offers the opportunity to explore some of the demands that may be placed on real world applications of the system: on the remote guidance of machine and equipment repair or remote artefact and specimen examination, for example. The use of Lego meant that we could model a complex array of interactions and thus inform the development of future applications of the underlying technology. A central feature of interaction is the clear asymmetry between the roles of the participants. Essentially one of the participants is an expert guiding the task or providing expert assistance and support, and the other is a worker who has less knowledge about the artefact or the operations to be performed on the artefact to bring about some desired result (such as assembling a particular model with the Lego pieces or diagnosing a particular machine fault). As the design of Lego is such that the connection of pieces is rather intuitive, little effort is required to learn how to put the pieces together and the worker's attention is instead directed towards artefact manipulation (assembling the model), which requires quite a high level of skill and dexterity. The task expertise of the helper was rapidly generated by giving them a set of clearly designed instructions, which accompany all Lego kits. Of course, all instructions no matter well designed are 'essentially incomplete' (Suchman 1987) and the work of *articulating* just what they mean 'here and now' for just these parties in just this situation is, in many respects, the focus of the experiments. We restrict our account of this articulation work (Schmidt and Bannon 1992) to the nature and role of gesture. It should be said, however, that the use of gesture is thoroughly intertwined with standard conversational mechanisms (Sacks et al. 1974), though space necessarily restricts our treatment of this intertwining.

The standard approach to studying lab-based experiments is essentially quantitative in character and largely concerned with establishing performance parameters and other salient metrics (see Kirk 2004 for these results). As we are concerned to understand the potential of unmediated gesture in mixed reality ecologies to promote awareness in cooperative activity, we elected to complement the standard approach with qualitative study as well. Our approach is motivated by the observation that "an experiment is, after all, just another socially organized context for cognitive performance" (Hollan et al. 2000). Lab-based experiments might be studied ethnographically then (ibid.) to complement standard user modeling approaches with detailed observation and careful analysis of the cooperative work of the experiment. Ethnographic study in turn allows us to explicate the socially organized properties of the experiments, bearing in mind Schmidt's caveat that the social organization of cognition is not be found in some "internal realm" but in the intersubjective practices whereby "actors effortlessly make sense of the actions of coworkers" (Schmidt 2002). The social organization of awareness is to be found, then, in the "practices through which actors align and integrate their distributed but interdependent activities" (ibid.). Those practices are done in bodily ways and through bodily movements. Accordingly, and following Suchman (2000), when examining video recordings of the experiments we pay particular attention to the bodily practices that participants engage in to "mediate interaction" and "highlight objects for perception", thus promoting awareness in cooperative activity.

The Mediating Body in Cooperative Activity

Below we present a series of vignettes that illustrate the cooperative work of the experiments and the bodily practices organizing that work. In bodily detail the vignettes display the range of gestures implicated in the collaborative assembly of the Lego kit. A standard treatment of gesture in a design context is to borrow classification schemes (taxonomies or typologies) from the social sciences to organize findings and inform design (see Bekker et al. 1995 for a classic example). We make no effort to reconcile our findings with existing taxonomies, however, for reasons best articulated by Adam Kendon, a leading figure in the study of gesture:

> "The various typologies of gesture that have been put forward are in part attempts to classify gestures in terms of the information they encode, albeit at very general levels. These typologies are often logically inconsistent, in many cases formed on the basis of rather hasty observation with a good admixture of 'folk' categories thrown in ... gestures that consistently occupy extreme ends of these dimensions (with little weighting on the others) get distinguished as "types" - but I don't think a typological way of thinking is very helpful. Rather, it tends to obscure the complexity and subtlety [of gesture]." (Kendon 1996)

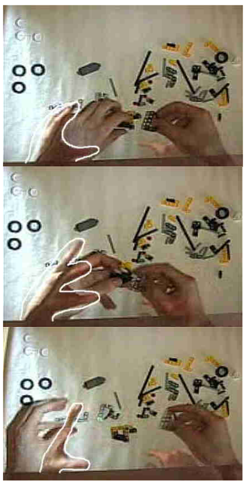
In order to develop a broader understanding of the potential of gesture to promote awareness in cooperative activity we replace a concern to cast our findings in terms of existing taxonomies, with a concern to understand the 'stroke of gestural phrases'. That is, to understand what gestures 'say' and 'do', what the gesture is 'meant for', or, more definitively, what the situational 'business' of the movement is. In addition to this, we wish to understand the 'content' of the stroke. As Kendon puts it,

> "It is often said that gesticulation is idiosyncratic, each speaker improvising his own forms. So far as I know, no one has ever really tested this claim. My own experience in gesture-watching suggests to me that people are far more consistent in what they do gesturally than this 'idiosyncrasy' claim would lead one to imagine … [There are] similarities in the patterning of gestural action and such patterns are socially shared - hence there is conventionalization to a degree affecting all kinds of gesturing." (ibid.)

The content of the stroke refers to the compositional character of gestural phrases - to the emergence of distinct patterns of gestural action. It is with an eye towards carefully describing *particular patterns of gestural phrase* and *the business or work that they do* that we approach analysis of the ways in which the body (and the hands in particular) mediates interaction and highlights objects for perception (Sudnow 1978). In turn, this approach might enable us to identify how gestures promote awareness as an integrated aspect of collaborative practice. Below we present a series of vignettes that articulate the patterns of gestural phrase 'at work' in our experiments and the ways in which they functioned (the helper's hands are highlighted to aid the visibility of her gestures).

'The Flashing Hand'

Before assembly begins the participants must first *align themselves* in the mixed reality ecology such that their movements and gestures might be understood in relation to the arrangement of (the Lego kit in this case) and each other's gestural activities. In other words, the participants must establish to their satisfaction that they share a common frame of reference that permits the reciprocity of perspectives. This is achieved through variants of the 'flashing hand' gesture:



The worker is picking up pieces of the kit and looking to see how they fit together. The helper moves her hand towards the worker's left hand.

As the helper's hand approaches the worker's left hand she says, "Is this your left hand?" The helper then starts to wiggle her fingers.

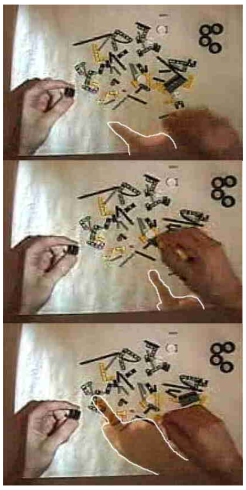
The worker then moves his hand into closer proximity with the helper's, copies the wiggling motion and says "Yeah."

Figure 2. The Flashing Hand Gestural Phrase

The 'flashing hand' gets its name from the wiggling movement of the helper's hand, which brings the helper's hand in and out of alignment with the workers and gives the impression that the worker's hand is flashing. Whilst simply done, it is used to establish the reciprocity of perspectives that is essential to mutual awareness and the coordination of task actions. Although indication of which hand is being referred to could be done by a simple pointing gesture, this form of gesture makes implicit reference between worker and helper to their comparative alignment to the artefacts. The mixed reality ecology enables the helper and the worker to effectively inhabit the same place and it is by this overlaying of hands in similar ways to the vignette above, and in the ways that follow, that the participants maintained reciprocity throughout the experiment.

'The Wavering Hand'

Having established reciprocity of perspectives, the participants begin the assembly task. The most obvious way in which gesture might promote awareness and coordination in cooperative object-focused activity is through an unfolding order of what is taxonomically referred to as deixis - 'pointing' in vernacular terms - at the particular items to be selected for assembly. Our experiments show that coordinating the selection of items for assembly is more subtle and complex than simply pointing, however. Whilst deixis does make up a large part of gesturing behaviour it usually occurs as a component feature of a larger gestural phrase. The 'wavering hand' illustrates the point. In the following vignette the helper is trying to get the worker to pick up a black L-shaped piece of Lego. Having been asked if he has "got an L-shaped piece" the worker scans the items in front of him and picks one up, but it is yellow (and therefore the wrong item). The helper responds as follows:



The helper reaches forward with his hand as he starts to look himself for the black L-shaped piece.

The helper's hand then wavers over the work surface, mirroring his visual scan over the pool of possible items.

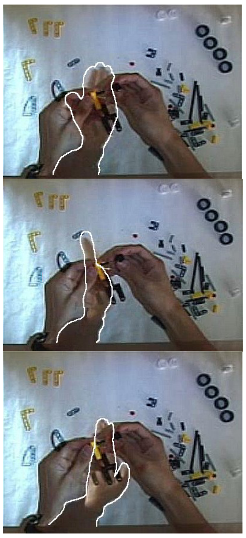
This lateral movement of the hand is followed by a final and decisive pointing movement over the required item, which is accompanied by the helper saying, "One of those I think."

Figure 3. The 'Wavering Hand' Gestural Phrase

Combined with talk, the 'wavering hand' makes the worker aware of the helper's search and location of the required piece, and in turn enables the worker to select the correct item. The ability to support cooperative activity through unmediated gesture reduces the amount of time required to secure a common orientation to material tasks of work. Systems that do not support deictic gesturing require that a great deal more verbal instruction be issued to achieve mutual awareness and the integration of tasks. Even though current systems have been built to support deictic gestures, mechanisms of projection still require a great deal of articulation work either to embed the gesture or to understand the specific meaning of a gesture. However, with unmediated gesture there is, at the same time, both an interactional *richness* and an *economy* that facilitates awareness and coordination on a moment-by-moment basis. Thus, and for example, as the 'wavering hand' moves from side-to-side it mirrors the visual scanning of the helper suggesting that the he is 'looking for the piece too' and promotes awareness of the search for the item. The local worker is made aware that the piece has been located when the 'pointing finger of the wavering hand' and the helper's utterance "One of those I think" together highlight its presence at a specific place in the worker's ecology. Without the economy and richness of movement that the unmediated representation of gesture affords, such use of demonstrative pronouns and deictic expressions would not be possible. Those affordances are provided by the mixed reality ecology, which aligns both the location and the representation of the remote helpers' gestures. In turn, this means that the local worker does not need to reconcile gestures dislocated from the actual task space (such as those presented on a separate screen, for example) or interpret the meaning of artificial representations of gesture embedded in the task space.

'The Mimicking Hand' (with One or Two Hands)

As the experiments unfolded it became apparent that different gestural patterns were implicated in the accomplishment of the different activities that make up the overall assembly task. As demonstrated by Fussell et al. (2004) those gestures that go beyond mere deictic reference are often the most important in terms of facilitating task performance. Whilst the 'wavering hands' make the worker aware of *just what pieces are to be selected* and coordinate selection, the 'mimicking hands' gesture is one of a range of gestures that are concerned with *ordering the assembly* of selected pieces. The following vignettes illustrate the role of the 'mimicking hands' gesture, with one and two hands respectively, in the ordering of assembly. In the first vignette, the worker has picked up what the helper has called the "main construction type bit":



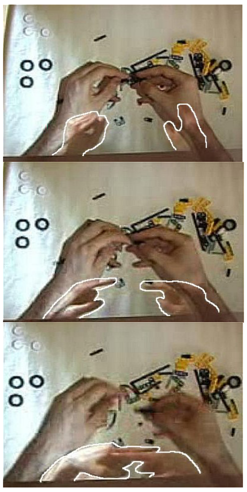
The helper then prompts the worker to rotate the piece prior to attachment. Her flat hand indicates the piece's current orientation.

The gesture unfolds as the flat hand is rotated to its side and the helper says, "If you flip it a hundred and eighty degrees like that."

The gesture is completed as the helper rotates her hand 180^{o}, and is then repeated for effect.

Figure 4a. The 'Mimicking Hands' Gestural Phrase (with one hand)

Here the 'mimicking hand' enables the helper to make the worker aware of the relative orientation of the Lego kit (what way up it should be, what way pieces should face, etc.) In the second vignette, the worker exploits hand gestures to show how the pieces should be manipulated and fitted together.



The helper places her hands at the edge of the table watching the worker assemble two pieces. The worker moves the pieces around as if unsure of how they connect together.

The helper says, "So they lie next to each other", extending her fingers to mimic the primary axis of the pieces.

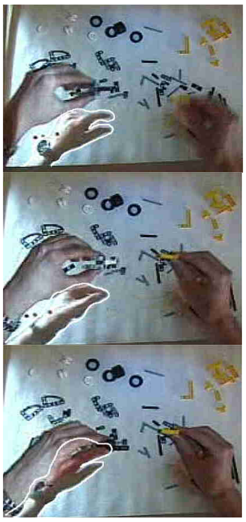
The gesture comes to a close as the helper indicates the direction of the movement required to fit the pieces together by docking her hands and saying, "Like that".

Figure 4b. The 'Mimicking Hand' Gestural Phrase (with two hands)

The 'mimicking hands' make the worker aware of the ways in which pieces fit together. This requires the arrangement of subtle and complex movements of the hands and fingers to articulate the particular way in which particular pieces should be oriented in 3D space, the spatial relationships that hold between pieces, and the ways in which they should be manipulated and moved so that they fit together. In the above vignette, for example, we can see that positioning of the helper's hands enables the worker to see the proper relation of the two pieces. This in turn enables the worker to orient the two pieces correctly and the 'docking hands' shows how they should fit together given that orientation. While unsophisticated, the technological arrangement at work here nevertheless allowed helpers to use their hands in intuitive ways to articulate the complexities of assembly.

'The Inhabited Hand'

Of course, ordering the assembly of a complex 3D object did not always run smoothly. Practical difficulties of orientation frequently occurred and workers could not always understand just how pieces were meant to fit together. To remedy this the helper would perform the 'inhabited hand' gesture. In this vignette the helper seeks to clarify instructions and to help the worker move a piece he has been struggling with into the right orientation and in the right direction:



The helper places her hand on top of the worker's, forms it into the same shape and says, "If you rotate."

The helper then rolls her hand forwards. Saying "Rotate your hand like that, yeah."

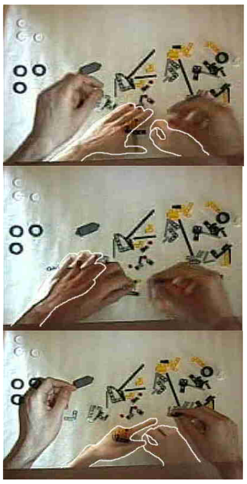
The helper then brings her hand back to its original position before repeating the gesture .

Figure 5. The 'Inhabited Hand' Gestural Phrase

The 'inhabited hand' makes the worker aware of the fine-grained movements that need to be done to align pieces and make them fit together. This is achieved by placing the hand in the same position as the worker's and making the same shape of the hand, a specific movement that indexes the verbal instruction to it. Through this movement the helper models how the worker is to hold the piece and shows the desired angle of rotation of the hand, making the worker aware of just how he needs to manipulate the piece to assemble it. It is not simply a case of showing the worker how the piece should be rotated however, which can be achieved by showing a representation of the piece in initial and final states, but is a literal instruction on the actions required to achieve the final state (which in this instance is to hold the piece in the left hand just “like that” so that it can be easily inserted into the piece held in right hand). The helper thus demonstrates just what is to be done with the hand to obtain the correct orientation of the piece and make it fit with its partner. Here we can see that the mixed reality ecology enables a level of interaction not easily achieved via other means, effectively allowing the helper to embody the hands of the worker to synchronize the task to hand.

‘The Negating Hand’

Other practical difficulties populated the assembly of the Lego kit, particularly the selection of wrong pieces. Such mistakes were highlighted and corrected through the ‘negating hand’ gesture. In the following vignette the remote helper has instructed the worker to put two particular pieces together. The worker goes to pick up the wrong piece, however:



The helper lays her hand flat on the desk over the wrong piece and says, “Forget about this.”

The helper moves her hand in a sweeping movement, emphasizing which piece is to be ignored.

The helper then points at the correct piece, which is now in the worker’s right hand and says “Just this piece.”

Figure 6. The ‘Negating Hand Cover’ Gestural Phrase

The ‘negating hand’ gesture makes the worker aware of his mistake and highlights the correct piece for assembly by combining covering, sweeping, and pointing movements of the hands and fingers. Effectively, the gesture says ‘not that, but this’. Although rapidly accomplished such gestures are complex and while laser dots, drawn lines, or virtual embodiments may be used to refer to and highlight particular objects in a shared ecology, fluid interaction and the ability of the recipient to make sense the situational relevance of the gesture are dependent upon the alignment of both the gestural representation *and* its spatial position within the ecology. The advantage of using gestures projected into the task space is that it allows the ‘spatial reference’ of a gesture to be held intact, as gestures are presented relative to their objects of work, readily enabling workers to see and repair their mistakes. The use of unmediated representation also allows gestures to retain their natural temporal characteristics, being both rapid and fluid and reconstituted on an *ad hoc* basis and avoiding excessive temporal residue such as the cluttered screen that results from a series of sketch-based gestures (though this may be a double-edged sword to some extent, as there may be some benefit to be gained from a degree of permanence on certain occasions).

‘Parked Hands’

It will come as no surprise to say that assembly activities were oriented around turn-taking, especially as one party had the assembly instructions and the other was obliged to follow them. In addition to employing standard conversational mechanisms (Sacks et al. 1974) the participants developed a particular gestural pattern to signal and accomplish the taking of turns. We refer to this pattern as the ‘parked hands’ gesture and it is illustrated in the following vignette. Through employing the gestures described above the helper has managed to guide the worker through the assembly of a particular section of the Lego kit and the task now is to assemble its partner:

The helper points out a piece and says,” Assemble that exactly the same as the other one.”

The helper then withdraws his hands and parks them at the edge of the ecology.

The worker assembles the section and the helper says, “Yeah, okay, and then put that on here.” The helper points to a specific location and then parks her hands again.

Figure 7. The Parked Hands Gestural Phrase

The ‘parked hands’ gesture indicates that a turn has been completed and that it is now the turn of the worker to undertake the instructions delivered. Moving the hands out of the parked position indicates that helper is about to take another turn, issuing new instructions accompanied by appropriate gestures. This simple but elegant gesture makes the worker aware when a turn is about to be taken and when it has been completed and enables the worker to coordinate his actions with the helper’s instructions.

Promoting Awareness through Gesture

The gestural phrases that we have identified in our study elaborate a corpus of patterns of gestural action that integrate awareness in practice in the mixed reality ecology (Table 1). They do not represent general taxonomic elements and their coincidence with, divergence from or elaboration of existing types is irrelevant. Emphasis is instead placed on the ability of gesture to promote awareness and (thus) on the situated 'business' or function of particular gestural patterns in action. As Schmidt (2002) points out,

> Awareness is an attribute of action. Doing one thing while taking heed of other relevant occurrences are not two parallel lines of action but a specific way of pursuing a line of action, namely to do it heedfully, competently, mindfully, accountably. In a CSCW context 'awareness' does not refer to some special category of mental state existing independently of action but to a person's being or becoming aware of something.

The patterns of gestural phrase we have identified make it visible how participants promote awareness of the tasks they need to accomplish and come to integrate and align their activities using non-verbal behaviours as well as speech. Each gestural phrase provides a way for the mediating body to highlight objects for perception and to make what Crabtree et al. (2004) describe as "a host of fine-grained grammatical distinctions". These marry utterances (such as verbal instructions) to specific actions (such as the selection and orientation of pieces and the manipulations required to fit them together, etc.), which in turn provides for the coordination of tasks.

Gestural Phrase	**Business of Phrase**
Flashing Hand	Establish reciprocity of perspectives
Wavering Hands	Indicates search for and location of items and coordinates selection of correct pieces.
Mimicking Hands	Orders assembly of pieces by indicating how pieces should be oriented for assembly and how pieces should be joined together
Inhabited Hand	Shows fine-grained movements that need to done to align pieces and make them fit together.
Negating Hand	Repairs mistakes and clarifies instructions
Parked Hands	Orders turn-taking

Table 1. Corpus of gestures that promote awareness in the mixed reality ecology

The corpus indicates that a rich grammar of gestural action is implicated in the organization of interaction around shared artefacts. This grammar enables participants to 'project' awareness of the tasks to hand and to integrate their actions accordingly. Its effective expression is afforded by the mixed reality ecology, which aligns the participants' distinct ecologies and their gestural phrases. The alignment of ecologies and expression of a rich grammar of gestural action that it affords *resolves* the distortion effect and fracturing of interaction between ecologies.

While existing systems support the expression of gesture, they still effect a separation between the ecologies of the participants. The GestureMan system (Luff et al. 2003), for example, fractures interaction as it fails to comprehensively reflect the remote helper's actions. Whilst extensions to the system have sought to address the problem of projectability of action, the mediated expression of gesture still does "not operate ideally" (Kuzouka et al. 2004). This is, in part, attributed to the temporal character of expression – to the timing of gestures and their placing within collaborative interaction. However, it is also recognized that representing the "movement, shape and motion" of gesture is also critical to the enterprise (ibid.). The DOVE system (Ou et al. 2003, Fussell et al. 2004) seeks to overcome these problems by overlaying sketches onto the worker's ecology. However, these gestures are removed from the working surface and their relation to task artefacts must be extrapolated by the worker if they are to gain practical purchase, again fracturing interaction.

The effort to align separate ecologies in ways that resolve distortion and the fracturing of interaction, revolves around the reconciliation of space, time and motion in direct relation to a shared object of work. The spatial, temporal and motional coherence of projected action is maintained in the mixed ecology as the ecology 'forces' worker and helper to adopt the same orientation. From this position they see things from the same perspective and see what the other sees *from same point of view that the other sees it*. By 'forcing' orientation, the spatial, temporal and motional coherence of projected actions is preserved then. Furthermore, the richness and economy of projected action – of unmediated gesture - means that participants do not have to 'decode' abstract representations of gesture, but can effortlessly make sense of the actions of coworkers. Embedding the remote helper's gestures into the local worker's task space in direct relation to his local orientation unifies separate ecologies and promotes awareness between the two by enabling a rich texture of grammatical phrases to be coherently articulated and expressed. It is this 'phenomenal coherence' of gesture – i.e., the *situational relevance and intersubjective intelligibility* of gesture that the mixed ecology promotes. Phenomenal coherence prevents distortion and the fracturing of interaction between separate ecologies and it is the ability of technology to support phenomenal coherence that represents a long-tern challenge for design.

Designing Mixed Ecological Arrangements

In this paper we have introduced the core notion of a mixed reality ecology and explored the ways in which it might be exploited *by users* to support interaction between a remote helper providing advice and guidance to a local worker. The low-tech prototype we have constructed to explore the potential of mixed reality ecologies exploits the direct unmediated projection of gestures to support phenomenal coherence. It provides a common frame of reference that promotes awareness between remote ecologies and enables participants to align and integrate their collaborative activities. Our analysis has shown that the mixed reality ecology provides an expressive medium allowing participants to exploit a subtle and complex range of naturally occurring awareness practices, which we have articulated in terms of a corpus of 'gestural phrases'. But what relevance do our experiences have for design more generally and to the development of technological arrangements supporting *remote users engaged in cooperative work on physical objects* in particular?

First and foremost our experiences suggest that there is a need for a shift in design orientation, particularly in the ways in which we consider the use of technology to support remote collaboration on physical tasks. The primary orientation to design at the current time places technology in the role of linking two distinct ecologies and essentially focuses on repairing the discontinuity or 'fracture' between them. In contrast, we think it is necessary to design for phenomenal coherence from the outset and see the role of the technology as one that is concerned to develop a shared environment that blends and mixes interaction between ecologies, thereby enabling participants to construct a 'world known in common' - a 'world' that is intersubjectively intelligible to participants and which provides for the situational relevance of gesture. Essentially it can be argued that current approaches to design support only one of the two key features of phenomenal coherence. The class of systems represented by DOVE, for example, support intersubjective intelligibility but not situational relevance (gestures must be made relevant by the local worker to the task). And on the other hand, the class of systems represented by GestureMan support situational relevance but not intersubjective intelligibility (that the robot is pointing at 'this' or 'that' is clear but what the pointing means in the absence of other subtle gestural cues is not so clear).

The design of mixed ecologies requires us to think carefully about key features of interaction that require support. Reflecting on our analysis of the mixed reality ecology, we would suggest a number of key questions become critical to the design of mixed ecological arrangements more generally:

- *How are participants' gestures placed within a mixed ecology?* Our arrangement directly projects gestures, overlaying them onto the shared task space to create a common frame of reference. Three key elements are central to this achievement.
  - *Aligning the orientation of participants* so that their gestural actions are interpretable in terms of a common orientation to the object of collaboration. This entails projecting the remote helper's gestures so that they share the same bodily relation to a physical object as that of the local worker's Thus, the remote helper's view of the shared ecology is identical to the local worker's.
  - *Aligning the effects of remote gestures and local actions* so that they are understood within the particular context of the activity taking place. Projecting remote gestures into the local

helper's task space and in direct relation to the physical object being manipulated situates interaction at a hybrid surface that not only aligns the participants orientation but marries remote and local actions together to create a shared, mutually intelligible context for collaborative action.

– *Projecting gestures to coherently reflect arrangements of collaborative activity* so that the display of gestures supports the arrangements of cooperation underpinning the task. In the current case our display strongly reflects the asymmetric nature of the task at hand in that a remote helper is giving instruction to a local worker. The local worker sees the remote gesture directly projected on the surface in front of him while the remote helper sees the remote surface through a monitor, thus providing for the mutual intelligibility of gesture between participants.

- More broadly we think it important for design to consider *how participant's actions and gestures are captured as part of a mixed ecology*. We would stress that the capturing of gesturing should be as lightweight and distract as little as possible from the activities to be supported by the mixed ecology. In our case we have used simple video capture and transmission rather than having the remote helper explicitly manipulate a secondary gesture interface. This arrangement meant that participants did not have to manage gestures separately and that their gestures are situated within the cooperative activities to hand. This blending of gesture and action helps promote a mutual sense of a shared ecology of action.
- We also think it important to consider *how gestures are represented and commonly understood in a mixed ecology*. In particular are the representations of participant's gestures readily intelligible within a mixed ecology or do they require a significant reinterpretation? A number of technologies have been used to represent remote gestures, the most notable of these being video sketching and remote pointing technologies. Our use of projected hands reduces the cost of interpreting remote gestures and provides for the situational relevance and intersubjective intelligibility of gestural actions in our mixed ecology.

Approaching the design of remote gesture systems by first tackling these key questions will allow designers to construct more effective mixed ecologies. By designing with the view to making disparate ecologies as mixed and overlapped as possible the purpose of the technology will no longer be to repair fractured interactions. It will instead be to support strong intersubjective understanding and awareness of remote collaborator's joint activities in a shared environment.

Conclusion

The mediation of awareness in remote interactions is of central concern to CSCW and is particularly germane when considering object-focused interactions (which have been plagued by the inability to adequately support intersubjective awareness in collaborative activities on 3D artefacts). Early efforts explored the communicative potential of 'media spaces', exploiting audio and video technologies to promote awareness amongst geographically dispersed parties located in remote locations or ecologies of work. Use of these technologies highlighted the distortion effect, where the salience of users' natural awareness practices - and of their gestures in particular - was obscured by the technology, thus undermining the foundations of socially organized interaction and resulting in the fracturing of interaction between ecologies.

Recent efforts have suggested that the distortion effect may be remedied by developing support for the 'projectability' of action and of gesture in particular. Technologies that have sought to do this have been designed with aim of reducing the fracture brought about by linking two distinct ecologies together. For different reasons fractured interaction is still a major issue to be contended with in both the DOVE system and the GestureMan systems, however. Our approach has been to utilise low-tech prototyping to explore how a system can be designed from the perspective of creating a mixed ecology rather than attempting to repair a fracture. Using video projection we have created a mixed reality surface at the level of the task space, allowing a remote helper to project their gestures directly into a local worker's environment. The 'forced' orientation and unmediated projection of gesture in the mixed ecology enables users to exploit natural awareness practices and align and integrate their object-focused actions and interactions.

The ability of a mixed reality ecology to effectively promote awareness was explored through lab-based experimentation that involved participants in a relatively complex collaborative physical task where there was an asymmetry of worker roles and functions. Our ethnographic approach to the analysis sought to explicate the ways in which the body mediated interaction and highlighted objects for perception. The pattern and function of gestural action was discerned, revealing a corpus of 'gestural phrases' that were integral to interaction and which promoted collaborative awareness and the coordination of tasks. Each gestural phrase performed a different function, enabling the participants to establish and maintain the reciprocity of perspectives that is essential to interaction, to select pieces for assembly, to orient pieces for assembly, to manipulate pieces and fit them together, to repair mistakes, and to signal turn taking and turn completion. In this the mixed reality ecology successfully conveys and preserves gesture's spatial, temporal and motional or 'phenomenal' coherence with reference to shared objects of work. Its ability to do this is dependent on the technological arrangement used to facilitate interaction. We have mooted several key questions which address the technological character of mixed ecology systems and which might drive the development of future systems supporting remote collaboration. Such systems might unify fractured ecologies, providing support for seamless object-focused interaction between remote parties and for the natural awareness practices that secure real world, real time collaboration.

Acknowledgement

The research on which this article is based was funded by UK Engineering and Physical Sciences Research Council 'Equator Interdisciplinary Research Collaboration' (EPSRC GR/N15986/01), www.equator.ac.uk

References

Bekker, M., Olson, J. and Olson, G. (1995) "Analysis of gestures in face-to-face design teams", *Proceedings of DIS '95*, p. 157-166, Michigan: ACM Press.

Bly, S.A. and Minneman, S.L. (1990) "Commune: a shared drawing surface", *Proceedings of Office Information Systems1990*, p. 184-192, Cambridge, Massachusetts: ACM Press.

Crabtree, A., Rodden, T. and Mariani, J. (2004) "Collaborating around collections", *Proceedings of the CSCW '04*, pp. 396-405, Chicago: ACM Press.

Dourish, P. Adler, A., Bellotti, V. and Henderson, A. (1996) "Your place or mine?", *Computer Supported Cooperative work: The Journal of Collaborative Computing*, vol. 5 (1), p. 33-62.

Fussell, S. et al. (2004) "Gestures over video streams to support remote collaboration on physical tasks, *Human-Computer Interaction*, vol. 19 (4), pp. 273-309.

Gutwin, C. and Penner, R. (2002) "Visual information and collaboration", *Proceedings of CSCW '02*, p. 49-57, New Orleans: ACM Press.

Heath, C. and Luff, P. (1991) "Disembodied conduct: communication through video in multi-media office environment", *Proceedings of CHI '91*, p. 99-103, New Orleans: ACM Press.

Heath, C. and Luff, P. (1992) "Media space and communicative asymmetries", *Human-Computer Interaction*, vol. 7, p. 315-346.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S. and Greenhalgh, C. (2000) "Object-focused interaction in collaborative virtual environments", *ACM ToCHI*, vol. 7 (4), p. 477-509.

Hollan, J., Hutchins, E. and Kirsh, D. (2000) "Distributed cognition: toward a new foundation for human-computer interaction research", *ACM ToCHI*, vol. 7 (2), p. 174-196.

Hughes, J. et al. (1994) "Perspectives on the social organization of work", *Field Studies and CSCW* (COMIC Deliverable 2.2), p. 129-160, Lancaster University.

Ishii, H. and Kobayashi, M. (1992) "Clearboard", *Proceedings of CHI '92*, p. 525-535, Monterey: ACM Press.

Kendon, A. (1996) "An agenda for gesture studies", *Semiotic Review of Books*, vol. 7 (3), p. 8-12.

Kirk, D.S. (2004) "Performance effects of using a mixed reality surface for collaborative physical tasks", *Equator Technical Report*, University of Nottingham.

Kirk, D.S. and Stanton-Fraser, D. (2005) "The effects of remote gesturing on distance instruction", to appear in *Proceedings of CSCL 2005,* May 30-June 4, Taipei: ISLS.

Koleva, B., Benford, S. and Greenhalgh, C. (1999) "The properties of mixed reality boundaries", *Proceedings of ECSCW '99*, p. 119-137, Copenhagen: Kluwer Academic Publishers.

Kuzuoka, H. et al. (1999) "Agora: a remote collaboration system that enables mutual monitoring", *Proceedings of CHI '99*, p. 190-191, Pittsburg: ACM Press.

Kuzouka, H., Kosaka, J., Yamazaki, K., Suga, S., Yamazaki, A., Luff, P. and Heath, C. (2004) "Mediating dual ecologies", *Proceedings of CSCW '04*, p. 477-486, Chicago: ACM Press.

Luff, P. et al. (2003) "Fractured ecologies", *Human-Computer Interaction*, vol. 18 (1), p. 51–84.

Ou, J. et al. (2003) "Gestural communication over video stream: supporting multimodal interaction", *Proceedings of ICMI '03*, p. 242-249, Vancouver: ACM Press.

Sacks, H., Schegloff, E., and Jefferson, G. (1974) "A simplest systematics for the organization of turn-taking in conversation", *Language*, vol. 50, p. 696-735.

Schmidt, K. and Bannon, L. (1992) "Taking CSCW seriously: supporting articulation work", *Computer Supported Cooperative Work: An International Journal*, vol. 1 (1), pp. 7-40.

Schmidt, K. (2002) "The problem with 'awareness'", Computer Supported Cooperative work: The Journal of Collaborative Computing, vol. 11 (3), p. 285-298.

Suchman, L. (1987) *Plans and Situated Actions: The Problem of Human-Machine Communication*, Cambridge: Cambridge University Press.

Suchman, L. (2000) "Embodied practices of engineering work", *Mind, Culture & Activity*, vol. 7 (1), p. 4-18.

Sudnow, D. (1978) *Ways of the Hand: The Organization of Improvised Conduct*, London: Routledge.

Tang, J.C. and Minneman, S.L. (1990) "VideoDraw:", *Proceedings of CHI '90*, p. 313-320, Seattle: ACM Press.

Tang, J.C. and Minneman, S.L. (1991) "VideoWhiteboard: video shadows to support remote collaboration", *Proceedings of CHI '91*, p. 315-322, New Orleans: ACM Press.

Tang, A. et al. (2004) "Embodiments and VideoArms in mixed presence groupware", *Technical Report 2004-741-06*, Department of Computer Science, University of Calgary.

H. Gellersen et al. (eds.), ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work, 18-22 September 2005, Paris, France, 23–43.

A Design Theme for Tangible Interaction: Embodied Facilitation

Eva Hornecker
Interact Lab, Dept. of Informatics, University of Sussex, GB
eva@ehornecker.de

Abstract. This paper presents parts of a design framework for collaboratively used tangible interaction systems, focusing on the theme of *Embodied Facilitation*. Systems can be interpreted as spaces/structures to act and move in, facilitating some movements and hindering others. Thus they shape the ways we collaborate, induce collaboration or make us refrain from it. Tangible interaction systems provide virtual and physical structure - they truly embody facilitation. Three concepts further refine the theme: Embodied Constraints, Multiple Access Points and Tailored Representations. These are broken down into design guidelines and each illustrated with examples.

Introduction

Tangible User Interfaces (TUIs) have become a hot topic in HCI. Until recently, research was mostly technology-driven, focusing on developing new systems. A special issue of 'Personal & Ubiquitous Computing' on 'tangible interfaces in perspective' (Holmquist, Schmidt and Ullmer, 2004) marks a change in focus towards conceptual analysis. Yet, there is still a lack of theory on *why* tangible interaction works so well (Dourish, 2001). Cooperation support might be the most important, domain-independent feature of TUIs, but this issue has attracted even less explicit attention. Many researchers agree that TUIs are especially suited for collocated collaboration and build systems aimed at group scenarios (e.g. Stanton et al, 2001; Ullmer and Ishii 2001). Nevertheless, conceptual papers (as in the mentioned special issue) tend to brush over this issue by briefly mentioning visibility of actions and distributed loci of control as collaborative affordances. User studies focusing on group interaction are still scarce, even though we know