**Access from the University of Nottingham repository:**
http://eprints.nottingham.ac.uk/10292/1/DSK-PhDThesisComplete.pdf

Turn It <u>This</u> Way: Remote Gesturing in Video-Mediated
Communication

David Stanley Kirk, BSc (Hons), MSc

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

December 2006

# Abstract

Collaborative physical tasks are working tasks characterised by workers 'in-the-field' who manipulate task artefacts under the guidance of a remote expert. Examples of such interactions include paramedics requiring field-surgery consults from hospital surgeons, soldiers requiring support from distant bomb-disposal experts, technicians inspecting and repairing machinery under the guidance of a chief engineer or scientists examining artefacts with distributed colleagues. This thesis considers the design of technology to support such forms of distributed working. Early research in video-mediated communication (VMC) which sought to support such interactions presumed video links between remote spaces would improve collaboration. The results of these studies however, demonstrated that in such tasks audio-video links alone were unlikely to improve performance beyond that achievable by simpler audio-only links. In explanation of these observations a reading of studies of situated collaborative working practices suggests that to support distributed object-focussed interactions it is beneficial to not only provide visual access to remote spaces but also to present within the task-space the gestural actions of remote collaborators. Remote Gestural Simulacra are advanced video-mediated communication tools that enable remote collaborators to both see and observably point at and gesture around and towards shared task artefacts located at another site. Technologies developed to support such activities have been critiqued; their design often fractures the interaction between the collaborating parties, restricting access to aspects of communication which are commonly used in co-present situations to coordinate interaction and ground understanding.

This thesis specifically explores the design of remote gesture tools, seeking to understand how remote representations of gesture can be used during collaborative physical tasks. In a series of lab-based studies, the utility of remote gesturing is investigated, both qualitatively, examining its collaborative function and quantitatively exploring its impact on both facets of task performance and collaborative language. The thesis also discusses how the configuration of remote gesture tools impacts on their usability, empirically comparing various gesture tool designs. The thesis constructs and examines an argument that remote gesture tools should be designed from a 'mixed ecologies' perspective (theoretically alleviating the problems engendered by 'fractured ecologies') in which collaborating partners are given access to the most salient and relevant features of communicative action that are utilised in face-to-face interaction, namely mutual and reciprocal awareness of commonly understood object-focussed actions (hand-based gestures) and mutual and reciprocal awareness of task-space perspectives. The thesis demonstrates experimental support for this position and concludes by presenting discussion of how the findings generated from the thesis research can be used to guide the design of future iterations of remote gesture tools, and presents directions for areas of further research.

## Published Works

At the time of submission, several sections of work from this thesis have previously appeared (or are scheduled to appear) in peer-reviewed publications as long papers. In the following list the full references for these publications are given.

- Kirk, D. S., Rodden, T. & Stanton Fraser, D. (2007) Turn It <u>This</u> Way: Grounding Collaborative Action with Remote Gestures. In *Proceedings of CHI Conference on Human Factors in Computing Systems,* 28th April-3rd May, ACM: San Jose, CA, pp. 1039-1048

- Kirk, D. S., & Stanton Fraser, D. (2006) Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. In *Proceedings of CHI Conference on Human Factors in Computing Systems,* 22-27 April, ACM: Montreal, Canada, pp. 1191-1200

- Kirk, D. S., Crabtree, A., & Rodden, T. (2005) Ways of the Hand. In *Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work (ECSCW)*, 18-22 September 2005, Springer: Paris, France, pp. 1-21

- Kirk, D. S., & Stanton Fraser, D. (2005) The Impact of Remote Gesturing on Distance Instruction. [1] In *Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL) 2005*, LEA: Taipei, Taiwan, pp. 301-310

The research results of this thesis are presented through chapters 4 – 7. The experimental comparison of remote gesturing versus non-gesturing communication (chapter 4, section 4.2) is highlighted in Kirk, Rodden and Stanton Fraser (2007). The analysis of the learning effects of remote gesture system use (chapter 4, section 4.3) is discussed in Kirk and Stanton Fraser (2005). The analytical comparison of differing gesture formats and locations in gesture systems (chapter 5, section 5.3) is presented in Kirk and Stanton Fraser (2006). The role of hand gestures in remote collaboration (chapter 6) is detailed and discussed in Kirk, Crabtree and Rodden (2005) and the impact of remote gesturing on collaborative language (chapter 7) forms the main focus of Kirk, Rodden and Stanton Fraser (2007).

---

[1] Nominated for Best Student Paper

# Acknowledgements

There are several people I would like to acknowledge who have helped me in various ways whilst I completed this thesis. First off I would like to thank the Mixed Reality Lab for hosting me for three and a half years and for giving me the space to work when I needed to take over chunks of the lab to run experiments. I would also very much like to thank the Equator IRC for giving me the funding to pay for these experiments, to do the research generally and to fly around the world presenting the results.

Academically, I would like to thank Dr. Mike Fraser, for pointing me in the direction of some useful literature in the very early days, and for uknowingly letting me use his thesis for three years as a general guide on 'how to write a phd'. I'd also like to thank Dr. Andy Crabtree for broadening my horizons by introducing me to different perspectives on my work and also for introducing me to stellar terminology such as 'phenomenal coherence'.

There are however, two academics to whom I owe particular thanks. Firstly, I wish to thank Dr. Danaë Stanton Fraser for being my first supervisor and getting me started in my research, and for keeping up an interest in me and my work even after the move to Bath, the new projects, the promotions and the new family. She has been a very supportive supervisor.

Secondly, I wish to thank my principal supervisor Prof. Tom Rodden, truly the smartest man I know, who has helped me in inumerable ways. He has taught me what it means to be an academic and a researcher, gifted me with a strong set of research ethics and without his advice, encouragement, discussion of critical concepts and timely 'critique' of the work, this thesis would not be what it is now.

Personally, I wish to thank my family for always being there for me, in particular my Mum and Dad, who may not have always understood what I was doing but were always keen to see me do well none-the-less, in part this thesis is for them.

Lastly and most importantly I wish to thank Lizzy, my wife-to-be, for putting up with me for soo long. She knew far better than me when I should actually be doing some work, and perhaps more importantly at times, when I should actually stop. Without her encouragement and making me plan I would probably have gone mad and simply never have managed to finish this thesis. To her I owe everything.

# Contents

# Detailed Contents

## Chapter 1 – Introduction

## Chapter 2 – Literature Review

## Chapter 3 – Research Methodology and Disposition

# Chapter 4 – Some Effects of Remote Gesturing

# Chapter 5 – How Best to Construct Remote Gestures

## Chapter 6 – The Communicative Functions of Gesturing

## Chapter 7 – How Gesture Interacts with Language

## List of Figures

## List of Tables

There was speech in their dumbness, language in their very gesture.
*The Winter's Tale (First Gentleman at V, ii)*
*Shakespeare*

# Chapter 1 – Introduction

## 1.1 Introduction

The pervasive nature of information and communications technology means that we are living in an increasingly networked world. Consequently the sphere of influence of any individual is increasing exponentially, at any given time a person can be present in some form at multiple global locations and advances in telecommunications technologies allow that sense of presence to be felt in richer and more diverse ways. A worker's regular environment, their 'working ecology' or 'Activity Landscape' (as Kirsh, 2001, frames it) is now likely to include telecommunication and computing devices that will link disparate people, spaces and resources to support the proliferation of knowledge and expertise within global enterprises. A necessity for common current working practices (Hinds and Kiesler, 2003).

That the development of communication devices should be towards making remote interactions richer, increasing a sense of remote presence, sits well with an understanding of human communication from an information theory perspective. Referring to 'An Ecology of Communication' the information theorist Abraham Moles (1920-1992) originally defined communication as:

> "The action of making an organism or system located at given point R partake in the experiences (Erfahrungen) and stimuli of the environment of another individual or system located in another place and time, by using the items of knowledge they have in common." (Moles, 1975, p. 49)

He also stated that:

> "To transmit a message is to make more complex the space-time surrounding the point of reception; it is to produce a micro-replica of the complexity created at the origin of transmission." (Moles, 1966, p.196-197)

Such a view of communication is supported by more recent work which has explored the situated nature of communicative behaviours in co-located interactions (Hutchins, 1995, Robertson, 1999). This body of work has clearly demonstrated the importance to shared activity of a whole host of contextually embedded physical representations of non-verbal behaviours and artefact manipulations used in conjunction with speech. These actions can embody and imply a plethora of system state properties and communicative intentions, forming an integral part of the collaborative development of task-focussed situational awareness, and becoming crucial for smooth interaction and common understanding.

The intuitive belief that visual access to others was important for helping to understand them was perhaps then the driving force behind the development of Video-Mediated Communication technologies. The benefits of these technologies, which are increasingly

becoming part of our 'Activity Landscapes', have however been demonstrated to be inconsistent at best, with different studies showing different advantages and limitations of the technologies (Finn, Sellen and Wilbur, 1997). For example, Williams (1997) demonstrated that visual access improves understanding when collaborators come from different linguistic backgrounds and a raft of studies of 'media-space' video communications arrangements have suggested that visual access can provide for new forms of interaction and increase sense of presence between remote sites, with positive outcomes (see Dourish and Bellotti, 1997 for overview). However, experimental studies of video-mediated communication have demonstrated that video access between remote spaces does always positively enhance task outcome (Sellen, 1997). In certain situations video-based communication devices are inadequate. Consider for example the scenario below.

---------------------------------------

**A Collaboration Scenario**

The Paramedic arrives at the scene of the accident; jumping out of the ambulance he tries to survey the scene. The air is filled with an obscuring oily smoke making it hard to make out what lies ahead. As the Paramedic advances he notices twisted car wreckage littering the highway, occasionally illuminated by small patches of burning fuel. Already, there are Fire crews frantically running between the wrecks dealing with the fires and trying to deal with the mounting tide of casualties. Up-ahead a Firefighter pulls a person from the wreckage of a car, laying them on the grass at the side of the road. The Paramedic runs to the Firefighter and the patient to see if he can help. The patient is bleeding heavily from an open chest wound. The Paramedic knows from experience that pressure or dressings will not stem the tide of blood and the patient will bleed to death in a matter of minutes unless there is something they can do. There is something that could be done. If the Paramedic could only open the chest wound slightly and locate the ruptured arterial structure and then clamp it, they could keep the patient alive for long enough to get them to a hospital for more significant surgical intervention. The Paramedic's training however did not cover such a complicated invasive procedure; they need a consult from a surgeon. Logistically it makes most sense for the surgical team to stay in the hospital and receive incoming patients rather than travelling themselves to the site of the accident. So the question becomes, how can the surgeon be in two places at once?

Existing practice in such a scenario might find the Paramedic talking to a Surgeon via mobile phone technology. The Surgeon will have to use the Paramedic's eyes to survey the situation and she will have to talk to the Paramedic to guide both his eyes and hands. Increasing development of technology has however meant that high-bandwidth, streaming video-enabled phones, can give the Surgeon remote eyes, letting her see the situation for herself. This may or may not help depending on how good at describing the Paramedic already is, and depending on environmental factors which might make the video image less than clear. But in this situation the real problem arises when the Paramedic must use the clamp. The rupturing has

occurred to the underside of one of several closely located branches of the exposed arterial structure. In the confusion the Surgeon must carefully use the Paramedic, she might have visual access to the patient but this doesn't necessarily help to guide the Paramedic's actions. The Surgeon's instructions must be precise, easily interpreted and quick; mistakes at this point in the process could be fatal. Unsure of the instructions and unable to understand the correct alignment for applying the clamp the Paramedic loses valuable time systematically moving the clamp through various orientations asking 'do you mean insert it like <u>this</u>? Or like <u>this</u>?' waiting for the Surgeon's confirmation or feedback, all the time the patient is bleeding and fading more. Finding the slow progress frustrating the Surgeon wishes that rather than having to reiterate her instructions she could get the Paramedic to move the clamp as she intends by merely saying 'Turn it <u>this</u> way' whilst confidently and observably motioning with her hand to show the correct angle.

---------------------------------------

The scenario above is just one form of collaborative task for which the use of communication focussed on artefacts in the real world and the manipulation of those physical artefacts are the overriding concerns. Other relevant examples could include bomb disposal experts receiving external support and advice, scientists in-the-field examining finds or specimens with the aid of remote colleagues or maintenance staff repairing intricate equipment and machinery with the support of an expert engineer. The common ground between all of these collaborations is the fact that whilst one worker is *in situ* with the task artefacts, the collaborative colleague is elsewhere and in many of the situations given above the person who is remote to the task space is the possessor of expert knowledge about the task or artefacts. A principle component of these tasks however, is that they possess an inherently physical nature, they are not software based tasks and therefore mutual and concurrent access to the artefacts for manipulation cannot be granted, there is an inherent asymmetry to the interaction that is created by the very corporeality of the task artefacts and the distributed nature of the working arrangement. And as the scenario presented above demonstrates, this poses certain difficulties for current technologies when it comes to adequately supporting communication. Whilst a video link can provide visual access it falls short of projecting the forms of situated and embedded communicative non-verbal behaviours which have been shown to be of such importance in co-located interactions.

The work of this thesis then, set to explore such forms of interaction and the design of technologies to support them, is situated within the sphere of Computer-Supported Cooperative Work (CSCW), an area of research within computing and the social sciences which has traditionally striven to understand how technology can be designed to adequately support collaborative endeavour (Baecker, 1993). More specifically within this field this thesis is concerned with the study of Video-Mediated Communication, and in particular adds to the body of work seeking to explore how Video-Mediated Communication systems can be

improved upon to support distributed interactions in specifically *collaborative physical tasks*. This thesis is an exploration of how to develop technology that *will* support the forms of interaction described above, studying the design and potential implementation of technologies which allow for the remote representation of non-verbal behaviours and artefact focused actions in addition to providing visual access between spaces.

The rest of this chapter provides the research background and context of the thesis. It discusses *remote gesturing technologies* as tools to support collaborative physical tasks, introducing the current *state-of-the-art* systems and briefly highlighting criticisms of their design. On the basis of these criticisms and the perceived failings of current approaches a research problem is constructed and the thesis's hypothesis for resolution of that problem is outlined. The rest of the thesis is then sketched out detailing the structure for the remaining chapters, explaining how they address the central research questions and the chapter concludes by detailing the thesis's contributions.

## 1.2 Research Background

The experimental work of Chapanis (1975) and Kraut et al (1996) systematically investigated the performance effects of varying communication media used by dyads engaged in collaborative physical tasks. These investigations presented the somewhat counter-intuitive findings that audio-video links are rarely more effective in terms of collaborative task outcomes than audio-only links between remote sites. And intriguingly neither form of technology-mediated communication between spaces can replicate the efficiency and fluency of natural face-to-face interactions. The inherent problems of video links have been consistently demonstrated in relation to the construction of collaborative physical action (see Heath and Luff, 1992 and Gaver et al, 1993). The conclusions drawn from this research usually suggest that the great failing of video technology in supporting collaboration over physical artefacts is its inability to adequately represent naturally occurring deictic (pointing) behaviours. The classic example of this is an observation made during the MTV (Multiple Target Video) study by Gaver et al (1993). In this study the experimenters noticed that whilst watching and directing action in another room over a video link, participants would continually (unconsciously) point at items on their video screen, whilst using deictic pronouns to refer to objects such as '*this* one here', when trying to direct the attention of a remote collaborator. Of course the remote collaborator was unaware of what the other was pointing at, as they had no visual access to the pointing behaviour. That humans express such a strong desire to use non-verbal communication comes as no surprise when one considers that studies of collaborative working practices have revealed the subtle ways in which highly situated communicative behaviours are used to structure interaction and guide task awareness (Hutchins and Palen, 1997). In many working situations gesturing behaviour is used in communication as it allows participants to construct simpler sentences (Clark and Brennan,

1991), which in conjunction with the expressive nature of the gesture itself aids the development of common understanding and the grounding of conversational references (McNeill, 1992, Clark, 1995).

Research has therefore been conducted to extend the functionality of video-mediated communication systems so as to adequately support collaborative physical tasks, by facilitating the remote representation of gestures during interaction. These new *remote gesture tools* have been developing along differing lines in different research labs but all conform to the central tenet of supporting the generation, and embedding, of some form of gestural simulacra within remote task spaces, increasing the presence of remote collaborators within those spaces. Some of these remote gesture tools are discussed further below.

Growing out of, and enriched by, a developing body of work concerned with understanding the construction and use of collaborative shared visual environments (e.g. Krauss and Fussell, 1991, Fussell et al 2000, Kraut et al 2003, Gergle et al 2004) one significant strand of research (see Ou et al, 2003 and Fussell et al, 2004) is the development of the Drawing Over Video Environment (DOVE) at CMU. This remote gestural simulacrum allows a remote expert's sketches to be pasted over a live video feed of a worker's task space. Research has demonstrated that such remote gesture tools can significantly improve performance in collaborative tasks over that achievable by audio-video only links (Fussell et al, 2004). However, these benefits have not always been replicable, even with the same system (Kramer et al, 2006). When critically considered, the DOVE system, has certain features which would arguably limit its benefits. The system uses a digital pen-based representation of gesture which potentially has a lower bandwidth for the expression of non-verbal communication than the use of hands for gesturing. Also, the DOVE system's output of gestures, provides the remote worker with a view (a separate VDU display) of a mixed reality environment, situated externally to the immediate task space. Through this view the worker can see a representation of what the remote expert sees of the working task space, and they can see the expert's gestures being sketched over this live video feed. Whilst this approach ensures that the worker is implicitly aware of the remote expert's perspective on the task space, the worker has the difficulty of perceiving gestures drawn over a video view of their work space which is potentially at a subtly different orientation to their own perspective on the space. The Worker must then have to record and translate this information, making it relevant to their perspective rather than the representation of it, a translation process which arguably carries with it a performance cost. The relative impact of these issues on performance has not yet been established.

Another strand of research has witnessed the construction of increasingly novel technological solutions to the remote gesturing problem, including GestureCam, GestureCar, GestureMan and GestureMan with a Pointing Stick (see Kuzuoka et al, 2000, & 2004). These systems all utilise human-proxy robots, physically located in a remote working space, carrying and

embodying the video link to the remote expert. They facilitate remote gesturing by allowing the expert to remotely operate a laser pointer attached to the robot that allows a remote gestural simulacrum to be physically embedded in the actual working task space. Whilst these technological solutions in themselves have been inherently interesting they make certain assumptions about the success of the technologies without empirical support, they have yet to demonstrate any actual performance benefits of their approach. Again a critical review of the systems would highlight the use of a laser dot pointer as the primary gestural representation. This must have the lowest bandwidth for representation of gestural information out of any of the currently used techniques, given its small presence, artificiality (we are at least relatively used to using pen drawn lines to annotate and guide attention) and lack of permanence within the task space. The Kuzuoka work has however managed to make the interactions far more mobile than any systems such as DOVE, which is possibly important given the possible applications of such devices. Later developments of the laser pointer approach such as the WACL system (Sakata et al 2003) have begun to explore the true value of mobile and light-weight remote gesturing systems, but have still been constrained by the limited use that can be derived from such a simple representation of remote gesture. Again, the relative ability of such low-bandwidth expressions of gesture to adequately support collaboration has not been evaluated.

Critiques of the effectiveness of remote gesture technologies in supporting artefact-centred interactions have focussed on the concept of *fractured ecologies* (e.g. Luff et al 2003, Kuzuoka et al 2004 and Kirk et al 2005), in some respect acknowledging the role of remote gesture representations in establishing 'ecologies of communication' which exist between distributed working partners. This concept postulates that key aspects of the design of remote gesture tools create unsurpassable barriers to a coherent understanding of intentionality and obscure the projectability of action between remote collaborators, fracturing the process of interaction between them.

As discussed previously, with DOVE style systems that promote the use of externalized VDU's, the site of gestural interaction is removed from the site of artefact manipulation, thus causing a fracture as the Worker is required to resolve the discrepancies between gestural instruction and their own task perspectives. Whilst laser pointer systems have traditionally avoided this problem, by projecting into the task-space, they are themselves fracturing interaction by the limited bandwidth capacity they have for the adequate expression of intention through gesture. Understanding of an Expert's orientation to and gestures toward task-artefacts is severely impaired by such systems. It is clear therefore that remote gesture tools as currently constructed are not without their problems, and despite the proposal that they should improve performance in collaborative physical tasks beyond that achievable with standard forms of video-mediated communication, this has not yet been proven conclusively. Equally the myriad design options for constructing such systems have not been adequately

compared and in the face of significant criticism it is clear that re-design is potentially necessary.


## 1.3 Problem Statement and Research Hypothesis

This thesis therefore seeks to address the problems highlighted in the previous section. The fundamental research question can be phrased as - how can technologies be built to improve remote collaborations for physical tasks, that don't fracture ecologies between remote spaces, but make the interactions as close to the presumed optimal standard of face-to-face communication as possible? Specifically, this research question can be broken down into several sub-issues, which the thesis seeks to address. Firstly, it seeks to understand and evaluate how and why *remote gesture tools* can benefit performance in *collaborative physical tasks*, exploring the ways in which such communication devices might be superior to standard video-mediated communications. The thesis also seeks to understand what creates a 'fractured ecology' of communication, exploring how interaction breaks down and how remote gesture tools influence this process. The thesis also strives to explore the relative benefits of the various system design choices that can be made, assessing whether location of gestural output or format of gestural representation influences the efficacy of the system. In doing this the thesis also develops a fuller understanding of the role of remote gestural action in collaboration addressing the issue of how communicative behaviours influence task performance.

In addressing these issues a research hypothesis is proposed and evaluated. Previous research has argued that the presence of dichotomous ecologies in such working collaborations is inevitable (Kuzuoka et al 2004), and the role of communication tools is to mediate between the ecologies without fracturing interaction. Referring back to the quotes of Abraham Moles (page 1), this thesis rejects such a notion. Moles' conception of communication argued that for effective communication one must make another 'partake in the experiences (Erfahrungen) and stimuli of the environment of another' and that to do this one must 'make more complex the space-time surrounding the point of reception', it is with these points in mind that this thesis proposes the notion of the 'mixed ecology'. A mixed ecology approach to communication device design assumes that rather than linking and mediating between spaces the technology should seek to construct a unified environment in which both parties can collaborate.

When collaborators are remotely engaged in communicative acts concerning some object-focussed interaction it is hypothesised that their performance will be optimised if they communicate using a mixed or shared ecology communications arrangement. The mixed or shared ecology supports communication by using technology to give collaborating partners access to the most salient and relevant features of communicative action that are utilised in face-to-face interaction (thereby conforming to Moles' desires for communication), namely mutual and reciprocal awareness of commonly understood, yet richly complex object-focussed actions (hand-based gestures) and mutual and reciprocal awareness of task-space perspectives.

It is proposed that a mixed ecology therefore has more ability to successfully relay those contextually embedded physical representations which have been shown to be of importance to collaboration in shared ecologies.

## 1.4 Thesis Overview

The following thesis chapters address the research problem discussed above. The ensuing section briefly outlines the content of each of these chapters demonstrating how they evaluate the design of remote gesture tools, explore the role of gesture in remote communications and how they consecutively build an argument for a mixed ecologies approach to designing communications support for collaborative physical tasks.

*Chapter 2 [Literature Review]* focuses on reviewing previous research in this area, taking the study of workplace communication, and in particular video-mediated communication, as a starting point, and drawing out the development of remote gesture tools within this context. The chapter describes in detail the *state-of-the-art* in remote gesture tools and discusses the evaluatory studies that have been performed with them. The chapter reveals that these studies have eventually lead to the realisation that remote gesture representation and shared access to views on task-spaces is important but have also highlighted that attempts to provide these things do not always work and can lead to a fracturing of the interaction between collaborators. Observations from this literature review are used to articulate areas for further research which form the basis for the specific research questions of the thesis.

*Chapter 3 [Research Methodology and Disposition]* forms a hypothesis on the basis of evidence from the literature review that the best way to support collaborative physical tasks is to create mixed ecologies, which are environments that project key features of face-to-face interaction, mutual and reciprocal awareness of hand-based gestures and mutual and reciprocal awareness of task-space perspectives. The chapter highlights the specific research questions which must be addressed to evaluate this hypothesis and discusses the appropriate methodologies for approaching the subject. The chapter concludes by presenting and discussing the 'mixed ecology' remote gesturing prototype which formed the basic system used for the experimental studies reported in later chapters.

*Chapter 4 [Some Effects of Remote Gesturing]* presents two experiments which demonstrate how remote gesturing can improve aspects of performance in collaborative physical tasks when compared to standard video-mediated communication links. The first experiment examines base performance metrics, including task completion time and cognitive effort, whilst the second experiment demonstrates the positive impact on learning of gesturing during remote instruction. Taken together the studies also demonstrate some subtle effects of remote gesturing on the relative perceptions of first time collaborators. The studies in particular highlight that the use of views of the hands embedded in the task space seems beneficial as a

gestural representation, discussing this in terms of a mixed ecologies approach but stressing the need for direct comparison with other methods of representing and locating gestures.

*Chapter 5 [How Best to Construct Remote Gestures]* presents two further experiments which address the issues of how to locate and represent gestures during remote collaboration, evaluating the relative benefits of the differing system configurations employed by current systems. The first study examines the impact of changing orientation on gesture insertion into a space, demonstrating that this counter-intuitively has minimal impact on collaboration. The second experiment addresses the issues of gross gesture location (presented within the task space or external to it) and gesture format (digital sketch vs. unmediated view of hands vs. hands and sketch). The studies demonstrate support for a mixed ecologies approach and highlight a key issue of designing for reciprocal views of tasks spaces which is discussed in detail.

*Chapter 6 [The Communicative Functions of Gesturing]* moves the argument of the thesis onto the examination of exactly how gestural representations influence collaborative performance. By performing a fine-grained video-analysis of scenes of interaction from the earlier experiments a praxiological account of gestural representations is revealed. A qualitative understanding of the gestural phrases used is developed and the varying methods of gestural communication, for each specific medium (hands and sketches), is elaborated, creating a taxonomy of gestures and gestural uses. Through a comparative critique of alternative gestural representations the strengths of using unmediated views of hands as the gestural representation are articulated.

*Chapter 7 [How Gesture Interacts with Language]* extends the analysis of the functions of gestural interaction to investigate how gesture use affects collaborative language. Again utilising fine-grained analysis of video data from previous trials, this time utilising a conversation analytic strategy combined with quantitative analysis of language patterns, earlier work is re-examined. The analysis reveals both the various means by which gesturing aids the achievement of grounding during collaborative discourse and also its role in structuring the interactions. This further reveals the importance of remote gesturing in collaborative physical tasks and provides important evidence of how gesturing influences the temporal course of grounding behaviours. This influence of gesturing on the time course of interactions is discussed in detail as it has significant implications for any future deployments of remote gesture technologies.

*Chapter 8 [Conclusions]* concludes the thesis by summarizing and evaluating the evidence for a mixed ecologies approach to designing support for collaborative physical tasks and presents answers to the research questions posed. It then discusses the implications of this for the design, deployment and development of remote gesturing technologies, articulating a program of future work to address issues raised by the thesis research.

**1.5 Thesis Contributions**

Having articulated the structure of the rest of the thesis and discussed how the thesis will address the research area it is pertinent to conclude this introductory chapter by detailing the overall contributions that the thesis makes. The main contribution of this thesis is a thorough understanding of human factors as they relate to the design and use of remote gesture tools. Specific contributions include:

- A thorough discussion of the requirements of studying remote gesture tools, including an evaluation of appropriate methodologies

- A set of guidelines for *deploying* remote gesture tools, covering environmental, task-focused and participant-oriented factors

- A set of guidelines for *designing* remote gesture tools, focusing on the identification of key criteria for collaboration, and the elimination of fractures in interaction

- A set of experimental comparisons of different remote gesture tool designs, illustrating relative impact on both physical performance and communication

- A taxonomy of remote gestures (in various media) and their communicative uses

- A deeper understanding of the (potential) role of remote gestures in collaborative physical tasks, focussing on their integration with naturally occurring collaborative speech patterns

- A discussion and evaluation of a *mixed ecologies* rationale for designing communications devices

- Indication of areas of further importance for future research and development

These thesis contributions have directly extended the body of research in the design and development of remote gesture tools. In a continuing process the work has been disseminated to a wider audience through presentation and publication.

The thesis work has thus far been presented for discussion at:

- The Doctoral Consortium of the 9th European Conference on Computer-Supported Cooperative Work (ECSCW) 2005 (Paris, France)

- A conference workshop entitled 'Giving Help at a Distance: Ubiquitous Computing to Support Problem-Solving' at Ubicomp 2004 (Nottingham, UK)

- An agenda setting workshop on 'Collaboration, Co-Laboratories and e-Research' as part of the UK e-Social Science program (invited talk)

- University of Bath, Department of Psychology, Seminar Series (invited talk)

And the work has also been published in peer-reviewed conference proceedings at (see Published Works section before acknowledgements for full references):

- The Conference on Computer-Supported Collaborative Learning (CSCL 2005) (Kirk and Stanton Fraser, 2005)

- The European Conference of Computer-Supported Cooperative Work (ECSCW 2005) (Kirk, Crabtree and Rodden, 2005)

- The ACM Conference on Human Factors in Computing Systems (CHI 2006, 2007) (Kirk and Stanton Fraser, 2006 and Kirk, Rodden and Stanton Fraser 2007)

The publications are based directly on the key study findings taken from various sections of the ensuing thesis chapters.

## Chapter 2 – Literature Review

### 2.1 Introduction

The purpose of this literature review is to provide some background to the ensuing discussions and investigations concerned with the development of remote gesture tools. The chapter begins by first highlighting a growing concern for the understanding of how collaborative environments are constructed to represent embodied collaborative actions and then continues by describing what other research has been performed in efforts to support communication and in particular communication around collaborative physical tasks. In doing this the chapter presents the evolution of remote gesture tools from their basis in simple extensions of video-mediated communication through to the state-of-the-art systems that are currently being explored. The evaluation of these presented studies highlights the areas of inadequacy of current approaches to remote gesture tool design. The discussion highlights current critiques of these systems and begins the process of articulating where and why existing literature is lacking, in turn suggesting areas for further research, in an effort to raise answerable research questions in chapter 3.

### 2.2 Ecologies of Communication in the Workplace

There is a growing body of research work which takes as its focus the uncovering of the fine-grained processes of interaction and coordination that take place in modern workplaces. Old models of work-flow and task analysis have been marginalised as they have rightly been critiqued for their lack of applicability owing to their failure to engage with and represent actual lived in working practices as they occur in actual working contexts (Bannon, 1991). From diverse disciplines there is a growing concern to understand how the embodied practices and actions of workers as they are physically presented in a collaborative working environment constitute a communicative act that is at once both a fundamental aspect of a worker's own activity and a resource for the development and manipulation of collaborative task awareness. Although the language used to describe these activities may differ by the disciplinary orientation of study authors, the principles of understanding the situated nature of embodied cognition and its relevance to collaborative work remain the same. Examples of relevant works include Suchman's (1996) study of an airline's operations control room, Heath and Luff's (1996) study of the London Underground control rooms and Nardi et al's (1997) study of the practices of neurosurgery teams. Additionally, of particular importance is the work of Ed Hutchins in developing the concept of distributed cognition (Hutchins, 1991, 1995), which as a framework sought specifically to redress the imbalance of traditional cognitive science paradigms which focused purely on cognitive processes as internal phenomena. Through his discussion of distributed cognition Hutchins attempted to develop the notion of cognitive

processes as being embedded in task artefacts, state representations and collaborative actions. The studies in Hutchins (1995) of ship navigation teams and Hutchins and Klausen (1997) of airplane cockpit crews supported the growing understanding that the processes of communication in a collaborative physical task are far more subtle and complex than might otherwise be presumed. Hutchins and Palen (1997) studied training sessions in an aircraft simulator and after observing the complexity of the communicative ecology of the 'cockpit' remarked:

> "Gestures and the space inhabited by speakers and listeners are normally thought of as providing context for the interpretation of speech…. space, gesture and speech are all combined in the construction of complex multilayered representations in which no single layer is complete or coherent by itself." (pp. 23-24)

And added further that, awareness of physical embodiments and cognitive representations within the space

> "…demonstrates the creation of a complex representational object that is composed through the superimposition of several kinds of structure in the visual and auditory sense modalities. Granting primacy to any one of the layers of the object destroys the whole." (pp. 38-39)

For Hutchins and Palen (1997):

> "Communicative behaviors *are* the representations by which a socially distributed cognitive system does its work." (p. 24)

This belief in the development of multi-layered communicative environments which embody cognitive processes resonates strongly with the embodied cognition work of Toni Robertson (1997a, 1997b) and her study of the embodied practices of working design teams. Robertson demonstrated that the very physicality of the designers, embodied within the workspace, was a cultural and communicative artefact of the workspace, awareness of which was of primary importance to collaborators' understanding of current task progress and communicative intent. The work of Robertson is particularly interesting as her motivation is the desire to support these design activities remotely, her taxonomy strives to articulate those embodied practices which are critical to supporting the design process. Efforts to successfully design tools to support collaborative physical tasks in other domains (such as those presented under the scope of this thesis) would do well then to consider which aspects of embodied practices it is sapient to support in distributed working arrangements.

The following sections of this chapter explore some of the avenues that have been investigated in efforts to construct exactly these kinds of richer communicative environments.

**2.3 Studies of Video-Mediated Communication (VMC)**

The bedrock of this thesis is an exploration of Video-Mediated Communication (VMC), as this is an integral aspect of most remote gesture tools, and to a certain extent remote gesture tools could be referred to as an advanced form of VMC[2]. In essence VMC technologies are tools that provide collaborators with visual access to remote spaces. The technologies of VMC have been iteratively developed over many years, with the earliest explorations occurring in the early 1970's (e.g. Chapanis et al, 1972). A good overview of the research in the area is provided by Finn (1997), itself a chapter within the definitive work on VMC by Finn, Sellen and Wilbur (1997) which presents studies from the leading strands of research within the field. This section of literature review attempts to provide a brief overview of the technologies encountered in the field, and the analytical approaches to evaluating them that have been adopted, discussing some of the conflicting findings that work within the area has generated and attempting to distil some conclusions about the overall efficacy of VMC as a tool for supporting groupwork.

**2.3.1 Technologies for VMC**

A pertinent point to start this overview of VMC is to familiarize oneself with the technologies used to provide the visual access to spaces. Angiolillo et al (1997) provide an in-depth study of the technical components in VMC systems, briefly discussing how technological factors may impinge on their usability. But rather than focus specifically on the technological requirements, as given the exponential growth in processing power of computers, they rapidly alter, it is perhaps more sapient to consider the general technological forms of VMC.

There are roughly six approaches to VMC which have evolved thus far and been evaluated in research studies (the first five are discussed in Finn, 1997), showing a natural progress and development over time. These forms are:

- *Fixed line, CCTV (closed caption TV) based systems* (used primarily for experimental purposes in early studies of VMC)

- *Video-conferencing systems* (supporting formal group meetings)

- *Desktop based video-conferencing systems* (supporting both formal and informal contact through video links presented on one's desktop)

- *Media-spaces* (which incorporated multiple reconfigurable video links between distributed people, spaces and resources)

---

[2] This is not to say that all remote gesture tools are based entirely on the principle of using video technology, as some clearly use non video-based methods for the remote presentation of gestures. However, a video feed of the remote task space will always be included in the apparatus for the Expert to view what is happening at the remote site and to guide their own gestural actions, so there is at least an asynchronous video link between spaces.

- *Video-as-data technologies* (essentially these could be considered as a regression to simpler communication links but actually represent a fundamental re-think about the role of visual resources in the communicative process based on observations from previous research)

- *Mixed Reality (live video in virtual worlds)*

These differing forms of technology shall be considered each in turn.

The early research work which utilised CCTV ironically had higher fidelity links than much of the later work performed with VMC systems. Because of the hardwired nature of the links however, they were constructed purely for exploration as a future development of technology and not evaluated as a deployed communication tool. Therefore the studies associated with such technologies are largely experimental lab-based studies which sought to compare various facets of performance under differing media conditions (e.g. Chapanis, 1975, Short, Williams and Christie, 1976, Williams, 1977, Rutter, Stephenson and Dewey, 1981).

Later work moved on to consider 'videoconferencing' systems which sought to support formal 'round-table' meetings. Typically in these systems each conference room was equipped with a large screen monitor and one camera (usually held above the monitor). On the monitor a group of colleagues could see the other office to which they were connected and therefore the other group of colleagues at that site. Examples of such systems include the ISDN and LiveNet systems reported in O'Conaill, Whittaker and Wilbur (1993, and see also O'Conaill and Whittaker, 1997), and the video teleconference rooms discussed in Tang and Isaacs (1993). Such systems did become adopted by large multi-site multi-national corporations and in many respects became the *de facto* form of VMC for many users (for example the XTV system at Xerox, discussed by Sellen and Harper, 1997).

Beyond the studies of supporting large group meetings a focus began to be drawn on desktop videoconferencing, providing video-based access to multiple participants at a variety of different locations. One particular system, the Hydra model (Sellen, 1992, 1995, see figure 2.1), extends the use of videoconferencing to multiple sites, whilst striving to keep intact processes of spatial awareness. In the Hydra system each collaborator was presented on a dedicated unit, which combined a small video screen with an integrated camera, this enabled spatially relevant information concerning focus of attention to be represented by the head movements of collaborators as they turned to focus on each participant.

Figure 2.1 Hydra system (taken from Sellen, 1992)

Contrary to the somewhat unique approach of the Hydra model most desktop based video conferencing systems employed a strategy of Picture-in-picture (PIP) presentation of collaborative participants (see figure 2.2, as seen in the DVC prototype of Isaacs and Tang 1993, Tang and Isaacs, 1993 and Isaacs and Tang, 1997 and the PIP component of the study in Sellen, 1995). This is commonly referred to as the 'talking heads' model of VMC, wherein only the upper portion of each collaborator's torso and head are viewed on the video link. Incidentally this model was also used in the video-conferencing systems mentioned above and also in the early CCTV linked studies. Interestingly respondents in the Sellen (1995) study reportedly claimed they preferred PIP because of the smoother turn-taking (the lack of inappropriate interruptions, whilst providing good support for selective listening and attending to others). As an extension to this talking heads model however, as systems such as the DVC prototype mentioned above were located as a part of the desktop PC system, it became possible to directly incorporate data sharing applications, and other collaborative editing software (an obvious limitation in the Hydra concept). This moved communication away from being purely discursive, towards supporting more object-focussed interactions. Equally as the location for VMC had changed, so too did the parameters under which it was used, whereas videoconferencing had previously been a formal activity taking place in a dedicated room, the provision of desktop VMC increased the potential for more 'informal' interactions (see Isaacs, Whittaker, Frohlich and O'Conaill, 1997 for discussion of the notions of informal communication). This move towards a more informal base for VMC recognises the research studies which had suggested that there was a potential for video-based technologies to support informal interactions, which were seen to be extremely common and a driver for collaboration in the workplace (Fish, Kraut, Root and Rice, 1992, Fish, Kraut and Chalfonte, 1990, Root, 1988, Kraut, Root, Fish and Chalfonte, 1990, Kraut, Galegher and Egido, 1990).

Figure 2.2 the DVC prototype of Isaacs and Tang (1993)

It is this notion of supporting the informal aspects of everyday communication which was behind the next conceptual step in VMC technology, the media space (see figure 2.3). Media spaces were attempts to integrate video connectivity into the very architectural construction of working spaces, providing ever-present and rapidly re-configurable video links between distributed spaces, people and resources. Several systems were constructed that explored this model of interaction including the Cruiser system at Bellcore (Fish, Kraut, Root and Rice, 1993), CAVECAT at Toronto (Mantei et al, 1991) including the later work of the Ontario Telepresence Project (Moore, 1997), the Media space at Xerox PARC (Bly, Harrison and Irwin, 1993) and EuroPARC's RAVE project (Ravenscroft Audio-Video Environment) (Gaver et al 1992). These systems frequently employed connections of many different types to many different locations, connecting individual offices to networks of other offices or establishing relatively permanent 'office-shares' (Dourish et al, 1996) or in some cases providing large video windows between the common areas of distributed workplaces (Harrison et al., 1997). To help boost the connectivity of users many systems employed modifications which allowed informal glances to be made into video-linked spaces, sometimes on a random basis (e.g. Portholes – Dourish and Bly, 1992), other times user controlled (e.g. Montage – part of the DVC system at SunSoft, Tang and Rua, 1994, Tang, Isaacs and Rua, 1994). Systems developed in this manner clearly have strong implications for privacy and in many cases this was studied and suggestions for modifications to the technology were mooted (Bellotti and Sellen, 1993).

Figure 2.3 A media space (showing two connected nodes)

The use of media spaces has not however, become common place. This may be for several reasons, despite the fact that those who have used them seemingly have come to love them (retrospective analysis tending to exhibit some nostalgia, Bellotti and Dourish, 1997), the potential investment in technology required to establish a media space infrastructure may be a limiting factor. Equally the systems themselves as presented in the earlier works tend to have certain limitations concerning the scope of access that is provided to remote spaces. It has been argued that in many instances what is required of a video link between spaces is not the talking heads communication link, that many of the media spaces supported, but also access to objects of interest (Heath, Luff and Sellen, 1997). Similar to the extensions made to the desktop-video conferencing models what was required of media spaces was access to shared artefacts, but the apparent problem was that users required access to physical objects in spaces, or at the very least shared views of physical objects. Research has suggested that increasing access to a remote space by increasing numbers of camera views within a given space (such as having dedicated object-oriented views does not improve collaboration (Gaver et al, 1993, Heath, Luff and Sellen, 1995), as such multiple views gives rise to discontinuities in orientation. This concern however, with ensuring that views of not just collaborators but objects of interest are being shared, marks the change from media space research which was concerned with an understanding of using technology to support social networks, to developing technology to support tasks, using shared video as data. Specific examples of this use of video collaboration can be seen in Nardi et al (1993, 1997) with their studies of neurosurgery teams. This notion of using the verbal channel as the primary conduit for interpersonal communication and the video channel as a secondary conduit for supporting shared access to a task-space is an underpinning feature developed in many applications concerned with supporting distributed collaborative work which will be discussed in later sections of the literature review.

The most recent developments in VMC have moved towards an integration of the physical and digital. The basic aspects of using video to link spaces have not progressed but the notions of

how this can be integrated within a working space have come under scrutiny. In particular the work to develop a Mixed Reality Architecture seen in Schnädelbach et al (2006, see figure 2.4) has striven to explore how multiple video-linked nodes can exist within a virtual space creating social networks and space for informal interactions mediated through access to a virtual world. Equally the development of Mixed Reality Boundaries (Benford et al 1998, Koleva et al 2000, 2001) has demonstrated how links between virtual environments and physical environments can be constructed and then traversed, extending the notion of how video-mediated communication links spaces.
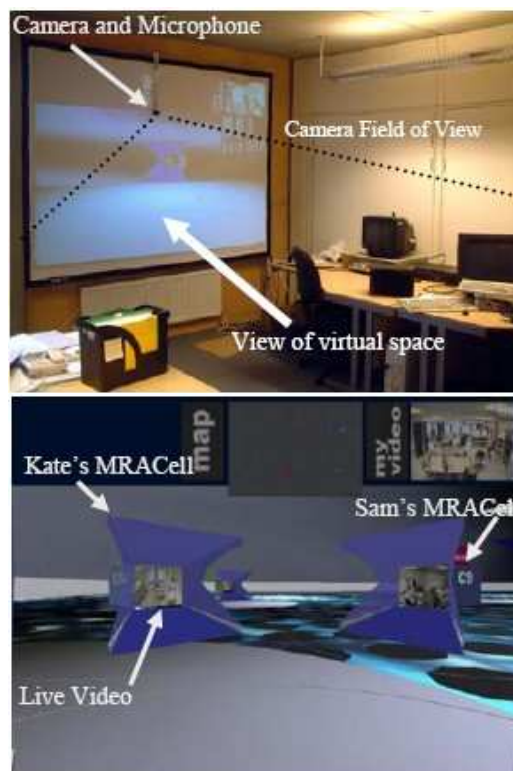


Figure 2.4 Mixed Reality Architecture (Boundary – from Schnadelbach et al 2006)

### 2.3.2 Analytical approaches to VMC

Along with the many different technical approaches to establishing VMC there have been a variety of analytical approaches taken to their evaluation, showing changes in both focus of the research and types of questions that were asked. This has often been tied to the form of technology that has been investigated. Sellen (1997) argues that there are principally four main approaches to the evaluation of VMC systems that have been encountered.

- *Experimental studies*

- *Living with technology*

- *Field studies*

- *Hybrid approaches*

### 2.3.2.1 Experimental studies

The earliest adopted of these analytical traditions in the study of VMC was the experimental analysis. Studies that adopted this approach were often derived from psychological perspectives on data collection and analysis and could reasonably be described as reductionist in approach, requiring firm control over variables and therefore being suited to lab-based analysis and the forms of VMC that utilised fixed link CCTV systems as discussed above. Mostly the studies in this area aimed to establish the base efficacy of VMC in measurable ways, often comparing it against face-to-face communication, or contrasting alternative system designs, such as the provision of audio versus video connections (Chapanis, 1975) or different qualities of video provision (O'Conaill et al., 1993). The experimental studies can be broadly split into three groups, those that focused on the task outcome benefits of VMC, those that focused on the effects of VMC on communication process and those that took a multidimensional approach.

Of those studies that focussed on the task outcomes of VMC use most demonstrated little support for the role of video in remote collaboration. The Chapanis studies (Chapanis et al 1972, Ochsman and Chapanis, 1974, Chapanis, 1975), the BT (British Telecom) works of Short, Williams and Christie (1976) and Williams (1977) and the work of Gale (1989) all failed to generate significant performance enhancements from the provision of video links between spaces as collaborators were engaged in collaborative tasks. From their manipulations of the modality of communication the studies all firmly believed that the audio channel was the communicative conduit of most importance in collaboration. An interesting note however, from the Gale (1989) study was that despite its lack of observable impact on performance and despite participants saying that they never used the video channel during communication, the study observed that users did in fact heavily utilise the video medium, and frequently focussed attention on it, if only in micro-glances, the author suggesting that use of a video channel was perhaps so pervasive that users were unaware that they were using it.

A large number of studies have alternatively focussed on how VMC apparatus affects communication process and structure during collaboration. O'Conaill, Whittaker and Wilbur (1993) considered specific aspects of VMC system design demonstrating that use of a VMC technology (when compared to face-to-face interaction) leads to more formalised turn-taking, fewer interruptions, giving a more lecture like interaction, these findings are extended and confirmed in O'Conaill and Whittaker (1997). These studies argue that even when video quality is extremely high there are likely to be differences between face-to-face and mediated communication, with VMC unable to replicate the fluent interactions of face-to-face meetings.

The studies demonstrated however, that higher quality VMC improved the process of communication making it more like face-to-face interaction. These findings support and extend the earlier work of Cohen (1982) who compared face-to-face communication with a PicturePhone Meeting Service (PMS) system. The results of this work also demonstrating that more mediated communication lead to more formalised turn-taking, and suggesting that participants preferred face-to-face interactions (slightly) as it was better for discussions facilitating more speaker exchanges. Sellen (1992, 1995) compared different forms of VMC with both face-to-face and audio only communication. She explicitly compared PIP, Hydra and LiveWire (which used audio-based video switching – so participants were shown the image of the current speaker only) VMC systems. Where Sellen noticed higher levels of interruption in face-to-face interactions she has argued that rather than being problematic (as users tend to prefer face-to-face communication) they are indicators of interactivity and therefore are a sign of more fluent interaction.

Overall then these studies which have focussed on the communicative process impact of VMC have remarked on how it fails to replicate the speech patterns observed in face-to-face interaction. But they have tended to remark on the general efficacy of VMC as a tool, suggesting that higher fidelity visual information improves collaboration making it more like face-to-face meetings.

Several studies of VMC from the experimental tradition have however taken heed of the comments of Monk et al (1996), who suggested the need for multidimensional analysis in CSCW, considering that both task outcomes and communicative processes should be examined to successfully determine the adequacy of VMC technologies. A primary example of this multidimensional approach can be seen in the body of work represented by Anderson et al (1994, 1997) and Doherty-Sneddon et al (1997). These studies which used the 'video-tunnels' VMC technology of Smith et al (1991) investigated the use of VMC technologies in collaborative problem solving. They studied both task performance and dialogue, demonstrating that dialogues in VMC are more like face-to-face than audio-only dialogues, suggesting that VMC users didn't need to provide verbal feedback of understanding, as this was presented visually as it would be in a co-present interaction. In line with the other studies of communicative process discussed above, the Anderson et al studies also demonstrated that VMC leads to more interruptions than audio-only interactions (thus demonstrating VMC's improved support for fluency). Improving VMC connections to include full eye contact did not however make interaction the same as face-to-face communication, key interactional aspects that face-to-face communication retains were still absent. Degrading video quality and introducing audio-video delays was shown to significantly impact performance, but it was the delays in the audio channel that were observed to have the most impact. Consequently, when these dialogue effects were combined with the analysis of task outcomes, it was demonstrated task outcome was unlikely to be effected by the use of a VMC connection. Audio channels could provide equally high quality collaboration, but the pattern of language to achieve the

same results would differ. One of the conclusions that the Anderson et al work suggests is that task oriented video views may have had significantly more impact on their study results, the talking heads model that they employed being of comparatively little benefit.

In another multidimensional study Olson, Olson and Meader (1997) again tested various communication conditions, measuring outcome, satisfaction and process. The results of the study however seem somewhat confused, with face-to-face interaction sometimes being worse and sometimes being better than remote collaboration in terms of outcome success. The results also apparently suggested that there was no advantage to adding remote video to remote audio connections in terms of outcome success and a video channel appeared to have little impact on the structuring of task processes, but the presence of video did impact on user satisfaction.

Williams (1997) expanded the area slightly by demonstrating how the utility of video connections could differ by the level of conflict involved in a task and also discrepancies in the linguistic background of collaborators. The results considered both aspects of visual behaviour and subjective preferences, showing in particular that a loss of visual presence in a connection can make it harder to establish understanding in collaborations with collaborators of differing linguistic backgrounds.

Daly-Jones, Monk and Watts (1998) studied VMC comparing audio-video and audio-alone conditions but eschewing the conventional measures of task outcome, opting instead for measures of conversational fluency and interpersonal awareness. Importantly they included a shared editing tool for the task, and extended the collaboration to consider not just person-to-person communication but pair-to-pair collaboration, wherein there would be discussion both between sites and within sites. The authors argued that video results in more fluent conversation especially when there are more than two people at each end, although this is somewhat obvious given that the video will inevitably support the remote representation of awareness and make it explicit that collaborators at a remote site are talking amongst themselves. In dyadic interactions, it appears that auditory cues suffice, for mediating fluent interactions. Measures of presence and awareness of attentional focus were rated as much higher in the video conditions.

These experimental studies have therefore yielded a variety of often conflicting results that have at times suggested the importance of the video channel to remote collaborations but at other times denied its importance. The results of the studies can however be difficult to compare as they do often engage the users in a variety of different experimental tasks, which potentially utilise very different aspects of interaction. The results do however seem to consistently suggest that in most cases regardless of task, the audio channel is of primary importance to successful synchronous collaboration.

*2.3.2.2 Living with technology*

Another analytical traditional in the study of VMC technologies is very much tied in with the development of the media spaces discussed above. In most instances these heavily pervasive technologies were deployed and evaluated at the site of development. They were playthings in the research labs of those scientists who were constructing them, and as such the longer term situated evaluation has tended towards the ethnographic and more sociological methods of analysis, exploring the theme of the co-evolution of users and technologies over extended deployment (in most cases over several years). This is perhaps in line with the general research aims of these systems as discussed above which were distinctly focused on the development of social networks and a reinvestigation of what it meant to construct a working environment linked through video technologies. Explicit measures of task outcome were therefore at odds with the research goals (Bellotti and Dourish, 1997). Such an approach to evaluation can be seen in the studies presented in Bly, Harrison and Irwin (1993), Adler and Henderson (1994), Harrison et al, (1997), Moore (1997), Mantei et al (1991), Buxton, (1997), Dourish (1993), Dourish and Belloti (1992), and Gaver (1992). In all of these studies there is a desire to report the experiences of working in what is considered a new form of working environment. Other analytical traditions, such as the experimental approach, had presumed a model of VMC where it extended existing working practices, merely facilitating distributed access to current practices, which therefore meant that direct comparisons with other models of communication such as face-to-face or audio only were perfectly acceptable. For the investigators of media spaces, approaching the evaluation from a living with technology perspective however, the media space afforded interactions and working practices which were markedly different from existing models of interaction and were therefore considered to be incomparable. But as suggested previously a research goal such as examining how a media space can foster a sense of co-presence in a distributed environment does not easily lend itself to experimental analysis.

*2.3.2.3 Field studies*

This notion however, of living with a developing technology is rightly critiqued by Sellen and Harper (1997), who demonstrate that the very fact that those investigating the media spaces had a vested interest in the work they were presenting. Being the developers of the technology they obviously had a certain impetus to portray the work from their own lab in a positive light, it is far from objective. But perhaps the most pertinent point made by Sellen and Harper (1997) is the observation that the media spaces were being deployed and evaluated from within tech company research labs. These were largely not systems deployed and evaluated in actual everyday working environments (although the Ontario Telepresence Project stands as an exception and did offer evaluations of deployed technologies in non-research lab contexts, Moore, 1997). Therefore many of the natural tensions and resistances to technological intervention that might otherwise be encountered in a working environment and which might

impinge on the usability and adoption of media spaces was never fully explored. This is perhaps a pertinent reason why such technologies were never widely adopted throughout the corporate world. The study presented in Sellen and Harper (1997) does attempt to redress this imbalance by relocating the site of evaluation of a media space technology in what could be described as a field study, evaluating a media space deployment in a working group outside of a research lab. This field study demonstrated interesting differences between the use of media spaces and more formal video-conferencing rooms which had hitherto not been considered. The study explored the different cultures of practices observed with each VMC environment and observed the organisational tensions which drove adoption and use of these systems.

This move to a more field-study based analytical approach brings with it a greater ecological validity than that observed in the more *lived with technology* studies. There were however early studies in non-media space environments which could also be characterised as field-study approaches to VMC evaluation. The work of Isaacs and Tang (1993, 1994), in particular, demonstrated the ongoing development and evaluation of the DVC prototype as it was used by working groups at SunSoft. Whilst it could be argued that field-study evaluations could be critiqued because of the potential lack of applicability to situations outside of the working context studied, there are benefits to the approach. There are the above mentioned benefits as compared to the lived with technology studies, and in comparison with experimental approaches there are benefits in that such studies have increased ecological validity and recognise the impacts of many different social processes on the users' perceptions of the technology, and potentially also tend to evaluate more realistic working tasks. The work of Isaacs and Tang (see Isaacs and Tang, 1997 for overview) demonstrates a natural understanding of these tensions and successfully combines the tight control of the experimental approach to data collection and analysis with the ecological validity of evaluation in a field study setting, as Sellen (1997, p.100) terms it, using the workplace as a 'living laboratory'.

In an exemplar study, Tang and Isaacs (1993) demonstrated that their DVC prototype did not increase overall levels of interactive communication, but it did impact on the process of communication. They showed that patterns of usage in experimental analysis of actual working teams showed reductions in the numbers of email messages sent, reductions in phone use and a possible reduction in face-to-face meetings. This use of the DVC prototype was however observed to be entirely dependent on the presence of the video channel being accessible. When the DVC was used it was noted that it facilitated interactions more like face-to-face interactions than those observed during use of video-conference room meetings. Interaction through the DVC prototype was observed to be more fluid, with interruptions more common, and a more informal attitude being taken, with participants being more likely to attend to additional tasks such as checking and reading emails. Some of the experimental observations did observe however that high quality audio was far more critical than high quality video, to establishing coherent communications. These forms of experimental findings perhaps

demonstrate more reliable results than the earlier lab-based experimental work, as they are tightly controlled studies, but of actual working technologies being evaluated *in situ* in actual working groups.

### 2.3.2.4 Hybrid approaches

The final analytical approach considered by Sellen (1997) is the hybrid approach which combines psychological and sociological analyses. She includes in this category the conversation analytic techniques of the work of Heath et al (1997) and Gaver et al (1993) on media space environments, which takes as its focus a much more specific behavioural analysis of communication, focusing less on the social world and more on behaviour at a local level in a media space interaction.

The strength of this micro-analytic approach is that common behavioural practices during interaction could be observed and compared with existing understanding and observations of comparative behaviours in other non-technology mediated settings. It is the most detailed analysis method for understanding the process of naturally occurring communication, and through its application to VMC use developed awareness of the processes by which collaborators organised their interactions through a VMC medium and the processes by which they established mutual awareness and negotiated practices of engagement (Heath and Luff, 1991).

It is this hybrid approach which utilised conversation analytic methodologies which was perhaps the first body of work to fully understand the impact of gesture (realised not just through hand gestures, but also through gross postural shifts, head nods etc.) on the accomplishment of grounded understanding, and interaction structuring, in VMC environments. But the work also highlighted an important awareness of the asymmetries in interaction that VMC engendered, which were not otherwise present in other interactional mediums such as face-to-face interaction. These asymmetries it was argued arose because of two key factors, firstly 'recipients having limited and distorted access to the visual conduct of the other' and secondly that 'an individual's limited and distorted access to the other and the other's immediate environment undermines the individual's ability to design and redesign movements such as gestures in order to secure their performative impact' (Heath et al 1997, p.336). This was particularly clearly expressed in the MTV (Multiple Target Video) studies of Gaver et al (1993). Because of their close analytical approach to understanding the mechanisms of interaction they were able to discern and articulate the difficulties that users were encountering in the MTV I and II prototypes. With MTV I patterns of usage demonstrated that in the task focussed interactions the view of the collaborators face was rarely used, being eschewed in favour of more object focussed camera views. However, despite this tighter focus on objects for manipulation it was also apparent that there was a loss of orientational awareness and difficulties for collaborators in tracking trajectories of attention

whilst the participant remote to the site of action switched between multiple views of the task space on one monitor, presumptions about reciprocity of perspectives could not be made which was reflected in extended verbal processes of establishing and re-establishing engagement after views were changed. In MTV II multiple monitors replaced the switching mechanism, and this revealed a wider pattern of camera view usage, and much more frequent 'switching' demonstrating that the physical process of switching views was hugely relevant to the tasks examined but had been made too costly in the earlier prototype. MTV II still suffered limitations however, as analysis of the language used during use indicated that it was still giving rise to difficulties in ascertaining relative mutual perspectives on the task space and it failed to adequately support mutual awareness of gestural actions. These breakdowns in interaction which could be decoded by detailed analysis of the video-footage of use of the VMC tools allowed a much richer understanding to be developed of how interaction was structured during collaboration, an understanding that was potentially unachievable in the more traditional experimental analysis techniques, or the more broadly defined social implications research of other analytical approaches.

### 2.3.3 Conflicts and conclusions for VMC

Some studies have tried to explain how VMC works or is limited in effect by referring to concepts such as 'social presence' (Short et al, 1976), 'Cuelessness' (Rutter and Robinson, 1981) or 'media richness' (Daft and Lengel, 1984). O'Conaill and Whittaker (1997) argue that cuelessness and lack of social presence can be explained by disruptions 'in basic conversational processes' (ibid, p.127) brought on by limitations of technology such as half duplex audio and delays in transmission. They argue that media richness is determined by access to these conversational processes. But a reading of the conflicting findings of the works detailed above would suggest that a simple statement of the efficacy of VMC or an attempt to describe how it works in terms of media richness as a medium for expressing 'basic conversational processes' through a visual medium, is insufficient. Perhaps the most compelling discussions of the efficacy of VMC centre on an understanding of what it is that video is used to communicate. The studies above demonstrate that in many instances when available at low cost, video will be used by collaborators. Subtle social processes will be engaged in and negotiated using visual cues concerning hard to communicate factors such as emotional engagement and attentional focus or relative orientation to task artefacts. Whilst there is a natural preference for the ability to guide actions using these visual cues it is rare that this has a significant impact on collaborative performance. The studies above through and through demonstrate the minimal requirement for any successful remote collaboration in a synchronous task is the provision of high quality audio connections. The lack of efficacy of video for performance outcomes was perhaps most succinctly demonstrated in the Chapanis (1975) work, and it is worth noting that in those studies it was the talking heads model of

VMC that was utilised, sensitivity to subtle social enhancements of visual communication had little bearing on the task at hand. However, if the visual channel had focussed on the task space then maybe the results would be different. The studies presented above suggest a divide in terms of whether VMC is useful based on the task properties engaged in during collaboration. For tasks or interactions primarily social in nature video links need only be of the talking heads kind, but when collaboration is object focussed the video-as-data model of VMC appears to show increased efficacy for a video channel in communication.

The limitations in this use of video-as-data have however already been demonstrated in the works of Gaver et al (1993) and Heath, Luff and Sellen (1997). Primarily the asymmetric access to the video representations and the problems this engenders for supporting awareness of mutual orientation to and mutual interaction with critical aspects of the video data are the key downfalls of the video-as-data model of VMC.

## 2.4 Shared Visual Spaces

Parallel to the work on the development of VMC technologies has been an ongoing investigation into the efficacy of providing shared visual spaces for collaborative tasks.[3] Work by Krauss and Fussell (1990, 1991) concerning the development of mutual knowledge and the construction of shared communicative environments for increasing communicative effectiveness, sought to explore the applications of a developing understanding of the processes of achieving grounded conversation to the design of communications technologies. Through their experimental analyses Krauss and Fussell began to understand how task-focussed language evolved during its interactive use during collaborative tasks. The evolution of referring expressions and the developing awareness of common referents was demonstrably shown to be significantly effected by the resources used to establish communications. If a shared visual environment was enabled it was often observed to be of significant support to the smooth establishment of such critical communicative processes. From the foundations of this work a new research focus was derived that sought to understand how best to construct shared visual environments for collaboration.

Studies such as Fussell, Kraut and Siegel (2000), demonstrated that whilst a shared visual context was important in collaborative tasks, current video-communications technology was potentially inadequate to establish such environments, at least at sufficient fidelity to support interaction to levels observed in face-to-face communication. In a study of interactions concerning remote help in computing tasks, Karsenty (1999), extended this argument by demonstrating that to support any given task it was crucial to determine which features of the

---

[3] Note that this is qualitatively different from VMC, although it of course is concerned with the presentation of visual information it is more akin to the video-as-data approach in VMC and is more concerned with the effects of providing visual access to salient features of collaborative tasks.

visual environment were critical to support. In Karsenty's study so much of the interaction was based on screen focussed activities that a shared representation of a user's VDU screen was sufficient to improve communication beyond that achievable by audio-only means (a feat shown to be un-achievable in other studies, e.g. Chapanis, 1975). In further efforts to understand the science behind how people are supported in collaborative tasks through the use of shared visual spaces Darren Gergle extended the body of work at Carnegie-Melon through several timely studies. For the completion of these studies Gergle developed a puzzle task paradigm (see figure 2.5 below) which required a Helper to guide the actions of a Worker in the assembly of a puzzle piece diagram.
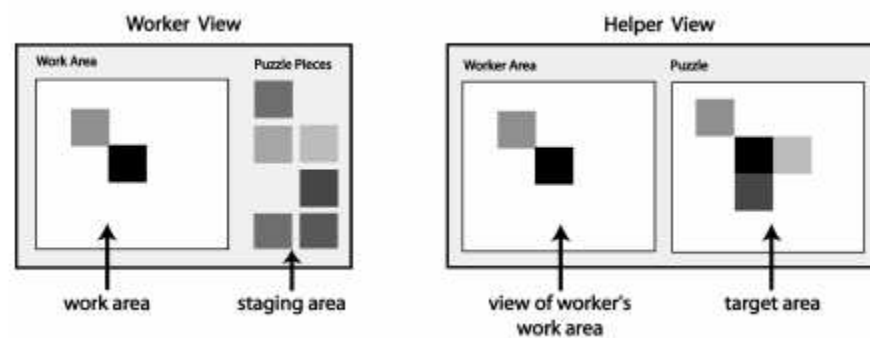


Figure 2.5 The Puzzle task developed by Darren Gergle (from Gergle et al 2006)

A task such as the 'puzzle task' is a form of referential communication task, heavily adopted in various investigations of language use (see Glucksberg et al. 1966 for the first use of this technique). This approach was reportedly taken to allow systematic manipulations to be made to the shared visual environments such that various parameters of their construction could be empirically compared.

In their early work on the subject (Kraut, Gergle, and Fussell, 2002, Gergle, Kraut and Fussell, 2004a) the CMU group demonstrated that the presence of a shared visual space significantly improved performance on the collaborative puzzle task. The presence of delays in the visual feedback received by the Helper and the difficulty in the task they were completing (influenced by how easily shapes in the space connected and whether the colours of the pieces remained consistent or 'drifted') determined success in the task. Delaying the visual update reduced the benefits of the shared visual space and degraded the performance and the shared visual space was shown to be of more use when the shapes were more visually complex. Gergle, Millen, Kraut and Fussell (2004) extended this finding by demonstrating that when the talk in collaborative tasks is mediated by text-based chat (such as Instant Messaging), persistence of the text messages improves task performance but less so than access to a shared visual space. When access to the shared visual space is denied, the role of persistence of text messages becomes even more significant, especially also when objects in the task are hard to describe.

The results suggested further that a shared visual space is the optimum route for efficiently establishing grounded interactions. In an effort to explain this finding later work (Gergle, Kraut and Fussell, 2004b) demonstrated, in a complicated sequential analysis, how visual actions within a shared space can be used to replace elements of dialogue that would be necessary in the absence of visual feedback. In efforts to ground verbal instructions, Helpers require confirmatory feedback that instructions have been understood, carried out and more importantly carried out successfully. In the presence of a shared visual space much of this explicit checking and confirming work (often carried out through direct questioning and back-channelling of semi-verbal responses) is dropped, in favour of a reliance on the visual feedback. Such behaviours conform to the principle of least collaborative effort (Clark and Brennan, 1991).

In the CMU group's most recent work (Gergle, Kraut and Fussell, 2006), studies have been presented which have shown the differential impact on performance of varying levels of delay to visual feedback in shared visual spaces and the influence of the dynamics of the visual environment when interacting with such delays. Put simply the research work demonstrates that serious time delays prevent collaborators from establishing situational awareness of the task, they are not mutually aware of the current state of task artefacts and this inhibits task performance. However, a small amount of visual delay was not problematic. The point at which visual delay did cause a problem was seen to vary as a function of how complex the visual environment was, increasing complexity (generated by dynamically changing the colours of the pieces being manipulated) resulting in increasing delays in feedback affecting performance much sooner.

A significant off-shoot of this shared visual space work can be seen in two papers, Ou, Oh, Yang and Fussell (2005) and Ou, Oh, Fussell, Blum and Yang (2005). These works use Gergle's puzzle task paradigm to analyse the movements of the Helpers' eyes during collaborative tasks. Working to potentially extend the functionality of the DOVE system (Ou et al 2003) by incorporating automatic camera view switching, determined on the basis of parsing Helpers' language use during collaboration, meaning that the system does not need physical manipulation to change camera view by the Helper during use. These studies showed some (limited) support for the notion that patterns of eye-gaze were highly systematic during the puzzle task and could be predicted on the basis of what the Helpers were saying at any given point. Such a finding supports the notion that different aspects of a task are supported by different elements of a shared visual space, which can vary by the dynamic visual environment of the task, but also by the very stage of the interaction that is to be supported. Given the constraints on the usability of multiple task views (see Gaver et al, 1993) and the bandwidth intensive nature of such set-ups it is perhaps an advantage to be able to automate and dynamically present multiple feeds of video information. Such a system could dynamically create a shared visual environment that feeds to a Helper, the optimum visual resources at any given time, reducing the costly need to search between multiple screens and the costs of

supporting such data intensive communication. This is at least the conclusion drawn in the Ou et al studies. However, this largely ignores the complexity of actually parsing spoken language, and brushes over the large amount of inaccuracy that the presented system demonstrated. The technology design also fundamentally assumes that visual saccades and general visual attention follows changes in speech pattern and not the other way around, which unless empirically tested and demonstrably shown to not be an issue of concern is potentially going to significantly hamper use of such a technology.

Despite being an interesting exploration of the ways in which shared visual environments should be constructed, this work on shared visual spaces is, however, fundamentally flawed. Quite acknowledgedly the work takes a reductionist approach to communication, hoping to distil key properties of communicative environments that influence behaviour. The approach taken creates a highly artificial working / communication task, which has significantly little similarity to any current collaborative tasks in which users might wish to engage. A primary point of contention is the use of the term collaborative physical task. Original conceptions of the term (Kraut, Miller and Siegel, 1996) were concerned with tasks which were inherently 3-Dimensional in nature, tasks which resolutely occurred in the real world. This term it appeared was used to differentiate between the types of technology required to support these tasks, with the already researched technology, to support more 2-Dimensional software based tasks. The puzzle task paradigm used is clearly a 2-Dimensional software based task, so not at all similar to the types of tasks referred to previously as collaborative physical tasks. Despite this the results of the studies are discussed in relation to the development of technologies to support such non-software based collaborations. Stepping aside from this issue for a moment, if one takes the studies at face value, the results as presented are also somewhat expected. Findings which demonstrate that visual delay impairs performance, were also predicted by the research literature (e.g. Clark and Brennan, 1991) but are also supported heavily by common sense. Explaining the reasoning behind this may be of interest to some but is fundamentally something which most technology designers would assume as a given, and try to avoid. And this issue of avoiding the problem of visual delay is not actually a significant one anyway, considerable research effort in other fields over many years has lead to the rapid development of increasingly high bandwidth communications technologies, as such problems of visual delay in communications channels are just not a significant issue. Equally, the findings that a more complex visual environment interacts with this problem, are again common sense. The ways in which this complexity was generated for the studies however, has significant lack of validity. In the studies above the puzzle elements being assembled dynamically changed colour during the task, occasionally on a high frequency rotation. What process this represents in the real world is somewhat questionable, physical artefacts for collaboration not normally changing significant visual properties to the extent that it is difficult to describe what they are during use. As such these discussions of the parameters of shared visual environments which

effect performance appear to be devoid of significant implication for the actual deployment of technologies to support collaborative physical tasks.

## 2.5 Collaborative Design

Having previously considered the extensive research into video-based communication and the provision of shared visual spaces it is clearly apparent that there are certain deficiencies in such modes of communication, when they are intended primarily to facilitate the coordination of group working activities. Research activity was expanded from the late 1980's into the mid 1990's to understand how systems could be designed to facilitate synchronicity in actual remotely located group work. One sphere of the working world that appeared to need such technological developments most, was the design world, where increasingly, within large international companies, design experts were required to collaborate despite being based in a variety of diverse company locations. Considering the visual nature of design work, and the importance of collaboration in the creative process, design teams therefore posed a particularly salient focus for CSCW research. In the following sections I will outline the observational work that was conducted to elicit the working practices of co-present design teams and then discuss the technological innovations that were proposed to meet the requirements of remotely located design teams.

### 2.5.1 Observation studies of design teams

Many of the CSCW systems that were created to support collaborative design were based on the work of John Tang, and his observational studies of design teams which formed the basis for his PhD thesis (Tang, 1989), and can be seen written up in several papers (see Tang & Leifer, 1988 and Tang, 1991). Tang's work, which utilised video-based interaction analysis methods, analysed small design teams (3 to 4 co-present designers) as they attempted to complete one of several designs tasks, all of which focussed on the human-machine interface design for an interactive computer-controlled system, whilst using a shared drawing artefact such as a large notepad or white board. The interaction analysis methods used (based on Goodwin, 1981 and Heath, 1986) focus on the analysis of the interactions among participants and the artefacts in their natural working environments. Tang's approach to the research was to analyse the interactions using a predetermined framework of actions and functions. The three actions were Listing, Drawing and Gesturing and the corresponding three functions were Information storage, Idea expression and Interaction mediation.

From his observations Tang noticed several key processes in co-present design activity which have a bearing on the design of collaborative design tools; a) collaborators use hand gestures in a significantly complex system which allows them to encode and convey a variety of different types of information; b) the process of drawing images is often more important than the result,

and conveys meaning in its' very act; c) the drawing space itself, becomes a tool for the mediation of communication and collaboration processes within the group; d) there are a variety of concurrent, different activities that take place within the drawing space and e) the literal spatial layout of the drawing space in relation to the collaborators has a role in structuring their activity.

This seminal work of Tang has been extended by further research, which is reviewed in a paper by Bekker, Olson & Olson (1995). In a series of studies (see Olson, Olson, Carter & Storrøsten, 1992, Olson, Olson, Storrøsten & Carter, 1993, Olson, Olson & Meader, 1995) extensive observational data of design teams was collected. Bekker et al (1995) use data from these studies in an analysis of the role of gestures, specifically to inform the design of groupware systems for designers. Using a coding system derived from the work of Ekman & Friesen (1969) and McNeill (1992), Bekker et al (1995) assigned the gestures they witnessed to 4 categories, Kinetic (related to modelling an action), Spatial (related to an indication of size, distance, location etc.), Point (a form of deixis) and Other (all other gestures not fitting in the above categories). The studies demonstrated that gestures rarely occurred in isolation and were often sequenced into patterns, 4 common patterns were identified. Walkthrough's (sequences of kinetic gestures), List sequences (commonly associated with pointing gestures and similar to written bullet points), Contrast sequences (also associated with pointing, but used to separate speech items conceptually) and Emphasis sequences (largely composed of the Other gestures, where emphasis was needed for a speech item).

Bekker et al. (1995) observed several key characteristics of gesturing in design meetings, which were:

- Many gestures are very brief

- Gestures are often unconsciously synchronised with speech

- Gestures often occur in sequences

- Gesturing is often procedurally linked to activities such as drawing

- Gesturing sometimes occurs whilst the gesturer is mobile and acting through an interaction sequence

- Gestures often have complex 2-D or 3-D trajectories which are important to their meaning

- Gestures are embodied in their spatial environment in relation to other people and artefacts and a knowledge of the spatial environment is often relevant when decoding them

- Gestures sometimes refer to imaginary objects, which can then exist throughout a meeting, and may be referred to and interacted with by third parties

- Gestures can refer to gestures in the past

Having observed and acknowledged the prevalence of gesturing in design meetings, Bekker et al (1995) go on to discuss the implications of this for the construction of systems to support designers. They consider several different forms of technical support for design meetings. The first being electronic device support for meetings, in which participants are co-present. Bekker et al argue that when designers must use their own interface to view a shared object many of the critical social processes of gesturing are impeded (for further discussion of this issue see Tatar, Foster and Bobrow, 1991). If a designer wished to point at something on the design their hand gesture would be visible to only themselves, to counter this many of the available systems have tele-pointing capacity (see Hayne, Pendergast and Greenberg 1993 for a brief review of such systems), however Bekker et al, argue that this is a weak form of gesturing as many of the kinetic and spatial movements possible with hands are not possible with a tele-pointer. To counter these limitations Bekker et al suggest the use of collaborative electronically supported public displays such as electronic whiteboards, which add computer support to the design process but do not impede the benefits of co-present interaction.

In remote design sessions where participants are not co-present video-conferencing is sometimes used. Bekker et al argue that this is difficult because of the loss of spatially relevant information between participants, but they argue that virtual reality techniques perhaps stand to alleviate such problems by reintroducing spatial relationships to remote meetings. Bekker et al however are unclear as to the specifics of how virtual reality technology might affect such spatially significant activities as gesturing. Later work by Fraser (1998) however has extensively considered this issue.

Clearly Bekker et al feel that gesturing is of vital importance in collaborative work, which they take as a given fact considering their evidence of its prevalence in design meetings. They argue that for any groupware system to be adopted successfully by design teams it must suit the way they work and consequently support the adequate transmission of gestural information.

### 2.5.2 Commune: A shared drawing surface

One early system which was developed in an effort to support such gestural activity in collaborative design work, when collaborators are remote from one another, was the Commune system (Bly and Minneman, 1990 and Minneman and Bly, 1991). Commune (see figure 2.6 below) was based on the understanding (derived from Tang, 1989 and Bly 1988) that the process of creating, referring to and using drawings was as important to the design process as the resultant images themselves. The system was therefore built to provide designers with access to a shared drawing space, utilising the metaphor of a drawing pad. Each collaborator had a stylus which could be used for cursor-based gesturing or for making pen-style marks on

the shared surface, natural verbal interaction was maintained through the use of telephone links. This approach was shown to be of benefit to collaborators in design meetings, effectively facilitating some of their primary requirements in collaboration. Bly and Minneman noted that even such a relatively simple system allowed the fluid interweaving of gesture, talk and drawing interactions. Problems observed with use of the system however, centred on the use of such a simplified tool (i.e. a cursor) as the primary medium for gesturing. Cursors, it was reported, were unable to represent the complexity of gesturing behaviour observed with hands and fingers. Equally it was not always possible to disambiguate between incidental movements of the cursor and actual intended gestures, and perhaps for these reasons, in several instances naturally occurring hand-based gestures were used, despite the fact that such behaviours could not be transmitted to the collaborating parties.
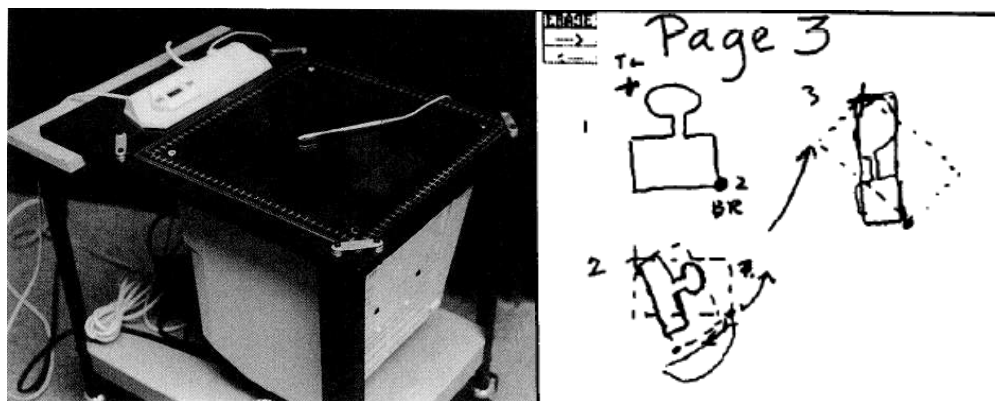


Figure 2.6 Commune Drawing surface from Bly and Minneman (1990) (left – equipment, right – resultant sketch appearing on surface)

Initial instantiations of Commune were improved by increasing the possible number of collaborators from two to three users (Minneman and Bly, 1991). It was expected that such an extension would reveal new interaction problems, given that little was understood about the differences between triadic and dyadic collaborations. These worries were however, unfounded, as there were no observed problems with extending the range of users, all collaborators being able to easily identify who was sketching or gesturing at any given moment (cursors and lines were of course different colours for each participant and most drawing space activity was coordinated with concurrent language). An interesting observation of Commune use, centres on the inclusion of face-view video links between the remote sites. Although observably not used directly for the task, there was anecdotal evidence that the presence of video links actually improved engagement with the task and the collaborative action. When video presence was not enabled it appeared that collaborators felt increasingly able to move within themselves and to not actively participate and interact with the other collaborators.

**2.5.3 VideoDraw: A video interface for collaborative drawing**

In concurrent research also being conducted at the Xerox Palo Alto Research Center (along with the Commune project), during the early 1990's, the VideoDraw system was developed. VideoDraw (Tang and Minneman, 1990, 1991a) grew directly out of John Tang's thesis work and took an alternative approach to supporting design activity to that of the Commune project. Working exclusively in video collaboration, VideoDraw sought to create a shared drawing surface that allowed the remote representation of not just sketched images but also the hands and arms of the sketcher as they were producing the drawings (see figure 2.7 below).
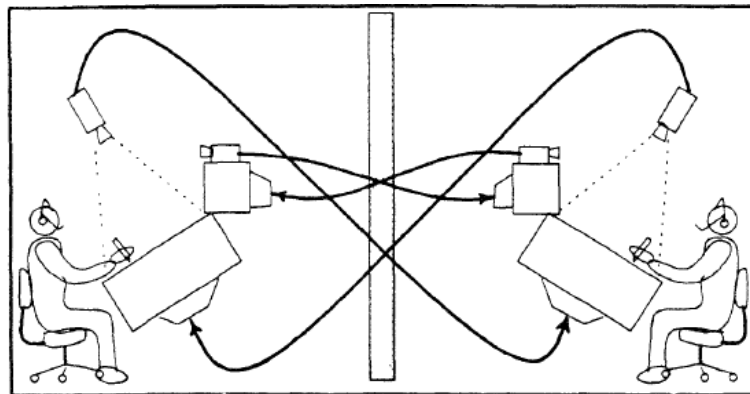


Figure 2.7 Schematic of VideoDraw system from Tang and Minneman (1991a)

By allowing collaborators to view a live video feeds of one another's workstations and consequently draw over those video images (these resultant sketches in turn then being captured and passed back to the linked workstation) collaborators could not only produce and share drawings but also collaboratively construct them. The communication environment was made all the richer for the ability to use naturally occurring forms of hand-based and pen-based gestural behaviour. This approach conveyed most of the benefits of a system such as Commune but improved on the paucity of the gesturing medium achieved in that system. Problems did however occur with use of the system. The relative thickness of pens and small size of the screens used meant that the drawing space was rapidly filled and previous content had to be repeatedly removed. The removal process was hampered by the uni-directional access that collaborators had to the shared sketches, each collaborator could only remove or indeed really interact with, the elements of the shared sketch that they themselves had produced. Coupled with this access issue is the fact that at no point was any computing technology involved, so many useful features of computer-aided design, such as the ability to save images or open and include designs from existing files, were not available, limiting the scope for use of such a system.

### 2.5.4 TeamWorkStation: Towards a seamless shared workspace

Extending the work of VideoDraw was a Japanese system for collaboration known as TeamWorkStation (Ishii, 1990, Ishii and Miyake, 1991). Interested in developing technologies for collaboration which would situate themselves comfortably within existing working practices, Ishii, sought to explore how technology could be designed to negotiate the cognitive seams that highlighted separations between private and shared objects and tools. Ishii based elements of the design of TeamWorkStation on the principles espoused by Grudin (1988), with his belief that if users were forced to utilise unfamiliar tools to access technologies then those technologies would never be successfully adopted. To this extent TeamWorkStation was built as a tool to facilitate group interaction and collaboration, as and when necessary, which could allow people to engage in ad-hoc collaborative design work whilst retaining use of their favourite tools for design, be they computer software based, or paper based. TeamWorkStation is essentially a bricolage of technologies, in which users have there own private PC monitor for digital content but also a second monitor, seamlessly linked to the first, which is a shared space for all collaborators. Content can be dragged and dropped directly from private space to the public space. The public space also supported face-view video feeds of all the current collaborators and contained a facility to present images from a desktop camera (held over a sketch pad) on each desk. This video feed could then be overlaid in the shared space with others' video feeds or images of digital content or applications opened by other collaborators in the shared space (figure 2.8 below shows some examples of TeamWorkStation).
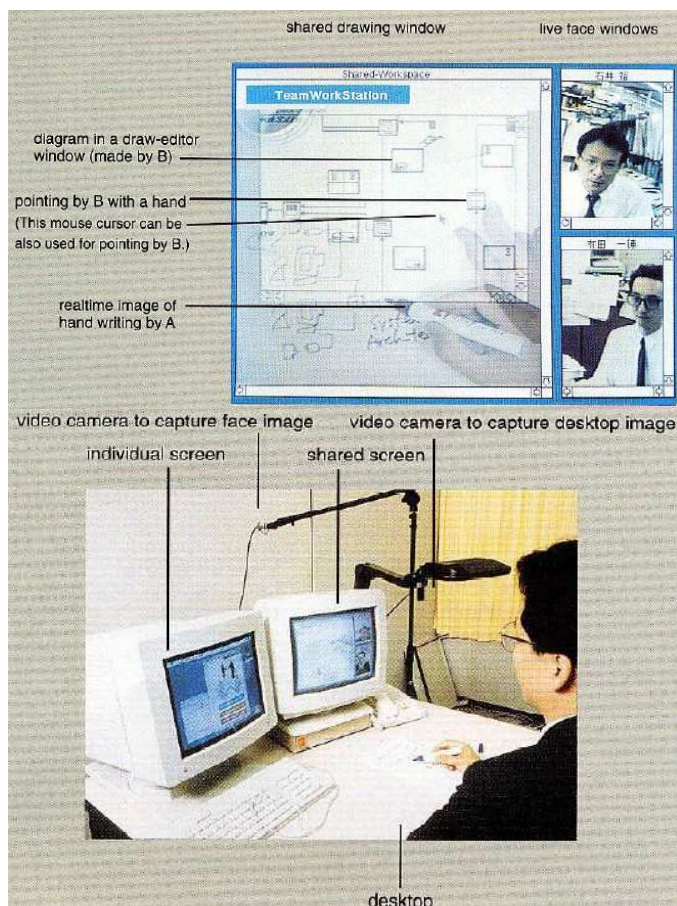
Figure 2.8 The TeamWorkStation of Ishii and Miyake (1991)

TeamWorkStation clearly is an advancement to the VideoDraw system in that it retains all of the function of that system, yet situates it within a more realistic collaboration environment (i.e. it is held alongside existing desktop working arrangements, rather than being a stand-alone unit for collaboration) but extends the functionality to incorporate digital content as a shareable media, that can actually be, in some limited form at least, integrated with non-digital content. Evaluations of the TeamWorkStation (Ishii and Miyake, 1991) have however, highlighted certain limitations. Despite the ability to record and store the resultant shared images that can be created, as images are produced there is not equal access to the information. Similar to VideoDraw elements that are collaboratively produced are held as layers in a collaborative construction, with individuals only having access to manipulate those aspects that they themselves produced, this is not an optimum arrangement. Equally evaluations reported difficulties engendered by the poor quality of the video links and the fact that gesturing or sketching behaviour when performed collaboratively had to be coordinated by watching feedback of sketching actions on a video monitor rather than at the actual site of sketch production (as per VideoDraw).

**2.5.5 VideoWhiteboard: Video shadows to support remote collaboration**

In an effort to improve on some of the observed problems with the VideoDraw system, namely the discomfort from use, the issues of parallax of drawn images making their alignment difficult and the fact that the screen was too small, Tang and Minneman (1991b) built and explored use of the VideoWhiteboard system.
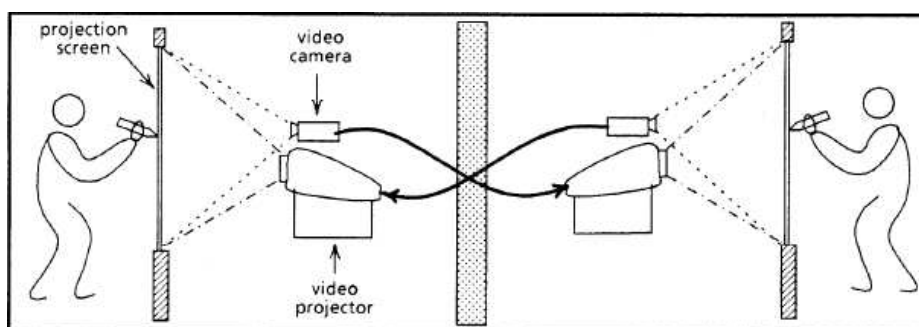


Figure 2.9 Schematic of VideoWhiteboard from Tang and Minneman (1991b)

VideoWhiteboard (as shown above in figure 2.9) is a system very similar in principle to VideoDraw, which provides collaborators with a shared design space in which they can equally interact, utilising the back projection and video capture of interactions at a large whiteboard surface. The result is a faithful presentation of collaborative sketches on to each collaborator's whiteboard but also the inclusion of shadowy representations of each collaborator's torso. VideoWhiteboard offered a larger collaboration space, allowing not only more work to be done, but also more collaborators to engage in the space. The colour representation of gestures and drawn images was however significantly reduced in quality compared to VideoDraw. But the limitations of this reduced quality of information, and the difficulties therefore engendered in disambiguating between images of different collaborators and abilities to represent gestures in 3-dimensions, were considered by Tang and Minneman to be outweighed by the advantages the system conferred. Additional advantages included a reduction of the parallax problem, no head blocking of projected images was possible and gestures presented were not restricted to the hands, as whole body non-verbal behaviours could be transmitted. An interesting phenomenon that was observed during use of VideoWhiteboard plays on this ability to provide full-body non-verbal behaviours. The system purportedly increased the collaborators' sense of having another presence in the room. In one cited example when one collaborator could not hear another, rather than moving closer to the off-set speaker system, she put her hand to her ear and moved closer to the video shadow of her colleague. The shadowy representation of this full-body gesture presented at the other side of the collaboration was enough to allow her colleague to understand that he must repeat what he had said, but louder. Clearly VideoWhiteboard was able to facilitate high levels of non-verbal

interaction, which are highly naturalistic, and potentially therefore reduce the requirements to negotiate activities through artificial verbal means.

### 2.5.6 Clearboard: A seamless medium for shared drawing and conversation with eye contact

This issue of adequately representing key non-verbal behaviours was one of the major driving forces behind the next generation of collaborative design tools. Additional functionality provided to the TeamWorkStation system, namely the ClearFace extension (Ishii and Arita, 1991), demonstrated that most remote design tools created an artificial seam between the shared work space and the interpersonal spaces used for communication (Ishii and Kobayashi, 1992). Noticing that in co-present design interactions there is a seamless movement between awareness of the non-verbal (often facial) gestures of collaborators and interaction at the drawing surface, Ishii developed the ClearBoard system (Ishii and Kobayashi, 1992, Ishii, Kobayashi and Grudin, 1993, Ishii, Kobayashi and Arita, 1994). ClearBoard was developed using a new metaphor for collaboration. Whereas other systems had employed literal translations of existing design practices by supporting either tabletop or whiteboard style interfaces, ClearBoard utilised a "through a glass window" metaphor (Ishii and Kobayashi, 1992, p.527), which saw full colour representations of collaborators working on a shared video surface, presented as if collaborators were sat on either side of a pane of glass (see figure 2.10 below for system architecture and illustrative image).



Figure 2.10 ClearBoard in use from Ishii, Kobayashi and Grudin, (1993)

A key functionality of the ClearBoard approach was its ability to support gaze awareness during collaborative design tasks. Head and eye movements of each collaborator could be monitored by their partner, and fine-grained judgements could be made about exactly where in the shared space the other was focusing. Early iterations of the system (ClearBoard-0, ClearBoard-1) however suffered from the difficulties first demonstrated in the VideoDraw system (Tang and Minneman, 1990) of sketching over a purely video based medium, with

none of the inherent functionality of computer-aided design tools being present. A later iteration of ClearBoard (ClearBoard-2, Ishii, Kobayashi and Grudin, 1993) incorporated a groupware painting package as the drawing/sketching tool (in conjunction with a digital stylus for enacting the sketching). By incorporating digital content, limitations of systems such as VideoWhiteboard (Tang and Minneman, 1991b), were overcome. One specific limitation that had plagued most of the systems discussed above was the layering of information and the unequal access to the shared images. Whereas previous uses of video sketching had meant that collaborators were only really free to modify content that they themselves had produced, the introduction of a groupware tool meant that each party had equal access to the shared constructions, a vast improvement on earlier designs. In overcoming these technological limitations however and in efforts to make the system easier to use, it would appear that the early desires of Ishii to provide seamless interaction have perhaps been less successful than he would argue. In Ishii, Kobayashi and Arita (1994) two facets of seamlessness are highlighted. One is the smooth transition between functional spaces e.g. switching effortlessly between awareness of non-verbal interactions and task-focussed design activities, this is well supported by ClearBoard. The other facet however, is the continuity with existing working practices. Whilst TeamWorkStation potentially achieved this with its bricolage of technologies situated at a normal working desk, ClearBoard perhaps does not. ClearBoard has become a standalone workstation with its own practices of use and its own protocols for interaction. The immense size of the set-up and the fact that it does effectively limit comfortable use to two users means that those problems discussed by Grudin (1988) and which Ishii sought to avoid may in fact limit the technology's adoption.

### 2.5.7 The DigitalDesk

The bricolage approach is returned to and explored further in the DigitalDesk prototype by Wellner (1993). In essence trying to create a device that possesses this notion of seamless integration with existing work practices Wellner demonstrated that computing technology could be used to enhance (after a ubiquitous computing fashion cf. Weiser 1991) interactions with current desk objects (such as paper). Applications supported by the DigitalDesk included the 'DoubleDigitalDesk' which enabled an image of a collaborator's desk to be projected onto a space on one's own desk, supporting the creation of shared designs and the sharing of gestures (focussed on and around the shared images of primarily paper documents). This collaboration scheme rejected the notion espoused in the TeamWorkStation studies (Ishii and Miyake, 1991) that views of non-task-focussed non-verbal behaviour, such as face images were critical to task-oriented communications. The approach essentially paired down those elements of communicative behaviour that were critical to the completion of task-relevant activities. Again however, from a collaborative design perspective the system is somewhat limited by its restriction to use by two collaborators, and was never a fully realised system.

### 2.5.8 VideoArms, Digital Arm Shadows and Mixed presence Groupware

Some of the latest contributions to this discussion of digital support for collaborative design have specifically sought to address the issue of engaging multiple users in the design activity. In particular Tony Tang's work on Mixed Presence Groupware (Tang, Neustaedter and Greenberg, 2004, and Tang, Boyle and Greenberg, 2004) has focused on devices which simultaneously support both co-located group activity and remotely located group activity. Heavily utilising ideas generated from the studies discussed above Tang et al have demonstrated how multiple collaboration surfaces including tabletop displays and digital whiteboards can be connected and augmented with embodied representations of collaborators arms. This is done to enhance feelings of co-presence and to replicate naturally occurring gestural behaviours. The efficacy of this approach has been discussed (ibid) in relation to two instantiations of mixed presence groupware, namely VideoArms (see figure 2.11 below) and Digital Arm Shadows (see figure 2.12 below).
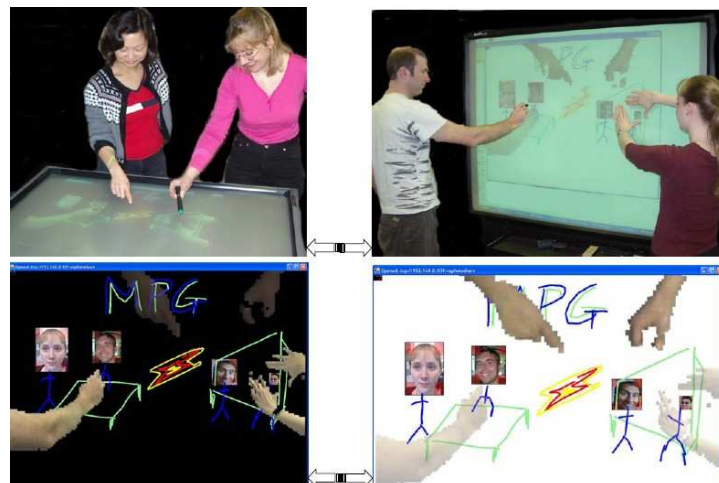


Figure 2.11 VideoArms from Tang, Neustaedter and Greenberg, (2004)



Figure 2.12 Digital Arm Shadows from Tang, Boyle and Greenberg, (2004)

Whilst the VideoArms prototype did manage to utilise full colour full detail representations of arms, the resulting images were clearly not very well constructed, although Tang et al admit that not being computer vision researchers, maximum fidelity of representation was not their aim. Where previously Tang, Neustaedter and Greenberg, (2004) had argued that mixed presence groupware embodiments have requirements such as,

> "the remote embodiment of this input (the arm) is presented at sufficient fidelity to allow collaborators to easily interpret all current actions as well as the actions leading up to them" (p.2)

and,

> "To support bodily gestures, remote embodiments should capture and display the fine-grained movements and postures of collaborators." (p.2)

Later mixed presence groupware systems that they present, utilise much lower fidelity 'Digital Arm Shadows', as the primary gesturing media. These shadow representations are highly stylised and would presumably offer extremely little in the way of communicative content. Despite the Tang et al work being an admirable attempt to explore how to extend previous collaborative design systems to facilitate multiple party interactions there are a variety of limitations inherent in the work. Firstly there is no real motivation for the argument that contrasting display formats need to be linked. There is no sense given that this is a common desire, but is presented in the papers (ibid) as a motivating factor for the research. A distinctly unfortunate side-effect of this desire to link diverging interaction surfaces is the requirement that participants at the tabletop interface cannot interact with their surface on all four sides (which is presumably one of the keys reasons for choosing to use a tabletop display) as this would doubly overlay embodiment information being presented with those using the whiteboard surface (who are physically forced to keep a North-Up orientation to their interface). A resultant effect of this is that those using the tabletop display are forced to collaborate with a sketch surface which is effectively upside down for them. Any attempts by either party to provide written input will be unintelligible to their collaborators, and for someone any shared sketch will always be in the wrong orientation. The work of Stacey Scott (2005) suggests that such orientational issues can be fundamentally important to the usability of tabletop interfaces.

### 2.5.9 Conclusions from collaborative design technologies

The attempts to support collaborative design have arguably demonstrated the importance of the representation of presence and object-oriented interactions. The representation of remote gesture in particular has been repeatedly suggested to be of importance for smooth interaction, and to a large extent the general representation of bodily orientations such as mutual gaze has been advocated. In all of the studies discussed above, the evaluations have however, been

relatively limited in scope and, potentially, reliability. Often claims were made in the studies about the importance of various features of interaction, such as gestures or gaze or the ability to share digital content, but these assumptions whilst purportedly based on observations of use are less than rigorous in their presentation. Hard and clear answers as to which elements of communicative behaviour are truly integral to the task and worthy of support in distance collaboration are less than forth coming. Such answers could presumably only really be provided by further empirical analysis, which made some attempt to understand the efficacy of differing approaches. These anecdotal investigations of use do however provide a good coverage of alternative approaches which could be investigated when trying to design optimal solutions for supporting collaborative interaction.

## 2.6 Designing Remote Gesture Tools for Collaborative Physical Tasks

Having considered how remote gestural representations became an integral aspect of collaborative design technology it is pertinent to begin to consider the developments that have been made in attempts to support collaborative physical tasks with remote gestural simulations. There are two main strands to this research, the development of the GestureMan system by Hideaki Kuzuoka, which has also spawned the WACL system of the University of Washington, and the DOVE system of Ou et al (2003) at Carnegie-Melon University. Each of these systems is considered in turn, the section finishing with a discussion of the potential role of Augmented Reality and Tangible user interfaces as an additional route to remote gesture tools design, although no systems in current development realistically deploy these tools in support of collaborative physical tasks.

## 2.6.1 Developing the GestureMan

The initial explorations of advanced configurations for communication for Kuzuoka and colleagues came with the investigations of the SharedView system (Ishii et al 1990, Kuzuoka 1992, Kuzuoka and Shoji, 1994). In the 1992 CHI paper Kuzuoka introduced to the HCI community the notion of designing tools and deriving requirements for supporting collaboration in 3-Dimensional space, a form of interaction which he referred to as 'spatial workspace collaboration'. The initial motivations for SharedView arose from a desire to support collaboration in active manufacturing contexts, in particular in situations of remote instruction. Observing key practices of face-to-face interaction in such contexts led to the realisation that there were a variety of common physical interactions, such as pointing, which would not be well supported by simple video communication links. Extrapolating to a more abstract task, to facilitate easy experimental analysis, which utilised collaborative placement of 3-Dimensional objects, Kuzuoka demonstrated that remote gesturing (in a video-draw style set-up) could lead to empirically measured faster performance. The work also demonstrated

some of the earliest observations that the very structure and content of collaborative language could be manipulated by the presence of remote deictic technologies. The observations from these experiments led to the presentation of several 'communication system requirements' (ibid, p.537):

- Variability of focal point to optimally accommodate viewing intentions.

- Ability to share a focal point; thereby minimizing differences in directional expressions.

- Capability to use superimposed gestures.

- Since the focal point should be variable, the operator's display showing applicable instructions should also be variable.

- Possess the ability to confirm an operator's comprehension and the object's actual manipulation.

To support these system requirements the SharedView System was constructed (see figure 2.13 below).
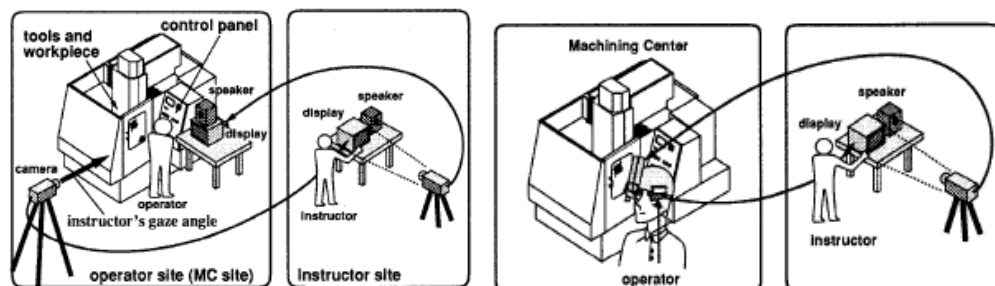


Figure 2.13 SharedView system from Kuzuoka (1992)

This system was observably of benefit to remote instructors and ensuing discussion of its use argued that the mobility of the system was its key strength, allowing interactants to successfully change focal point of discussion at several key points during their interaction, and demonstrating extreme ease in initial set-up for collaboration.

SharedView was however critiqued in later work. Kuzuoka et al (1994) discuss how longer-term use of the SharedView system proved impracticable. They discuss claims made by users that the system offered too narrow a view of the collaborative workspace, which created difficulties in negotiating a natural work flow. This was further compounded by the Instructors restriction to view only what the Operator wished them to see (because of the Helmet mounted camera). Without an independently operable camera it was observed that Instructors would become frustrated as they had to stop their discourse flow to elaborately re-orient the focus of

attention of the operator. To counter these problems a new system was designed, called GestureCam (ibid, see figure 2.14 below).
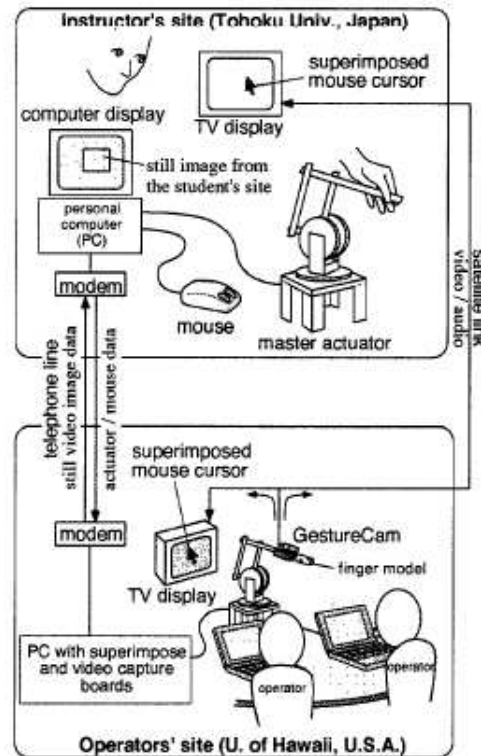


Figure 2.14 Schematic of GestureCam System from Kuzuoka et al (1994)

In GestureCam, a master-slave system was incorporated that allowed a remote instructor to manipulate an otherwise static camera in the operator's workspace. Attached to this camera was a laser pointer to aid the projection of pointing behaviours (and a finger to supposedly aid the generation of sympathetic feeling toward the robot arm - to make it seem more lifelike). Observational analysis of the use of the system demonstrated that its success was largely hindered by the poor quality of the video link between the spaces used for the collaborative exercises. The primitive communication links are a product of the capacity of the satellite technologies used in 1994, and would obviously be overcome in more recent attempts to replicate such a system. Regardless of this however, the study of GestureCam demonstrated that a significant problem in use was the decision that was necessarily made between using a broader context view of a workspace and a narrow focused view of specific artefacts for manipulation, an issue already raised in Gaver (1992). A benefit of the system however, was an observation of its epiphenomenal uses such as the embodiment of physical presence of the instructor. Kuzuoka et al draw specific reference to the acts of one instructor when 'bowing' with the camera in greeting. The reports of the ease with which such embodiments can be performed and the extent to which they can help smooth complex inter-cultural social

interactions highlight a turning point in the thoughts of Kuzuoka towards moving beyond mere representation of communicative content toward the use of technology to construct literal embodiments of remote collaborators, a theme which would recur heavily in later work.

In the 1995 ECSCW paper (Kuzuoka et al 1995), this notion of the GestureCam as a surrogate for the remote instructor is increasingly investigated. Drawing on observations of the previously highlighted problems of video communications systems, (i.e. the problem of static cameras (Gaver 1992), the problem of establishing gaze awareness and the problem of supporting remote pointing) Kuzuoka et al explored the use of GestureCam as a means for alleviating these issues. These experiments demonstrated as predicted that GestureCam could adequately support remote pointing behaviours and could be used by the collaborator who was interacting with it to infer the gaze direction (to some extent) of the remote instructor. There were however, reported difficulties with the use of the system. The master actuator was reportedly difficult to use for fine pointing tasks, so direct hand gestures in conjunction with the touch-sensitive CRT were preferred, but overuse of this element of the system often lead to the users losing sense of where the camera was focussed and having difficulty in repositioning it to focus on different areas of the task space. Equally, sketches overlaid on the video image were often presented in a different perspective to the task space than that enjoyed by the operator (the person being instructed), this difficulty in resolving relative gestural orientations causing some significant difficulty in interpreting those gestures. In those situations where the operator had no monitor view of the instructors view, they were more inclined to look at the GestureCam itself and interact with it as a surrogate. When a monitor view was available however, the users appeared to prefer to watch this view, suggesting that they desired to have an explicit understanding of exactly what the instructor was looking at. The ultimate conclusions of the study highlighted the necessity for easy and natural gestural production for the instructor and the importance of using the technology to support gaze awareness, such that the operator could infer the relative attentional perspective of the instructor.

Further to this a study by Kato et al (1997) explored the set-up of communications technologies such that they directly followed a body metaphor. Arguing against the traditional approach of linking remote spaces by using a face-to-face video communication set-up the study authors suggested that by placing monitors offering views of different elements of a remote instructor (i.e. face views and hand-task artefact gesture views) on separate monitors a more natural orientation to the external collaborator could be established. The study demonstrated some benefits of this approach but clearly illustrated the existing problems of collaborators orienting themselves towards multiple monitors and sources of communication (as discussed in Gaver et al 1993). Moving beyond this issue of multiple monitors led to the construction of the Agora system (Yamashita et al 1999, Kuzuoka et al 1999, see figure 2.15 below).
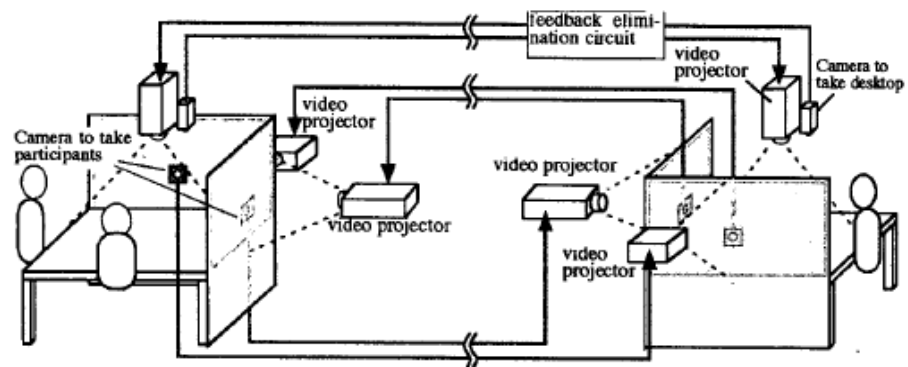
Figure 2.15 Schematic of Agora System from Kuzuoka et al (1999)

The Agora system provided further tabletop interaction but was based on a conception of significantly more equal involvement in a task and moved away from the instructor / operator paradigm. Its use of natural gestural forms was praised by the authors and discussed as a distinct advantage along with its ability to establish interactions between multiple collaborating parties. This use of actual physical presence and embodiment through video representations was not maintained however as the work of Kuzuoka swung further towards the use of physical and digital surrogates to represent remote behaviours (several examples of technologies working as 'digital but physical surrogates' can be seen in Kuzuoka and Greenberg 1999). The systems of the GestureCam were returned to in later research with the presentation of the GestureLaser and GestureLaser Car systems (Yamazaki et al 1999, see below in figure 2.16).
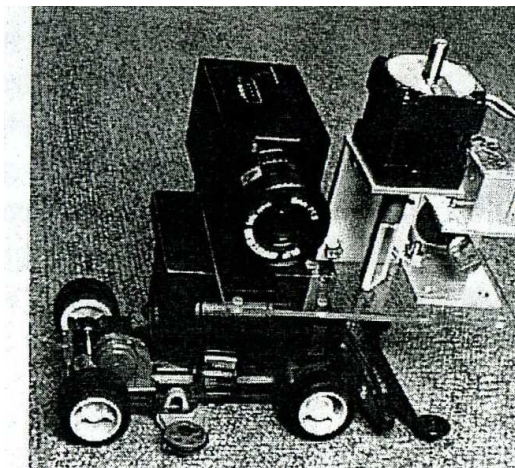


Figure 2.16 GestureLaser mounted on GestureLaser Car from Yamazaki et al (1999)

The GestureLaser systems extend the GestureCam systems by focussing more on the utility of the Laser pointers employed, making them function independently of the camera, such that

camera view is no longer explicitly tied to gestural action, theoretically allowing increased range of viewing and gesturing function for the user. The addition of a small remotely controlled car (operating on a fixed length track) in the GestureLaser Car system further increases the range of action achievable by the system, effectively allowing the laser pointer and camera to be moved so that obstructions can be seen around and pointed behind.

From an analysis of collaborative work studies (e.g. Goodwin, 1994, Heritage, 1997, and Heath, 1997) and on the basis of reflections on the use of the previously discussed systems, Yamazaki et al derived four key features for consideration for the design of remote instruction systems. The first requirement is that any system should represent adequately the relative orientations of collaborators to one another and the relevant task artefacts. The second requirement was that systems should adequately express gestural communications. The third requirement is that systems should facilitate adequate representation of the sequential flow of interaction and task-focused action such that interactions could be sensibly constructed and interpreted by the collaborators. The fourth and final system requirement was that it should allow for the interaction of multiple parties (i.e. more than two collaborators). This final point in particular seems largely unfounded in that no clear argument is really given in Yamazaki et al (1999) for why one should necessarily always strive to support such large scale interactions. In the majority of the situations studied by the Research Group of Technology and Interaction, there is a strong instructional paradigm which does not necessitate multiple party interactions. Regardless of which a strong element of the Yamazaki et al study is the explication of the notion of 'embodied spaces' or "virtual environments that can embody participants' behaviour" (ibid, p.256, see also Kuzuoka et al 2001). In this they are returning to their focus on the use of technological mediation to represent the actions of a remote collaborator with sufficient fidelity to make it as though the remote participant were almost co-present.

The study neatly discusses some of the pitfalls of using other gestural representations (such as projection) in acutely real-world (non-tabletop) interactions, given their difficulties in working with deformed surfaces and variable lighting conditions, but the authors are forced to accept that the use of laser pointers is a restricted means through which to express gestural action, and carries with it certain safety implications (lasers are after all a dangerous tool to be pointing at people). Additionally, the work overplays the potential benefits of laser pointers in establishing multiple user interactions, as it focuses solely on their being one remote gesturer. Obviously if there were multiple remote gesturers, as would be afforded by a more Agora style system, then multiple laser pointers would become increasingly difficult to differentiate.

It is perhaps through this analysis of the GestureLaser system, that the Research Group of Technology and Social Interaction became so focussed on the nature of embodying the gestural action in a remote space, and developing 'instructor' surrogates as a solution to facilitating smooth interactions during remote collaborations, situated within 3-Dimensional contexts. The benefits derived from the separation of camera and laser and the increased

mobility of the system leading to the final incarnation of the 'laser-pointer' systems in the realisation of the GestureMan system (Kuzuoka et al 2000). GestureMan is illustrated below in figure 2.17:



Figure 2.17 GestureMan from Kuzuoka et al (2004)

The GestureMan system, as discussed by Kuzuoka et al (2000), is a robot based tool for human-human communication. The camera and laser components of earlier systems (Yamazaki et al, 1999) are simply mounted on a robotic body which is remotely controlled by a remote instructor. The Kuzuoka et al (2000) paper discusses the construction and design of the system and elaborates on some initial explorations of its use, highlighting its general utility but also observing some problems that were encountered with the robot's ability to adequately represent the remote instructor's focus of attention.

The concerns highlighted above were significantly elaborated on and discussed in later works (e.g. Heath et al, 2001 and Luff et al, 2003). Drawing inspiration from field studies of communication such as Robertson (1997), Latour (1992), Goodwin (1995) and Hutchins (1995), and their observations of the extent to which communicative acts are embodied and embedded within a specific ecology, these evaluative studies sought to explore the embodied nature of communicative action, when communication was channelled through a mediating robot.

Heath et al (2001) observed that:

> "Action is *transposed* and embedded within the immediate environment; the participant's talk and gestures, their interaction and collaboration are inseparable from particular objects and artefacts, and the ways in which they, at some particular moment, are constituted as relevant. The reflexive relationship between action and the environment is a critical feature of the participants' conduct and collaboration." (p. 32)

And therefore concluding that:

"The embeddedness of action in the environment allows participants to discover why and what others are doing." (p. 34)

Taking these observations as the basis for developing their perspectives on the ensuing analysis, the studies (Heath et al, 2001 and Luff et al, 2003) evaluated how collaborators achieved certain key processes of communication during a collaborative task, using the GestureMan as a communication tool. The key processes that they observed were, how an instructor located an object for another, how they secured a common orientation to an object and how the person situated with the robot made sense of disembodied gestures. In the Heath et al (2001) paper, these observations were also then compared with observations of what were termed 'everyday interactions' taken from other contexts, used to highlight further the importance of understanding context to disambiguate and interpret collaborators communicative acts. The analysis of these vignettes of interaction led Heath et al (2001) to claim:

"However, once we begin to create new environments to enable people to interact and collaborate with each other, we fracture the relationship between action and the relevant environment, and thereby engender difficulties, which may render even the most seemingly simple form of activity problematic." (p. 34)

With specific reference to the gesturing capacity of GestureMan (which is presumably, after all the previous research, its most significant feature, and its reason for being used), Heath et al (2001) also state that:

"Despite efforts to provide 'common spaces', 'symmetric environments' or resources for pointing and reference, these technologies can be seen to inadvertently fracture the relationship between conduct and the environment in which it is produced and understood." (p. 27)

The analysis of the use of GestureMan had demonstrated that conduct became disembodied. A simple understanding of key aspects of interaction in an object-focussed collaboration, such as orientation and reference to the task artefacts and one's collaborator, was being in some way fractured. It is this concept of the 'fracturing of ecologies' which is the most significant contribution of these works (Heath et al, 2001 and Luff et al, 2003). The idea of the fractured ecology stems from this belief (as demonstrated above) that all communicative acts are constructed (when face-to-face) in a shared ecological context. How one constructs, delivers, understands, shows awareness of and responds to communicative acts is based on shared and equal access to a physical space (or ecology). When a mediating communicative tool is inserted into the loop, there is less of a shared environment. The easy access to certain features of the communicative cycle such as demonstrating understanding through back-channels and visible 'correct' orientations to items in focus can become fractured. This leads to the suggestion that technology design should support the transmission of elements of

communicative behaviour that might otherwise become fractured and therefore require increasing verbal effort to support. As Luff et al (2003) state:

> "The problem as we have demonstrated is not simply how we can detect and identify particular objects, but rather how they can establish and maintain a relevant connection or relation between the co-participant (even an avatar) and the environment in which that person (or representation) is located." (p.81)

An extension to the GestureMan analysis, as discussed in Kuzuoka et al (2004a, 2004b), discussed the notion of 'Dual Ecologies', making an explicit division between the remote and local sites. In this analysis they demonstrated how modifications to the design of the GestureMan, such as the provision of a pointing arm and the addition of more face like features on the robot (see Kuzuoka et al 2003 for further description) enhanced the interpretability and projectability of intention between the two sites. The studies in the paper also demonstrated that by modifying the experience of a remote instructor by reducing the field of view of the distant task space (by reducing their number and breadth of view of monitors) they could force the remote instructor to make more explicit (through the actions of the robot) salient features of their orientation to task artefacts, which in turn aided their collaborators in interpreting their actions. Whilst Kuzuoka et al (2004) discuss this approach in relation to the adequate establishment of mediation (through a communication tool i.e. a robot) between two distinct and remote ecologies, they ignore the extent to which they are in effect making the communicative ecologies increasingly similar. If one considers the 'ecology of communication' arguments of Abraham Moles (1966, 1975), then it becomes apparent that for successful communication 'to transmit a message is to make more complex the space-time surrounding the point of reception; it is to produce a micro-replica of the complexity created at the origin of transmission' (Moles, 1966) and therefore one must make one's collaborator "partake in the experiences (Erfahrungen) and stimuli of the environment of another individual" (Moles, 1975). Given this perspective, clearly the strength of the modifications of the GestureMan is not that it makes communication between sites smoother per se, but that it makes the experiences of the collaborators at each remote site increasingly similar.

This notion of creating environments in which participants share equally is perhaps then a fundamental building block of the Agora system, returned to by the Research Group of Technology and Social Interaction in their Luff et al (2006) paper. This work demonstrates a turn to the perspective of the importance of shared environments rather than mediating between dual ecologies (which in itself may reflect the fact that Agora as a system is now more an exploration of shared document handling, with its implications of highly similar collaborating environments, rather than an effort to support what Kuzuoka (1994) originally referred to as 'spatial workspace collaboration', which in turn is often characterised by highly *dissimilar* collaborating environments). This shift in focus leads Luff et al (2006) to argue that to adequately support interaction one must "provide multiple access to another's remote

domain, even if this results in multiple images or representations of the conduct of another;" (p.569) and one must "provide resources to allow others to see and recognize trajectories of conduct, from their outset." (p.569). Such remarks whilst based on their observations of use of the system, are also based on the authors observations of communicative conduct in other situations (as per Heath and Luff, 1996), and to a certain extent are open to criticism. This notion of the importance of understanding all gestures from the point of origin, and the implicit awareness of others activities being integral to successful communication is in part based on observations of control room activity. A situation which is highly contextualised and has a set of operating practices which one must assume are highly idiosyncratic. Regardless of this there is an extent to which the work presented above has taken the ethno-methodologists' stance to CSCW (and HCI in general) that current unmediated forms of interaction are *de facto* superior. Largely arguing that existing work practice must not be changed by intervening technology, the technology must be changed to accommodate practice. The tone of much of the development work of GestureMan and Agora, is that insurmountable obstacles are created by the fracturing of ecologies due to the technology. But what the studies clearly demonstrate is the infinite adaptability of the collaborating parties to accommodate difficulties that the technologies engender. Talk is a medium through which almost all inconsistencies in task perspectives can be negotiated, and the fact that people are always eventually successful in presented GestureMan and Agora work vignettes is testament to this fact. Given that people can always accomplish these collaborative tasks using what is presented as slightly 'broken' technologies for communication, surely the important line of research is to establish the relative benefit of various approaches to technology design. The argument made by Luff et al for the presentation of what might be determined, in their own words, as *redundant* information is highly questionable. The key failing of the GestureMan / Agora work could arguably therefore be the fact that it fails to motivate a discussion of the key requirements for improving communicative action, which could surely only come from a more quantitative comparison of key features of communication tool design. The studies presented above currently argue for progress towards supporting technology which re-creates full blown co-present interaction, in itself a highly complex interaction to support, without exploring the performance benefits that can be derived from more satisficing technologies.

### 2.6.2 Developing the WACL (Wearable Active Camera/Laser)

Extensions to the system presented in the GestureMan studies (see Heath et al 2004) have lead to the creation of the Wearable Active Camera with Laser Pointer (WACL) device (Sakata, Kurata, Kato, Kourogi and Kuzuoka, 2003, and Kurata, Sakata, Kourogi, Kuzuoka and Billinghurst, 2004a). This device is a shoulder mounted camera, which relays context (task-space) views to a remote expert. The camera is independently operable by a remote Helper, and has a laser pointing device attached. This facilitates simple gesturing behaviours and

allows the Helper to manoeuvre the camera and view the task space independently of the Worker's (wearer's) orientation. Essentially the system is a replication of the GestureLaser and GestureLaser Car systems (Yamazaki et al 1999), modified for shoulder mounting, and with alterations to the software such that the camera view and laser pointer action are stabilised to counter the natural instability caused by being worn during use.

In an evaluation of the WACL Kurata et al (2004b) compared performance in a remote collaboration (an assembly task) between WACL users and users of a head mounted display. The study failed to show performance differences between the two systems but did report (somewhat inevitably) that the WACL system was preferable as it was more comfortable to wear and hindered sight of the task-space far less than a head-mounted display. The study clearly demonstrates the problems engendered by use of a head-mounted display and camera in a collaborative physical task, findings highlighted previously in studies such as Kuzuoka (1992). The evaluation of the WACL system failed however, to address many of the critical issues of the GestureLaser style approach, discussed in Luff et al (2003). As such it is obvious that any attempt to use the WACL system would suffer from similar problems to the GestureLaser systems, such as the inability of the gesture recipients to adequately interpret the meaning of gestural projections (owing to the limit bandwidth for expression of a laser dot), and equally the inability of the Worker to adequately disambiguate the Helper's current focus of attention from camera angles and incidental laser pointing gestures.

### 2.6.3 Developing DOVE (Drawing Over Video Environment)

Initial investigations by Kraut, Miller and Siegel (1996) sought to explore how people engaged in tasks could be supported with the help of remote experts. By engaging study participants in a bicycle repair task (for which they were provided with an instruction manual), the effects of performing the task alone or with the support of a remote expert could be observed. The study manipulated both the presence of the remote expert and, if present, the means by which they communicated with the Worker performing the repair task. Contrary to their expectations Kraut, Miller and Siegel (ibid) observed that providing a video link between the Helper and the Worker (such that the Helper could see the Worker's task space) did not improve performance beyond those levels observed when the Helper and Worker communicated via audio link alone. The presence of the visual link between the spaces did however, have an effect on the pattern of communication between the collaborators but this altering of communication was not reflected in a change to performance times, whether or not the Helper was a current part of the Worker's task however, did influence the performance times incurred, greatly improving success.

This original study was reprised by Fussell, Kraut and Siegel (2000), a study which sought to extend the earlier work, by taking account of the lack of an adequate control condition in the original study, by introducing a side-by-side remote expert help condition. Further, attempt

was made to counter any bias that may have crept in from the experts previously using almost scripted language. An additional variable was also included, the variation of the expertise of the remote Helpers, to ascertain as to whether the level of expertise of the Helper made any significant difference to the interaction and the effect of the various communications technologies. The results of the study demonstrated that regardless of Helper expertise (which did not affect performance in the task) side-by-side collaboration was faster than other remote collaboration conditions; this was achieved without a reduction in the quality of the work achieved. The assessment that was made of the work quality, along with the quality of Worker and Helper communication was made by expert observers, this may however, have led the results to be open to experimental bias. Despite this, conclusions were drawn that side-by-side dialogues are significantly more efficient. On the basis of the experimental results four limitations to video-mediated visual spaces were suggested. 1) Workers' queries suggested that they were uncertain of the field of view of the Helper. 2) Helpers' views were in fact less than optimal – important features of the work space were often held external to their normal view. 3) Helpers' had no access to Workers' faces – this may have hampered the Helpers' understanding of the Workers' comprehension of verbal instructions. And finally 4) Workers' views of the Helpers were limited to upper body images thus preventing the Helper from effectively gesturing at shared objects. These observations concerning the limitations of the existing visual space lead to the creation of several suggestions for video system design including 1) the provision of better feedback to the Worker about what is perceptible (in terms of view) for the Helper. 2) The provision of a wider field of view for the Helpers. 3) Provide Helper's with feedback of worker's attentional focus. And finally 4) support Helper's in gesturing within the shared visual space.

These first two seminal studies were re-presented and evaluated in a further paper by Kraut, Fussell and Siegel (2003). Again considering bicycle repair as an exemplar of a task that might require expert support, the paper decries the fact that most groupware systems support activities that can be performed without reference to external objects and the external spatial environment. The paper argues that the "Development of systems to support collaborative tasks involving physical objects has been much slower." (p.15). On the basis of this the paper introduces the notion of collaborative physical tasks as:

> "Tasks in which two or more individuals work together to perform actions on concrete objects in the three-dimensional world." (p.15)

And specifically in this instance:

> "Collaborative physical tasks can vary along a number of dimensions, including number of participants, temporal dynamics, and the like. The task on which we focus here, a bicycle repair task, falls within a general class of 'mentoring' collaborative physical tasks, in which one person directly manipulates objects

with the guidance of one or more other people, who frequently have greater expertise of the task." (p.15)

The research interest of the CMU group is defined as being primarily concerned with the provision of and support in tasks with *visual information*. They argue that this can be used to improve situational awareness of a task (Endsley, 1995) and to aid conversational grounding (Clark, 1996). An interesting argument is put forward that situational awareness and conversational grounding are developed in face-to-face settings using a variety of behavioural expressions and interpretations (the work of Robertson, 1997, is perhaps the most literal interpretation of how such physicality is construed in the structuring of collaborative environments). Kraut, Fussell and Siegel (2003) argue that due to the constraints of bandwidth and the difficulties of representing all of this information coherently (as per Gaver et al 1993), such elaborate environments cannot be constructed, therefore they claim:

> "Our approach is instead to try to identify the critical elements of visual space
> for collaborative physical tasks and to design video systems that support these
> critical elements." (p.16)

Their analytical strategy was to take a decompositional approach, and systematically evaluate the various elements that might influence communicative behaviour when engaged in collaborative physical tasks. However, given that the original interests of the group had been the exploitation of video-mediated communications it is apparent that this early work shows the first hint of a locking in of the use of video windows as a permanent fixture in their communication system infrastructure, the work becomes an effort to extend the functionality of video-mediated communication, rather than a direct exploration of techniques to link spaces. The work is resolutely situated within the theoretical framework provided by Clark and Brennan (1991) and the affordances of communication media affecting the ease of maintaining task awareness and establishing common ground. Given the argument that various media hold different costs for the grounding process, assumptions were made about the suitability of various elements of shared visual spaces to support collaboration. The work of Kraut, Fussell and Siegel (2003) highlights the importance of object-centred shared visual spaces in object-focussed tasks (as typified by collaborative physical tasks), referencing Karsenty (1999 with her study of shared computer screens for problem solving), Gaver et al (1993, with their analysis of the usability of multiple video windows demonstrating that object views were viewed most often) and Nardi et al (1993, whose study showed how Nurses use monitors during surgery to view the current stage of surgical procedure to find tools in advance).

In the type of interactions studied by Kraut et al the role of Helper was seen to be constructed of several phases of action, the first phase was the determination of what help was needed. Secondly the help must then be provided, during which the Helper must coordinate their utterances with those of the worker, the workers actions and the current state of the task. From their observations of this process being enacted with either the Helper being side-by-side with

the Worker or at a distance (but linked through varied communications media) several general conclusions were drawn. The first was (as stated previously) that the provision of expert help is a positive enhancement to performance. However, despite differences in the articulation work that is performed when a video image is shared, a video representation fails to effectively improve performance over audio-only support. On the basis of this conclusion strong claims were made that side-by-side collaboration is superior because of the way in which it supports natural deictic communication, by actively supporting gestural behaviour. This it was argued must be supported in later systems so as to improve the time required to achieve grounding. In more mediated conditions more time and resources were spent acknowledging (back-channelling) instructions. In the side-by-side collaboration communication from the Helper was far more directive, no understanding was provided by Kraut et al however, of the relative perceptions that participants have of this more directive approach and the impact that this might have on longer term patterns of collaboration.

The paper goes on to suggest that video-audio links may have failed to improve performance beyond that achievable by audio-only links because of a lack of a head shot of the worker, meaning that Helpers found it harder to determine whether Workers had understood instructions. The paper counters this supposition by citing Whittaker and O'Conaill (1997) who had previously shown that such information was rarely useful to collaborators. Their final conclusion then rests with the limitations observed in the Helper using a camera mounted on the Worker's head, which necessarily restricts their field of view to that which the Worker is looking at. Party to this is a limited understanding from the Worker of exactly what of their view the Helper can actually see. This reciprocal awareness of mutual perspectives being considered a sizeable problem to be overcome, which the authors argue may be answerable in part by providing enhanced access (as perhaps is provided in side-by-side collaboration) to collaborators' gaze patterns. Some of these various issues were addressed in subsequent papers.

Fussell, Setlock, and Parker (2003) used eye tracking techniques to assess where Helpers look as they are providing assistance to a Worker during collaborative physical tasks. The results of the study suggested that Helpers did not look at the Worker's faces but did look heavily at their hands, the pieces being manipulated and the developing assembled piece. Whilst the results provide value for those wishing to develop technologies to support remote collaboration there are several problems with the study. Firstly, it is not made clear if the pair are co-present or using some intervening technology. Multiple video windows rather than side-by-side collaboration may have led to more use of face views. Secondly, Worker responses to Helper instructions were also scripted, making for a highly unusual interaction which would not conform to most standards of free flowing collaborative task focussed discourse. Finally, a large proportion of glances were reportedly made toward the instruction manual, but if the Helper were a true expert then this resource may not be used and therefore a significant amount of 'gaze time' would be needed to be distributed elsewhere and this may end up being

focussed, in the absence of other requirements, on the Workers face so as to more securely confirm understanding. Whilst not strictly critical to the task, it may be *preferable*.

In a further study to test the benefits of the provision of simplified remote gesturing behaviours, Fussell, Setlock, Parker and Yang (2003) compared performance in side-by-side, video-audio and video-audio plus cursor instruction conditions. The results demonstrated that performance is better in side-by-side, and that the addition of cursor information does not improve performance over video-only presentation. The self reports of participants however suggested that the use of a cursor made the identification of objects easier. However, side-by-side collaboration was still rated as the easiest format. The fidelity of the pointing achieved with a cursor on a video view may however be responsible for its lack of success. Pointing in the 3-D world is considerably more accurate and easily interpretable than pointing in 3-dimensions over a 2-D representation.

To further understand the visual requirements of the Helper in a collaborative physical task, Fussell, Setlock and Kraut (2003) compared collaborative performance when using scene oriented and head-mounted cameras. Five distinct collaboration conditions were compared, side-by-side, audio only, head camera, scene camera and finally scene camera plus head camera. The performance results illustrated that side-by-side collaboration is fastest (faster than all other conditions). Performance when using the Scene camera was faster than audio only, but was not significantly faster than performance when using the head mounted camera, despite the conclusions the authors attempt to draw. The proposed difference between the head camera and the scene camera is somewhat controversial, with the only real difference being that the head camera views a subset of what is available in the scene camera (which could presumably be rectified by changing the head mounted camera for a more extreme wide-angled lens). It was interestingly noted however that the head camera may also provide some epiphenomenal information about current gaze awareness, as the centre point of the camera shot is clearly aligned with the Worker's facing direction, which is potentially why the authors expected the head camera plus scene camera views to be superior, as they would provide context views with an indication of current attentional focus. But obviously this plays into the trap of dividing the attention of the Helper between multiple windows (as discussed by Gaver et al 1993). That the head-mounted camera appeared to offer no real advantage is perhaps not surprising given that its ability to provide orientational information and a tight focus on task artefacts was potentially rendered ineffective. The very set-up of the scene camera may have incidentally supported the implicit development of awareness of the Worker's orientation as part of their head was captured in the image, the angle of the head therefore providing some gross orientational information, and considering the large nature of the pieces required for assembly, a fine-detailed close focus was not generally necessary. When situations did require a close focus Helpers could negotiate this deficit by having Workers hold items up to the camera.

Reported subjective preferences were for the side-by-side condition and then, secondly, for the scene camera, as the next best alternative, over a variety of measures. This is an interesting issue however as user preference is not necessarily the best indicator of performance, depending on the context of use, actual progress made in the tasks potentially being of considerably more value. There is a clear increase in communicative efficiency with use of the scene camera over the head camera. This has been discussed in previous research and is clearly predictable owing to the fact that with a head camera more work must be done to re-orient the visual image so it is suitable for the Helper.

The ultimate finding of the above studies was that along with various considerations concerning the adequate establishment of a shared visual space connecting the Worker's workspace to the Helper, the most primary reason behind the inability of video connections to facilitate collaboration to the levels witnessed in side-by-side collaboration was the conspicuous suppression of naturally occurring gestural behaviours. Time and again observational analysis during these studies demonstrated that Helper's wanted to be able to point at objects on their video feed, but clearly what was required was more than simple deictic behaviours as this had been shown to be of limited value during the interactions. As an answer to the problems highlighted in the above papers a system was built to perform the critical function of presenting remote gestural information whilst providing the wide angled scene oriented views of the task space which were demonstrably required. The system was referred to as the Drawing Over Video Environment (DOVE) and was first presented and discussed in Ou et al (2003a, 2003b). Figure 2.18 below illustrates the DOVE system.
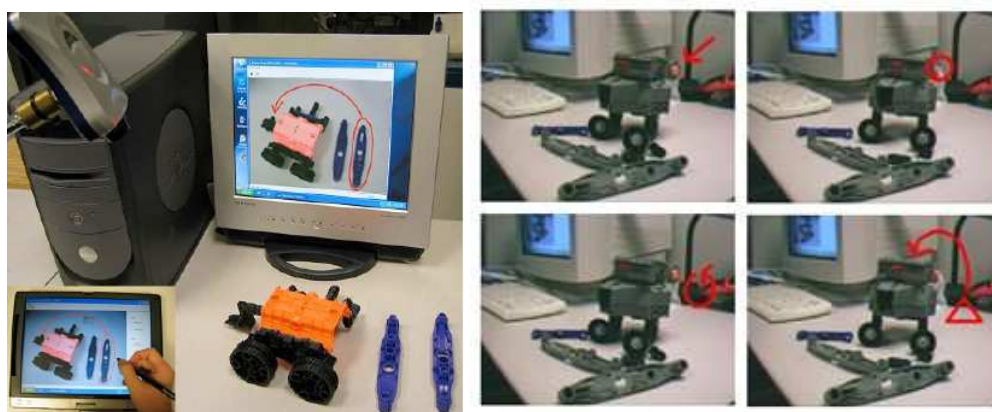


Figure 2.18 The Drawing Over Video Environment (DOVE) from Ou et al (2003a,b)

The DOVE system works by capturing a live video feed of the Worker's task space via an IP camera. This video feed is then relayed to a remote Helper, who views the live images on a tablet PC. With the tablet PC the Helper is then able to write or draw, making marks over the live video feed with a digital pen. These resultant 'gestural sketches', along with the video feed

are then passed back to a monitor (VDU) located at the edge of the Worker's task space. By looking up from their task artefacts and towards the monitor the Worker can then see the video image of their own task space with the gestural sketches overlaid. In some iterations of the system the sketches made by the Helper are normalised and corrected by software on the Helper's tablet PC, to conform to standard shapes (such as arrows and circles). The removal of the sketches from the video feed can only be effected by the Helper and can either be achieved manually by pressing a button on the tablet PC, or in later versions of the system is enacted automatically after a period of 3 seconds.

Fussell et al (2004) provided the first full evaluations of the DOVE system. In a review article they presented two experiments, the first is a re-write and evaluation of the findings from Fussell, Setlock, Parker and Yang (2003) (see above) and the second is a comparison of performance in a collaborative physical task using DOVE versus video only communication. To help ground the studies in some research context Fussell et al (2004) cite studies such as Flor (1998), Goodwin (1996) Kuzuoka and Shoji (1994) and Tang (1991), which argue that speech and action are intricately related to various external elements (people, objects, activities) within the collaborative environment. On the basis of this and their own earlier work, the authors make a call for the inclusion of gestural information in technologies to support remote object-focussed collaborations, discussing how gestures can be used to enhance spoken messages (as observed in Bekker, Olson and Olson, 1995 and McNeill, 1992).

To understand the context further of how gestural activity is situated within the context of a collaborative physical task Fussell et al (2004) break down the structure of a common interaction in this class of task:

> "First, collaborators come to mutual agreement upon or 'ground' the objects to be manipulated using one or more referential expressions. Next, they provide instructions for procedures to be performed on those objects. Finally, they check task status to ensure that the actions have had the desired effect." (p.277)

Of course this is a simplified structure which could be elaborated further, there being iterative cycles of interaction at each stage, with each stage having significant potential to suffer communicative break-down, should the interaction be insufficiently grounded, and thus require a process of repair to be enacted (see discussion in chapter 1 for more detailed breakdown of a common task lifecycle).

Having broken down the common structure of a collaborative physical task the paper then demonstrates that gestures can be used at a variety of the points of this cycle. It argues for the usefulness of concrete representational gestures rather than abstract representations (McNeill, 1992), and highlights the already discussed role of deictics, also arguing that some supposedly non-communicative gestures such as beats (Argyle, 1988) are not worth supporting (it is worth baring in mind that the argument that a beat is non-communicative is not strictly clear cut, with some evidence arguing that regardless of intention all gestures if enacted publicly are in some

way communicative, cf. Robertson, 1997). The paper goes on to claim that amongst the concrete gestures there are three forms of particular relevance that should be supported, iconics (i.e. gestures where the hand literally represents something tangible such as an assembly piece), gestures representing spatial/distance information (e.g. hands showing how far apart two items should be or how a given object should be moved) and finally kinetic/motion gestures (demonstrating through use of the hands what action should be performed on an object).

Fussell et al (2004) make the argument that:

> "Concrete representational gestures may facilitate conversational grounding in collaborative physical tasks by allowing speakers to communicate multiple pieces of information about the task simultaneously (*citing Clark, 1996 and McNeill, 1992 to reinforce this point*)." (p.280) (*my italics*)

It is interesting to note at this point that they discuss the importance of direct hand gestures but then limit the capacity of their designed system to adequately produce them.

From this point the Fussel et al (2004) paper tries to make an argument for the use of surrogate gesture methods. Citing systems such as ClearBoard as examples of systems which utilise unmediated representations of hand-based gestures, the authors claim that such systems are problematic because of the inherently costly specialised equipment that they require. This is a naive point of view, ClearBoard requires such a complicated set-up because it needs to produce a fully symmetric interaction, which is not the same when supporting collaborative physical tasks, those of the nature discussed above. In such instances there is a much more asymmetric interaction which presumably predicates a simpler system. Because of this assumption about the difficulty they opt for this notion of using gestural surrogates, citing evidence that such surrogates "incorporate visible *embodiments* of gesture" (p.281); and are therefore just as good as representations with a higher fidelity. Noting the work of Kuzuoka et al (2000) they argue that laser pointer systems can only be used for deictic activity and as such are too limited in their ability to project representational gestures. In a search for a method for creating such representational gestures they discuss the use of cursors as a collaborative tool (e.g. Greenberg et al, 1996, Gutwin and Penner, 2002) but eventually decide that sketch-based tools and hand-written text may be the best solution for adequately supporting this higher level of gesturing. They cite old studies such as Bly and Minneman (1990) with their study of Commune, which illustrated the comfort with which collaborators will use sketch tools for the purposes of gesturing (although this is clearly a wildly differing context of use). With this in mind Fussell et al (2004) settle on the use of sketching over remote video.

The first study of the paper is a more detailed write-up of the experiment presented above in Fussell, Setlock, Parker and Yang (2003), and is concerned with the evaluation of a cursor pointing tool. The results demonstrated (as stated above) that the use of a cursor was not sufficient to elevate performance in a video-mediated collaboration to the levels of side-by-

side interaction. There are some issues which should be considered however, which may have a bearing on the results. Firstly, as the authors acknowledge there is a potential problem in that the pointing information is provided over a video feed which is taken from an angle which is not the same as the Worker's visual perspective. As such all gestural 'pointing' information must be extrapolated from the video feed and be mentally re-situated to align with the Worker's own perspective, an activity which surely carries with it some cognitive cost, which one would assume would carry over into the task completion times. Fussell et al (2004) argue that this did not have an effect as none of the participants complained of difficulties in performing this activity. Perhaps of more concern however is the fact that the instruction manual used in the side-by-side and video-mediated conditions (the manual used by the Helper) was not consistent. They had different formats, presented on screen for the video-mediated conditions and printed out for the side-by-side conditions. Research evidence would suggest that reading from and skimming information on VDU screens is slower than in paper presentation, or at the very least there can be significant differences between reading from a screen and reading from a page (Muter, 1996) which would necessitate that such factors should be controlled in an experiment. Equally, the paper does not make clear how, in the side-by-side condition, the Helper who was obviously directly adjacent to the Worker, managed to shield their view of the paper instruction manual from the Worker. If the Worker could glance at the manual (either surreptitiously or otherwise) then this was clearly an unfair advantage and a confounding variable. Regardless of which, the use of a side-by-side condition does not clarify which aspects of the side-by-side collaboration are responsible for the increase in performance, whilst Fussell et al would argue that the pointer did not work because of its inability to adequately represent more complex gestures there is clearly also a difference between the view of a workspace afforded by a 'scene-camera' and that afforded by being sat next to the Worker at the workspace. The effects of such a variable should also be taken into consideration as they may far out-weigh the contributions achievable from the use of a simple deictic representation.

The more important work of the Fussell et al (2004) paper comes in their presentation of experiment 2 (some brief details of the study were originally presented in Ou et al, 2003a). In this study they experimentally compare performance in a collaborative physical task (a standard robot assembly task used as the collaborative physical task for several of the previously discussed studies) in each of three conditions, video-only, DOVE (with auto-erasure) and DOVE (with manual erasure). The DOVE system, as described above, facilitates the production of more complex representational gestures, and the authors hypothesised that this would lead to superior performance. Equally they hypothesised that the use of the manual erasure facility would allow Helpers to produce more complex multiple gestural sketches which also improve performance.

The results of the study demonstrated that performance (i.e. how fast a model was completed) was significantly faster when collaboration was facilitated with the DOVE (auto-erasure system). This was faster than both the video only condition and the DOVE (manual erasure

condition). Whilst the performance times showed that DOVE (manual erasure) was faster than video only, it was not significantly so. Fussell et al (2004) go on to argue that the DOVE (auto-erasure) performance levels are comparable to the performance levels witnessed in the side-by-side condition of the first experiment. This assertion is not however, statistically compared, which is somewhat surprising given the rigorous nature of the rest of the analysis. A problem with the analysis of performance times is that they fail to demonstrate the accuracy with which the models were completed. As such faster performance may have lead to a decrease in the quality of the work achieved. The results presented cannot tell us whether this is indeed the case.

Further analysis of the experiment focused on users' preferences and the nature of the collaborative language used during the sessions. Measures of the collaborators self-reported coordination, their evaluations of each systems' ease for identifying referents, general preference for the systems used, efficiency of communication and use of local versus remote language all showed that DOVE use was preferable to video-only collaboration; but failed to show any differences between the manual erasure or auto erasure systems. Such findings demonstrating that participants did not feel that such a facility impacted on their performance, and showed no effects on the language that they used.

A later stage of analysis considered the form of the drawing activities that were use during the study, by the Helpers. This demonstrated that the majority of the performed actions were pointing gestures, with a smaller number being considered 'directional'. Given the findings of the first study, that pointing per se was not helpful, the authors conclude that these, smaller number of more complex representational gestures, are in fact crucial for improving collaborative performance in these forms of task.

In conclusion the broad results suggest that there is little difference between video alone and DOVE with manual wipe, the increase in performance is purportedly from using DOVE configured with an auto-wipe capacity. The authors argue that communication is more efficient when this system is used. This conclusion is clearly however, not supported by the results and is in fact directly contrary to the findings, use of the manual erasure system did not lead to decreased communicative efficiency. The results did however successfully suggest that more complex forms of gesture, rather than simple deixis are responsible for the performance benefits, and therefore gestural systems should be designed to represent more complex forms of gesture. There are areas of interest and further conclusions drawn however which are not clearly supported by the data. The first of these is concerned with the notion of presenting gestural data pasted over a live video feed on an externalised monitor as opposed to projecting it into the task space. The conclusions provided are based on the fact that the participants did not appear to complain about this, and they make overly ambitious statements about the 'misalignment' issue being unproblematic (this issue has received much attention in the development of other comparable systems cf. Luff et al 2003). Clearly any true answer as to

which is the most appropriate way of presenting the remote gestures can only be derived from empirically comparing performance under the two conditions. In addition to this the authors claim that the two task views (Worker and video-feed) could be aligned further by using a head-mounted camera, but obviously this would not in fact work, as clearly as soon as the Worker moved their head the gestural sketches would cease to be sensibly aligned with the intended task artefacts. Equally the second set of conclusions which are somewhat controversially drawn surrounds the notion of the use of gestural surrogates as being preferable to unmediated representations of hands. The evidence being largely based on the fact that to construct and transmit unmediated representations is more expensive (both economically and computationally) and gesture surrogates perform the functions of remote gesturing just as well. The assertion of comparative costs is however, unfounded and not thoroughly investigated, and the issue of performance benefits can surely only be adequately answered if surrogate and unmediated representations are directly, experimentally, compared. A final issue of contention with the observations made by Fussell et al (2004) lies with their insistence regarding the potential usefulness of the gesture normalization aspect of the system, despite the users' rejection of this facility. Whilst gestural sketch normalization may indeed be a technically interesting issue, it was seen to be of little utility for the actual users, presumably in certain instances the vagaries of sketches may be intended and integral aspects, and as such user control is paramount.

Whilst these issues have not been resolved, some considerations such as the reported desire of Helpers to be able to manipulate their camera views have been investigated in later evolutions of the DOVE system, such as DOVE-2. The results of these findings have however, not yielded positive benefits and have therefore not been reported in the academic literature (Ou, 2006, personal communication). A final problem that has been encountered is observable in the most recent report of the use of the DOVE system. In a paper by Kramer, Oh and Fussell (2006) which is actually reporting on the use of linguistic features as a tool for measuring presence in CMC, use of the DOVE tool demonstrated significantly inconsistent results. Whereas previous studies have shown clear advantages of the DOVE system over mere video links, in this tightly controlled study, utilising 38 pairs of users, the DOVE system failed to yield any performance advantage over a video only link. Equally it made no significant difference to the pattern of language used or the users' self reports of presence or coordination. The DOVE system clearly failed to offer performance beyond that achievable by simple video links between spaces. So ultimately despite some of the grander claims made in earlier works, about the benefits of the use of gestural surrogates in linking spaces during collaborative physical tasks, the results previously shown in reports of the use of DOVE are not replicable, even by the team who conducted the original studies. This significantly draws into question the validity of the previous results, suggesting that any conclusions drawn in earlier work on the DOVE system should actually be considered as inconclusive.

**2.6.4 Developing collaborative augmented reality and TUIs**

A final alternative strategy to linking spaces and / or supporting object-focussed activities can be found in the approaches of Augmented Reality (AR) and Tangible User Interfaces (TUIs).

A paper by Billinghurst and Kato (2002) discusses the potential for AR to be used as a medium for collaboration in both of two contexts, face-to-face and remotely. Billinghurst and Kato discuss how with current collaborative systems for co-located groups there is an artificial barrier created between the 'real world and the shared digital task space', arguing that this limits the potential for naturalistic interactions between participants, that utilise natural communication behaviours. Equally for remote working interactions current systems leave the participants with a very limited sense of presence, and vital perceptual cues for normal communicative gestures (such as turn-taking indicators like changing eye directions) can be obscured.

Billinghurst and Kato (2002) discus how AR systems can be constructed to facilitate group work, drawing on the earlier work of Schmalstieg et al (1996) which identified five key attributes of collaborative AR environments:

- *Virtuality* – Objects that don't exist in the real world can be viewed and examined.

- *Augmentation* – Real objects can be augmented with virtual annotations.

- *Cooperation* – Multiple users can see each other and cooperate in natural ways.

- *Independence* – Individual users control their own independent viewpoints.

- *Individuality* – Displayed data can appear in different forms for individual viewers depending on their personal needs and interests.

One key factor that Billinghurst and Kato (2002) highlight is the seamless way in which the task space and communication space are held together, unlike other forms of CSCW interface. They have also discussed how through use of AR in remote working situations system can be built that will posit the remote workers into each others working environments essentially allowing participants greater presence in one another's environments, which in turn holds great benefit for various social processes. The conclusions that are drawn about the future of AR systems are that the technology for their construction is becoming rapidly accessible, but what is not understood is how such systems can best enhance various forms of CSCW communication.

An alternative to this approach has been the use of AR to directly support physical tasks (such as assembly tasks) by rather than supporting a shared visualisation or a medium for communication with an expert, but by being used to strongly support the provision of knowledge resources directly to the person performing the work (Tang, Owen, Biocca and Mou, 2002, 2003). Tang et al (2003) show clear results that the use of AR technologies can provide situated and spatially relevant instructions in assembly tasks which greatly improve

worker performance (over receiving instruction from printed manuals). Such an approach seeks to circumvent the requirement of expert support, but is severely limited in that it can only be used in highly conventionalised tasks. Where a task context is dynamic or elements of the task itself are potentially highly idiosyncratic, and fluctuate over time, overlaid AR images in a task space are likely to be of limited applicability.

Equally such augmented environment solutions to collaboration difficulties, like tangible interfaces and '*synchronized distributed physical objects*' (Brave, Ishii and Dahley, 1998), whilst offering novel ways of ensuring that experience in a task is similar at both ends of a remote collaboration, are unlikely to be able to adequately support what have been defined as collaborative physical tasks. This is unless the tasks themselves are highly routine (ensuring that sufficient synchronized distributed physical objects can be provided at both points of action). Of course the studies of developing common ground (Clark, 1996) make it obvious that the deployment of such technologies, wherein the physical manipulations of an object are dynamically experienced by a linked object at a remote site, is probably unnecessary in support of well grounded routine interactions. When the actions and interactions are highly conventionalised and well understood by all collaborators elaborate connections are not necessary, in many instances an audio link will satisfice.


## 2.7 Summary and Conclusions

This chapter has thus far presented a detailed description of the various strands of research which are important for understanding and framing the work that is presented in the rest of this thesis. Having detailed the research studies within this space and articulated the development of various remote gesture technologies and the critiques that have been applied to them, it is worthwhile concluding this chapter by briefly summarising the evidence so far discussed.

- Studies have suggested that communication in co-present working situations is reliant on more than the expression of verbal actions. Communication happens in a rich multi-layered environment in which verbal expressions are combined with a host of non-verbal actions. The non-verbal behaviours can be crucial for interpreting the verbal content of communication but can additionally provide a host of interaction and work practice structuring components (such actions can be enacted and interpreted both consciously and unconsciously).

- In efforts to support more 'distributed' working environments, in which collaborators are not co-present (and to make them as 'successful' as co-present interactions), technological solutions have been offered which seek to increase collaborators' mutual awareness of various non-verbal behaviours, transmitted between their remote spaces. A commonly explored strategy to facilitate such communication is the provision of video links between spaces.

- Research findings have been frequently inconsistent about the actual benefits of providing video links during remote collaboration. There is some evidence however that video links are of benefit in particularly object-focussed interactions, where colleagues are collaborating around some common objects of interest. In these situations it has been shown that shared visual access to the objects of interest (as opposed to shared visual access to epiphenomenal aspects of interactions such as collaborators' faces) can be of particular benefit to task completion. Studies show however, that providing such links is unlikely to improve collaborative performance to the levels observed in co-present interaction.

- In one form of collaborative activity, design work, gesture has been demonstrated to be of significant importance. When attempting to support this activity remotely (which is an inherently object-focussed activity) many CSCW systems have utilised very un-modified representations of gesture (often utilising overlapped direct video feeds of disparate working spaces – which supports shared drawing activity). Such approaches have developed an awareness of the potential benefits to be gained from remote representations of gesture in collaborative tasks and have consequently influenced the development of remote gesture tools in other areas of application.

- One particular area of application (which forms the focus for this thesis) is the collaborative physical task, in which two or more people collaborate in a task which is inherently physical in nature, and in which the task artefacts cannot be equally accessed by all remote participants (often involving a mentoring scenario in which one collaborator has more knowledge than the local worker). Studies have shown that in such scenarios, when representations of remote gesture are provided between remote sites, allowing remote experts to gesture towards (and around) task artefacts at another locale, levels of performance can almost match those of co-present interaction and can (in some reports) exceed the benefits provided by standard techniques of VMC. In the development of systems to provide such remote gesturing capability, in support of collaborative physical tasks, there have been a variety of systems built, the GestureMan systems and the DOVE system being the most prominent and thoroughly explored.

- Whilst some studies have supported the benefits of remote gesturing systems, some have demonstrated their significant short-comings, demonstrating that in many situations their benefits are combined with significant limitations. In some instances it has been demonstrated that a limitation in the design of a remote gesture tool has unintentionally 'fractured the interaction' between collaborating parties, meaning that a gesture representation is misguiding attention, or failing to attract it in the first place. In such instances research shows that increased effort is required to establish smooth communication.

- Whilst there has been some consideration of how interaction can be fractured with the GestureMan systems, such critique has not been addressed by the DOVE systems. And the existence of at least two such distinctly different remote gesturing arrangements means that any researcher who wished to build a remote gesture system would have very little guidance as to which approach is best. The relative merits and indeed the relative impacts on the fracturing of interaction of each of the different systems have not been considered.

- A further limitation to the studies within the field stems from the lack of understanding of exactly how remote gestures are used either independently or in conjunction with discourse during collaborative physical tasks. To fully establish the benefits of a remote gesture technology and to understand how this process of fracturing ecologies arises a firmer understanding would be needed of the actual functional utility of gesturing, which in turn would inevitably shed light on the best ways in which gesture systems could be constructed and indeed, be deployed.

## Chapter 3 – Research Methodology and Disposition

### 3.1 Introducing Mixed Ecologies of Communication

A wealth of evidence has been presented, which demonstrated that in co-present interactions collaborators rely on verbal communication being embedded within a rich and complex environment in which there is mutual awareness of task relevant actions and gestures. The processing of this gestural information, it is argued, is used to help interpret the verbal interactions. Efforts to re-create such forms of gestural interaction, to extend the capabilities of VMC by developing remote gesturing systems, have however, encountered certain difficulties, which have limited the potential benefits of these technologies. Some authors have talked of how these technologies in effect fracture interaction, as features of their design inhibit smooth interaction in what should be a common ecology of communication. And prior work (discussed in section 2.6.1 in particular) has explored a notion of dual ecologies and the use of remote gesture tools as technologies that mediate interaction between spaces. But as technologies designed under this rubric have developed it has become increasingly clear that rather than supporting the mediation of communication between two separate ecologies the changes to the technology have sought to increase the *similarity* between the two distinct spaces (such as creating tethers between a remote expert's head direction, as she scans multiple screens, and the movements of a robot's head in a local task space). In effect the technologies are striving toward the creation of a 'mixed ecology' in which the two spaces are increasingly blended to develop a unified working environment in which each collaborator has a presence within the task space, whether remote or local to it.

To a certain extent however, the previous studies which have explored these notions of ecologies of communication have often made an explicit assumption that the actions exhibited in face-to-face interaction are *de facto* superior. An assumption has been made that perhaps technology should be designed to make remote interactions exactly the same as co-present interactions. But the benefits of such an approach have not been established. Given the constraints on communication bandwidth, the re-creation of a co-present interaction at a distance, at an indistinguishable fidelity to reality, might not be possible. This would suggest that if one did wish to construct remote interactions that are as fruitful as co-present interactions, one must first understand which specific aspects of co-present interaction are the most important to transmit between remote spaces. This is affirmed when considering that for successful communication we wish to make a recipient 'partake in our experiences' and 'make more complex the space-time around the point of reception' of a communicative action, as per Moles' view of ecologies of communication (see page 1).

From the discussions presented in previous work which sought to construct remote collaboration technologies it has become apparent that such devices have significant potential

to improve collaborative performance but in some instances the design of the technology can actually become counter-productive and hinder smooth collaboration. Analysis of the development of remote gesture tools (see chapter 2) has suggested two key features of remote collaboration which can be identified as being potentially crucial for the establishment of coherent interaction. These features are mutual and reciprocal awareness of commonly understood, yet richly complex object-focussed actions (hand-based gestures) and mutual and reciprocal awareness of task-space perspectives. The presence of aspects of these features such as the ability to perform remote gesturing and providing collaborators with access to shared visual spaces improves collaborative performance in remote interactions. Studies show that when these issues are mismanaged such as using low fidelity gesture representations e.g. laser dots or by creating unnecessary disparities between collaborators' views of the task space, such as using an external monitor to represent feedback of artefact-gestural actions, the interaction can become fractured, and the achievement of common understanding is impaired.

This thesis therefore argues that when collaborators are remotely engaged in communicative acts concerning some object-focussed interaction (e.g. a collaborative physical task) their performance will be optimised if they communicate using a mixed ecology communications arrangement. The mixed ecology supports communication by using technology to give collaborating partners access to the most salient features of collaborative interaction, as discussed above, namely the ability to see remote working spaces and to remotely gesture into them, but also the sense of *reciprocal awareness* of these and related gestural actions which underpin co-present interaction. A proposal that the technology supports the construction of a mixed ecology rather than purely transmitting information between spaces is most pertinent to this last point, namely the issue of reciprocal awareness of these various object-focussed actions (gestures) and orientations (views). A remote gesture tool which is designed sensitive to the goal of constructing a mixed ecology, it is argued, supports interaction by supporting these forms of reciprocal awareness.

## 3.2 Research Questions

The fundamental desire of this thesis is to explore the question of how technologies can be built to improve remote collaborations for physical tasks, that don't fracture ecologies between remote spaces, but make the interactions as close to the presumed optimal standard of face-to-face communication as possible. The arguments presented above hypothesise that a tool designed from a mixed ecologies perspective, fundamentally a remote gesturing tool, will achieve this. To fully evaluate this hypothesis several key questions must be addressed.

Previous research demonstrated evidence that the provision of remote gesturing improves performance in collaborative physical tasks. However, it is not properly understood how or why this occurs. Whilst performance can be seen to improve, what is it that is actually changing and at what level does gestural action act? Of equal importance is an understanding

of what element of gesturing is responsible for the effect and what elements of gesturing behaviour are required for these effects to be witnessed. The work of Fussell et al (2004) clearly demonstrated that the true benefits of gesturing were not just based in a simple replacement of verbose verbal descriptions with succinct deictic gestures, but were in fact somehow related to the ability of their DOVE system to generate more complex forms of gesture. In that paper they argued that it was the ability of the more complex gestures to simplify or represent complex functional descriptions of dynamic interactions of how task-artefacts should be assembled. In face-to-face settings however studies of gestural use (e.g. Sacks et al 1974, Duncan and Fiske 1985, and McNeill 1992) have demonstrated that gestures can have a variety of other descriptive and interaction structuring uses.

Understanding how gesturing is used in a remote collaboration and seeing how the process of the fracturing of interaction is caused or mediated by the introduction of remote gesturing technologies will facilitate an exploration of this notion of designing from a mixed ecologies perspective. To firmly establish how to build these potentially useful remote gesturing systems, and to understand how their design impacts on their use and potential deployment, the meta-level research questions for this thesis have therefore been framed as:

1. How and why does a representation of gesture improve remote communications (and consequently performance) in collaborative physical tasks?

2. What creates a fractured ecology, how does interaction breakdown and how can a remote gesture simulacrum overcome this problem?

3. What does an understanding of answers to the above questions mean for the design, development and deployment of such technologies?

A significant part of understanding and answering the above questions relies on an in-depth study of the technology design of remote gesturing devices, studying how the performance of users is impacted by their construction. Key structural configurations that can be manipulated that should further elucidate responses to the main research questions include the actual format for representing gestures and the location for their representation. As can be seen amongst the previous research a variety of methods have been used for the representation of gesture, laser pointers, digital sketches and unmediated projection of hands all being used in different devices. Equally some researchers have chosen to project gestures directly into a task space and some have chosen to externally represent them. No one has yet comprehensively compared these different strategies to see which works best, and importantly understand which methods have the greatest potential for fracturing interaction. Further structural points of consideration include the orientation to the task space that is provided (should both parties have the same orientation to the task), and what happens when the device is used in a mobile context?

Fundamentally therefore, one of the key aspects of addressing question 1 is to investigate exactly what are the key advantages and disadvantages of different remote gesture

representations (and to a lesser extent, how they're situated within the working environment). In particular a focus is necessary on richer methods of remotely representing gestures, research already suggesting that for collaborative physical tasks it is a richer expression of gesture which is of the most utility. Comparing and contrasting differing representations of gesture will help to explain how remote gesturing influences performance. To consider the 'why' element of question 1, a specific research question must be addressed which focuses on the relationship between the use of such remote representations of gesture and theories of collaborative action. By understanding how the potential effects on collaboration of remote gesturing can be understood in terms of existing theories a richer understanding of remote interaction can be developed.

To consider question 2, determining what creates a fractured ecology, the focus needs to be on this issue of how the gestures are situated within the environment, relative to task artefacts and collaborators' task-space perspectives, previous research having suggested that it is perhaps this issue of relative orientations and perspectives which is of most importance to the fracturing of interaction. Naturally it follows that specific sub-questions then should also address the relative qualities of different output devices (VDU versus projection etc.) and the relative orientation of gestures and perspectives in the task-space. These different gesture output methods should differentially affect users' abilities to coordinate their actions, and the interrogation of performance and breakdowns in performance during comparative use of these should help to explain this fracturing process. Again specifically determining how the use of gesturing tools can be understood in relation to theories of collaborative action may also help to further elucidate this investigation of question 2 (the fracturing of interaction). The ensuing answers to these various sub-questions consequently complementarily feed into the discussion of research question 3, offering specific guidelines on the design, development and deployment of remote gesture technologies. These specific sub-questions raised are listed below.

a.  What advantages/disadvantages do different gestural representations (i.e. mediated or unmediated gestures) have?

b.  What are the most appropriate output devices for representing remote gestures (VDU, projection, HUD) and therefore should the gestures be embedded in or held external to the task-space?

c.  What is the most appropriate orientation to the task space?

d.  How do collaborators use remote gesturing in collaborative physical tasks, and how can this be understood in terms of theories of collaborative action?

**3.3 A Choice of Research Methodologies**

Having summarised both the key problems to be addressed (section 3.2) and the previous research attempts to investigate elements of these issues (Chapter 2), and consequently developed a set of research questions, it is appropriate to discuss the choice of methodologies which are available for investigation in this research area.

In the field of CSCW (that research domain to which this thesis most readily applies itself) there has been a surge in interest in the qualitative methodologies of ethnographic research, perhaps stemming in part from the seminal work of Suchman (1987) with her thesis on situated plans and actions. Ethnography as a tradition is primarily associated with anthropological research and has been much adopted by the social sciences and the discipline of sociology in particular. It is an immersive observational approach to research in which the primary aim is to extensively document the practices, customs and interactions of those being researched as they are performed in their natural habitats (Robson, 2002, Hughes et al 1994). One specific field of enquiry within the bounds of ethnographic research that has been gaining prominence in the CSCW community is ethnomethodologically-informed ethnography (see Garfinkel, 1967, Sacks 1992, Crabtree 2003), which relies on an in-depth analysis of conversational behaviour to derive the social nature of work and its ensuing impact on the use of technology during that work. Such an explicit analysis of worker interactions has been welcomed within the CSCW community as practitioners have begun to move from studying human factors to studying human actors (Bannon, 1991) and as such there have been a wealth of studies that have adopted this approach (e.g. Heath & Luff, 1992).

Ethnography as a whole has been postulated to be of potential application at a variety of stages within the design life-cycle, indeed Hughes et al (1994) discuss four forms of ethnography that can be applied in the systems design process at various stages of the project. These are concurrent ethnography, quick and dirty ethnography, evaluative ethnography and the re-examination of previous case studies. All of these approaches attempt to provide for system designers a better understanding of working context in the belief that an inherently better understanding of the context of use will facilitate superior system design. Indeed Hughes et al (1994) suggest that (in reference to their ethnography of Air Traffic Control),

> "What the ethnography especially provided was a thorough insight into the subtleties involved in controlling work and in the routine interactions among the members of the controlling team around the suite; subtleties which were rooted in the sociality of the work and it's organisation. The vital moment-by-moment mutual checking of 'what was going on' by the various members of the team had been missed by earlier cognitive and task analytic approaches to describing controlling work."
> (p.432)

There are however certain limitations to the use of ethnographic research. Traditionally the complaints have centred on the reliability, validity and generalizability of ethnographic findings (as with all qualitative methodologies). The issue of reliability is addressed by the

insistence that observation and subsequent recording of data is performed as objectively as possible, although it is apparent that the ability to view something without a cultural bias in one's interpretation is extremely difficult. Shapiro (1994) in particular, pointing out the difficulty that purist ethnomethodologists have in structuring accounts of action without reference to pre-existing sociological theory. In an interesting further point in reference to the structure of ethnomethodology, Shapiro (1994) comments "*It does, therefore, specify not only how to look but also what to find*" (p. 419), bringing into question the reliability and objectivity of findings. In contrast the ecological validity of research is obviously extremely high given the situated, context-bound, nature of the research, however this strength increases the lack of generalizability of the findings. The results of any given ethnography must by their very nature be applicable largely to only that situation which is the focus of the specific analysis.

Other critiques of ethnography in CSCW have considered its utility, Hughes et al (1994) questioning efficiency of the timescales used for situated in-depth research,

> "We also learned that there was a declining rate of utility for the fieldwork contribution to the design." (p.432)

And also considering the deliverables from the research, arguing that there is a difficulty in extrapolating from the produced 'thick description' to actionable design recommendations (ibid.),

> "While the fieldworker learned a great deal in the study that was just discussed, certainly much that is useful for a sociological study, it proved difficult to hang this onto clearly formulated design objectives." (p.434)

An alternative approach to the more qualitative methodologies of ethnography is to take a positivist approach and gather quantitative data. The quantitative approach has a long tradition within psychology research and has therefore been adopted by the HCI community, which has significantly borrowed from psychology's predominant cognitivist paradigms when attempting to develop HCI theory. It is perhaps this overt cognitivism, with its potentially limited capacity to account for the social nature of action, which has lead to the increasing use of sociological methods in the area of CSCW. In conjunction with the sociological and ethnographic treatments of data comes a rejection of the empiricist experimental approach. Attempts to analyse the inherently messy and complex interactions in social settings uncover the limits of experimental methods. The very nature of the experimental method is to reduce the effects of confounding variables such that key independent variables can be manipulated and effects on dependent variables therefore measured. However, in these complex social settings there are often too many variables to adequately control for confounding factors. Equally, in a fashion perhaps similar to Heisenberg's uncertainty principle, in the very act of constraining an

interaction to observe it (perhaps by removing it from a naturally occurring setting and resituating it within a lab) it is being fundamentally changed.

Despite these criticisms the use of the experiment is a firmly established methodology within human behaviour research, and this is largely due to the benefits that such an approach confers. The first and foremost of these benefits are reliability (replicability) and objectivity (lack of bias in interpretation), both of which are strong arguments against qualitative methodologies. The use of the experimental method also allows certain forms of question to be answered in an easily interpreted fashion that would be notoriously difficult to approach from an ethnographic perspective. Examples of such questions include '*will collaborative performance be faster if I project remote gestures onto a shared work surface or present them on a separate monitor*?' Such a question would be extremely difficult to answer without collecting some form of empirical data and comparing the results statistically. The use of statistics to support experimental results is a standard measure and integral to the reliability of the method, as statistics are normally used to give clear indication of the probability behind the validity of any assertion made on the basis of the experimental results.

There are therefore acknowledged strengths and weaknesses to both quantitative and qualitative research methodologies. The answer then to which approach best suits the research questions proposed herein perhaps lies in the style of the research questions that are being asked. It is quite apparent that an ethnographic study will allow the situated analysis of work practices in a social context so if the research questions were to understand how people use remote gesturing systems in work practice, then ethnography would be a useful tool to evaluate a deployed technology. However, if one wished to understand the performance effects of changing various key aspects of a remote gesture system set-up (as stated above) then ethnography would be an unwieldy tool. A series of experiments would however enable a rapid and reliable assessment of the effects of changing system configurations whilst other potentially confounding variables could be held constant – so as to distil the pure effects of system change.

The research questions generated actually ask for analysis of remote gesturing systems at a variety of levels, considering both functional aspects of system set-up and more global issues of practical use of remote gesturing in collaborative action. What is needed therefore is an alternative methodology, a methodology which encompasses elements of both qualitative and quantitative traditions. Guidelines for such an approach can be found in Clarke's (2004) discussion of mixed methods in psychology and Shapiro's (1994) conception of Hybrid Methodologies. In a particularly cogent argument Shapiro explains the need to match methodology to research question and calls for social scientists to step forward from their entrenched positions and accept that effective research in an inherently multidisciplinary area such as CSCW will benefit from incorporating research from a variety of different perspectives, both the quantitative and qualitative and both the cognitive and sociological.

This thesis therefore takes a broadly ergonomic perspective to the research. As such the choices of methodology for addressing the research questions are open to a certain degree of latitude, considering both the proliferation of the experimental method in the early research and traditions of ergonomics and the new found move towards more social science qualitative methodologies. The research comfortably takes a pragmatic approach to the subject matter (in the literal rather than philosophical intention), and this is underpinned by a focus on user-centred technology design. This pre-eminence of the user-centred approach allows one to effectively *cherry-pick* research techniques as befits the research questions, building a hybrid methodology. For the purposes of this thesis therefore the experimental method is used to answer those questions where there is a need to test key factors of system design and ethnographic methods are used to further elucidate the nature of joint action and the role that gesture plays in structuring collaborative object-focussed interactions. This use of the ethnographic research method also extends to helping inform theory development.

## 3.4 Frameworks of Data Analysis

Having described the research questions which will be perused and articulated the methodological approach to the data collection it is perhaps germane to continue by briefly describing the theoretical basis and disposition that will be used to frame discussions of the research findings and to give strategy to the overall analysis.

Any thorough investigation of the nature of remote gesturing behaviour will find that there are several theoretical viewpoints to which one could subscribe when attempting to explain research findings in relation to relevant theory. As has been discussed this thesis exploits both quantitative and qualitative methodologies as appropriate when answering various questions. In a similar fashion, which is felt to be consistent with the overarching nature of this thesis and its interdisciplinary hybrid methodology approach to the research subject, accounts of gestural action are discussed from several separate perspectives, in the hope that consideration of the different approaches together will further elucidate the data.

One approach to the interpretation of the role of gesture in object-focussed interactions can be informed by an ethnographic sociological perspective. Drawing on the pre-existing CSCW literature, it focuses on the use of gesture as an awareness generating practice (Schmidt, 2002). And considers the various gestural practices or phrases that can be observed as ways for the mediating body to highlight objects for perception and make what Crabtree et al. (2004) describe as "a host of fine-grained grammatical distinctions". These actions, conjoining utterances (such as verbal prompts and instructions) to specific actions, in turn provide the coordination of tasks, the gestural phrases promoting projectability of intention and action. In an ethnomethodological sense there is also a desire to reject notions that gestures conform to specific categories or classifications of gesture (as equally suggested by prominent members of the psychology community, see Kendon, 1996). Such an approach to the role of gesture allows

one to analyse how a rich grammar of gestural action is implicated in the organization of interaction. This grammar enables participants to 'project' awareness of the tasks to hand and to integrate their actions accordingly. This consideration of awareness practices promotes the consideration of how collaborative work ecologies are structured and the impact that this has on the workers' action's situational relevance and intersubjective intelligibility, referred to as the *phenomenal coherence* of collaborative action.

An additional theoretical viewpoint is to approach the data from a more cognitivist perspective (in line with traditional HCI but falling out of fashion within the CSCW community). Accepting the critique of the use of cognitive theory in CSCW (Button et al 1995) that there is a lack of sociality in its perspective one could be drawn to the theory of Distributed Cognition, which has been proposed as a solution to this very problem. The nature of Distributed Cognition is to consider that cognitive activity does not reside solely within the individual but occurs in the functioning of task artefacts and the interactions of working colleagues. If one considers Hutchins' (1995) discussions of Distributed Cognition and descriptions of information representation passing and propagating between individuals and their task artefacts there is a suggestion that in group situations it is only through this flow of information that complex tasks can be achieved. It is therefore arguable that information is easier and quicker to access if the changes in representative state have been kept to a minimum and the translational overhead introduced by any mediating technology is kept to a minimum. This suggests that when remote gesturing is utilised in a collaborative physical task there would be different levels of translational overhead.

Without remote gesturing a helper can see items in a task space but cannot point to them. This means that they need to translate their visuo-spatial instructions into a verbal code which must be transmitted to the party who is in physical proximity to the task artefacts and then be decoded, introducing a significant overhead. In the presence of remote gesturing capability visuo-spatial references are kept intact. The helper can make gestural references (in a myriad of fashions, deictic, structural and dynamic and functional), which are aligned with their collaborator's visual perspective on the task. Therefore, references can be kept in a visuo-spatial medium when presented remotely reducing the requirement for complex verbal encodings. This reduction in the amount of processing required for the translation of information reduces the effort required in establishing conversational grounding (Clark, 1996, Fussell et al., 2004). This concept of considering collaborative interactions with a sensitivity towards the process of conversational grounding (Clark, 1996) is not exclusively a distributed cognition issue and consequently draws in a third perspective on understanding remote gesturing, this could in essence be characterised as a linguistic perspective on the analysis. An analysis from a linguistic perspective facilitates a more structural analysis of how gesture use is interleaved with patterns of language use during interaction, focusing on how the grounding of understanding is achieved but without necessarily relying on recourse to cognitive models of the phenomena.

We therefore have different perspectives on how to interpret the data that the research uncovers. From Distributed Cognition we gain the perspective of understanding how information is propagated through a communicative system, this leads us to interpret how system changes will affect the efficiency of information flow between the collaborators. From the ethnographic sociological perspective we can focus on the change in awareness that structural system changes bring between the parties and from a linguistic analysis we develop an understanding of how gesture influences language and supports the grounding of common understanding and communicative intentions. Essentially therefore these approaches allow us to consider a) how information is passed and manipulated and b) how that process influences people's understanding of that information, and at each level the thesis seeks to understand how the structure of the gesturing technology is influencing these processes. Therefore gaining a more thorough understanding of the nature of how gesture can be used to structure object-focussed interactions and how this is supported by the notion of the construction of mixed ecologies for communication.

A final caveat for consideration is that these differening perspectives suggest differing methodologies. As discussed in the previous section, the approach taken to data collection is as per the requirements of the research questions posed, but the appropriation of different perspectives on the data gathered is used to draw on the strengths of the different interpretative frameworks, not necessarily their adoptive research methodologies. It is their language and capacity for explanation of different levels of the data that is of particular use not their over-reliance on specific methodologies. Additionally, as remote gesturing tools are being studied, in this context, as generic tools not tied to specific working practices a highly contextualised investigation as befits an ethnography would not be suitable. Hence certain approaches to data collection such as the cognitive ethnography of distributed cognition have not been adopted as this has not necessarily suited the research questions being asked, but the explanatory frameworks it provides have been used to help eludicate some features of the observed data. As the research has become focused on different levels of analysis certain interpretative frameworks (such as the distributed cognition perspective) have inevitably receded from prominence as the focus shifts from observing flows of information to areas less comfortably explained by cognitive terms such as the mediation of awareness and the interaction of gesture and language at a structural level and the implications of this for the design, development and deployment of the remote gesture technologies.

## 3.5 A Remote Gesture Technology for Experimentation

### 3.5.1 Basic system set-up

To begin to explore the potential role of remote gesture in collaborative physical tasks, and to begin to test the benefits of designing remote gesture tools from a mixed ecologies perspective, it was clearly necessary to construct some form of remote gesturing device.

Considering the nature of the program of research that was intended it was decided that the technology need not be constructed so that it was mobile (despite acknowledgement that future deployments of the technology would need this functionality). As these early explorations were focussed far more on understanding the base effects of remote gesturing and the impact that fundamental changes in system construction might provide, it was decided that a desk-top based system would be built. The desk-top based system was constructed in a low-tech prototype form factor which facilitated lab-based experimentation and the rapid prototyping of system configuration changes.

As an element of the mixed ecologies approach suggests the use of unmediated direct representations of gesture, digital video capture and projection was used to construct the links between working spaces. The technology was therefore designed similarly to the Digital Desk system discussed in Wellner (1993), a system primarily used for interaction for 2-dimensional tasks but with the unexploited potential for adaptation to 3-dimensional use. The proto-type therefore consisted of a closed circuit system of digital video cameras, TV monitor and digital projector, as appropriate (schematics of the basic system design can be seen in figures 3.1 and 3.2).
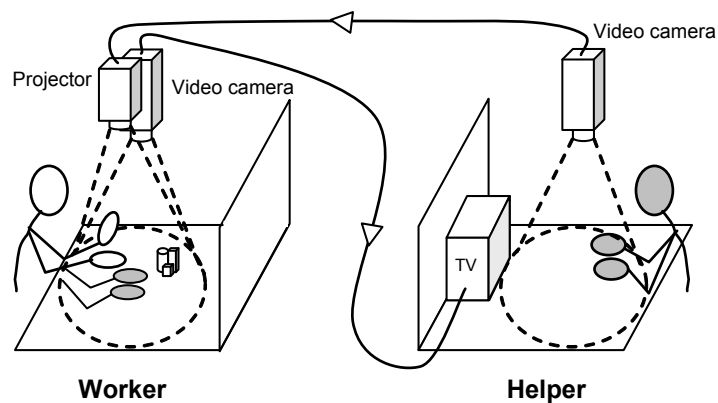


Figure 3.1 Voice + Projected Hands
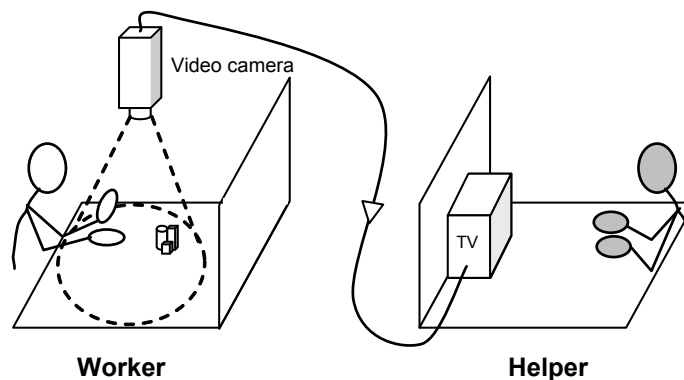


Figure 3.2 Voice Only Communication (Helper retains visual access to external workspace)

A working space was created for each participant (on a desk 60x80cm in size). A Sony MiniDV video camera (DCR-TRV900E) was held 90cm above the 'Worker's' desk, focused to capture the entire desk area. This resulting video image of the Worker's desk, their hands and anything they were manipulating, was passed via composite video cable to a 14" TV monitor (22x30cm, with standard TV resolution), on the 'Helper's' desk (see figure 3.1), the Helper being sat approximately 60cm from the TV monitor.

To allow the Helper's to project their gestures a second Sony MiniDV camera was positioned 90cm above their desk (again capturing the entire desk area). This video image was passed from the second video camera, via S-Video cable, to a digital projector (an A5 sized Sharp Digital Multimedia Projector PG-M10s with an SVGA resolution of 800x600). This was held 90cm above the Worker's desk, projecting a video image (approx. 40x53cm) of the Helper's hands and anything else on/over their desk space onto the centre of the Worker's desk (see figure 4.1 for illustration). The Helper's hands were therefore projected onto the Worker's task space, the Helper being able to guide their hand movements in relation to task artefacts on the Worker's desk by viewing video feedback from the Worker's desk presented on their TV monitor. To construct comparisons with non-gesturing, but still video-linked environments, the video camera above the Helper and the projector above the Worker could both be disabled, giving a connection which worked as shown in Figure 3.2.

The system was constructed such that both participants would be in the *same room* during use, but only had visual access to each other and each other's desks through the mediating technology – partitions ensuring that direct visual access was blocked (see figures 3.3 and 3.4 below). This enabled the use of *full audio* in all studies without having to use any audio communications technology. The role of differing qualities of audio connection being evaluated in other studies and having been shown to have insignificant impact on performance (Kraut et al 1996); it was therefore felt that further evaluation of audio links would be of little benefit, and given the desire to use low-tech prototyping to promote rapid evaluation of technology, it was considered to be of little priority to further control the audio set-up.

Wooden frames were constructed to hold the recording and projecting equipment. These frames inadvertently became much more than props for holding equipment, as they simultaneously blocked visual access between each participant and clearly delineated the task space, as all assembly and gesturing had to be conducted within the confines of the frames. Therefore there was an explicit construction of a public-private space divide for the collaborators, which was artificially derived from the constraints of building a working technological solution to the gesture representation problem.

Figures 3.3 Frame 2

Figure 3.4 Frame 1 (left) and the back of Frame 2 (right)

In figure 3.5 below, a representation can be seen of the working surface as it appeared to the Worker during interaction. The projection of the Helper's desk can be seen overlaying the task artefacts that the Worker is being instructed to manipulate. This view is also correspondingly exactly the image presented to the Helper on their own TV monitor, as they viewed the task progress and coordinated the movements of their hands in relation to the task artefacts.

Parts for assembly
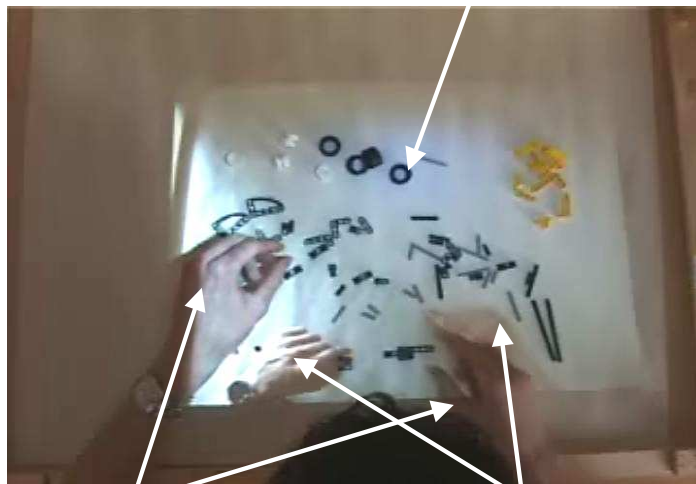


Local Worker Hands          Remote Helper Hands

Figure 3.5 Gesture Projection System in use

One feature of note that can be observed in figure 3.5 is that the relative orientation of the collaborator's hands suggests that they are effectively sitting on top of each other. This is in itself an unfamiliar arrangement, and dissimilar to comparative co-present working interactions. Other options were available in which re-orientation of the remote collaborators' gestures could be achieved by the relative rotation of the projection and capture devices, causing the projected hands to enter from either a lateral or 'across the table' orientation. The

effects of different approaches were left for experimental comparison and a decision had to be made about which approach the basic system set-up utilised. To keep the system relatively consistent with other related remote gesturing devices, this common orientation (overlapped hands) was chosen as the default setting.

An additional concern that needed to be addressed in the basic system design was how the remote Helper was able to align their gestures with task artefacts (i.e. how the task artefacts were presented to them). Whilst some might consider that it would make most sense to allow Helpers to align their gestural actions with a projection of the Worker's task space onto their own desk space (as seen in systems such as AGORA, Kuzuoka et al 1994) a decision was made to have Helpers align their gestures with the Worker's task artefacts by viewing their hands on the live video feed. This decision was motivated by two factors. The first was a desire to stay true to the ideal of providing a low-tech prototype to motivate design and experiment (projecting onto both desks requiring technological intervention to eliminate costly problems of video-feedback loops). The second factor was a realisation that from a mixed ecologies perspective the implicit ability to be aware during production of how your gesture will be perceived may have a significant impact on how you produce that gesture. When we gesture whilst side-by-side we are implicitly aware, through an understanding of relative perspectives, of how our collaborator will view our communicative gestures, if the collaborative environment is split, such as in a remote interaction, there is potential for these implicit links to be inadequately re-established. When gestures are to be received in a two-dimensional format, such as when they are projected after video-capture, it is potentially of benefit for the Helper, to be able to produce their gestures also in a two dimensional environment, such as by watching their hands on a TV monitor. Of course this is in itself conjecture, and would need to be adequately established by research. But the issue remains that there are two potential methods for setting up gesture production at the Helper's end of the interaction, with no conclusive evidence as to which is best. Left therefore with a decision as to whether projection should be to one desk only or to both, as the production of gestures was not to be considered as an explicit experimental variable a decision was made to opt for one strategy of gesture production only.

### 3.5.2 System re-configurations

Later stages of the experimental work of the thesis required that modifications be made to the basic system set-up to allow for the experimental comparison of different system configurations. When constructing remote gesture technologies there are various factors which can be modified to affect how the device works. Two key factors of remote gesture system design were identified as target areas for experimental modification.

The first factor was the location of the gesture output (to the person located with the task artefacts), with the remote gestures being either projected directly into a task space or

alternatively presented to the remote worker on an external VDU showing an overlay of the workspace and the gestural information (these represent the two current trends in remote gesture tool design). A second factor that can be modified is the format of the gesture. For the experiments in this thesis three different formats of gesture, unmediated representation of hands, unmediated hands with a sketch facility and digital sketch only were explored. Mentioned further in section 5.3.1 the basic premise for the inclusion of these formats being that this thesis has chosen on the basis of prior research evidence to focus more closely on rich forms of gesture representation, which have been shown to be of particular benefit in collaborative physical tasks (consequently excluding from consideration simplified gesturing systems such as the laser dot systems sometimes explored). The combination of these location and format factors gives six different possible system configurations, displayed in table 3.1 (these six combinations represent the specific conditions that were compared in subsequent experiments).

| *Gesture Format* | *Gesture Location* | |
| --- | --- | --- |
| | *Projected* | *TV* |
| *Hands only* | Projected hands | TV hands |
| *Hands and sketch* | Projected hands & sketch | TV hands & sketch |
| *Digital sketch only* | Projected sketch only | TV sketch only |

Table 3.1 Comparison of possible gesture locations and formats

To construct these various system configurations low-tech prototypes were assembled, modified from the basic set-up discussed above (as shown in figure 3.1).

The modification that was incorporated to allow the workers to see gestures on a separate video window, as opposed to the previously used projection, was the removal of the video projector, and the inclusion of a second TV (one for the Worker as well as the Helper). To create the effect of remote gesturing, a 4-channel video mixer unit was incorporated (Videonics MX-1) and using the manual T-bar control, a 50% transition image was created of both video feeds overlapped. This mixed live video feeds from above both the Worker and Helper desks (see figure 3.6a for schematic and 3.6b for an annotated image of a typical TV overlay view). This enabled the Helper's to guide their hand movements and gestures relative to the shared task artefacts by looking at a video window, exactly as in the original system set-up.
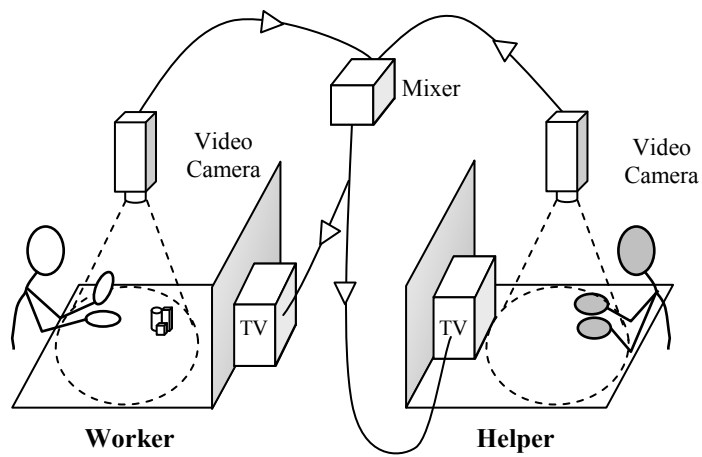
Figure 3.6a Video presented Hands (schematic)

Parts for assembly



Local Worker Hands          Remote Helper Hands
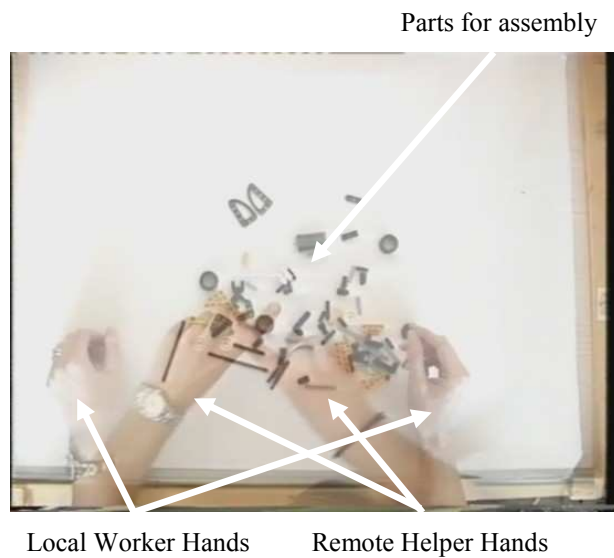
Figure 3.6b Video presented Hands (screen capture)

The system was extended for analysis of the 'Hands & Sketches' condition, by facilitating remote sketching by giving the Helpers a board marker pen, and allowing them to write on the surface over which they were gesturing (which was a dry-wipe whiteboard). The resultant sketches either A) being projected onto the workers desk (see figure 3.7) or B) video mixed over their video feed of the task space (see figure 3.8). Example illustrations of the gesture output for each of these systems can be seen in figure 3.9.
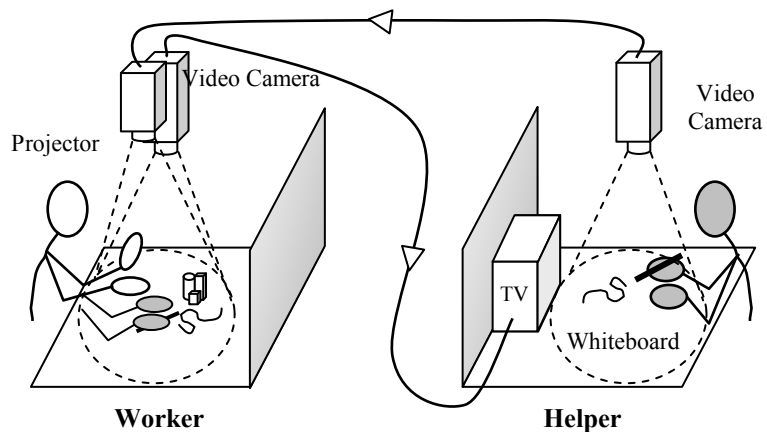
Figure 3.7 Projected Hands & Sketches
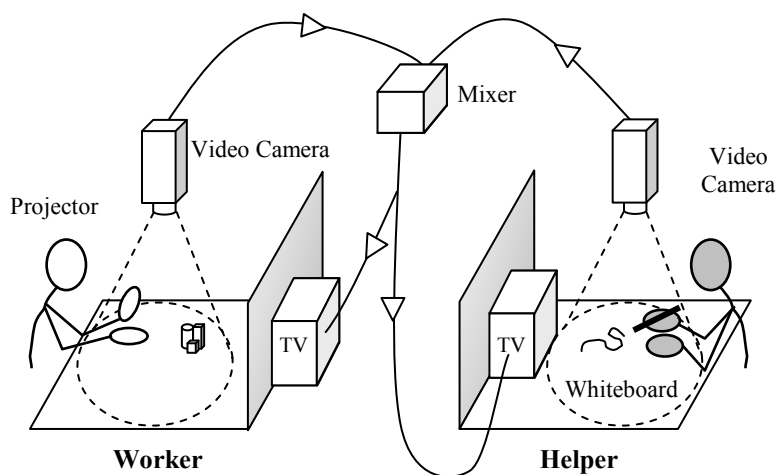


Figure 3.8 Video presented Hands & Sketches
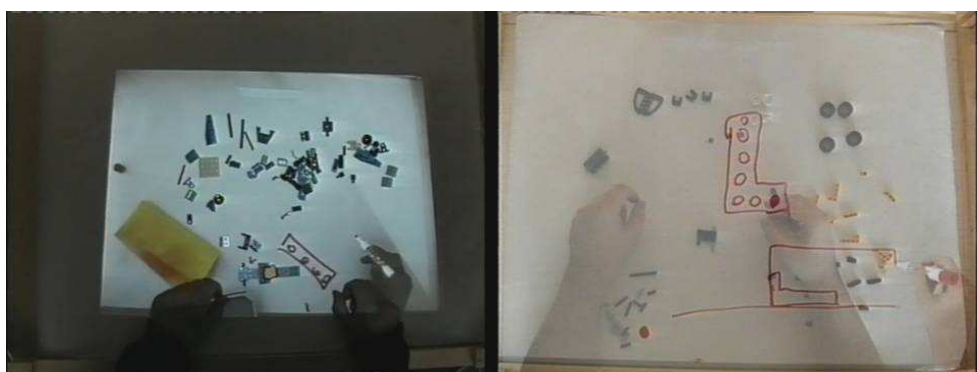


Figure 3.9 Projected (left) and Video (right) presented Hands & Sketches (in each case image captured from Helper's TV view)

To produce a 'Sketches only' configuration a significant change was required in the technology. As a video feed from above the Helper's desk was no longer required the camera that normally occupied this position was removed. The gesturing / sketching surface that had previously been used was also replaced with an A2 sized Wacom Tablet. The tablet was connected to an IBM Pentium III PC (with 547 MHz CPU, 256MB RAM, and a 32MB Matrox Millenium G400 graphics card) running MS Paint V.5.1 on Windows XP. The PC was set to an output resolution of 1024x768 pixels in 32bit colour. The output of this paint program was then presented to the Worker through a projection onto their desk (see Figures 3.10a+b) or presented mixed over a live video feed of their task space (see Figure 3.11a+b).
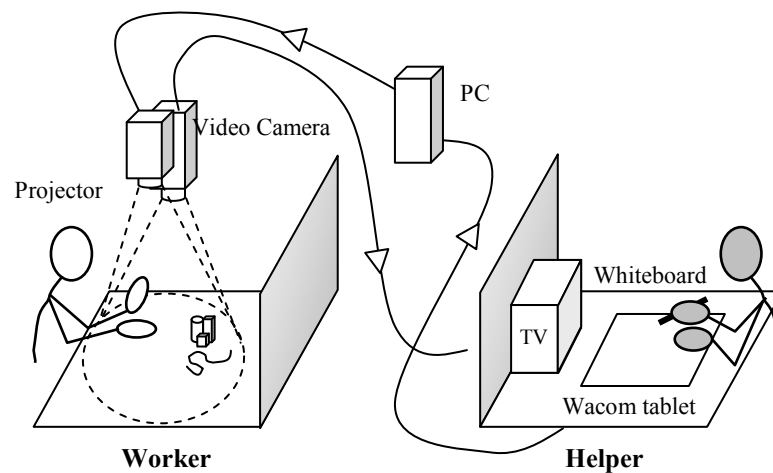


Figure 3.10a Projected Sketches only (schematic)
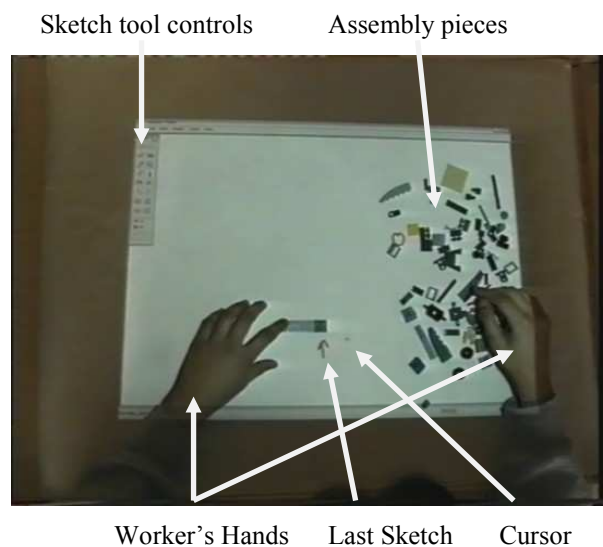


Figure 3.10b Projected Sketches only (screen capture)
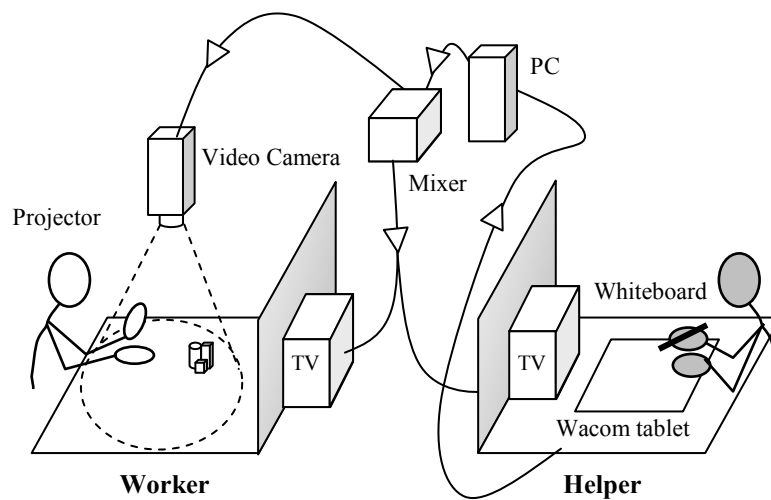
Figure 3.11a Video presented Sketches only (schematic)
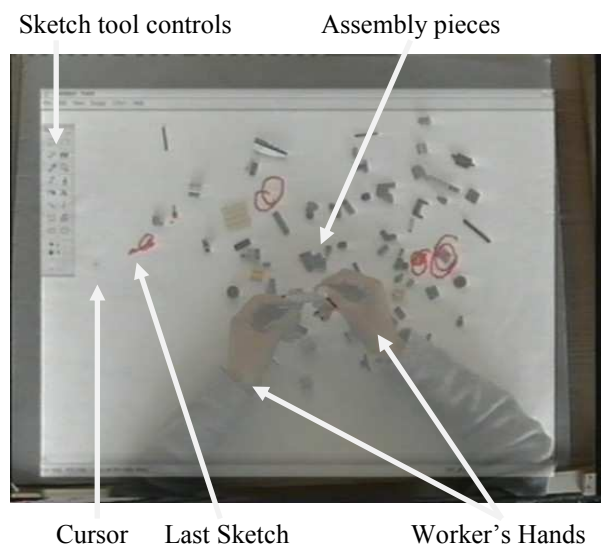


Figure 3.11b Video presented Sketches only (screen capture)

To achieve the video mixed configuration the PC output was altered accordingly using a SCAN converter (not shown in the diagram) before being passed to the video mixer unit. In the case of the projection configuration a video camera above the Worker's desk picked up the projected MS Paint interface and presented this along with whatever else was in the task space on the Helper's TV. In the video window configuration both collaborating parties saw exactly the same video mixed images of live video feed of the task space and MS Paint interface, presented on their TV's. In both cases the MS Paint interface was presented to the Helper with enough resolution to enable them to manipulate the required controls to use the paint package with the tablet and pen (by guiding their actions as per feedback from their TV monitor – this is in essence exactly how the Wacom tablet is used in a standard PC / VDU set-up). During

use, and as a further aid, a prompt sheet (with an enlarged image of the interface) was given to the Helper, with a series of common interface manipulations highlighted (see appendix 3.1).

With the Sketch only configurations another system design consideration presented itself, namely the removal of sketch information from the workspace. Previous work has demonstrated the potential benefits of automatic remote sketch erasure (Fussell et al 2004). Whilst installation of an auto-wipe system for the Sketch only system would have been possible, implementing it in the low-tech prototype Hands and sketch system would have had significant design implications. A design decision was therefore made that a manual wipe should be adopted for the Sketch only configuration to keep consistency between the features of the various systems for experimental comparison. Therefore removal of sketches during the use of the Hand and sketch configurations was achieved by manually wiping the desk surface, and in the Sketch only configurations this was achieved by using the digital pen to highlight an area for removal and then using a button on the pen to select a 'clear' function. Pilot-testing did demonstrate that there was still a difference in the relative ease of use of the two deletion methods. Acknowledging that previous research had suggested method of deletion to be of possible interest it was felt that should experimental work reveal a significant performance difference between the Hands and sketch and Sketch only configurations, the system could be further re-configured to facilitate specific comparison of the relative effects of auto-wiping versus manual erasure techniques.

## 3.6 Chapter Summary

This chapter has formed a hypothesis on the basis of evidence from the literature review that the best way to support remote collaborators in the accomplishment of collaborative physical tasks is to design remote gesture tools that seek to create mixed ecologies (or are at the very least designed sensitive to this notion of designing for the construction of mixed ecologies), as an approach to alleviating the problems of current forms of VMC fracturing interactions. On the basis of this hypothesis a series of testable research questions have been generated, which seek to interrogate specific aspects of this notion of mixed ecologies and the role of remote gesturing in supporting them and the process of remote collaboration itself. The chapter has also provided an in-depth discussion of the methodological disposition that has been taken to the process of data collection, and explored the differences between possible research methodologies, and on the basis of this made an argument for the methodologies chosen. Given the interdisciplinary nature of the research and the varying styles of research question that must be addressed to provide in-depth understanding of the central issues of the thesis, a framework of hybrid methodologies has been adopted, utilising both quantitative and qualitative techniques as appropriate.

A large portion of the chapter has also been given over to an explanation of the remote gesturing system that has been constructed for use in the following empirical work of the

thesis. Reasons behind the various system design choices that have been made have been explored and the ways in which the basic system set-up can be modified to experiment with alternative system configurations have been presented and discussed.

## Chapter 4 – Some Effects of Remote Gesturing

### 4.1 Introduction

This first chapter of empirical work explores the efficacy of remote gesture tools and in particular remote gesture tools designed from a mixed ecologies perspective. The chapter presents two experimental studies of user performance with the basic remote gesturing system presented in section 3.5.1. The chapter provides some basic data about what actually happens to collaborative performance when a remote gesture tool is used. This is important because it begins the process of developing an investigation of one of the fundamental research questions namely how and why a representation of gesture might improve remote communications.

Prior research highlighted in section 1.2 of chapter 1 has already demonstrated some of the performance benefits of remote gesture technology, suggesting that when a remote gesture tool is used, collaborative physical tasks can be performed at a faster rate. Part of the work of this chapter is to confirm such a finding, to further demonstrate a simple collaborative performance enhancement of these technologies. The evidence for this prior assertion, that remote gesturing leads to faster performance is somewhat limited, only the DOVE system of Fussell et al (2004), being actually tested in this experimentally verifiable method. The work of the GestureMan studies (Kuzuoka et al, 2004b) not ever actually seeking such direct analysis of performance enhancement. If similar performance benefits to the DOVE studies can be demonstrated with the significantly different remote gesturing arrangement proposed in section 3.5.1 then the argument that remote gesturing improves performance becomes significantly bolstered. This is further justified because an ability to demonstrate the benefits of remote gesturing with this alternative arrangement will also demonstrate that it is the principle of remote gesturing per se, rather than some epiphenomenal aspect of the DOVE system that currently improves collaborative performance. To elaborate this point further one needs to consider the general tone of the research findings of the two major lines of research presented in chapter two, namely the GestureMan and the DOVE studies. With the DOVE studies because of the inherently experimental nature of the work, the results as presented, consistently strive to affirm the benefit of their particular approach. However, with the GestureMan studies, the analysis often comes from an approach of sociological critique which is naturally more biased towards investigating the faults of the system. If taken at face value without a critical overview, one could possibly surmise that the GestureMan style of remote gesturing is of little benefit whilst the DOVE system is somewhat infallible.[4] The work of this chapter and in particular the first experiment strives to demonstrate comparable performance effects with a

---

[4] Although a reading of Kramer et al 2006 (see p. 63 for discussion) ensures that this is not true.

significantly different remote gesture tool, to demonstrate the overall benefit of remote gesturing.

As a part of attempting to understand how remote gesture 'benefits performance' it also becomes important to understand what this means. Prior work in this area has generally focussed on issues of how remote gesturing makes collaborative performance quicker[5], but this is not the only facet of performance that might be affected by an intervention in the structure of the mediating communications technology. The two experiments presented in this chapter further lend themselves to examination of other ways in which remote gesturing improves remote collaborations.

In the first experiment there is a comparison of performance in a collaborative physical task in two communication conditions, using a simple (non-gesturing) VMC link and using a VMC link with an enhanced gesturing facility (as presented in 3.5.1). This obviously allows investigation of how remote gesturing can speed up performance, therefore achieving the aims discussed above, but this approach also lends itself to the investigation of more cognitive aspects of collaborative performance. By evaluating the perceived (self-reported) mental workload during the use of each communications technology the first experiment of this chapter begins to map out the cognitive performance benefits of remote gesturing technologies.

If however, this chapter focused solely on the physical performance and the cognitive performance benefits, of remote gesturing, it would not be fully exploring the potential of remote gesturing technologies and would leave a significant aspect of collaborative performance unaddressed. It is clearly evident from the prior research literature that 'collaborative physical tasks' can be framed as a class of 'mentoring' task (see Kraut, Fussell and Siegel, 2003). In many of the potential applications of remote gesturing technologies it is likely that collaborators will have a disparity of knowledge levels, and in essence the interaction will become a tutoring one in which an expert is guiding the actions of a novice, essentially attempting to *impart* knowledge. In these situations then, a remote gesture tool is being used as an instructional aid. No prior research has considered however, whether or not such a device has an impact on the learning that takes place during interaction. Clearly if the technology either aids learning (i.e. the long term retention of task related knowledge) or indeed, importantly, if it hinders learning, then it has an impact on what could be termed, collaborative performance.

The second experiment presented in this chapter provides an experimental comparison of task-knowledge retention after instruction in either of two conditions, again comparing the simple (non-gesturing) VMC link and the VMC link with an enhanced gesturing facility. This second experiment therefore seeks to explore this further basic facet of collaborative performance,

---

[5] This being said, some work has focused on the impact of remote gesturing on the use of collaborative discourse, as this is a somewhat substantive issue, it is addressed at length in its own chapter (chapter 7).

namely the influence of remote gesture technologies on the learning of a physical task. In doing this it further develops understanding of the various ways in which remote gesture tools affect collaborative performance.

The rest of this chapter is given over to a discussion of the specific methodologies and results of each of the two experiments in turn, and concludes by discussing the overall implications that they have for understanding the basic effects of remote gesture tools on collaborative performance.

**4.2 Comparing Remote Gesture vs. Voice Only Communication**

**4.2.1 Study methodology**

**4.2.1.1 Experimental design**

The study was conducted using a within-subjects repeated-measures design, which had one independent variable, communication method, consisting of two levels, audio-visual only communication (comparable to a simple video link between spaces) and audio-visual plus gesture communication (a video link with the enhanced capability of projecting remote gestures). Each pair of participants experienced both conditions[6] whilst they collaboratively performed a Lego construction task (in which one participant had the instructions and the other the Lego pieces). Lego was chosen as it represents a generic object-focused task and is comparable to the tasks utilised in previous work in other studies (see Fussell et al 2004 and Clark and Krych 2004). To accomplish any given stage of a Lego model requires a variety of actions including assemble, disassemble, rotate, align, search and select. Participants were randomly assigned to one of two roles 'Helper' or 'Worker'. To create expert status in the Helper they were given a diagrammatic instruction manual of how to construct their given Lego model (they had the manual only and no access to their own set of reference Lego blocks). They were instructed that whilst they could talk at all times to the Worker, providing both verbal and in the trial when possible, gestural instruction, they were not allowed to show the Worker any of the manual. As the nature of the Lego pieces precluded any guessing of how they should be put together, and the Worker had no visual guide of what the end model should look like, the Worker relied completely on instruction from the Helper.

The dependent variables included task performance speed, task completion rate and a subjective rating of mental workload. Each participant swapped roles between trials and a different assembly of the Lego pieces was utilised, order of trials was also counterbalanced across pairs, these measures were taken to counter the potentially confounding variables of learning bias and order effects.

**4.2.1.2 Participants**

A total of 48 participants took part in the study, 11 female pairs and 13 male pairs. Participants' ages ranged from 18-26 (mean 20.83, st. dev. 1.59), and they were mostly Computer Science (males) and Psychology (females) undergraduates (participants from arts backgrounds were also included). The first 12 participants were used as a pilot trial; their data was determined to be of acceptable enough quality for use along with the main group. Participants were paid £6 each for taking part in the study.

---

[6] the experiment was not however a full-factorial design with all participant's experiencing all trials, as this was felt to be inappropriate given the time constraints of running subjects.

*4.2.1.3 Equipment*

The study utilised the basic remote gesturing apparatus described in section 3.5.1, including both instantiations of the system (i.e. the 'with gesturing' and 'without gesturing' VMC arrangements, see figures 3.1 and 3.2 respectively, both p. 77).

*4.2.1.4 Materials*

The experiment used several copies of the manufacturer's instructions for assembly of a Lego kit (model no. 8441, in both possible assemblies, see appendix 4.1), plus various questionnaire materials including NASA TLX for subjective assessment of mental workloads (including both the subscales and the paired-comparisons forms) and a bespoke evaluation questionnaire (see appendices 4.2 and 4.3).

**4.2.1.5 Procedure**

Participants were paired (same sex pairs) and invited to the lab. Prior to the trials starting participants were asked to read an information sheet outlining the structure of the experiment (see appendix 4.4) and were also asked to sign a consent form (see appendix 4.5). Once this was completed they were shown the experimental equipment and were told how it works they were then allowed to practice with it – collaboratively constructing small model toys (each of which only had three parts to connect, see appendix 4.6). Participants were made to experience the gesture projection from both sides so that they would have experience of both creating and decoding projected gestures before the experiment started.

During the experiment participants were asked to collaboratively assemble a Lego kit. One participant being randomly assigned the Worker role and one the Helper role. The Helper was given the kit's instructions and could see the task space (via video camera - TV link), and was told they could not touch the pieces. In contrast the Worker was told that they could touch and manipulate the pieces but they could not see any instructions. Participants were instructed that they could talk freely to each other at all times. Each pair was given 10 minutes to assemble as much of the kit as they could (no pairs ever managing to complete a model within 10 minutes). This was done in both of two conditions; in one condition the set-up was as described above, but in the other condition the Helper could put their arms and hands out in front of them (over the surface of their desk, and within the bounds of the frame on their desk) thereby having their gestures picked up by the video camera held above them and having the image of their hands and arms projected over and into the Workers' task space. This projection was fed back via video link to the Helper so that they could see where their gestures were in relation to items in the task space and guide their movements appropriately. Participants swapped roles between conditions.

After each condition both workers completed a measure of mental workload (NASA TLX) and after both test conditions were completed they filled out the evaluation questionnaire and were debriefed.

### 4.2.1.6 Problems encountered

The experiment ran smoothly with few problems encountered. There was some worry that the potential learning bias may have swamped the experimental effect as participants progressed through the two trials, so after the pilot trial it was decided to incorporate a more formal introduction to the gesture system before the experimental trials began. This led to the brief toy assembly task outlined above (section 4.2.1.5). It was hoped that this exposure would result in a reduced lag time when getting to grips with using the technology for the first time (a consideration which was entirely redundant as will be discussed in further sections).

Some problems were encountered in a handful of trials as the projection equipment occasionally turned itself off during moments of overheating. Participants were informed of the possibility of this happening prior to the trials and were instructed to continue with their tasks if it did happen. On the occasions when the projector failed it recovered function within 5 seconds and resumed as normal, so it was felt that it would not have adversely affected the experimental results at any point. Certainly considering the nature of the device and the working ecologies that these studies are purporting to investigate it seems appropriate that communication delivery should not be seamless, as in most working situations there is a degree of 'interruption' which cannot be controlled, and therefore such interruptions were considered to be healthy 'noise' in the data.

In one instance a participant complained repeatedly that they were unable to see clearly what was going on, even after stopping briefly to put on their glasses, they felt that their vision was impairing their ability to perform the task. The data from this pair of participants was therefore excluded from the analysis and their trials were re-run using a new pair of participants.

### 4.2.1.7 Statistical analysis

Statistical analysis of the data consisted of a general trends analysis looking at mean scores for subgroup comparisons and where appropriate, as indicated by an observable trend, t-tests were conducted to assess statistical significance. T-tests were the only tests required as only two subgroups of data were ever being compared at a time. An ANOVA was not used as the design of the experiment was not a full-factorial, and would not therefore fit an ANOVA model. The measures of Mental Workload were first analysed using the weighting score for the NASA TLX (derived using the paired comparisons form), however, results for this were somewhat inconclusive and based on the evidence of Byers, Bittner & Hill (1989) and Fairclough (1991) it was decided that an analysis should also be conducted on the mental workload sub-scores. Again trends were analysed using measures of central tendency and those measures that

appeared most interesting were singled out for further statistical comparison across the key factor.

## 4.2.2 Results

There were a variety of different dependent variables measured for the study. Analysis of the raw data suggested that comparing the average final stage of completion for this experiment was offering little utility, whilst time to complete the first three stages of each model did prove to be interesting; consequently the analysis of task performance was based on this measure. Statistical analysis was also performed on the measures of Mental Workload (and its sub-scales), results for these analyses are presented in turn.

### 4.2.2.1 Performance times

Each pair of participants experienced two experimental trials; in one trial the feedback from Helper to Worker was voice only and in the other trial feedback was via voice <u>and</u> gesture. Order of presentation of the two trials was counterbalanced across pairs of participants, therefore half the participants experienced the voice and gesture condition in the first trial and the other half experienced it in the second trial. Table 4.1 below summarises the results for participants across both trials, showing the mean time in seconds to complete the first three stages of the model they were working on.

| Condition | First Trial | Second Trial |
|---|---|---|
| Voice Only | 227 (87.70) | 165.75 (86.84) |
| Voice and Gesture | 164.33 (41.31) | 164.08 (41.52) |

Table 4.1 Time in seconds to complete first three stages (standard deviation in brackets) (N= 24 pairs)

It is interesting to note that in the voice only conditions (in both first and second trials) the variability in performance times is over twice that of the performance in the gesturing conditions (for which performance was very consistent over the two trials). Figure 4.1 below illustrates the changes in performance between trials.

**Time to complete first 3 stages of model by trial order**
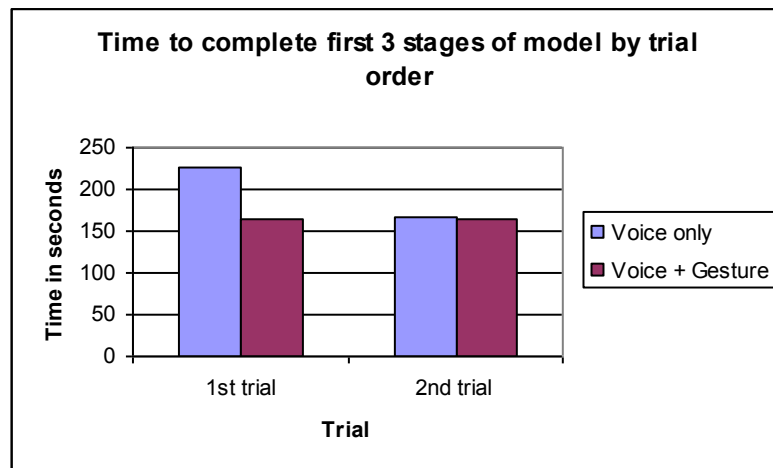
Figure 4.1

As the trials were appropriately counterbalanced the data can be collated from both first and second trials and used to compare performance across the two levels of the key condition, in a within-subjects comparison, this is represented below in figure 4.2

**Time to complete first 3 stages of model by communication method**
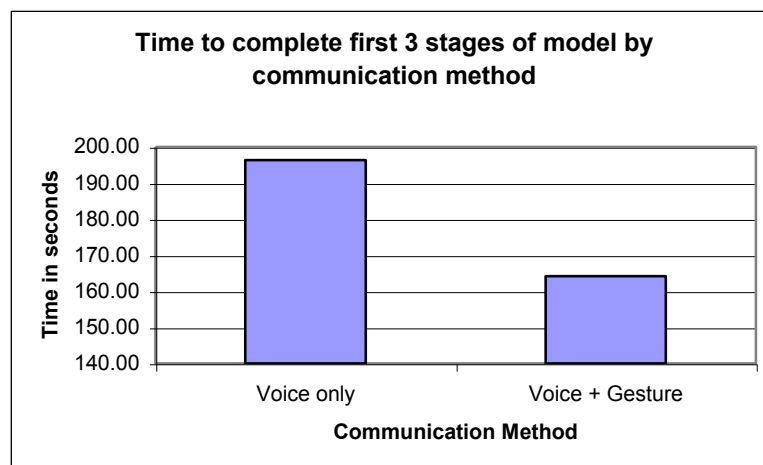
Figure 4.2

Figure 4.2 provides evidence for the presence of an experimental effect. Despite the counterbalancing, it appears that people were performing quicker when gesture projection was being used.

The difference between the two groups (voice only and voice and gesture) was found to be significant (one-tailed repeated-measures t-test ($t(23) = 1.87$, $p = 0.037$)). This was repeated for a between groups analysis of the first and second trials. In the first trial the difference between the two groups (voice only and voice and gesture) was found to be significant using a

one-tailed independent-samples t-test (t (22) = 2.24, p= 0.018). However, the difference between the two groups in the second trial was not statistically significant (using a one-tailed independent-samples t-test (t (22) = 0.60, p= 0.48)).

The results therefore show that for a measure of performance speed based on the time in seconds to complete the first three stages of any of the models used, performance was significantly improved when participants used the gesturing system. The results indicate that the difference is significant for the first trial but not the second. A reading of the results might suggest that that it takes a certain length of time to establish common grounding (Clark, 1996), and this can be seen from looking at the results associated with the voice only condition. It could be argued therefore that the use of a remote gesturing system appears to significantly reduce the amount of time it takes to develop a common grounding.

Having observed the significance of the experimental effect, it was thought appropriate to assess the data for any sign of learning bias over the trials. Several key variables had been counterbalanced in the design of the experiment to try to eliminate any such effects, because it was feared that the influence of learning the task may swamp the experimental effect. Table 4.2 below summarises the difference in performance times over the two trials, and this relationship is illustrated in Figure 4.3.

| Trial | Times |
|-------|-------|
| First | 195.67 (74.28) |
| Second | 164.92 (66.57) |

Table 4.2 Time in seconds to complete first three stages for the First and Second trials (standard deviation in brackets) (N= 24 pairs)
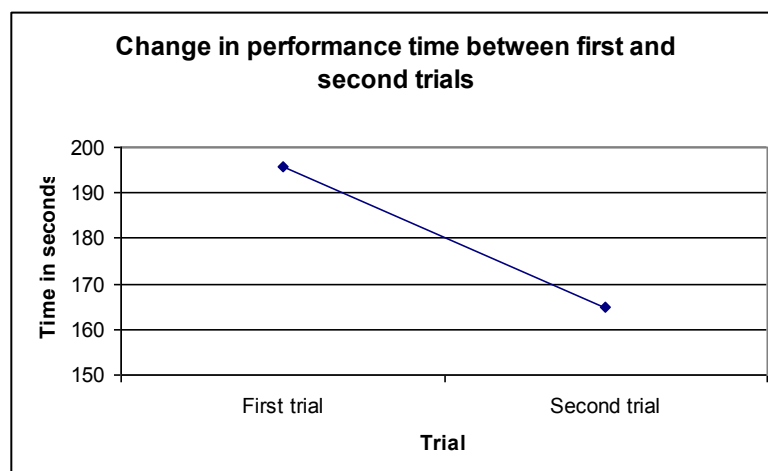


Figure 4.3

The difference between the two groups (First trial and Second trial) was found to be significant using a one-tailed repeated-measures t-test (t (23) = 1.78, p= 0.044). Clearly therefore, despite best efforts to counterbalance there was still a clear effect of learning on the task. Participants were becoming much better at coordinating their interactions and were becoming more successful with the task over time.

The final factor of interest regarding performance time differences was the effect of building the different models. The decision to time participant's performance over the first three stages of each model was based largely on the prevalence of clear audio-visual markers for the beginning of stage four of each model which facilitated the timing of the performance over the initial three stages. Due to this there were potential differences in the content of the first three stages of the Car and Forklift models, which consisted of types of pieces to be used, number of pieces to be used and types of manipulations to be performed. Table 4.3 below illustrates observed differences in performance for each model.

| Model | Voice Only | Voice and Gesture | Total |
|-------|-----------|-------------------|-------|
| Car | 155.42 (60.13) | 155.92 (45.6) | 155.67 (52.19) |
| Forklift | 237.33 (100.01) | 172.5 (34.65) | 204.92 (80.34) |

Table 4.3 Time in seconds to complete first three stages for each model in each condition (standard deviation in brackets) (N= 24 pairs)

As the construction of the two different models was counterbalanced across the voice only and voice and gesture conditions the time to complete three stages of each model can be compared. This demonstrates that the first three stages of the forklift appeared to take longer to construct than the same number of stages of the car. This is illustrated in figure 4.4.
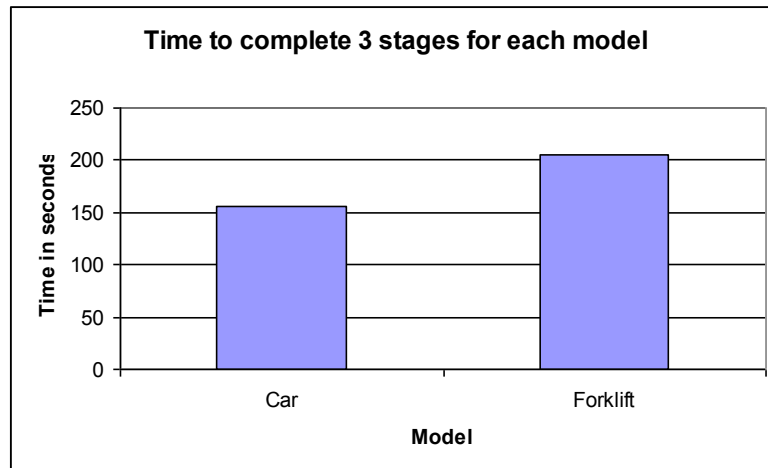
Figure 4.4

However, the figures would also suggest that the use of the gesture system made no impact on performance for the car, but did affect performance when assembling the forklift model. These differences can be seen below in figure 4.5.
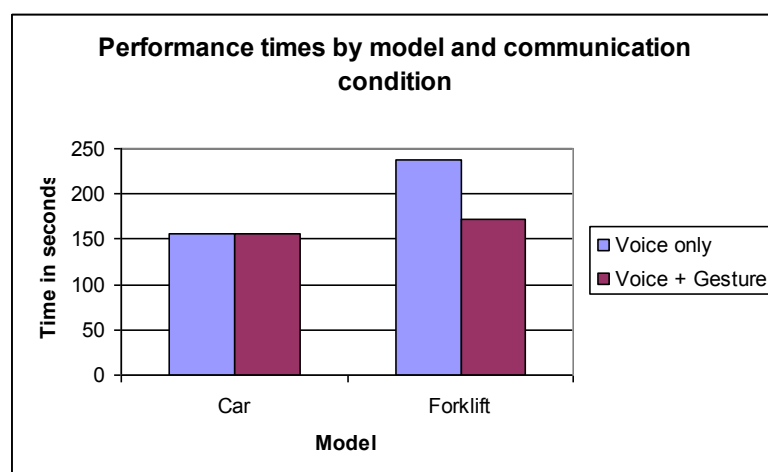


Figure 4.5

The participant's performance difference between the two models (Car and Forklift) was found to be significant using a two-tailed repeated-measures t-test (t (23) = -3.22, p= 0.004). The impact of the gesturing technology on performance for each model was then assessed. For the Forklift model the difference between the two groups (voice only and voice and gesture) was significant using a two-tailed independent-samples t-test (t (22) = 2.12, p= 0.045). However, the difference between the two groups for the Car model was not statistically significant using a two-tailed independent-samples t-test (t (22) = -0.02, p= 0.98). The statistical analysis

therefore confirmed that participant's behaviour had been influenced by the use of the gesturing system over the first three stages of model assembly for only the Forklift model. The reasons for this are not however clear, but the conclusion that should perhaps be tentatively drawn is that there is evidence to suggest that factors concerning task structure and requirements influence the necessity to use gestural support. Potentially it may be that in some situations language use alone leads to rapid orientation to the task, and there are no benefits to be had from gestural support.

*4.2.2.2 Mental workload analysis*

Having thoroughly investigated the data for performance times attention was focussed on the data from the assessment of mental workloads. As discussed Mental Workload was assessed via the administration of the NASA TLX, using both sections of the assessment the sub-group scales and the paired comparisons section. This measure gave a score out of 20 with 20 representing the highest possible level of mental workload. A breakdown of scores for various key comparisons can be seen in table 4.4.

| Sub-group | Mental Workload score |
|---|---|
| Helper | 12.08 (2.72) |
| Worker | 10.45(2.92) |
| First trial | 11.48 (2.91) |
| Second trial | 11.06 (2.95) |
| Voice only | 11.44 (2.88) |
| Voice and Gesture | 11.10 (2.99) |

Table 4.4 Mental workload scores for Helpers vs. Workers, first trial vs. second trial and voice only vs. voice and gesture conditions (standard deviation in brackets) (N=48)

The mental workload trends suggested that the Helpers found the task harder than the Workers (NB, it is of importance to remember that participants will have experienced both of the roles), the first trial was rated as more demanding than the second trial and the voice and gesture condition was reportedly less demanding than the voice only condition. These results are illustrated in figure 4.6.
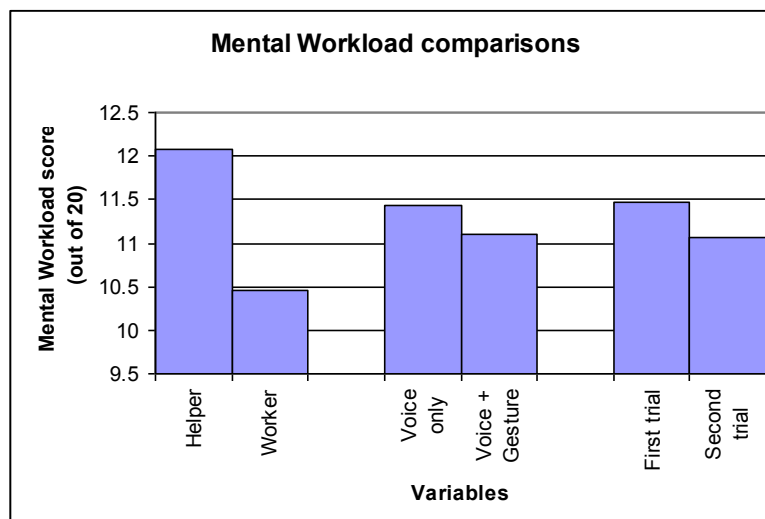
Figure 4.6

The differences between sub-groups in each pair were assessed for statistical significance before conclusions could be drawn. The difference in participant's ratings of mental workload were compared between Helpers and Workers, the difference between the two groups was found to be significant using a two-tailed repeated-measures t-test (t (47) = 3.33, p= 0.002). However, the difference between the voice only and voice and gesture groups was not significant. Equally, the difference between the first and second trials was also not significant.

As there was a significant difference found between the rated mental workload of Helpers and Workers but no effect was found using the gesture technology it was determined to be appropriate to compare the effects of using the gesture technology on each of the Helper/Worker sub-groups. For the Helpers there was no significant difference between those in the voice only condition and those in the voice and gesture condition using a one-tailed independent-measures t-test (t (46) = 0.41, p= 0.34). Again, for the Workers there was no significant difference between those in the voice only condition and those in the voice and gesture condition using a one-tailed independent-measures t-test (t (46) = 0.40, p= 0.34). These results suggested that the use of the gesture technology had no impact on the users' perception of mental workload.

Having considered the data concerning the individual ratings of mental workload it was felt sensible to analyse averaged mental workloads for each pair. These results were calculated and then statistically analysed. The difference between paired mental workloads in the voice only and voice and gesture conditions was found to be non-significant (one-tailed repeated-measures t-test (t (23) = 0.67, p= 0.25)). This result was repeated for differences between first and second trials (t (23) = 0.85, p= 0.20), and Car/Forklift comparisons (t (23) = 0.67, p= -0.42).

Given the evidence of Byers, Bittner & Hill (1989) and Fairclough (1991) it was determined that the process of using the paired comparisons form of the NASA TLX may be unduly removing important data from the analysis. The process of requiring participants to weight their responses to the mental workload sub-scales was demonstrably difficult during data collection and it was felt that the participants conscious evaluations of the task (which led to their weighting decisions) may have unnecessarily excluded information from the analysis which whilst it was of unobvious relevance to the participant could have been, unconsciously, a factor of extreme importance.

The un-weighted average scores for each sub-scale were therefore compared for the two main conditions of voice only and voice and gesture. This data is presented in table 4.5 below.

| Workload sub-scale | Voice only | Voice and Gesture |
|---|---|---|
| Mental Demand | 11.90 (1.42) | 11.82 (0.72) |
| Physical Demand | 6.46 (3.54) | 4.74 (1.65) |
| Temporal Demand | 10.47 (1.65) | 10.44 (1.42) |
| Effort | 13.17 (1.53) | 11.71 (1.04) |
| Performance | 8.98 (1.00) | 9.89 (0.81) |
| Frustration Level | 10.35 (1.82) | 9.60 (2.20) |

Table 4.5 Average mental workload sub-scale scores for voice only and voice and gesture conditions (standard deviation in brackets) (N=48)

The results shown in Table 4.5 are illustrated further below in Figure 4.7. The data suggested that the key sub-scales which were showing the largest differences as a result of use of the gesture technology were the Physical Demand and Effort scales (Effort was described to participants as a combined measure of both physical and mental demand).
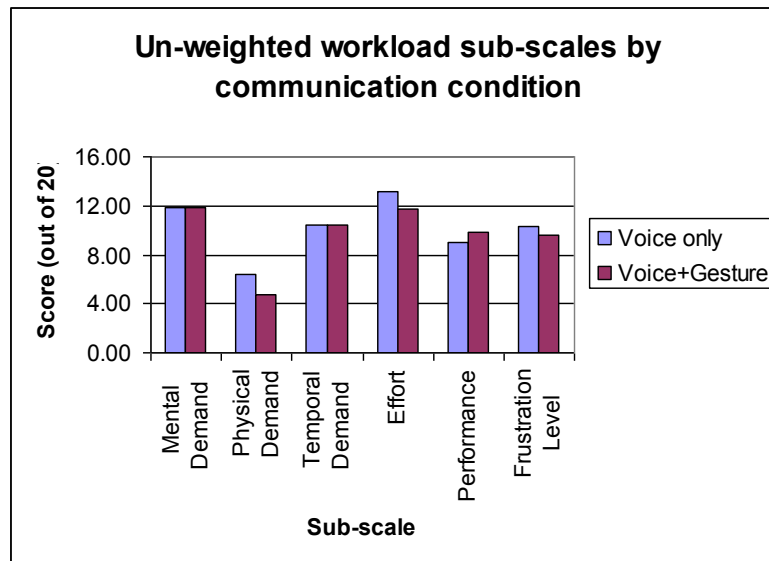
Figure 4.7

The difference between voice only and voice and gesture performance was therefore assessed statistically for these two key sub-scales. For the measures of Physical demand the difference between voice only and voice and gesture conditions was found to be significant using a one-tailed repeated-measures t-test (t (47) = 1.684, p= 0.049); and for the measures of Effort the difference between voice only and voice and gesture conditions was also found to be significant (one-tailed repeated-measures t-test (t (47) = 2.254, p= 0.01)).

It was therefore felt appropriate to further investigate these specific sub-scales looking in turn at the importance of each of the two subscales for the different roles in the interaction (Helpers and Workers), and then for each of the specific sub-scales to look at the impact of the use of gesturing technology for both the Helper and Worker sub-groups.

Table 4.6 and Figure 4.8 below demonstrate the difference in perceived levels of physical demand and effort for the subgroups of Helpers and Workers.

| Workload sub-scale | Helper | Worker |
|---|---|---|
| Physical Demand | 3.77 (3.96) | 7.43 (5.71) |
| Effort | 13.35 (3.27) | 11.53 (3.99) |

Table 4.6 Average Physical Demand and Effort sub-scale scores for Helpers and Workers
(standard deviation in brackets) (N=48)
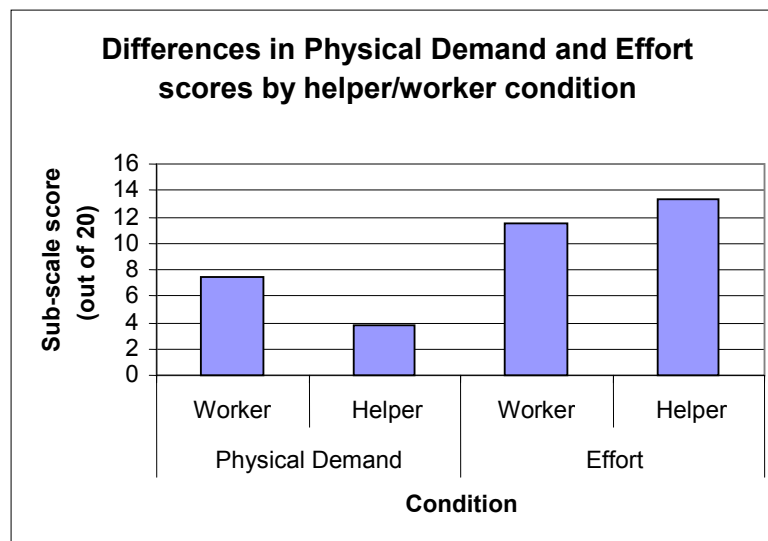
Figure 4.8

The results suggest that for both physical demand and effort there was a large difference between the Helpers and the Workers, with the Helpers finding the task much more effort than the Workers and the Workers finding the task much more physically demanding than the Helpers. If nothing else such results stand as a clear logic check demonstrating that the measures themselves were demonstrating logical results that one would hope to find. These relationships were also assessed statistically. For the measures of physical demand a one-tailed repeated-measures t-test (t (47) = 4.06, p< 0.0001) was found to be significant when comparing scores for Helpers and Workers; and equally for the same comparison for the measures of effort a two-tailed repeated-measures t-test (t (47) = -2.90, p= 0.006) was also found to be significant.

Subsequently an assessment was made of the Physical demand scores for the Helper and Worker sub-groups assessing the impact of using the gesture technology on perceptions of physical demand. The results of this can be seen in Table 4.7 and Figure 4.9 below.

| Participant Group | Voice only | Voice and Gesture |
|---|---|---|
| Helper | 3.95 (4.30) | 3.58 (3.68) |
| Worker | 8.96 (6.48) | 5.90 (4.43) |

Table 4.7 Average <u>Physical Demand</u> sub-scale scores for Helpers and Workers by Voice only or Voice and Gesture conditions (standard deviation in brackets) (N=48)
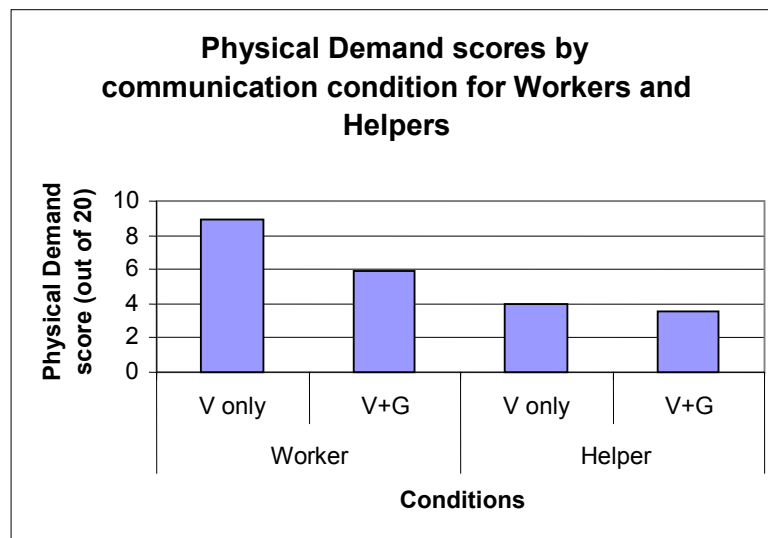
Figure 4.9

These results suggest that the ability to use gesturing support during task completion made large differences to participants' perceptions of Physical demand but only for the Worker sub-group. The Helpers did not appear to feel that the use of the gesture system influenced the level of physical demand that they experienced from the task. These findings were assessed statistically; showing a significant difference between voice only and voice and gesture conditions for Workers ratings of physical demand using a one-tailed independent-measures t-test ($t(46) = 1.90$, $p < 0.031$). The comparative test for the Helpers' data showed a non-significant result ($t(46) = 0.33$, $p < 0.37$).

Consequently the Effort sub-scale was assessed along similar lines, with the effects of using the gesturing support being assessed for the Helper and Worker sub-groups. The results for these comparisons can be seen in Table 4.8 and Figure 4.10 below.

| Participant Group | Voice only | Voice and Gesture |
|---|---|---|
| Helper | 14.25 (2.37) | 12.44 (3.81) |
| Worker | 12.08 (4.03) | 10.97 (3.95) |

Table 4.8 Average Effort sub-scale scores for Helpers and Workers Voice only or Voice and Gesture conditions (standard deviation in brackets) (N=48)
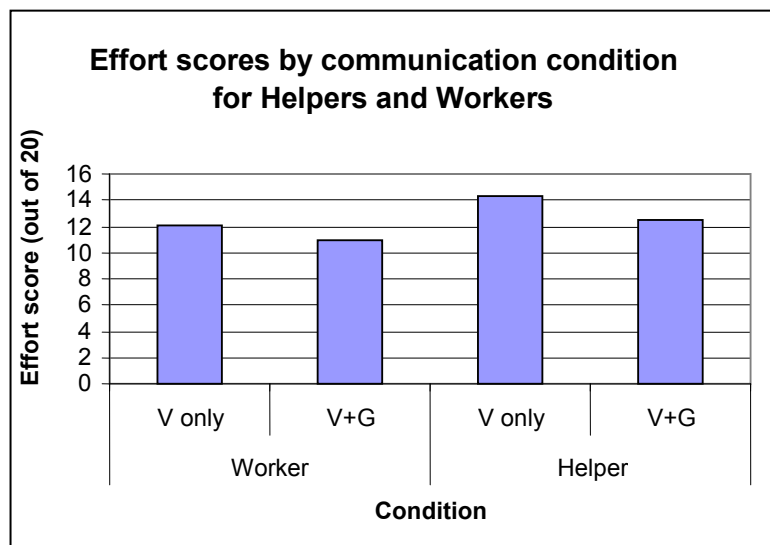
Figure 4.10

The results suggest that whilst there is some impact derived from the use of remote gesturing on perceived levels of task Effort amongst the Workers the largest impact is felt amongst the Helpers, who rate Effort much lower when they have been able to use the gesturing technology. These results were again assessed statistically. A significant difference between voice only and voice and gesture conditions for Helpers' ratings of Effort was found using a one-tailed independent-measures t-test ($t$ (46) = 1.98, $p < 0.026$). The comparative test for the Workers' data showed a non-significant result ($t$ (46) = 0.97, $p < 0.34$). The use of remote gesture therefore reduces perceptions of task related effort for Helpers but not Workers.

**4.2.3 Results summary**

The results of the first experiment have demonstrated a replication of evaluations of comparable systems such as DOVE (e.g. Fussell et al 2003). Whereas systems such as DOVE have utilised an abstracted form of remote gestures such as digital sketches overlaid on a VDU image of the task space, the system used for this experiment utilised more naturalistic representation of gesture. Designed initially with a motivation towards creating a mixed ecology, the system utilised direct projection of video images of actual hands positioned directly in the Worker's task space. The results of the study demonstrated that such an approach can also provide significant performance benefits in collaborative physical tasks. Whilst it was demonstrated that remote gesture improves performance it was demonstrated that this is most prevalent in early trials, during early stages of collaboration.

Some task differences were also noted, with one model used for the task benefiting significantly more from the use of remote gesturing during collaboration than the other model.

These results also demonstrated that the use of gesturing technology impacts on specific aspects of working practices depending on the role being performed in the collaborative activity. For workers who are in receipt of the remote gestures there is a significant reduction in the physical workload of the task. This is presumably due to the reduced necessity for them to perform certain actions, such as searching the task space for missing items and having to hold pieces and model actions to demonstrate understanding of instructions from the remote Helper. Equally for the Helpers a representation of gesture reduces their perceived levels of Effort expenditure (a compound category derived from measures of both Physical and Mental demand); they feel that their task is made easier, perhaps because of the reduced amount of explanation and guidance that they feel necessary to provide during their instructions.

**4.3 The Effects of Remote Gesturing on Distance Instruction**

**4.3.1 Study methodology**

**4.3.1.1 Experimental design**

Given the paucity of literature available on learning effects in remote instruction, the study of post-instruction performance was achieved by asking learners to complete a Lego assembly task on their own after being instructed. Testing post-instruction effects eliminated the possibility that learners were blindly following instructions without retaining task knowledge in their own right. The study was conducted using a between-subjects independent-measures design. It employed one independent variable, communication condition, which consisted of two levels, voice-only and voice-plus-gesture. One study assistant was trained in the task to allow them to provide all instruction to participants during the task. Each of the learners experienced only one form of communication condition. Presentation of the two communication conditions was counterbalanced across participants, to avoid the instructor developing a learning bias by becoming more familiar with one instruction method over the other. The dependent variables included assembly speed and assembly accuracy measured during instruction and post-instruction at 10 minute and 24 hour intervals, following a delayed post-test design. A further questionnaire obtained data on perceived instructor presence and interpersonal variables, which also acted (in conjunction with a simple number addition exercise) as a distraction task during the 10-minute interval after the instruction period.

**4.3.1.2 Participants**

A total of 18 participants took part in the study, 14 females and 4 males. Participants' ages ranged from 19-37 years (mean 23.5, st. dev. 5.16). They were primarily undergraduate students. Participants were paid a small fee for taking part in the study. One participant (a female student, aged 26) acted as the instructor for all trials, and was paid a larger fee for participation. The instructor had prior experience and training in using the gesture projection apparatus, and had received four hours training in constructing the model prior to the experimental trials. One female was excluded from the data analysis as her instruction phase was severely interrupted. Sixteen participants returned for the second self-assembly (with 2 dropping out), returning an average of 23hrs 54mins after the start of their instruction period.

**4.3.1.3 Equipment**

Again, see section 3.5.1 (see figures 3.1 and 3.2 respectively, both p. 77).

*4.3.1.4 Materials*

Again the experiment used a copy of the manufacturer's instructions for assembly of the Lego kit (model no. 8441 – used in its forklift assembly version only), a set of mathematical problems (randomly generated four figure additions) and a bespoke evaluation questionnaire (see appendix 4.7).

**4.3.1.5 Procedure**

The study examined the impact on learning of using a projected gesture system in remote instruction situations. In these situations the learner has physical artefacts to manipulate. The instructor has a video view of the task space and can communicate normally through audio channels. The instructor was not told the hypotheses of the study.

Participants were invited to the lab. Prior to the trials starting they were asked to read an information sheet outlining the structure of the experiment (see appendix 4.8) and were also asked to sign a consent form (see appendix 4.5). Once this was completed they were shown the experimental equipment and were shown how it works. During the experiment, participants were randomly assigned to one of two groups (either voice only or voice-plus-gesture). Each participant was then remotely instructed in how to assemble the final stages of a Lego™ forklift truck model. The majority of the model had already been completed so that complete assembly was achievable within the time limit and consisted of a recognizable end goal state. One group of participants experienced the instructions with the aid of projected gestures; the other group experienced the instructions in audio only. Prior to instruction, participants were made aware that they would be required to assemble the model themselves after instruction. The instruction in object assembly lasted until the model was completed (up to a total of 10 minutes). After assembling the model, participants were given a distraction task for 10 minutes, which included the completion of questionnaire on the experiment and then a large number of simple mathematical problems. Participants were then given a further 10 minutes to independently try and complete as much of the object assembly as they could from the same starting point. This attempt at self-assembly was then repeated approximately 24 hours later. All attempts at self-assembly were video-recorded, as was all instruction, using recordings from the video cameras integral to the technological set-up.

**4.3.1.5 Problems encountered**

There were no significant problems encountered, the gesture equipment maintained consistent performance throughout the experiment. The only minor difficulties stemmed from the non-return of two participants, for their 24 hour post-test component of the study. These two participants were from different experimental groups and therefore there non-inclusion was not a hindrance to statistical analysis.

*4.3.1.6 Statistical analysis*

The time required to complete instruction in how to assemble the model was recorded. Measures of time taken were then also recorded as participants assembled the model for themselves after 10 minute and 24 hour intervals. The numbers of mistakes made on each completed model were also calculated (on a simple scoring method with points derived for the correct piece of Lego™ being used in the correct place and in the correct alignment). The change in time taken to complete the model from instruction to 1st self-assembly and then to

2$^{nd}$ self-assembly was also calculated. Responses to the questionnaire items were also analysed. Where appropriate ANOVAs were performed to analyse differences between learning, first self assembly and second self assembly periods for both timings and accuracy levels. Also where relevant comparison was made between the gesturing + voice and voice only instruction method groups using t-test comparison of group means. As necessary a Cohen's *d* was calculated to demonstrate the effect size of differences between these groups.

### 4.3.2 Results

Table 4.9 details the average Time Taken to complete the model and the number of mistakes made in each of the three phases of the study, grouped by instruction method. The results indicate that the amount of time participants took to self-assemble the model on the first attempt was longer than their original instruction time. However, after 24 hours, learning had apparently consolidated and time taken to complete the model had dropped dramatically. The number of mistakes made followed a similar pattern. Differences in performance between the three phases of the study are statistically significant for both Time Taken (one-way repeated-measures ANOVA ($F(2,15) = 8.88$, $p \leq 0.001$) ) and number of Mistakes (one-way repeated-measures ANOVA ($F(2,15) = 9.25$, $p \leq 0.001$) ).

| | *Instruction* | | *1$^{st}$ Self Assembly* | | *2$^{nd}$ Self Assembly* | |
|---|---|---|---|---|---|---|
| | Time Taken | Mistakes | Time Taken | Mistakes | Time Taken | Mistakes |
| Voice only | 358 | 0 | 471 | 5 | 357 | 3 |
| Voice plus Gesture | 320 | 0 | 441 | 2 | 229 | 2 |
| **Average** | 340 | 0 | 457 | 3 | 297 | 2 |

Table 4.9 Time taken (in seconds) and number of Mistakes made during model construction in three phases, Instruction, 1$^{st}$ Self Assembly (after 10mins) and 2$^{nd}$ Self Assembly (after 24hrs), by Instruction communication condition (N=18)

The Time Taken to complete the assembly can be seen in Figure 4.11 and the pattern of mistakes over the experimental phases is shown in Figure 4.12.
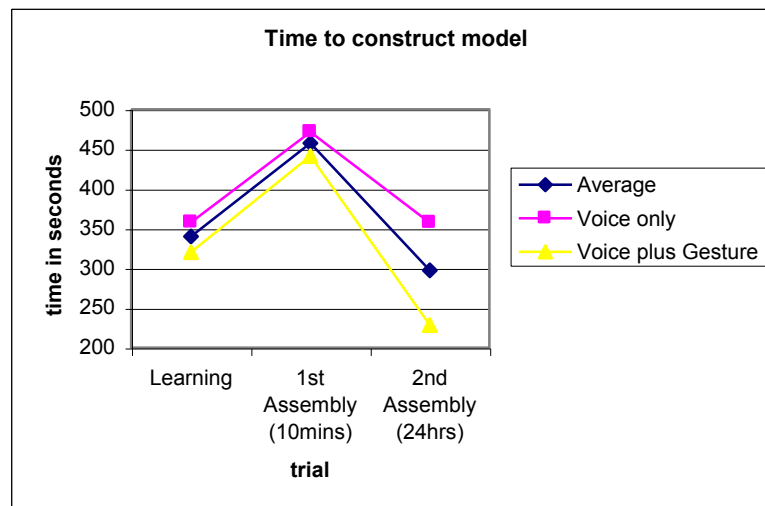
Figure 4.11 Time to complete model in each of three phases
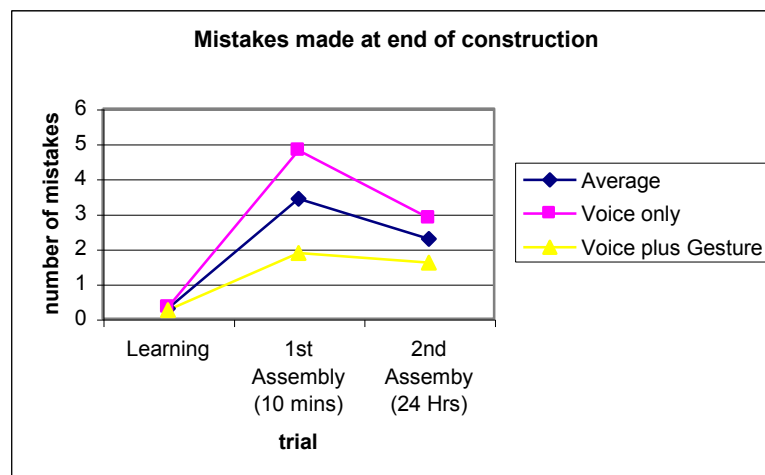


Figure 4.12 The numbers of mistakes made in each experimental phase

Analysis of the number of mistakes made in each condition showed no significant differences during instruction or during self-assembly 24 hours post-instruction. The number of mistakes made during self-assembly 10 minutes post-instruction did show a strong trend indicating more mistakes in the voice only instruction condition but the difference was only approaching significance ($p \leq 0.06$). An analysis was also carried out on the performance times in each of the three phases. Despite the trends shown there was only one significant difference found between the Instruction communication conditions. This was for the second self-assembly trial. After 24 hours it appeared that those participants who were instructed with the aid of remote gesturing were assembling their models significantly faster than those who had not experienced remote gesturing ($t(13)=1.73$, $p \leq 0.05$). Intriguingly, as demonstrated in Figure 4.11, the data also suggests that whilst those who were instructed by voice alone had a self assembly performance speed that returned to the level of their performance during instruction

those who were instructed with voice plus remote gesturing had a self assembly performance level on the second self assembly that was in fact better than their performance during instruction. The effect size for this difference was 0.89 using Cohen's *d*.

A further analysis was therefore conducted to consider the change in performance speed after initial instruction. This demonstrated that after initial instruction assembly times went up relatively equally regardless of instruction method, and after 24 hours assembly times dropped (see table 4.10).

|  | *After 10mins* | *After 24hrs* |
|---|---|---|
| Voice only | 114 | -98 |
| Voice plus Gesture | 121 | -215 |
| **Group Average** | 117 | -153 |

Table 4.10 Change in time taken to complete model after 10 minutes and then after 24 hours by Instruction communication condition (N=18)

The drop in assembly times after 24 hours appears to be most marked for those participants who were instructed using remote gesture, their assembly times dropping on average more than twice that of those instructed by voice alone. Those who experienced remote gesture instruction had significantly improved performance over the other group (t (13) =1.83, p≤0.045). The effect size for this difference was 0.95 using Cohen's *d*. The inclusion of remote gesturing during instruction therefore appears to produce better performance amongst participants in later attempts at self-assembly. Remote gesturing during instruction therefore appeared to improve task learning.

The study was complemented by a questionnaire administered to the participants whilst they were being distracted prior to the first attempt at self-assembly. The questionnaire consisted of 12 analogue rating scales. The scales used disagree-agree anchor points, and were used to provide a percentage value of agreement with each given statement. Data was computed by measuring the distance from the lower end of the (100mm) scale to the mark placed along the line by the participant. The statements centred on the participants' perceptions of the instructor and their interaction, gauging how much the learner liked / trusted / understood the instructor, how well they thought they did on the task / would be able to do it in future and how much the technology impacted on their ability to communicate with the instructor.

Figure 4.13 below illustrates the responses to these statements, by instruction method condition (voice only or voice and gesture). For several of the statements the percentage

agreement followed a trend suggesting an improved experience for those being instructed by remote gesture methods. However, the very large individual variability between subjects meant that it was very hard to find reliable statistically significant differences between the groups.
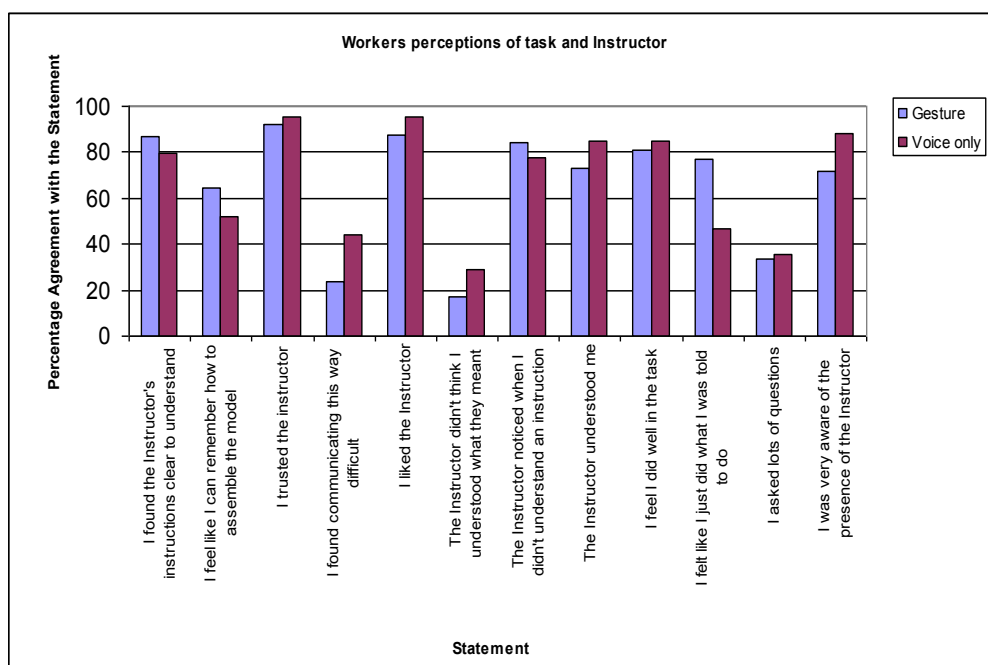


Figure 4.13

Two statements (highlighted in figure 4.14) however, were found to significantly differ by instruction communication group. These two statements though, presented a much different (but perhaps more interesting) expression of the impact of remote gesturing on the Worker's perceptions of the interaction. Those participants who had experienced instruction utilizing remote gesture actually rated the instructor as slightly less likeable (t (16) =-2.08, p≤0.05) and simultaneously were actually more likely to agree with the statement "I felt like I just did what I was told to do" (t (16) =2.65, p≤ 0.02), which demonstrates a perceived lack of involvement with the task, and a feeling that the task was less interactive and collaborative and more directive. Both of these suggest a particular orientation between the learner and the instructor with the learner less involved in determining the manipulations being undertaken and less of a positive rapport emerging during the instruction.
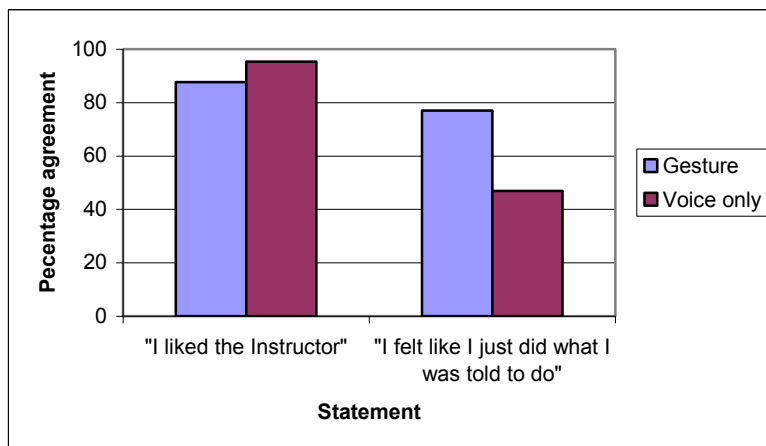
Figure 4.14 Responses to two statements by Instruction communication group

### 4.3.3 Results summary

In summary the results have demonstrated that immediately after instruction there is a refractory period wherein performance may be impaired (with potentially larger numbers of mistakes made by those instructed via voice only methods). After a period of consolidation, however, knowledge has been retained and performance in self-completion of the task improves (both in performance time and number of mistakes made). For remote instruction in the performance of physical tasks it has been demonstrated that learning can be improved through the use of a remote gesturing device. Using this method of instruction over audio-only methods significantly improves subsequent task performance. The results have also indicated that whilst performance is improved, learners may have inferior perceptions of the instructor, regarding them as more impersonal, and they feel subsequently less involved in the task as they are learning.

### 4.4 Discussion

The first two user studies of the remote gesture tool have demonstrated its basic performance benefits (both physical and mental) for collaborative tasks and also that the use of such a device during *instruction* in a physical task leads to significantly improved self-performance of the task post-instruction (and purportedly therefore *learning* of the task). Intriguingly, however, the study has also demonstrated that the relationship between the instructor/Helper and the learner/Worker is affected by the use of the technology. In the learning situation the ability of the instructor to develop a rapport with the learner was slightly impaired. However, this effect on the relationship does not have a negative impact on the quality of the learning, as performance is improved when remote gesturing is used during instruction.

One way in which we might seek to understand these performance benefits gained from the use of remote gesturing would be to consider Hutchins' (1995) discussions of Distributed Cognition and descriptions of information representation passing and propagating between individuals and their task artefacts. Hutchins' would suggest that in group situations it is only through this flow of information that complex tasks can be achieved. It could be argued that information is easier and quicker to access if the changes in representative state have been kept to a minimum and the translational overhead introduced by any mediating technology is kept to a minimum. It is suggested therefore that the two communication conditions tested in these studies (i.e. gesture enhanced communication and non-gesturing standard video linked spaces) reflect different levels of translational overhead.

**The overhead of "translating" representations**

In the voice only case, the Helper can see items in the task space but not point. This means that they then need to translate their visuo-spatial instructions into a verbal code which must be transmitted to the Worker and then be decoded introducing a significant overhead. This decoding process causes Luff et al.'s (2003) 'fractured ecologies' to become evident, as any mismatch between the perspectives on the task of the Helper and the Worker will render the process of decoding talk and then resituating visuo-spatial information within the Worker's ecology much harder.

Alternatively, a particularly close alignment of remote and local ecologies such as that used in the experiments provides direct visuo-spatial reference intact. The Helper can make gestural references, which are aligned with the Worker's visual perspective of the task. Therefore, references can be kept in a spatial medium when presented remotely. This reduction in the amount of processing required for the translation of information potentially reduces the effort required for establishing conversational grounding (Fussell et al., 2004). Such considerations are reinforced by the arguments that meaning in a dyadic interaction is derived in part from awareness of interpersonal behaviours such as gesture (Garfinkel, 1967; McNeil, 1992; Clark, 1996). However, further investigation of how remote gestures specifically influence the grounding process is required to understand this issue fully.

**4.5 Chapter Summary**

The two studies presented in this chapter have approached the problem of understanding how remote gesture tools affect performance in collaborative physical tasks. One specific aim of the studies was to provide further support for the previously observed finding that remote gesture tools will improve performance by increasing the speed of collaboration. This finding has indeed been confirmed. And importantly the evidence from the first experiment confirms that this effect is possible with a gesture representation which is significantly different in form to the representation utilised in the DOVE system. This supports the notion that it is a feature of

remote gesturing per se that is responsible for the performance enhancement rather than some specific design aspect of the DOVE system. This opens up the research space ensuring that there is room for other approaches to system design. The arguments presented in chapter 3 that perhaps such technologies should be designed from a mixed ecologies perspective have gained some support. Whilst it is clear that at this point the evidence so far presented is unable to make an argument for any design being better than any other (more comparative evaluation of system designs being required for this) it is apparent that a remote gesture tool which has been built sensitive to the approach of mixed ecologies (i.e. the gesture tool presented in 3.5.1) can offer performance enhancing benefits to collaborative physical tasks.

Additionally, the findings from the first experiment have broadened understanding of how remote gesturing can improve performance by demonstrating the cognitive workload benefits that such tools can confer. And importantly, the results of the second experiment further this extensively by demonstrating clear cognitive performance advantages in terms of the benefits to knowledge retention when remotely instructed with the aid of remote gesturing. These latter elements of the results presented in this chapter significantly broaden the scope and understanding of the effects of remote gesturing on basic performance issues in collaborative physical tasks, but also highlight as being impacted by remote gesturing some epiphenomenal aspects of interaction such as subjective interpersonal response. Issues which had not previously been much considered.

The results of the two experiments have also been considered in relation to aspects of distributed cognition theory, providing some first explorations of the actual process by which remote gesturing comes to influence collaborative behaviour. To begin the process however, of establishing that the approach of designing from a mixed ecologies perspective has the most utility, one must look to extend the program of research. As stated above, these findings demonstrate the base efficacy of such a technology but more comparative evaluation is required to establish the primacy of such an approach.

## Chapter 5 – How Best to Construct Remote Gestures

### 5.1 Introduction

Previous work has now demonstrated the utility of unmediated representations of hands being directly projected into remote task spaces. On the basis of the evidence presented however, it is not possible to make strong claims about the best ways in which to construct a remote gesturing technology. The experimental work presented serves only to demonstrate the base efficacy of remote gesture tools designed from a mixed ecologies perspective.

As articulated in chapter 3, there are however, a variety of different potential system configurations that could be employed when constructing a communication device to support collaborative physical tasks. The purpose of this chapter is to begin to explore some of these basic design issues, to develop an understanding of how best to construct these technologies, and to begin to understand the impact that different technology designs have on performance. To facilitate this, a comparative evaluation is required, which compares performance during use of different remote gesture tools (gesture tools which encompass the different system configurations found in other approaches such as DOVE and GestureMan). There has been relatively little previous work in this vein. Through understanding the differences between the approaches and by directly comparing them this should enable firmer conclusions to be drawn about the value of a mixed ecologies approach, as it is represented in some features of system design more than others.

In chapter 3 when the basic remote gesture technology was introduced, some of the key aspects of system configuration were discussed. The three most prominent of these (and those worthy of investigation for their potential to influence interaction) were:

- Gesture Orientation (relative to the local worker)

- Gesture Location (relative to the task artefacts – e.g. projected into the task space or presented external to it)

- Gesture Format (i.e. was the gesture a laser dot, a sketch or a video of hands)

These three factors in a remote gesture system can arguably all have an effect on the format and feel of an interaction, and are (based on the discussions presented in chapter 2) some of the main ways in which remote gesture tools can differ in configuration. It is on these aspects of system configuration, that this chapter therefore focuses. To do this, two experiments were conducted.

- The first study (experiment 3) addresses the issue of gesture orientation. It compares collaborative performance whilst the gestures of the remote expert are projected into

the task space at differing angles across the worker's table, simulating face-to-face, side-by-side and at-right-angled interactions across a tabletop.

- The second study (experiment 4) is concerned with analysing the impact of both changes to gesture location and gesture format[7]. Experimentally comparing the effects of presenting gestures embedded within task spaces as opposed to presenting them on external video windows, and comparing three methods of actual gesture representation, unmediated images of hands, abstract digital sketches and a mixed approach of unmediated views of hands with a sketch facility.

This chapter proceeds by describing and reporting on each of these two experiments in turn and then concludes by discussing the implications of the findings of these studies for an understanding of how best to construct remote gesture tools.

---

[7] The gesture location and format studies were combined for purposes of expediency and to facilitate a meta analysis of the results which allowed compound effects to be analysed. Alternatively one large experiment could have been conducted, which also incorporated the orientation conditions, but it was felt that the number of conditions and trials within such a study would have made it too complicated logistically and too difficult to report coherently.

**5.2 Gesture Orientations**

**5.2.1 Introduction**

When a remote gesturing system is built the way in which the remote gestures are inserted into the working environment is subject to possible manipulation. One of the key characteristics that can be manipulated is the angle to which the gestures are oriented relative to the person physically located in the workspace. Essentially there are three broad forms of relative orientation that can be adopted, face-to-face (as if collaborators are working across a table), lateral (as if the collaborators are at right-angles to each other) and overlaid (or side-by-side, wherein collaborators share a common orientation to the task artefacts). Figure 5.1 below illustrates some of these common orientations of collaboration that are used.



| Overlaid | Face-to-face | Lateral |
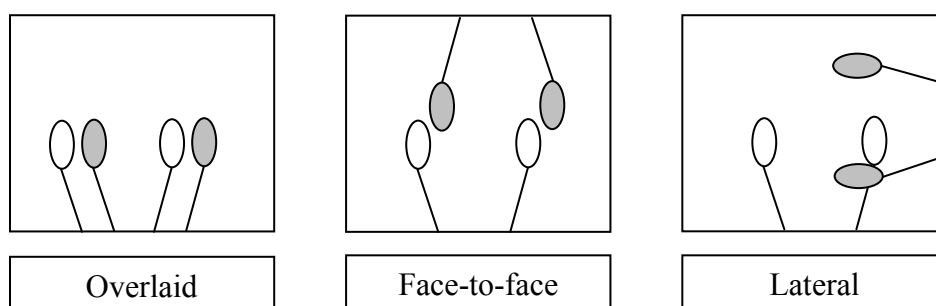
Figure 5.1 Relative orientations of Helper's and Worker's hands

Some of these forms of gestural orientation can be seen in existing systems. Traditionally, collaborative design systems (see section 2.5) operated with an overlaid orientation. This has been supported in more recent systems (which have moved away from collaborative design towards collaborative physical tasks), such as the DOVE and GestureMan systems (albeit in a slightly more side-by-side fashion). The desire to support this orientation stems from the common desire to support mutually coherent *visual* orientation to task artefacts and it makes logical sense for gestural and visual orientations to be aligned. This standard approach has however, been challenged, as collaboration systems have moved towards multiple operators, moving beyond dyadic interactions. Multiple party remote gesturing systems (which are normally based on a tabletop collaboration metaphor) have found, through the requirements of space, that they must support gestural orientations at multiple angles (see Agora, section 2.6.1, pp. 46-47, and Mixed Presence Groupware, section 2.5.8), incorporating all of the orientations shown in figure 5.1. Understanding the relative impact of forcing change in the orientation of interaction has not previously been attempted.

**5.2.2 Study methodology**

*5.2.2.1 Experimental design*

The study was constructed with a within-subjects repeated-measures design. Each pair of participants completed three Lego model assembly tasks, in each of three different gesture orientation conditions (the independent variables – see figure 5.1). These conditions were Lateral (e.g. gestures projected onto Worker's task space at 90° to Worker's seated position – the Helper's hands always entering from the right hand side), Face-to-face (e.g. Helper's hands projected onto Worker's task space as if participants were sat facing each other on either side of a table) and Overlaid (e.g. as per standard technology set-up, see section 3.5.1, similar to a side-by-side orientation, but more overlapped). Measures were recorded (the dependent variables) of time taken to reach a certain stage of each model, the final stage of completion of each model at the end of 10mins and whether any mistakes were made with the model up to the last completed stage (results were coded as either minor mistake(s) made or major mistake(s) made). A questionnaire was also administered at the end of all three trials to assess participant preferences amongst the three tested orientations. Models assembled were alternated between trials to limit task learning bias and presentation of models and experimental conditions was counterbalanced across pairs.

*5.2.2.2 Participants*

A total of 36 participants took part in the study, 18 self-selected pairs (15 females and 21 males). Participants' ages ranged from 18-34 years (mean 21.2 years, standard deviation 3.74 years), and they were mostly University undergraduates. Participants all had normal or corrected to normal vision and were paid £5 each for taking part in the study.

*5.2.2.3 Equipment*

The equipment used for the study was as per the technological set-up described in section 3.5.1, with some minor modification. To alter the relative orientation of gestures between trials the video camera above the Helper's desk was rotated by 90° or 180° as appropriate, so that the projection of gestures would show the Helper's hands entering the Worker's task space from the desired alternative angle. To ensure usability of the system and to keep gesturing effort consistent for the Helper between trials the monitor used by the Helper to view the remote task space was appropriately rotated between trials. This ensured that the Helper's view of their own hands was always such that their hands were presented going up the screen, effectively maintaining their natural orientation to their own gestural actions.

*5.2.2.4 Materials*

Three Lego model kits were used for the study (model numbers 8441, 7113 and 1354), accompanied by the three relevant sets of assembly instructions. A bespoke questionnaire was also provided (see appendix 5.1) for completion at the end of the study.

*5.2.2.5 Procedure*

Participants volunteered in pairs. They were then invited to the lab. Prior to the trials starting they were asked to read an information sheet outlining the structure of the experiment (see appendix 5.2) and were also asked to sign a consent form (see appendix 4.5). Once this was completed they were shown the experimental equipment and were shown how it works utilising gestures in whichever orientation was to be used in their first trial. Participants were invited to decide amongst themselves as to who would be the Helper and who would be the Worker (in the event that a decision could not be made the Experimenter chose randomly). Once this decision was made the participants briefly trialled the equipment, using a three piece simple toy assembly. Once the Experimenter was satisfied that the participants understood the nature of the task to be completed and the way in which the trials would differ, the experiment began. Participants worked through three trials in turn, being allowed 10mins for each trial, and being told to complete as much of the model as they could in that time. Participants were given a short break between each trial as the video camera and VDU were rotated to construct the environment for the next trial. They were not allowed to leave their seats between trials. After the final trial participants were given the bespoke questionnaire to fill out.

*5.2.2.6 Problems encountered*

The only minor difficulties encountered occurred in a small number of trials where the VDU unit used (a TV) suffered difficulties through operating on its side or when fully inverted. Such motion occasionally altered the colour balance of the screen, making all images (as seen by the Helper) tinged green. The problem was rectified after it occurred in the first three trials by giving the VDU longer to rest in its new orientation between trials. When the problem did occur the experiment was paused briefly, the participants informed of the nature of the problem and were then instructed to continue. In no cases did the participants feel that the VDU difficulties impinged on their ability to perform the task.

*5.2.2.7 Statistical Analysis*

Results were analysed as appropriate with a one-way repeated measures ANOVA, comparing the three orientation conditions. Responses to the questionnaire were categorised and analysed using Chi-squared tests.

**5.2.3 Results**

The first stage of analysis was to compare the final stages of completion for each pair of participants' three models. After each trial the model they had been assembling was inspected and the final the stage they were currently working on was recorded. Table 5.1 below illustrates the average final stage of assembly for each of the three orientation conditions.

| | Lateral | Face-to-face | Overlaid |
|---|---|---|---|
| Stage of Assembly | 8.39 (2.3) | 8.61 (3.2) | 9.22 (2.8) |

Table 5.1 Average final stage of assembly for each of the three orientation conditions (N=18 pairs) (Standard deviation in number of stages is shown in brackets)

The results have a clear trend suggesting that in the Overlaid orientation more stages of the models were being completed. This finding was statistically analysed using a one-way repeated-measures ANOVA, the finding however, suggested that the difference between the conditions was not statistically significant (F $(2, 34)$ = 0.792, p= n/s), presumably due to the relatively large standard deviation caused by variability in performance.

Having analysed the final stage of completion and being concerned that natural differences in the ability to complete each of the models (due to varying model complexity) might have affected the results a further analysis of performance was conducted. This analysis focussed on the time to complete a specified stage of each model. To counter model complexity differences the time to complete different stages was recorded, varying by model. The correct stage to use for each model's measure was calculated by assessing what the average stage of progress was at 5mins for each of the models. This average stage was then taken as the point to which timings should be made. This weighting meant that measures were taken to the completion of stage 4 for 1 model (serial no. 8441) and stage 6 for the remaining models (serial no.'s 7113 and 1354).

Table 5.2 below presents the average time in seconds for the pairs to complete the required number of stages of each model, split out by gesture orientation condition.

| | Lateral | Face-to-face | Overlaid |
|---|---|---|---|
| Time to reach required stage | 374.94 (149.82) | 357.33 (145.72) | 345.67 (154.42) |

Table 5.2 Time to reach required stage of assembly (in seconds) for each of the three orientation conditions (N=18 pairs) (Standard deviation in number of stages is shown in brackets)

Again a clear trend in the data was present suggesting that the overlaid orientation was leading to faster performance, with the face-to-face orientation coming second and with lateral

presentation of gestures leading to the slowest performance. The results of this analysis were again analysed using a one-way repeated-measures ANOVA, the finding however, again suggested that the difference between the conditions was not statistically significant ($F_{(2, 34)} = 0.435$, p= n/s). From an inspection of the raw data it was obvious that a ceiling effect had been present. Table 5.3 below illustrates the number of pairs who managed to complete the required number of stages for the above analysis.

| | Lateral | Face-to-face | Overlaid |
|---|---|---|---|
| Number of pairs | 14 | 15 | 16 |

Table 5.3 Number of pairs to complete (within 10mins) the required stage for analysis in each of the three orientation conditions (N=18 pairs)

The results shown in figure 5.3 would suggest that pairs completing the lateral orientation assembly were less likely to complete the required number of stages. This meant that for the lateral orientation condition there were 4 pairs whose time was fixed to 600 seconds, when in reality they may have required much longer. All scores that were at the ceiling of 600 seconds were therefore removed from the data. Re-analysis of this new data sample proved inconclusive, and given that the samples for each orientation condition were now of very different sizes, it was felt that the analysis would be unlikely to be able to find any new significant results.

Having observed only trends thus far and no firm statistical support for differences between orientation conditions, attention was turned to the numbers of mistakes made during completion of the assembly. Each model was inspected after assembly was ended at 10mins. The models were classified according to whether they were currently *Correct*, they suffered from a *Minor* mistake or a *Major* mistake, Table 5.4 below, highlights the differences in accuracy for model making as a product of gesture orientation condition.

|  | *Lateral* | *Face-to-face* | *Overlaid* |
|---|---|---|---|
| Correct | 9 | 9 | 10 |
| Minor Mistake | 4 | 4 | 5 |
| Major Mistake | 5 | 5 | 3 |

Table 5.4 Number of pairs to assemble their model (up to last completed stage) correctly or with mistakes in each of the three orientation conditions (N=18 pairs)

The results suggested little difference between the orientation conditions. There appeared to be a mild suggestion that in the overlaid condition, models were more likely to be correct, and if mistakes were made they were more likely to be minor mistakes than major mistakes. The high similarity between scores however, suggested quite strongly that statistical differences would not be found between the conditions, so a statistical analysis was not required.

The final stage of analysis for the orientation experiment concerned the responses to the final evaluative questionnaire that participants received after completing their final trial. The questionnaire posed two questions (supported by accompanying diagrams for clarification), firstly, which of the three orientations did the participants prefer using? And secondly, which orientation caused the most confusion?

Total responses to the first question can be seen in Table 5.5 below, with these results split out by Helper and Worker shown in figure 5.2 (also below).

|  | *Lateral* | *Face-to-face* | *Overlaid* |
|---|---|---|---|
| Number of pairs | 6 | 13 | 17 |

Table 5.5 Number of respondents to choose each of the three orientation conditions when asked 'Which orientation was easiest to use?' (N=36 respondents)
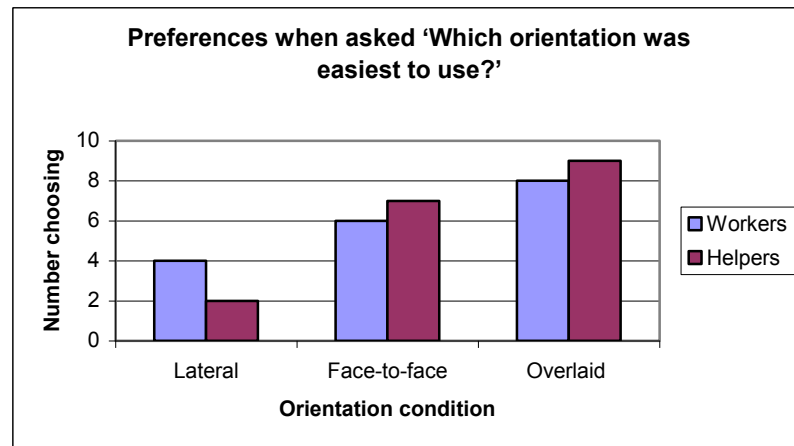
Figure 5.2

Figure 5.3 clearly demonstrates that the use of a lateral orientation for the presentation of gestures appears to be particularly frustrating for the Helpers (i.e. the producers of the remote gestures), and generally there is some preference for overlaid orientations. The numbers reported in table 5.5 were subjected to statistical analysis with a Chi-squared test, the result ($\chi^2$ (2) = 5.167, p=0.076), suggests that the difference is approaching significance, but does not quite meet the criteria for acceptance as statistically significant. The trend however, remains firm, in that a preference is observed for the overlaid orientation.

Responses to the second question 'Which orientation did you find most confusing?' are presented below in table 5.6 and figure 5.3.

|  | *Lateral* | *Face-to-face* | *Overlaid* |
|---|---|---|---|
| Number of pairs | 13 | 10 | 11 |

Table 5.6 Number of respondents to choose each of the three orientation conditions when asked 'Which orientation did you find most confusing?' (N=34 respondents – due to 2 abstentions)

**Preferences when asked 'Which orientation did you find most confusing?'**
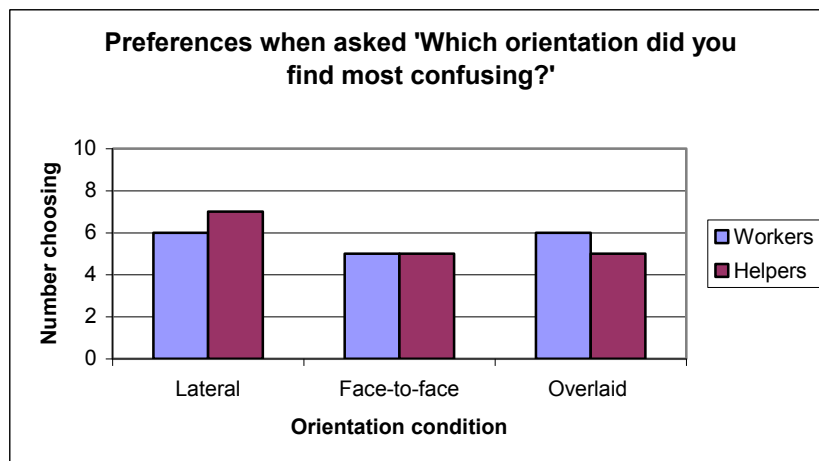
Figure 5.3

Comparable with other results the lateral orientation was perceived to be the most confusing by the largest number of respondents. There were very little differences between the ratings of face-to-face and overlaid orientations, with each receiving largely the same number of votes. Chi-squared analysis showed the differences between the three conditions, however, to be non-significant. Analysis of the reasons behind these decisions though are of some interest (full transcripts can be seen in Appendix 5.3). For the Workers, all of those who chose the overlaid orientation as more confusing claimed that this was so, because the representation of hands presented to them occasionally obscured items on their desk. It is interesting however, that this obscuring of task artefacts does not appear to have had any significant impact on performance. One participant in particular claimed that the projected hands had at times forced him to stop what he was doing and pay specific attention to the gestural movements of the Helper. The hands therefore appearing to, in some cases, actually enforce turn-taking and attention apportioning behaviours (by the participants own admission).

For those Helpers who chose the overlaid orientation as the most confusing, all of the reasons given centred on the arguments that the orientation seemed in some way 'unnatural'. None of the Helpers claimed that it had hampered their performance only that it did not seem like a form of interaction which would be physically possible, unless collaborators were sat on one another, and therefore they argued that it must be the most confusing. It could be argued that this is relatively weak criticism, and is a rejection based on the novelty of the technology, rather than any actual discrimination based on performance impact.

**5.2.4 Results summary**

In summary, the raw data appeared to offer some promising trends. On multiple occasions the data suggested that there was a possible advantage, in terms of speed of performance, progress, and accuracy in using an overlaid gestural orientation, as had been adopted in the mixed ecologies inspired remote gesturing prototype (although this was admittedly trend-line data

and was not statistically reliable). A key factor to focus on is the strength of opinion given by the users who appeared to claim they preferred the overlaid orientation. Of those critiques of the overlaid orientation that were given all appeared to stem from an uneasiness concerning the novelty of the arrangement, rather than an outright dissatisfaction with the usability of the technology. And measures of performance were at least, unaffected by using an overlapped orientation, suggesting that it could have no negative impact on performance. Whilst arguments for the merits of choosing between face-to-face and overlaid orientations are perhaps a little tricky, it is apparent that the data is giving a good clear indication that a lateral orientation is not liked, and is likely to impinge on performance, at least in this form of task.

**5.3 Gesture Format and Location**

**5.3.1 Introduction**

There are a variety of further ways in which a remote gesture tool can be constructed which might have a bearing on how it functions. In section 3.5.2, which discussed possible alterations which could be made to the basic gesture system, two key factors were identified, gesture format and gesture location.

Of the existing gesture systems a variety of gesture formats are used. The GestureMan systems principally adopt a method of laser dot gesturing, the DOVE system incorporates a model of 'digital sketching' (drawing digital lines over a live video feed image) and systems such as Agora and Mixed Presence Groupware use direct video captures of gesturing hands and arms (mediated to different extents i.e. presented with greater or lesser fidelity).

Likewise, gesture locations can also differ. Logically the location can vary between one of two places in any given system, gestures can either be posited directly in the task space itself or can be added to an external video feed of the space. This difference is best exhibited by comparison of the DOVE and GestureMan systems. In figure 5.4 below, the way in which DOVE provides an external sketch/gesture view of the task space and the way in which GestureMan's laser dot is embedded directly within it, can be compared.
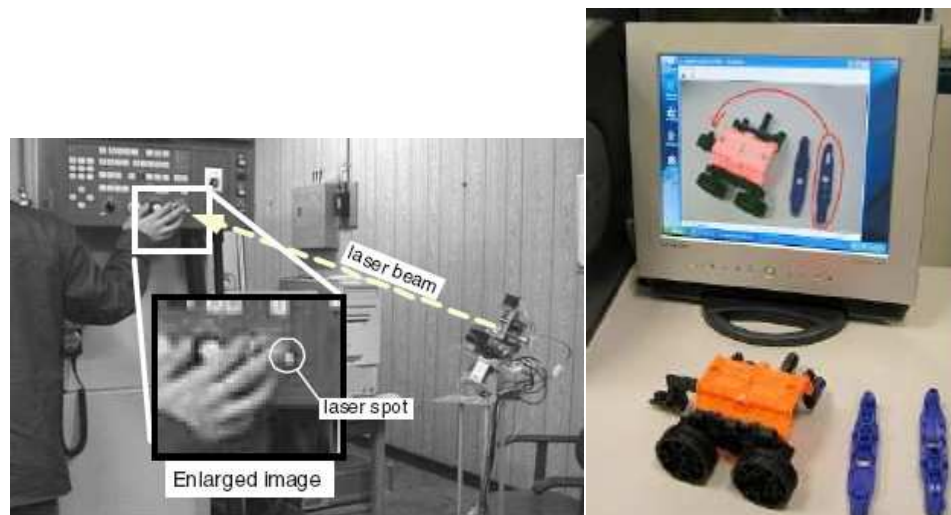


Figure 5.4 Comparing Embedded and Externalised Gesture Locations (GestureMan from Kuzuoka et al 2000, DOVE from Ou et al 2003b)

To a large extent the research work that has occurred previously (see chapter 2) has only compared modifications within each system. It is clear however that there are large differences between the different systems and these should be compared. Some previous work demonstrated the functional superiority of digital sketches over laser dots (Fussell et al 2004).

And consequently from this work there is a clear argument that richer formats of gesture have increased utility in specifically collaborative physical tasks (which is the theme of this thesis), therefore it was felt that comparison between richer formats of gesture was more important than extending the analysis to include formats with a particularly reduced capacity for expression such as laser dots, contrasting instead perhaps richer systems such as DOVE and Agora. In the following experiment the critical variables of gesture format and location were experimentally compared, to evaluate their relative impact on collaborative performance and therefore consequently system usability.

### 5.3.2 Study methodology

#### 5.3.2.1 Experimental design

To compare the different systems pairs of participants collaboratively performed a Lego assembly task (very similar to that utilised in experiment 1 reported in section 4.2). Again, Lego was chosen as it represents a generic object-focused task encompassing a variety of actions including assembly, disassembly, rotation, alignment, search and select.

The study was designed to be a series of three mini-experiments each employing a within-subjects repeated measures design. Each pair experienced two gesture location conditions (the independent variables), these were, video projection (i.e. gestures projected onto worker's desk) and video window (i.e. gestures mixed over a video feed of the task space and presented on a Worker's TV). The three experiments were conducted using the same design with the only difference between them being the method of remote gesturing used in each. This allowed an overarching between-subjects meta-analysis to be performed of the differences between the gesture format conditions (with the independent variables of Hands only, Hands & Sketch and Sketch only gesturing). By keeping the experiments consistently controlled a meta-analysis was enabled that allowed the comparison of the two gesture location methods (projection and TV) to be carried out over all three of the sub-studies. This arrangement allowed for the overall comparison of six different experimental configurations, these are presented in Table 3.1, (p. 81) and which are detailed in section 3.5.2.

In all cases the dependent variables were the progress made with the model after 10mins (measured in stages of the Lego kit completed), and the accuracy of the work achieved. A post-test questionnaire given after each condition, assessed a variety of inter-personal perceptions and opinions about task performance. Exposure to experimental trials was counterbalanced to control for order effects, and each pair constructed two different Lego kits so as to avoid practice effects between their trials – the Lego kits chosen for their comparable complexity but differences in colour ranges and predominant shapes. Pairs were given 10mins for each of their two models and were told to complete as much as they could (no pairs ever managing to complete a model within 10mins).

*5.3.2.2 Participants*

A total of 96 participants took part in the study, mostly volunteering in pairs (with a small number being paired by the experimenter). There were 48 mixed and single sex pairs in total, comprised of 44 males and 52 females (as pairs largely self selected no control was made of their gender ratios), with roughly equal numbers of males and females in each of the three experiments. Participants' ages ranged from 18-36 (mean 22.20, standard deviation 4.22), and they were mostly taught undergraduates / postgraduates. Participants were paid £5 each for taking part in the study.

*5.3.2.3 Equipment*

The equipment used for the experiment is described in detail in section 3.5.2 'System Re-configurations' (pp. 80-84). Figures 3.6 - 3.10 illustrate the main system changes, demonstrating the methods by which gesturing was projected or displayed on an external TV and the ways in which the three different gesture representations were generated.

*5.3.2.4 Materials*

Two Lego model kits were used for the study (model no.'s 8441 and 7103), accompanied by the relevant sets of assembly instructions. A bespoke questionnaire was also provided (see appendix 5.4) for completion at the end of the first trial, with a further questionnaire at the end of the second trial (see appendix 5.5).

*5.3.2.5 Procedure*

Participants volunteered in pairs and were invited to the lab. On arrival participants were thoroughly briefed about the nature of the experiment, they were provided with an information sheet (see appendices 5.6, 5.7 and 5.8) and encouraged to ask questions. After reading the information sheet participants signed a consent form (appendix 4.5) and were invited to self-select roles (Helper or Worker) for the ensuing task (being warned that they would not be allowed to change roles during the study). If participants encountered difficulties picking roles the experimenter assigned them randomly. Then participants were given training in how the system worked and were asked to perform a pre-study task involving a simplified assembly task using a non-Lego model, which was conducted until the Experimenter was satisfied that the participants understood the nature of the task they would complete and how the gesturing system could be used. Specific attention was drawn, where appropriate, to the methods of deleting sketched gestures, to ensure that participants were fully aware of how to perform this action. Participants were instructed about the use of verbal discourse during the interaction (i.e. talking was allowable at all times – but restricted to English only when participants were multilingual). Participants were then instructed that they had 10mins to complete as much of the model they had been assigned as they possibly could. After each trial participants were given a questionnaire to fill out, concerning their perceptions of the task and their performance, the second questionnaire including summative evaluation questions concerning both trials.

*5.3.2.6 Problems encountered*

Occasional problems were encountered with equipment over-heating or breaking down and as such small delays were worked into a non-significant number of trials. In all cases the trials were stopped whilst the problem was fixed. With all task related conversation banned and ensuring that participants did not leave their positions (i.e. get an opportunity to look at each other's working areas). Trials were resumed once the problem was fixed. For one pair technical difficulties necessitated that they leave after the first trial and return the next day to conclude the study.

*5.3.2.7 Statistical analysis*

The study utilised T-test comparisons for meta-analysis of the projected and TV output conditions and utilised a one-way ANOVA for comparison between the gesturing conditions. The study was not ran as a mixed design (with an ensuing mixed design ANOVA reporting main effects and interactions) because it was determined that the data was potentially inappropriate for such an analysis given that the within-subjects nature of the projected vs TV comparison was questionable. As participants changed conditions they also changed the Lego model they were using. It was felt that using t-tests and one-way ANOVAs gave a stronger more rigorous analysis of the data. It was also apparent that the raw data was demonstrating that interaction effects would be highly unlikely and therefore it was felt that to reduce the complexity of the discussion such results would be unnecessary.

### 5.3.3 Results

*5.3.3.1 Performance times analysis*

To calculate some measure of the speed of performance under the various experimental conditions a timing was made of each pair of participants' efforts to reach a given stage of each Lego model. Every pair of participants attempted to construct each of two Lego models for up to 10mins. Two models were used as this helped to control for performance effects. As there was a potential for one of the models to be easier to make than another it was deemed inappropriate to measure to the same number stage for each model. It was felt that a more accurate timing could be achieved if a different stage was used for the timing for each of the two models – therefore *weighting out* any of the potential inconsistencies in ease of construction. To calculate these different stages – the video footage of all of the experimental trials was inspected. For each trial the number stage that was being worked on, 5mins into the trial, was recorded. On the basis of this the average stage of construction at 5mins was calculated for each of the two differing Lego kits. The results for this can be seen in table 5.7.

|  | Lego Technic (Car) | Lego Star Wars (Speeder) |
|---|---|---|
| **Average stage at 5mins** | 5.06 | 5.77 |

Table 5.7 Average stage of construction at 5mins

Timings were then taken from the video recordings of each pairs two trials, timing to the completion of (using rounded figures) stage 5 for the Lego Technic model and stage 6 for the Lego Star Wars model.

The average timings for pairs in each of the main treatment groups (i.e. Hands only, Hands & sketch and Sketch only gesturing) can be seen in table 5.8. Showing the average timings for their first and second trials, then subsequently sorted by each model and each gesture presentation condition (i.e. projected and presented on TV).

|  | Trial | | Lego Model | | Output condition | | Total |
|---|---|---|---|---|---|---|---|
|  | First | Second | Star Wars | Technic | Projected | TV |  |
| Hands only | 374.94 | 349.38 | 376.25 | 348.06 | 356.50 | 367.81 | 362.16 |
| Hands & sketch | 409.88 | 403.56 | 433.94 | 379.50 | 427.25 | 386.19 | 406.72 |
| Sketch only | 377.81 | 427.06 | 432.25 | 372.63 | 387.44 | 417.44 | 402.44 |
| **Totals** | 387.54 | 393.33 | 414.15 | 366.73 | 390.40 | 390.48 |  |

Table 5.8 Average Total Seconds taken to reach specified stage of model by Gesture Condition group

The average time to complete the specified stages can also be compared across the three main gesture conditions and the two main gesture output methods (see Figures 5.5 and 5.6). Inspection of the graphs suggests that there was very little difference in performance times between models assembled using projection of remote gestures or TV presented remote gestures. However, the graph comparing the gesture formats, seems to suggest that performance with Hands only gesturing is much quicker than either of the gesturing with Pens methods, both of which seem to be similar in performance.

**A graph comparing effect on performance time of gesturing in various formats**

Figure 5.5

**A graph comparing performance times by gesture output condition**
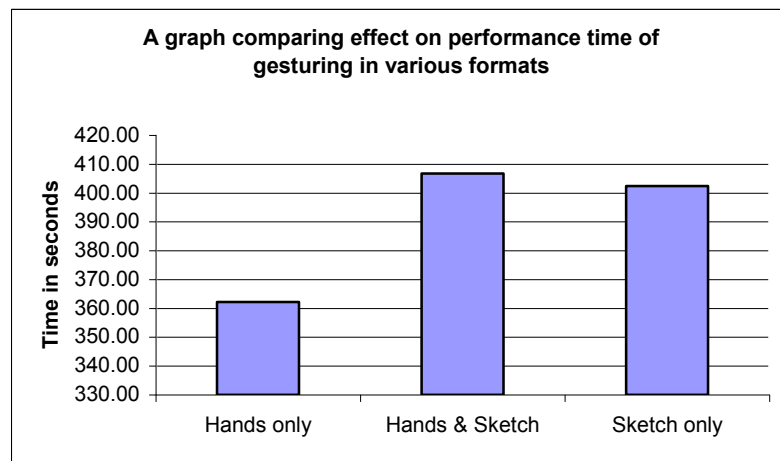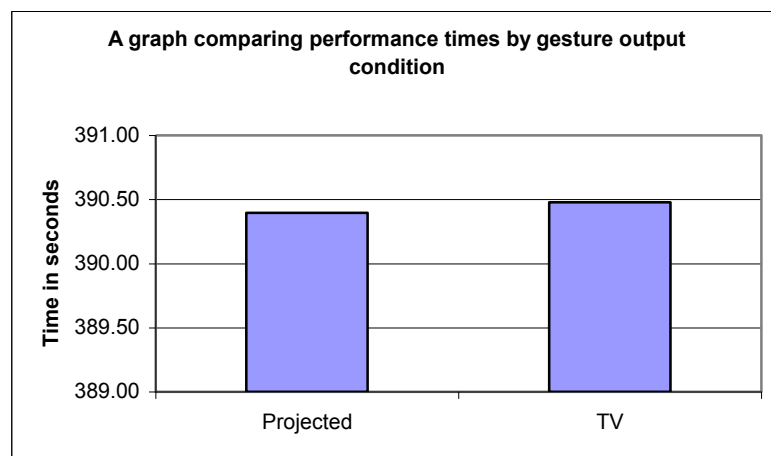
Figure 5.6

These results were then statistically analysed (see notes in the previous section on choices behind the statistical analysis). Aggregating results over the three trials (so including data from the tests on all three forms of gesturing, Hands only, Hands & sketch and Sketch only gesturing) a t-test comparison was conducted of the differences between projected remote gestures and TV presented remote gestures. The difference between the two groups was not found to be significant to the $p \leq 0.05$ level using a two-tailed repeated-measures t-test ($t (47) = -0.004$, $p=$ n/s). The average timings for the three gesture formats were also compared (including both the gesture projection trial and gestures on TV trial results for each pair). A between-subjects one-way ANOVA was conducted ($F (2, 93) = 1.21$, $p=$ n/s), but this failed to show any significant differences between the three conditions.

Analysis of the performance speeds therefore failed to generate any significant differences between the experimental conditions.

*5.3.3.2 Final stage analysis*

Having failed to find any utility in measuring performance times it was deemed appropriate to make an analysis from the perspective of the total amount of work accomplished during the trials. Interestingly there had been some suggestion in the first experiment that timings lasting only a few minutes, when constructing the Lego Technic Car model (model number 8441), had failed to show any benefits of remote gesturing (figure 4.4, p. 94). As such there may have been some unique features of the Car model that meant that more time and a further level of accomplishment was required with this specific model to begin to show the benefits of remote gesturing. As such this further contributed to the notion that by merely inspecting the stage of completion for each model at the end of the 10min trials a more robust comparison of the experimental conditions could be made. As there was an a priori assumption that there would be an inherent difference in the levels of completion of the two different models, the presentation of the models was appropriately counterbalanced across participant pairs and experimental conditions.

The analysis of final stages was performed by evaluating the video footage of all trials to ascertain which (numbered) stage each pair was working on as they were asked to stop after 10mins of assembly. The results of this analysis can be seen in Table 5.9, which shows the average final stage for each gesture format (Hands only, Hands & sketch and Sketch only) broken out by trial number, Lego model and gesture output condition.

| | *Trial* | | *Lego Model* | | *Output condition* | | *Total* |
|---|---|---|---|---|---|---|---|
| | First | Second | Star Wars | Technic | Projected | TV | |
| Hands only | 9.25 | 9.75 | 10.25 | 8.75 | 9.56 | 9.44 | 9.50 |
| Hands & sketch | 7.63 | 8.06 | 8.31 | 7.38 | 7.69 | 8.00 | 7.84 |
| Sketch only | 8.63 | 8.06 | 8.75 | 7.94 | 8.13 | 8.56 | 8.34 |
| **Totals** | 8.50 | 8.63 | 9.10 | 8.02 | 8.46 | 8.67 | |

Table 5.9 Average Stage of model being worked on at 10mins by Gesture Condition group

These interrelationships are also illustrated in figures 5.7, 5.8 and 5.9 below.
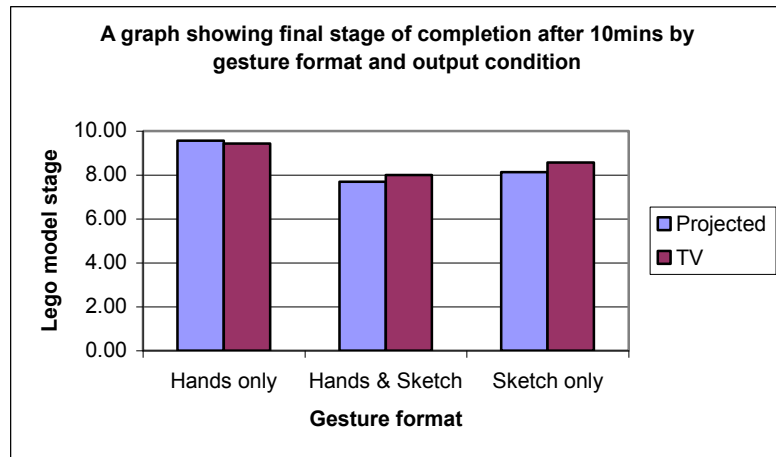
Figure 5.7
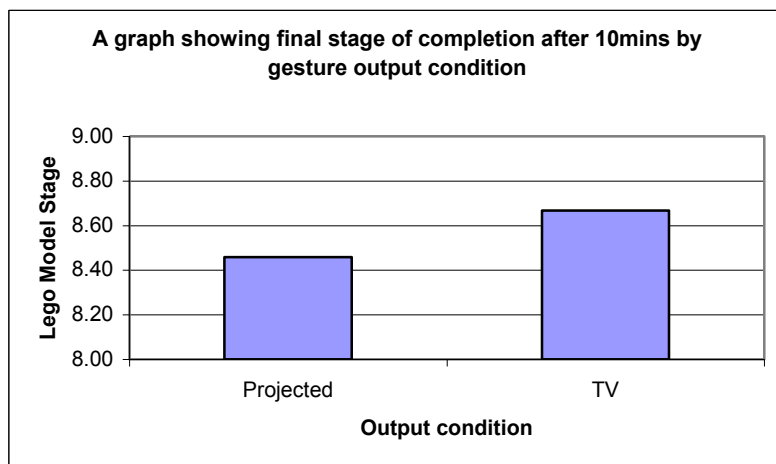


Figure 5.8
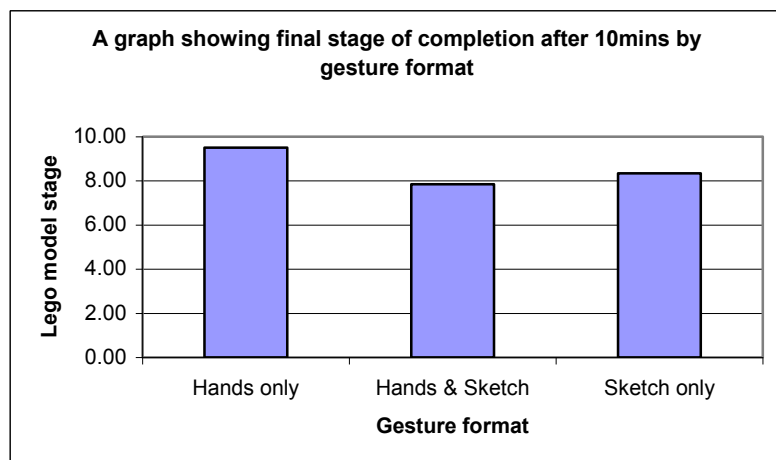


Figure 5.9

Trends on the graphs seemed to again suggest very little difference between the gesture output conditions, with Figure 5.7 potentially suggesting that this was most evident for the gesturing with Hands only format. Again globally, it appeared that the Hands only gesture format was leading to a better level of performance suggesting that more of the assembly task had been completed under this condition.

Having been aggregated the results were then subjected to statistical analyses. The two gesture output conditions (projected vs. presented on TV) were compared, incorporating data from all three gesture format studies. A two-way repeated-measures t-test ($t (47) = -0.59$, p= n/s) again failed to find a significant difference between the two gesture output conditions. Subsequently the three gesture format conditions were compared in a between-subjects one-way ANOVA (again incorporating data from both of the gesture output conditions for each of the pairs). The results of this analysis ($F (2, 93) = 4.04$, p= 0.02), demonstrated a significant difference (at the required $p \leq 0.05$ level) between the gesture format conditions. Inspection of Figure 5.9, clearly illustrates how remote gesturing performed with only a representation of hands led to a significantly larger amount of the Lego models being produced. The use of pens reduced performance capability. A series of pot-hoc comparisons were performed to define the cause of the effect, with Tukey's HSD and a Bonferroni analysis both demonstrating the significant difference between the conditions was stemming largely from the difference between the Hands only condition and the Hands & Sketch condition (difference was significant to p= 0.018). Differences between Hands only and Sketch only were also approaching significance (p= 0.13).

*5.3.3.3 Accuracy*

Having analysed base performance effects and noticed a significant difference in performance between the conditions it was deemed appropriate to perform some analysis of the accuracy of the work being produced. Whilst a conclusion could be drawn that remote gesturing with Hands only leads to more of an assembly task being completed than with Pen-based methods of gesturing, this result is of little use if the quality of the work is reduced significantly by this increase in speed. A method for analysing the accuracy of the work being completed was therefore derived.

Analysing how well someone has assembled a Lego kit is a rather difficult task, as each stage of a model has a different number of pieces that need to be attached and may also have differing levels of both complexity and significance for the model as a whole. After considerable rumination a scoring scheme was derived which, rather than totalling up the number of mistakes made (which unfairly biases against faster performance) a percentage accuracy for each pairs efforts was calculated at the end of each of their trials. Photographic evidence of what they had produced at the end of each 10mins was recorded and these were later inspected. Coupled with an analysis of the video recordings of the trials, an assertion could be made of which stage of any given Lego kit a pair was working on at the 10min limit.

Analysis using the Lego instruction books was then performed, assessing the completeness and accuracy of each supposedly completed model stage. If any stage was completed with total accuracy a whole 1 point was awarded for that stage. Within each stage every single part to be attached to the model was awarded a relative value, e.g. if a stage required two parts to be added each part was worth 0.5 of the marks for that stage, with 0.25 being awarded for the correct piece being selected and 0.25 for it being attached correctly. Therefore for every mistake there was a deduction relative to the number of pieces in that stage. All deductions were cumulated and subtracted from the total number of completed stages. This figure was then divided by the total number of completed stages to give a percentage accuracy score.

Analysis of the accuracy scores followed a largely similar pattern to the previous analyses. Results for accuracy scores can be seen in Table 5.9 and figures 5.10, 5.11 and 5.12 below.

|  | *Trial* | | *Lego Model* | | *Output condition* | | *Total* |
|---|---|---|---|---|---|---|---|
|  | First | Second | Star Wars | Technic | Projected | TV |  |
| Hands only | 89.24 | 89.07 | 90.09 | 88.21 | 90.26 | 88.05 | 89.15 |
| Hands & sketch | 84.85 | 93.76 | 93.07 | 85.54 | 84.79 | 93.82 | 89.31 |
| Sketch only | 85.54 | 89.55 | 86.67 | 88.42 | 86.69 | 88.40 | 87.55 |
| **Totals** | 86.54 | 90.79 | 89.94 | 87.39 | 87.25 | 90.09 |  |

Table 5.9 Average percentage accuracy of model after 10mins by Gesture Condition group
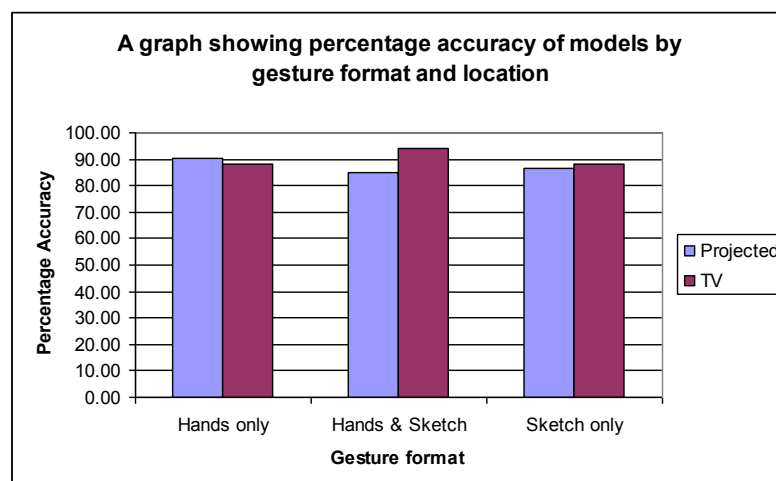


Figure 5.10
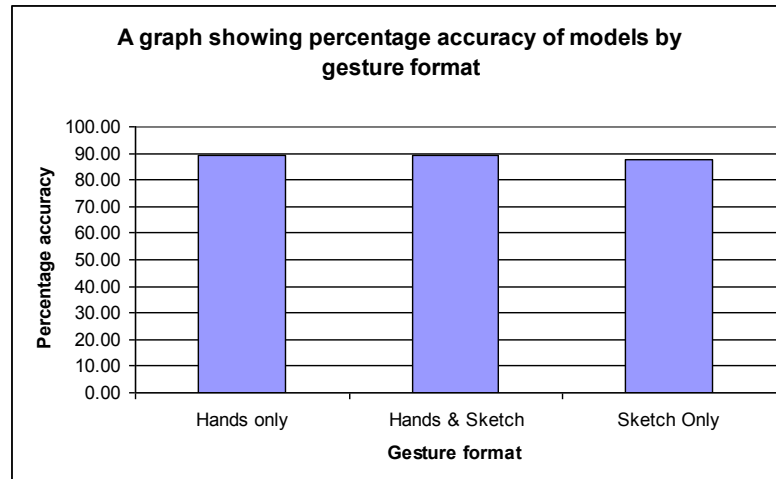
Figure 5.11
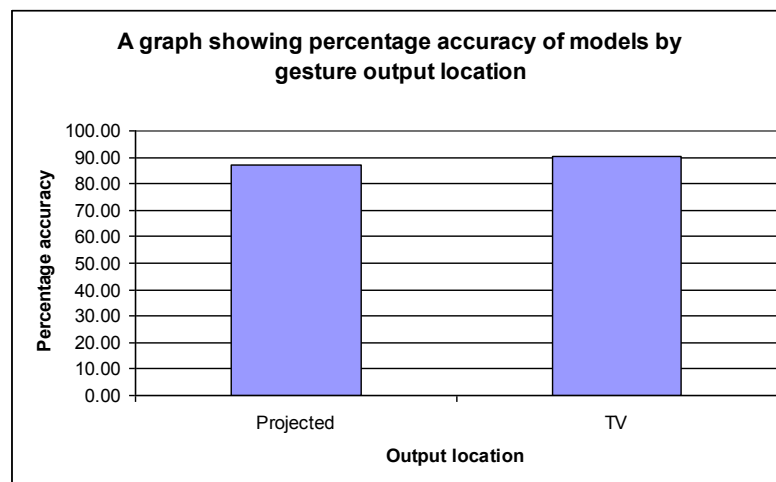


Figure 5.12

As can be elicited from the graphs above there was very little perceptible difference between any of the conditions in terms of accuracy of model assembly. To analyse this further a series of statistical analyses were again performed.

Comparing the gesture output conditions with a two-way repeated-measures t-test (t (47) = -1.04, p= n/s) showed there to be no significant difference between the conditions. Equally a between-subjects one-way ANOVA (again incorporating data from both of the gesture output conditions for each of the pairs) failed to find a difference (F (2, 93) = 0.13, p= n/s) between any of the gesture format conditions. The results therefore indicating that a systematic difference in accuracy could not be generated by manipulation of any of the independent variables.

*5.3.3.4 Questionnaire responses*

The final stage of analysis conducted was a study of the responses made to a series of four questions given to participants after each trial; the intention being that participants marked each question on a scale of 1 to 10 e.g. from 1 'Very Hard' to 10 'Very Easy' (as appropriate to the question). The questions given to the participants and summaries of the responses are given below in table 5.10.

| Hands only | Helper | | Worker | | Totals | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | TV | Project | TV | Project | TV | Project | Helper | Worker | All |
| a) How hard was the task to complete? | 5.25 | 4.75 | 5.38 | 5.19 | 5.31 | 4.97 | 5.00 | 5.28 | 5.14 |
| b) How did you find communicating this way? | 5.00 | 5.00 | 5.69 | 5.50 | 5.34 | 5.25 | 5.00 | 5.59 | 5.30 |
| c) How did you feel you did in the task? | 5.94 | 5.88 | 6.25 | 5.50 | 6.09 | 5.69 | 5.91 | 5.88 | 5.89 |
| d) Did you understand what your partner was saying? | 3.63 | 3.75 | 4.38 | 4.31 | 4.00 | 4.03 | 3.69 | 4.34 | 4.02 |
| **Hands & Sketch** | Helper | | Worker | | Totals | | | | |
| | TV | Project | TV | Project | TV | Project | Helper | Worker | All |
| a) How hard was the task to complete? | 5.19 | 4.69 | 5.81 | 4.13 | 5.50 | 4.41 | 4.94 | 4.97 | 4.95 |
| b) How did you find communicating this way? | 5.13 | 5.06 | 4.75 | 5.81 | 4.94 | 5.44 | 5.09 | 5.28 | 5.19 |
| c) How did you feel you did in the task? | 6.25 | 6.50 | 5.75 | 5.31 | 6.00 | 5.91 | 6.38 | 5.53 | 5.95 |
| d) Did you understand what your partner was saying? | 2.81 | 2.63 | 3.88 | 4.31 | 3.34 | 3.47 | 2.72 | 4.09 | 3.41 |
| **Sketch only** | Helper | | Worker | | Totals | | | | |
| | TV | Project | TV | Project | TV | Project | Helper | Worker | All |
| a) How hard was the task to complete? | 4.38 | 4.31 | 6.19 | 5.63 | 5.28 | 4.97 | 4.34 | 5.91 | 5.13 |
| b) How did you find communicating this way? | 5.94 | 5.81 | 4.81 | 5.81 | 5.38 | 5.81 | 5.88 | 5.31 | 5.59 |
| c) How did you feel you did in the task? | 6.06 | 6.44 | 6.88 | 6.00 | 6.47 | 6.22 | 6.25 | 6.44 | 6.34 |
| d) Did you understand what your partner was saying? | 3.50 | 3.50 | 3.47 | 4.31 | 3.48 | 3.91 | 3.50 | 3.89 | 3.70 |

Table 5.10 Average response to questions by gesture format, gesture output condition and participant role (Helper or Worker) ((on a scale of a) 0 Very Hard to 10 Very Easy, b) 0 Very Easy to 10 Very Difficult, c) 0 Very Badly to 10 Very Well, d) 0 Yes-always to 10 No-never))

A comprehensive and exhaustive series of statistical analyses compared average responses to each of the questions by gesture output condition and gesture format, for both total participants and then split out by Helper and Worker. None of these analyses yielded any significant

differences between the conditions. This leads to the conclusion that varying the format of remote gesture or the location at which it is presented has no discernable effect on participants self-reported ratings of task difficulty (question a), communication ease (question b), personal productivity during the task (question c) or understanding of partner's communications (question d).

*5.3.3.5 Gesture output preferences*

A basic question of preference was also asked of every participant at the end of their two trials so as to ascertain which gesture output method (projected or TV) they had preferred using. The details of their responses are shown below in Table 5.11

| Output preference | Helpers | Workers | Total |
|---|---|---|---|
| Projected | 17 | 22 | 39 |
| TV | 15 | 19 | 34 |
| Don't Know | 16 | 7 | 23 |

Table 5.11 Showing the relative preferences for gesture output condition amongst participants

The table neatly demonstrates that amongst the Helpers there was very little preference for one method of gesture output over any other. However, amongst the Workers, opinion was more polarised, with fewer having difficulty deciding between the two options (this is confirmed by a Chi-square test, showing a significant difference between preferences for workers, $\chi^2 = 7.88$, $p \leq 0.02$). Whilst there is almost an equal split in preference between those favouring Projected gestures and those favouring gestures presented on TV, the trend is towards the projected gestures.

The propensity for preference for projected gestures is seen most in the gesture formats using pens, as can be seen further in table 5.12 below.

| Output preference | Hands only | Hands & sketch | Sketch only |
|---|---|---|---|
| Projected | 13 | 11 | 15 |
| TV | 13 | 9 | 12 |
| Don't Know | 6 | 12 | 5 |

Table 5.12 Showing the relative preferences for gesture output condition by gesture format group

### 5.3.4 Results summary

The results of the three small studies can be (and have been) combined to form an overarching meta-analysis, each small study essentially acting as a separate between-subjects condition for the larger experiment. In summation the study clearly failed to demonstrate any actionable differences between the gesture output methods of projecting gestures directly into a task space and representing them externally on an adjunct TV monitor. Such differences in gesture output failed to have any consistent effect on collaborators performance, speed, progress or accuracy.

The results have however, indicated that there is a clear performance advantage based on the format of gesture representation that is used to convey the remote gestures. Whilst the timings analysis failed to show significant differences the trend patterns clearly showed unmediated hand based gesturing to be quicker than either alternative form of pen-based gesturing. This result was reinforced by the progress analysis which clearly demonstrated a significant performance benefit of again using unmediated representations of hands for the remote gesturing tool. Gesturing by 'hands only' meant that more of the assembly task would be completed after 10mins than in either of the other gesture format conditions. Fears that this increased speed in performance would be accompanied by a loss in accuracy and quality of work, were also disproved with statistical analysis failing to provide any significant differences in measures of accuracy between any of the independent variables.

### 5.4 Discussion

The purpose of this chapter was to begin to explore some of the basic design issues that arise when constructing remote gesture tools, to develop an understanding of how best to construct these technologies, and to begin to understand the impact that different technology designs have on performance. Through understanding the differences between the approaches taken by different systems and by directly comparing them it was hoped that this would enable firmer conclusions to be drawn about the value of a mixed ecologies approach.

Three key aspects of system configuration were identified as components of remote gesture tools that could easily affect their usability and performance capabilities, these aspects of system design were gesture orientation, gesture format and gesture location. The comparative benefits of different approaches to these system properties were addressed over the course of two experiments.

### 5.4.1 Gestural orientation

The evidence from the first study was, to some extent, inconclusive, but there were trends in the data suggesting a preference for the overlaid orientation adopted by most remote gesture tools (at least those working to support dyadic interaction). This inconclusiveness may arise however, due to the strength of humans in their ability to adapt to the constraints of disfluencies caused by poorly designed communicative media. Whilst altering the orientations of gestural streams within the task space had some impact on performance, its effects were significantly ameliorated by the ability of the participants to communicate 'around' these problems. Put simply altering the orientation of the gestures was effectively only a minor change to the usability of the systems. Analysis of preference did however, suggest that the use of overlaid gestures in a task space did suit participants engaged in a collaborative physical task. And the results showed some support for the 'fractured ecologies' critiques of such technologies in that the more coherent and easily interpreted gestural orientations of face-to-face and overlaid, were consistently rated as preferable to the more awkward lateral orientation. This suggests at the very least an approval for the choice of using an overlaid gestural orientation in remote gesture tools to support collaborative physical tasks.

### 5.4.2 Gesture format

The results of the second experiment argue that in a dyadic collaborative physical task gesturing using 'Hands only' in conjunction with speech works best as a medium for remote instruction. When compared to a system that uses digital sketches, remote gesturing with a video representation of the remote helper's hands actually makes performance quicker, without any loss of accuracy or increase in mistakes made. Interestingly if given the option to use both hands and pen-based gestures (the Hand & Sketch condition) performance is again impaired. Whilst some might argue that the advantage of unmediated hands over digital sketches is due to the potential unfamiliarity of participants with a digital sketching tool and therefore performance was unfairly biased against them it is quite clear that as the hands and sketch condition was also much slower than the hands only condition this cannot be the case. The hands and sketch condition combined the functionality of using a pen (as in the digital sketch condition) but was clearly as easy to use as the hands only gesture format – as it used exactly the same level of technological mediation between collaborating pairs (and most participants

would have been familiar on some level with using pen-based sketching to provide instruction or augment speech). Therefore the only clear difference between the conditions that the performance benefit can be attributed to is the lack of having pen-based gesturing.

It could be argued (in extension of a proposition put forward in section 4.4) that this performance advantage stems from the lack of translational overheads required when understanding hand gestures, which are commonly used in naturalistic face-to-face communications. The use of pen-based gesturing creates an unnecessary level of abstraction in the gesturing behaviour causing a fracture between the ecologies of the collaborators. By keeping the gesturing behaviour more naturalistic the communication device is essentially being designed from a more 'mixed ecologies' perspective – designing the gesture system such that it approximates natural interactional behaviours as closely as possible, reducing the requirements on the communicative recipient to interpret abstractions.

### 5.4.3 Gesture location

The second experiment also considered the role of gesture location during collaboration, comparing gestures which were embedded within the workspace with those pasted over an external view of the task space. The study demonstrated that gesture output condition does *not* necessarily affect collaborative performance, but the surprising result and the observation of the participants raised some significant considerations concerning the communication of mutual awareness.

It was originally hypothesised that using the projection of gestures would by its very nature be establishing a system from a more mixed ecology perspective, as it is a basis of normal face-to-face interaction that gestural communication and action occur, be perceived and acted upon, all within the same situated task space. By removing the gestures to a separate window and a separate external view of the task space it was felt that a fracture would be introduced to the collaborators' shared ecology. But no corresponding performance deficit was found[8], and this result must therefore be scrutinised.

A potential answer to the question of why there was no difference lies in some observations of participants' behaviour which are quite intriguing. As the system had no capacity for the Helper to refocus their view of the task space by zooming and so forth, and some of the task artefacts (Lego pieces) were quite small, the participants needed to develop strategies for coping with this inherent problem. The clearly obvious strategy was for the Worker to hold

---

[8] Perhaps as an aside though, some benefit is proffered to the projected gestures systems by a) the lack of any performance benefits of gestures presented on TV over projected gestures and b) the mild user preference amongst Workers of using projected gestures. This preference was in several instances (and as expected) attributed by Workers to the fact that they did not have to switch attention between two sources of action one for gesture and one for assembly – meaning that they did not miss the gestures as they happened when using the projected system.

pieces up to the camera above their desk – essentially performing a zoom function for the Helper (a function requested by participants in an analysis of similar technologies by Fussell et al 2004). When this action was performed by Workers who were experiencing the TV gesture output condition, they could clearly see on their own monitor how successfully they were holding their piece up to the camera and because of this they were *made implicitly aware of exactly what the Helper could see* of the task space and task artefacts. In the alternative condition, the Worker had to rely on verbal feedback to determine whether the Helper could adequately see the piece in question. This example demonstrates that when the gestures are presented on an external window it is possible to make the Worker implicitly aware of the bounds and limitations of the Helper's view (as they know they have a shared view), in a way that is possibly less easy to replicate with a projection of gestures. This is essentially a confounding variable, if the analysis is used to draw conclusions about designing from a mixed ecology perspective. This is because in naturalistic interactions the gestures are embedded within the environment (as in the projected gestures system) but the recipient of the gestures (being co-located) is also always implicitly aware (at least reasonably so) of the bounds and limitations of the gesturing parties perspective on the task space (as was created by the external TV presented gestures). Bearing this in mind therefore the desire to design from a mixed ecology standpoint is potentially still held intact. In the gesture output analysis the focus is held on trying to investigate one aspect of gesture location without realising that inherently tied to this issue is a problem about how you make the worker implicitly or explicitly aware of what the Helper can see of the task space. Had the system been designed from a truly mixed ecology perspective then it should perhaps have somehow combined the embedded (projected gestures) with some way of making it directly clear to the worker exactly what the helper was experiencing on the other side of the interaction – this is inherently more 'mixed ecology' in design perspective.

### 5.4.4 Implications

Ultimately the study has demonstrated that key decisions about system design for remote gesture tools are best approached from a mixed ecologies perspective. By making design decisions which push the technology towards supporting more naturalistic interactions (i.e. making them more like co-located interactions) eventual system use is made more efficient.

There are several key system design recommendations generated by the studies of this chapter.

- *Gestures should be oriented within a task space such that they can be easily interpreted*. The use of an overlaid gestural orientation is a commonly used and successful method that aligns perspectives on the task space with the dominant direction of gestural actions. This keeps the perspectives on the task space and the actions within it consistent between collaborators, promoting ease of interpretation of interpersonal action and meaning.

- *Unmediated video representations of hands should be used as the primary source of gesture representation*. Rather than the use of remote laser pointers or digital sketches, actual hands have an increased utility and can clearly adequately represent a variety of gestural requirements in object-focussed interactions. The use of hands speeds up performance without adversely effecting accuracy.

- *Gestures should be projected into the workspace*. There is a mild user preference for this approach and certainly no adverse research evidence which suggests the counter (that gestures be pasted over an externalized view). The observation, made by some participants, that attention must be split between multiple locations when gesture is not projected, is a logical argument for the inclusion of projection.

- *The system should make collaborators aware (explicitly or implicitly) of the respective task space perspective of their collaborator*. The study's inability to find a difference between the output conditions, the comments of the collaborators and the observations of practical use of the system all imply that in the video window possessed an intriguing quality in that it implicitly transmitted awareness of the Helper's visual task-space perspective, and when available this was used by the Worker to help coordinate action.

## 5.5 Chapter Summary

Chapter 5 sought to promote the benefits of designing collaboration and communication devices from a mixed ecologies perspective. Whereas chapter 4 had demonstrated the efficacy of such a system, the ability to argue that this approach had an increased utility when compared to other methods for system construction relied on an accurate evaluative comparison between a system designed from a mixed ecologies perspective and other systems being currently developed. By introducing modifications to the gesture tool design, a variety of the alternative design choices could be systematically evaluated. The studies presented in the chapter, questioned the utility of these various design choices. The orientation of presented gestures was compared, contrasting overlaid gestures with face-to-face and laterally presented gestures. The format of gesture representation was evaluated, comparing unmediated representations of hands with unmediated hands and sketch and also fully mediated digital sketch representations of gesture. And finally the gross location of gestural output was evaluated, comparing systems wherein gestures are represented on a VDU held external to the task space with systems where the gestures are directly embedded in the task space. The results of these studies provided a set of clear design guidelines for the construction of remote gesture tools, and these were discussed in relation to the notion of designing such tools from a mixed ecologies sensitive perspective. A further implication of the findings of this chapter was the observation of the importance of designing remote gesture tools to enhance mutual awareness of *both* relative orientations to task artefacts and actions and relative task perspectives. An

implicit understanding of what a collaborator can see of your task space and your actions is clearly crucial to interpreting their task-oriented actions.

This form of analysis of performance in a collaborative physical task did not however, provide much of a sense of how interaction might be fractured, or really begin to unpack what was driving successful performance when it did occur. The measures adopted, whilst very useful for ascertaining the relative benefits to performance of alternate technologies, don't really help to develop understanding of *why* interaction is affected the way it is when the above guidelines are followed and remote gesture tools are constructed from a mixed ecologies perspective. For this a deeper analysis is required, which unpacks in fine detail the interaction that takes place during collaboration.

## Chapter 6 – The Communicative Functions of Gesturing

### 6.1 Introduction

The presentation of remote gestures improves performance during collaborative physical tasks. In terms of performance outcome, the best method for presenting these gestures appears to be an unmediated video representation of the hands (as opposed to digital sketching or hands and sketching or laser pointing).

In establishing these findings the work of previous chapters and their analysis of remote gesturing, has failed to achieve a full understanding of exactly what information hands are conveying during collaboration. Previous research (e.g. Ochsman and Chapanis 1972, Kraut et al 1996) clearly demonstrates that collaborators can achieve successful task completion with voice only methods, gesture is not therefore essential. It merely enhances task performance, by somehow shifting some of the burden of articulation effort into another communicative medium. The work of Fussell et al (2004) suggests that the benefits of remote gesturing are not to be found in the simple replacement of verbal deictic references with pointing behaviours. Increasingly complex gestures are of equal, if not more, importance. So the question of what information remote gestures convey during collaborative tasks becomes quite pertinent, and given that there are multiple methods for conveying remote gestures, why is it that unmediated representations of hands appear to be superior to pen-based systems which present remote sketches? This leads to the question, what are the qualitative differences between gesturing with the hands and gesturing with sketches?

To begin to understand these differences and to begin to explore how understanding these differences enriches an understanding of the benefits of designing communications tools from a mixed ecologies perspective, a more fine-grained deeper level of analysis is required. The rest of this chapter is therefore based on an analysis of the video recordings generated during the studies presented in Chapters 4 and 5. It attempts to provide a qualitative understanding of the communicative functions of gesturing in object-focussed interactions. To achieve this the analysis takes as its focus an investigation, not just of the forms of remote gesture that are observed in collaborative physical tasks, but more importantly of the business or work of the gestural phrases that become common communicative currency during such interactions. The analysis draws specific comparison between hand-based gesturing methods and sketch-based gesturing methods in an attempt to differentiate the intricacies of their various uses and to understand on a deeper level their comparative benefits. Taxonomies of gestural use are generated and explanations for the results of chapter 5, which suggested the superiority of hand-based gestures over sketch-based gestures, are sought.

The chapter proceeds directly by presenting the methodology of the ensuing analysis and then by analyzing the use of gesture in each of the three studied remote gesture formats

(unmediated hands, hands and sketch, digital sketch only) each in turn. The chapter concludes by detailing the aforementioned taxonomy of gestural use and then discusses the contrasts between sketched gestures and hand-based gestures and the implications this has for communication.

## 6.2 Study Methodology

### 6.2.1 Study design

For the gestural analysis the data was generated from video recordings of collaborative interaction during several of the experiments presented and discussed in chapters 4 and 5 (specifically, section 4.2, which compared Remote Gesture vs. Voice Only Communication and section 5.3 which specifically analysed the performance effects of three forms of remote gesture production). The study design is therefore identical to that described in sections 4.2.1.1 and 5.3.3.1.

### 6.2.2 Participants

Participants are the same cohorts as presented in section 4.2.1.2 and 5.3.3.2.

### 6.2.3 Equipment

Equipment used was as reported in section 3.5 – including the various different apparatus required for the generation of differing forms of remote gesture as discussed in chapter 5.

### 6.2.4 Materials

Materials present during collaborative interaction were the same as detailed in the descriptions of the relevant experiments.

### 6.2.5 Procedure

As reported in sections 4.2.1.5 and 5.3.3.5.

### 6.2.6 Analysing the gestures

Having viewed the video recordings of collaborative action, common patterns of behavioural interaction and gesture use were observed and noted. The structure of these typical interactions is discussed in detail, and presented below as a series of vignettes, in each case describing the work of the gesture at that specific point and its' attempts to aid communication. This analysis demonstrates how bodily practices help to structure the organization of work. This form of analysis is conducted first for gestures performed during collaboration through hands only means of gesturing and subsequently with addition of a sketching facility and finally when only sketching was available (and hands were no longer visible). Given that the functions of gesturing presented significantly overlap, in terms of their prevalence amongst the differing forms of gesture production, greater emphasis is given in later stages of the analysis to those

aspects of gestural behaviour which are significantly unique to the specific gesturing medium in question.

Prior research has demonstrated that the most prevalent forms of gesture in collaborative activity are pointing gestures, but recreation of these simple pointing behaviours is not sufficient to significantly improve performance. It was therefore felt that to quantify the observed patterns of gestural use would lead to inappropriate assumptions being made about the relative contribution to task performance of the varying forms of gesture. To describe in detail the variety and possible form of these 'remote gestures' was considered to be of more benefit. Likewise it is an oft used methodology within such research, concerning the collaborative use of gesture, 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). It was felt that to make such comparisons would be largely futile. From an analysis of the video recordings it was clear that the majority of gestures utilised in an object-oriented collaboration are (as stated above) primarily deictic, and of those other gestures the majority are as McNeill would describe them, Concrete Iconics (McNeill, 1992), with a smattering of specific discourse structuring and perhaps incidental (i.e. not intentionally communicated) gestures such as Batons (Argyle, 1988) and Butterworths (McNeill, 1992). To reduce the gesture analysis to a quantification of the gestures and a resignation to use these pre-determined categories limits the interpretation of the function of the gestures. It does not help to explain what these gestures actually mean to the process of the discourse and their relative importance at various points in the communication, i.e. to know that a gesture is a concrete iconic gesture is all well and good but this does little to inform us of exactly what the gesture is trying to convey. As Adam Kendon argues:

> "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 role of remote gesture in cooperative activity the concern to characterize findings in terms of existing taxonomies was replaced with a concern to understand the 'stroke of gestural phrases'. That is, to understand what gestures 'say' and 'do', put simply, what the gesture is 'meant for'. Whilst some might contend that such an approach, to reject the use of pre-existing taxonomies of gesture type, will lead to the inclusion of discussion of highly idiosyncratic and therefore unrepresentative forms of gesture, it was felt that the approach of exhaustively analyzing the recorded data and analysis with

multiple observers present would enable more consistent patterns of gestural behaviour to be distilled. Despite certain idiosyncrasies between individual signalers some authors firmly believe that gestural communication can only work if there is some consistency between gesturing behaviours which enables their common interpretation. As Kendon (ibid.) 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."

The following sections present the series of vignettes that articulate the patterns of gestural phrase 'at work' in the previous experiments and the ways in which they functioned, considering multiple forms gesture representation.[9]

## 6.3 Functions of Hand-Based Gesturing

This section focuses on those gestures generated when unmediated representations of Hands were used as the means for representing the remote gestures. The ensuing analysis of common forms of gesture use takes as its structure a typical cycle of interaction for the assembly of a part of a Lego kit. During such a typical interaction there were several key stages which were invariably followed. The first stage is to unify awareness of perspectives on the task space and relative orientations, then, an item for assembly will be found and selected. Once a piece has been successfully selected the Helper directs the Worker to either find another piece or to attach the currently held piece to the parts already assembled. Once the Helper is satisfied that the pieces are correctly assembled the search begins for the next piece.

### 6.3.1 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 both artefact arrangements (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 reciprocity of perspectives. This is achieved through variants of the 'flashing hand' gesture:

---

[9] Video stills included in the vignettes highlight the remote helpers' hands to aid the visibility of gesture in the mixed reality ecology. The reader will appreciate the extreme difficulty in conveying patterns of movement in the absence of video. Attempt has been made however, in accompaniment with the video stills, to articulate in the body of the text the actual movements made to perform the gestures.

A              B              C
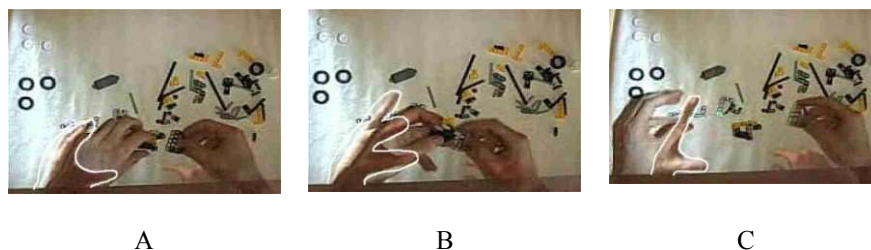
Figure 6.1 The Flashing Hand Gestural Phrase

In the initial stages of the flashing hand gesture (figure 6.1 - A), 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, already mimicking in some fashion the global shape of the Worker's hand. As the Helper's hand approaches the Worker's left hand (figure 6.1 – B) she questions, "Is this your left hand?" the Helper then starts to wiggle her fingers to draw attention to both her own hand and the Worker's hand in closest proximity. The Worker then moves his hand closer to the Helper's 'flashing hand', copying the wiggling motion and saying "Yeah" (figure 6.1 – C).

The 'flashing hand' clearly derives its name from the wiggling movement of the Helper's hand, which brings the Helper's hand in and out of alignment with the Worker's and gives the impression that the Worker's hand is flashing. Whilst simply done, it is used to establish the reciprocity of perspectives, 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 allows implicit reference between Worker and Helper of 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.

### 6.3.2 The 'Wavering Hand'
Having established reciprocity of perspectives, the participants begin the assembly task. At this early stage it is crucial that the correct items for assembly are identified. The Helper must understand their diagram and identify the correct piece from the collection of items on the Worker's desk. As the Helper cannot touch the pieces themselves, they must guide the Worker to the piece by use of description (thus allowing the Worker to perform the visual search) or (with the aid of remote gesturing) they engage in a process of deixis – 'pointing' in vernacular terms. Observations of these highly prevalent pointing gestures revealed that the deictic element is often a component part of a larger gestural phrase. The stroke does not always consist of a hand, with a pointed index finger, moving directly from rest position to the target and back again. The gestural phrase is often split into multiple elements which implicitly convey the cognitive processes in which the Helper is engaged.

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 by taking over the search process.



| A | B | C |
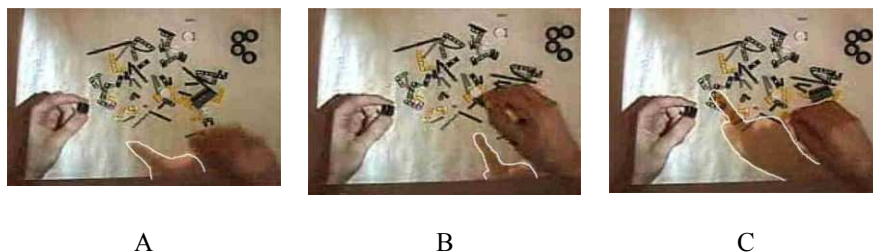
Figure 6.2 The 'Wavering Hand' Gestural Phrase

Initially the Helper reaches forward with his hand as he starts to look himself for the black L-shaped piece (figure 6.2 – A). The Helper's hand then wavers over the work surface (figure 6.2 – B), effectively mirroring his visual scans over the pool of possible items. Eventually 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 6.2 – C).

When combined with talk, the 'wavering hand' can perform a variety of functions. As illustrated above, not only is the movement deictic but it also demonstrates that the Helper is taking over the turn, by entering into the shared space, it demonstrates that they are searching through the items in the task space, with the wavering motions implicating the Helper's visual saccades, and at times the hand will be brought forward to point in error and then withdrawn at the last second, demonstrating that certain items offer some similarity to the target item. The global location, within the task space, of the wavering motions, also helps to refine the search space for the Worker, they can see in which area the Helper is expecting to find the correct piece, a function which might enable the Worker to disregard items in other areas of the workspace. It would appear that the use of an unmediated representation of a hand as the gesturing tool offers both a *richness* and an *economy* of function. Pointing with the hand is a simple gesture that is readily interpreted, but subtle patterns of movement within that gestural stroke can be clearly interpreted as conveying a much richer level of information than could be expressed by a simpler tool which merely affords the replication of the deictic component of the gestural phrase, such as a laser dot, utilised in other systems of remote gesturing.

### 6.3.3 The 'Negating Hand'

Of course even with successful pointing gestures, attention of the Worker could not always be successfully marshalled by the remote Helper, and inevitably there were mistakes wherein, despite best intentions, the Worker would select the wrong item. If not resolved through verbal prompting, mistakes where attention was directed toward the wrong objects were highlighted and corrected through forms of 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:



A                              B                              C
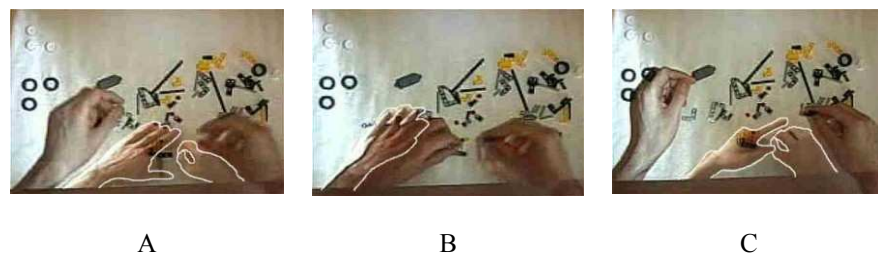
Figure 6.3 The 'Negating Hand Cover' Gestural Phrase

The first action (in response) for the Helper was for her to lay her hand flat on the desk over the wrong piece and say, "Forget about this" (figure 6.3 – A). The Helper then moves her hand upwards in a sweeping movement emphasizing which piece is to be ignored, by figuratively pushing it away from the site of action (figure 6.3 – B). Finally the Helper then points at the correct piece, which is now in the Worker's right hand and says "Just this piece" (figure 6.3 – C).

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 'decode' the situational relevance of the gesture are dependent upon the alignment of *both* the gesture 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 of the hands also allows gestures to retain their natural temporal characteristics, being both rapid and fluid and reconstituted on an ad-hoc basis whilst not leaving an excessive temporal residue such as the cluttered screen from a series of sketch-based 'scrubbing outs' or deletions.

**6.3.4 The 'Drawing Hand'**

In other instances when the ability to find the correct piece was failing and the Helper had decided that they would not be able to visually locate the piece and therefore distinctly relied upon the Worker to perform this action, greater emphasis was given to describing what the desired piece looked like. In these instances a variety of strategies could be used, the most basic of which was to provide richer description, this however was often accompanied by some form of gestural activity. Whilst the hand can be used to model pieces this approach does offer some difficulty (more on this later). In some instances therefore a gesturing practice was utilised referred to here as the 'drawing hand.' In the vignette below, becoming exasperated with the Worker being unable to accurately understand their verbal descriptions the Helper resorts to using the drawing hand gesture:



A                              B                              C
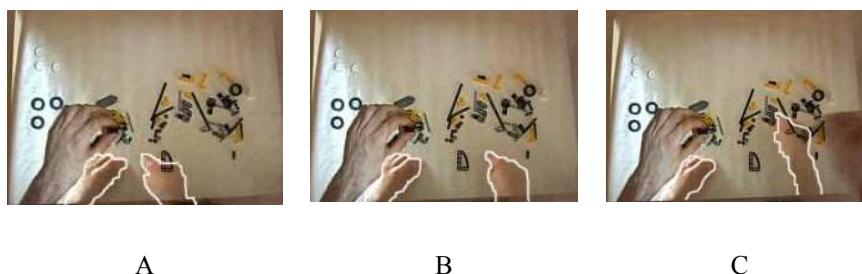
Figure 6.4 The 'Drawing Hand' Gestural Phrase

The gesture starts with the Helper forming an outstretched index finger which is placed on the work surface very determinedly to gather attention (figure 6.4 – A). The hand is then traced (still touching the work surface) horizontally across the desk to the Helper's right (figure 6.4 – B). Accompanying the final phase of the stroke of the gesture the Helper says "an L shape" finally moving their index finger at a 90° angle to the line they have just drawn, continuing to trace along the work surface (figure 6.4 – C).

That the Helper would choose to draw a shape even when no permanent trace of the gesture is left behind is a curious phenomenon that would suggest a possible desire to gesture or sketch with pens, to help clarify intention. The fact that the gesture is understood even without the permanent trace however, suggests that pens in themselves might not be necessary. It is unclear exactly how complex a gesture, sketched with a finger tip, can be, before it is unrecognizable to the recipient. Understanding that gesturers will signal to imaginary constructs that they have previously built, through gestures held at specific points in an empty 3D space (McNeill 1992), suggests that people are relatively comfortable using such non-visible constructs to aid discussion. What is perhaps of particular interest, and to be gleaned from the studies of conversationalists and their references to invisible objects which are held in common and mutually referred to, is the notion that despite the lack of a physical trace

presence, if a gestural construct is given sufficient focus and attention during its creation, it retains its spatial location (unless perhaps figuratively moved for some purpose) and as such retains as an objective entity for the length of it's desired use. It is also presumably wise to assume that signallers are capable of determining at which points a sketched invisible image is of such complexity that it will no longer be decipherable by the recipient, and on this basis make a judgement as to the utility of using such a device to aid communication. There was no evidence witnessed of Helpers producing elaborate finger tip sketches, they usually contained no more than four distinct movements and were invariably 2D.

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

As the experiments unfolded it became apparent (as mentioned above) 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 this selection, the 'mimicking hands' gesture is one of a range of gestures that are concerned with the sequential ordering of selected pieces as they are to be connected and demonstrating the relative orientations of these pieces so as to facilitate assembly. The following vignettes illustrate the role of the 'mimicking hands' gesture, with one and two hands respectively. In the first vignette, the Worker has picked up what the Helper has called the "main construction type bit":



| A | B | C |
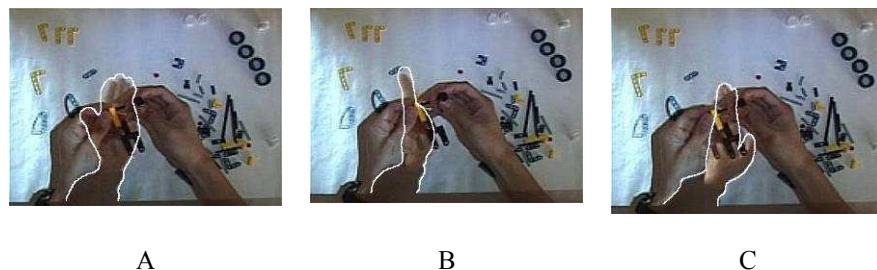
Figure 6.5 The 'Mimicking Hands' Gestural Phrase (with one hand)

The Helper prompts the Worker to rotate the piece prior to attachment, her flat hand indicating the piece's current orientation (figure 6.5 – A). The gesture unfolds as the flat hand is rotated to its side (adduction about the wrist) and the Helper says, "If you flip it a hundred and eighty degrees like that" (figure 6.5 – B). The gesture is completed as the Helper rotates her hand to the final $180^{O}$ point, and is then repeated for effect (figure 6.5 – C).

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.). The hand is physically overlapped with the piece in question so as to reinforce that the hand now 'represents' not a hand, but the piece that must be rotated. Despite the fact that the hand clearly is not formed into a comparable shape to the piece to be rotated, the placing of the gesture (in terms of both the overlapping with the model and the use of the left hand, same as the Worker's hand being used to hold the model) in accompaniment with the speech, means that the gesture is not misinterpreted. If the gesture were taken too literally the Worker would reject the accurate interpretation of the gesture. The dissimilarity of the presented hand shape with the object in question would confuse the interpretation, but this does not happen. A certain amount of latitude is clearly given to the interpretation of how a hand can represent an abstract shape, the dissimilarities are ignored and only the salient features are acknowledged. In this given instance the salient features being the relative rotation of the planar surfaces of the object in question.

In the second vignette, the Worker exploits two simultaneous hand gestures to show how the pieces should be manipulated and fitted together.



A                                    B                                    C
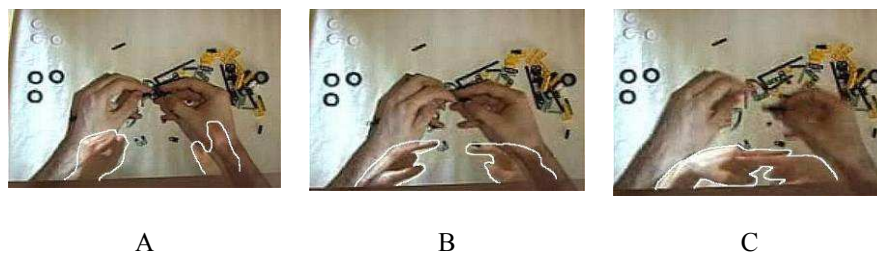
Figure 6.6 The 'Mimicking Hand' Gestural Phrase (with two hands)

Initially 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 (figure 6.6 – 5). The Helper then says, "So they lie next to each other", and begins to extend her fingers to mimic the primary axis of the pieces (figure 6.6 – 5). 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 6.6 – C), the final resting place of the fingers showing the relative orientation of the pieces.

One of the strengths of using hands as demonstrating tools in this context is the multiple points of information that the two hands can represent. In any assembly task, at any given point of assembly, it is unlikely that more than two pieces will be being connected, precisely because the person doing the assembly has only two hands, one hand to hold the current assemblage and one hand to connect the new piece. For the gesturer therefore the ability to use both of their hands to represent pieces offers a significant advantage, especially when both of the

pieces may require dynamic manipulation to expedite their connection. Through the arrangement of both subtle and complex movements of both the hands and fingers a variety of motions and relative orientations can be demonstrated in a 3D space, modeling the spatial relationships between the items being discussed. Whilst the technology used to support this form of interaction was technologically unsophisticated what has been clearly demonstrated is the utility of providing remote gesturing facilities which offer gestural representations which have sufficient degrees of freedom such that they can represent a significant variety of potential configurations of shapes and spaces that may need to be brought into connection with one another. The strengths of gesturing with hands specifically (rather than some other complex representation) however, being demonstrated to be the latitude the Worker gives the Helper, in accepting that the hand gestures shown are a representation only of the manipulated shapes. The Worker understands that the hands can never be entirely accurate and as such makes more of an effort to interpret what the gestures might mean in relation to the pieces with which they are working. The Helper is given leave therefore to use their hands in intuitive ways to articulate the complexities of assembly.

### 6.3.6 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. Despite seeing representations from the Helper as to the movements that a piece should make in order to be ready for connection sometimes the Worker would be unable to translate these concepts into a mental model for manipulating the pieces with their own hands. In remote collaborations where remote gesturing is not available such issues are inevitably resolved through increased discourse, often instructions are repeated and simplified and broken down into further, clearly articulated, stages. For the users of the remote gesturing system however, the Helper could perform a form of gesture, here referred to as the 'inhabited hand' gesture. In this vignette the Helper seeks to clarify instructions and help the Worker to move a piece he has been struggling with into the right orientation and in the right direction:



| A | B | C |
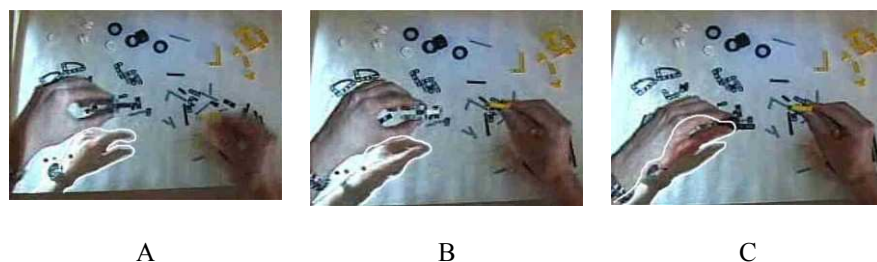
Figure 6.7 The 'Inhabited Hand' Gestural Phrase

In the first stage the Helper places her hand on top of the Worker's and forms it into the same shape to emphasise that she is referring to the Worker's hand, she then says, "If you rotate" (figure 6.7 – A). The Helper then rolls her hand forwards, saying "*Rotate your hand - like that, yeah.*" (figure 6.7 – B). The stroke of the gesture ends by her bringing her hand back to its origin, before repeating the gesture once more for emphasis (figure 6.7 – C).

The 'inhabited hand' makes the Worker aware of the fine-grained movements that need to 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 of 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.

### 6.3.7 'Parked Hands'

A focus on the work of gestures within collaborative action is not to deny the pre-eminent importance of language during the interaction. It is a well established fact that in the accomplishment of a collaborative physical task speech is the primary medium through which the task is structured and interaction enabled (see Ochsman and Chapanis 1972 for an early example of this finding). During the establishment of spoken interaction, as would be expected, a relatively consistent pattern of turn-taking behaviour was adopted. Using standard conversational mechanisms (e.g. Sacks et al. 1974) the collaborating partners attempted to take turns in their speech and activity, so as to reduce the overlapping and ensuing confusion as they tried to communicate. As gesture was being heavily used to facilitate understanding of spoken instruction it was expected that the use of the shared workspace would conform to this model of turn-taking. And this is indeed what was observed. A distinct pattern of gestural movement was observed amongst the Helpers to signal and accomplish the taking of turns. This pattern has been named the 'parked hands' gesture and it is illustrated in the following vignette. Through employing some or all of 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 an additional identical section (not therefore requiring further instruction):



|  A  |  B  |  C  |
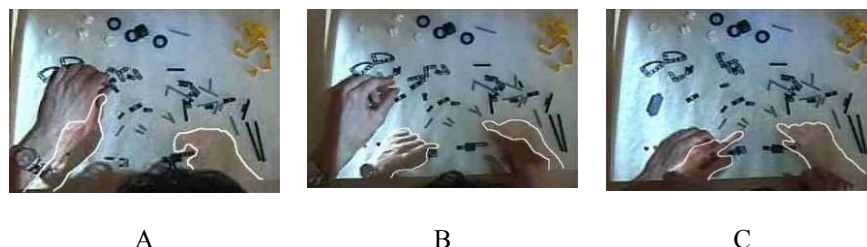
Figure 6.8 The Parked Hands Gestural Phrase

The movement begins when the Helper points out a piece and says "Assemble that exactly the same as the other one" (figure 6.8 – A). The Helper then withdraws his hands and parks them (visibly) at the edge of the shared task space (figure 6.8 – B). The Worker assembles the section and the Helper recovers control of the turn by saying "Yeah, ok, and then put that on here" whilst pointing to a specific location and subsequently parking her hands again and relinquishing the turn (figure 6.8 – C).

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. The 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. Clearly this process reflects the findings of other works (such as Duncan and Fiske 1985) which have demonstrated the importance of gesture as a means to negotiate turn-taking, albeit normally in conversational rather than task-focussed settings. This final point of drawing the hands back to the edge of the shared space also ensures that the Worker is aware of the continued presence of the Helper. In the system used to generate these remote gestures if the hands are completely removed and the Helper is not speaking they no longer have any presence in the shared space. Therefore the hands remaining visible are potentially of benefit in ensuring that the Worker is aware of the continued presence and observation of the Helper.

In the accomplishment of a part of the Lego model the necessary gestural action has ended. Having demonstrated that they are aligned in the space sympathetically with the Worker, and having managed to help them find and orient the pieces for assembly the Helper withdraws their gestural action and waits for the Worker to signal that the required assembly is complete.

**6.3.8 The 'Fluid Hands'**

As a final observation of the forms of use of hand-based gesturing there is one last consideration of the nature of such activities which is worth highlighting. Rather than fitting a specific point in the common lifecycle of gestural interaction in a collaborative assembly task, this aspect of hand-based gesturing is pervasive to the entire lifecycle. This aspect of gesturing is referred to as the 'fluid hands' and is demonstrated in the vignette below:
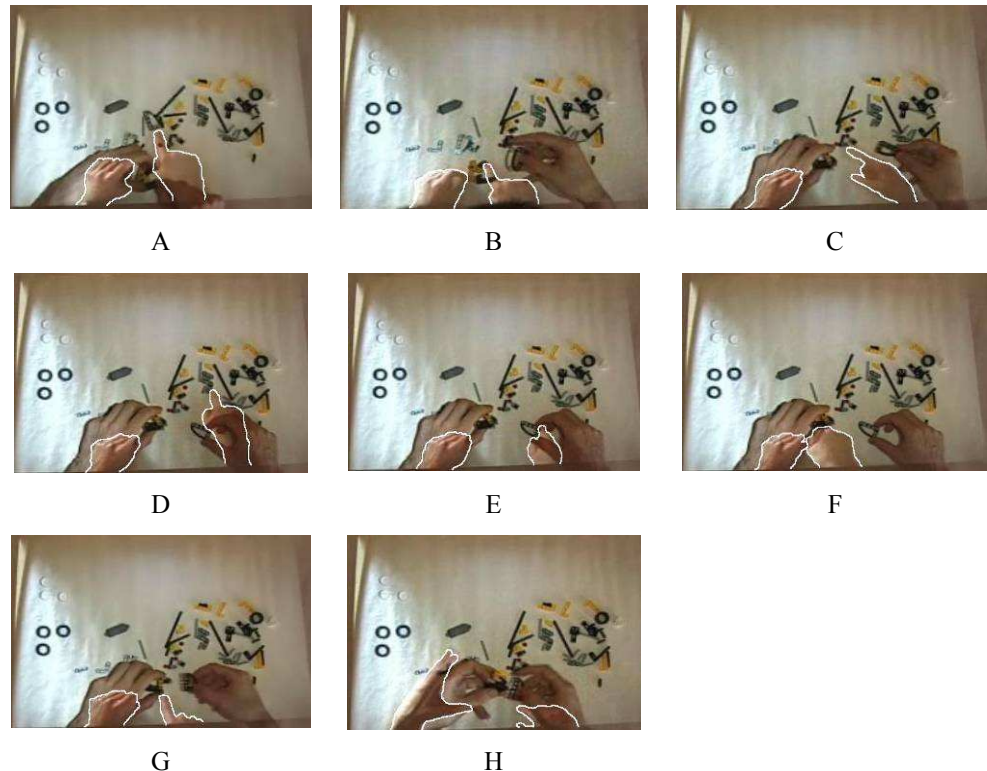


Figure 6.9 The Fluid Hands

The sequence (figure 6.9) demonstrates how the Helper first uses the gesture system to select an item for the Worker, they then point to specific locations on the assembled chunk to illustrate where the new piece should fit. The Helper then uses their hands to model the rotation of parts of the assembled chunk, resorting to using a hand to 'draw' the orientation when there is ambiguity (note how there are multiple methods of expressing the same concept when using hand gesturing), after this the Helper is again pointing to locations on the chunk and finally moves onto a form of embodied gesturing using the flashing hand gesture, all of this to aid in the resolution of an ambiguous problem. For a laser pointer to attempt such gesturing sophistication there would be a great deal of difficulty in distinguishing where one gesture ends and another begins. Even if the laser dot can be made brighter when there is an intention to gesture, its output is still ambiguous. Equally there is an extreme lack of permanence of the signal which can make the perceptual awareness of one gesture merge into

another. The DOVE system, would also encounter problems with rapid gesturing owing to the screen that is used to present the gestures becoming clogged with extraneous information that has ceased to be relevant. Clearly Ou et al (2003) have considered this and have tested the relative benefits of manual and automated wiping of screens to ascertain which confers the best performance. Even so, the presence of gesturing information is likely to be optimal in a projected hands system as all gestural information can be replicated with speed if it is purged from working memory and subsequently required for recall, leading to the increased fluidity of gesturing.

As touched on briefly in the discussion of the 'parked hands' gestures, a further area of some importance with regard to gesture systems is the co-ordination that the system engenders between the two remote collaborators. At the most simple level a gesture system must be able to allow two remote collaborators to coordinate action over finding a piece, one can spot the piece and indicate it's location for the other to select it, at a higher level action can be coordinated such that one collaborator can show the other how to assemble some pieces using gestures to indicate appropriate processes of assembly, and at perhaps the highest level a good gestural system will allow two (or more) collaborators to divide up a task and work on sub-processes of the task independently, whilst being aware of exactly what each other are doing and then coming together *sharing equally* in the task space to resolve any problems and ambiguities that have arisen. The complex and rapid switches between periods of gesturing activity and assembly activity that can be seen above in the detail of the fluid hands example demonstrates just this sought of high level interaction, further demonstrating the richness of information that can be expressed through well supported remote gesturing which accompanies the design of collaboration systems from a mixed ecologies perspective.

**6.4 Functions of Hands and Sketch Gesturing**

This section focuses on the gesturing behaviours observed when Helpers were given a pen to use in addition to their hands. Rather than focusing again on the lifecycle of common gestural interaction, which was discussed in some detail above, the analysis focuses specifically on the different observed sketches and their functions. Particular attention is drawn to the ways in which sketching gestures is different to performing hand based gestures and in subsequent sections some of the problems that this can create are discussed with examples from the video data.

The prevalence of differing forms of sketch was not quantified, as previous work (e.g. Fussell et al 2004) has already made sufficient progress in this area and it would only serve the purpose of corroborating already published data. Equally for the reasons stated in section 6.2 there is a potential for such quantification to lead to the drawing of inappropriate conclusions regarding the relative importance of some gesturing functions.

Whereas other work (ibid) has utilised systems for remote gesturing which possess automatic sketch wiping functions, the system used here did not. The reason that this was done was to attempt to further knowledge of how collaborators use existing sketches, to understand if there is an iterative process of returning to and annotating sketches during the process of collaboration and to understand how this influences the success of the collaboration and the interaction.

### 6.4.1 Sketches to highlight

Sketches were observed to perform a variety of functions, but one of the most prevalent uses appeared to be the highlighting of objects for perception, providing similar function to the 'wavering hand' gesture and fulfilling a deictic role. In some instances the pen was utilised as a tool and directly replaced the indexical finger used to point to items seen in unmediated hand-based gesture use, the narrow tip of the un-capped pen affording a particularly fine aspect for pointing to smaller items in the task space or on the assemblage. When Helpers marked the working surface with their pen however, to highlight something, there were a variety of potential sketches that they performed. Images of these sketches are presented below[10]:

---

[10] In the following diagrams the sketch in question is shown highlighted to disambiguate between other pre-existing sketches in the workspace.
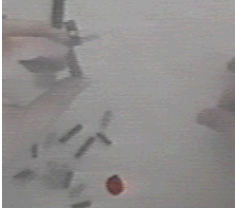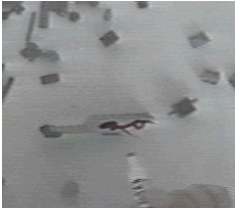
| | | |
|---|---|---|
| Circle |  | **H**: and these bad boys here<br><br>**W**: ok |
| Dot |  | **H**: These little black things yeah<br><br>**W**: Yeah<br><br>**H**: Yeah can you describe it |
| Underline |  | **H**: No but here (.) on the last two place's here |
| Shading |  | **H**: Is the the long one (.) on by two points |

Figure 6.10 Sketches used to highlight objects

These sketches presented either circling, dotting, underlining or shading (figure 6.10) to act as a medium to highlight either a piece for selection, a place on a model or (specifically unique to a sketch-based medium) a place on a pre-existing, more complex sketch. The sketches were used to facilitate the Worker's understanding whilst the Helper made a specifically deictic verbal reference, potentially reducing the amount of articulation work necessary to locate a key feature in the assembly task.

**6.4.2 The use of arrows**

One specific form of sketch which has received quite some attention in the psychology literature concerned with sketching is the arrow (Tversky et al 2000, Heiser and Tversky, 2006), and many instances of arrow usage were observed during collaboration. Examples of arrow use in the task space can bee seen below:

| | | |
|---|---|---|
| Highlight a Piece |  | **H**: You know what we passed through here (.) the sleeve we passed through here (.) look for one of those |
| Move |  | **H**: What you need to do is just move this yeah yeah |
| Flip |  | **H**: Just twist this one to here (.) no (.) err: flip it |
| Relative Orientation |  | **H**: Like this direction |
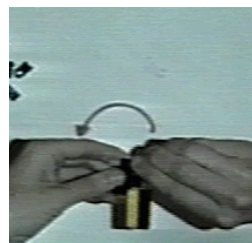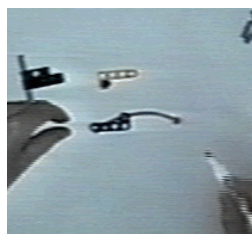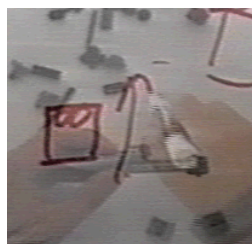
Figure 6.11 Uses of Sketched Arrows

The simplest use of the arrow sketch was the highlighting of objects (see figure 6.11) as discussed in the previous section. But as Heiser and Tversky (2006) argue, the application of an arrow to a diagram allows the reinterpretation of the diagram as one showing the functional relationships between its various elements. It is a reasonable assumption therefore that the use of sketched arrows, presented drawn over objects which are to be manipulated, will provide a level of functional information about the intended manipulation. This assumption was clearly often made by the Helpers, who would frequently use the standard recognized format of an arrow to convey motions through which pieces should be moved.

### 6.4.3 The use of drawn shapes

An obvious advantage of pen-based sketching is the ability to provide a series of drawn images which offer a significantly more complex form of gesture and which have an extended presence (and possibly usefulness) within the task space. Drawings always consisted of

representations of actual Lego pieces (as accurate in appearance as the Helper could manage). Examples of typical drawings constructed are given below:
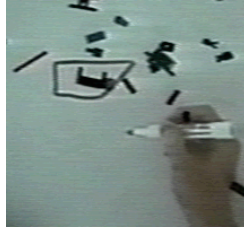
| | | |
|---|---|---|
| Single Shape |  | **H**: From the side it looks like (.) that |
| Multiple Shapes |  | **H**: If you imagine this is what you've got on the table (.) the yellow one needs to go like that (.) with the stick in there <br><br> **W**: Like that? |
| Shapes drawn onto models |  | **H**: You need a yellow piece (.) this one here <br><br> **W**: This one? <br><br> **H**: Yeah (.) and then you can fix it here <br><br> **W**: Very nice that's good yeah very good instructions |
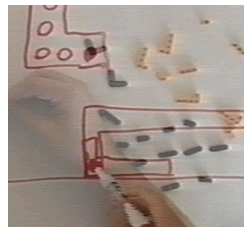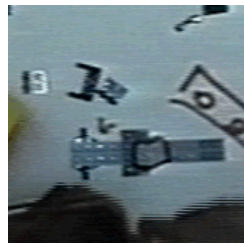
Figure 6.12 Observed Forms of Workspace Drawing

In the examples shown in figure 6.12 a variety of the functions of these drawings can be distilled. Clearly single shape drawings, performed in open space, were used to aid the search of particular pieces. In most instances actual drawings were only engaged in when articulation had become confused and the Worker was unable to decipher the meaning of the Helper. This with-holding of drawn information until all other options were exhausted, is presumably because of the greater investment in time required to construct a fully drawn image, and it should be borne in mind that the context of use of such sketches was one of a time pressured task. It is also worth considering that for some Helpers there may well have been a reluctance, as with all adults, to actually draw, as there is often a fear of ridicule (however unlikely), if the drawing is perceived to be poorly formed. In those instances where a specific item was perceived to be too unusually shaped to facilitate simple description the option to draw the shape became incentivised, and was often chosen.

When multiple objects were drawn (see figure 6.12 – multiple shapes) this was usually to show the relative orientations of the pieces in question at a specific point in assembly. So rather than as a function of searching for an item, drawings could be used to provide the sorts of

functional information that was conveyed with hand based gestures such as the mimicking and inhabited hands. The drawings were often annotated, to add emphasis, with specific functional descriptive tools, such as the already discussed arrow. Again such drawings were only entered into when dialogue was breaking down and the Helper began to doubt their ability to accurately describe what they felt to be a complex interrelationship of multiple parts. This appeared to happen when the pieces to be manipulated had several degrees of freedom and the final desired configuration required fine grained co-ordination of several moveable parts (more on this later).

An interesting third form of drawing was the representation of single items, drawn not in clear space but used as an annotation to the main assemblage of parts (see figure 6.12 – shapes drawn onto models). In these instances the Helper literally sketched directly over their visual image of the parts currently assembled. Such an action provided a variety of functions, the example shown in 6.12 being of particular interest. In this example the piece in question was unusually shaped so when the drawing was created it was being used to describe not only the shape to aid in its search but it was simultaneously being presented in it's relative orientation to the parts already assembled. The Helper clearly made an assumption that if the Worker had an understanding of how the piece fitted the assembly this would aid the understanding of the descriptive information given regarding its form, and this would enhance the chances of the correct item being rapidly identified. In such instances the information of how the item was then attached to the assemblage would not be required after the item had been found. This essentially demonstrates that a well formed drawn sketch presented at an opportune moment can provide significant economy of description.

Returning to the discussion of the drawing of multiple objects to aid description of the relative orientation of parts, mention should be made of the role of iterative construction of complex drawings. Such a process is illustrated below:



| A | B | C | D |
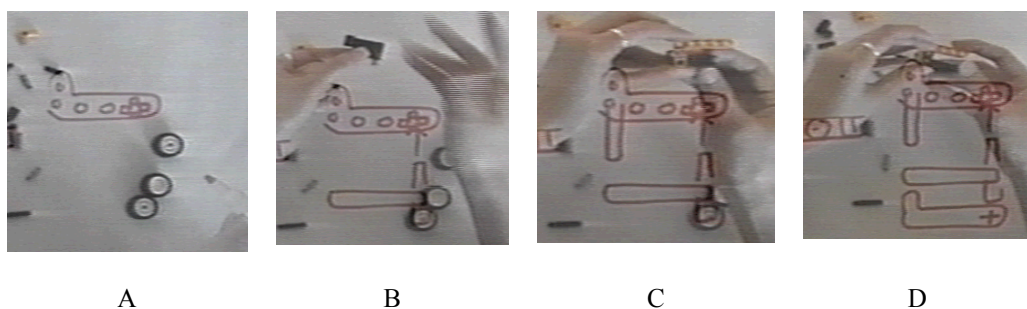
Figure 6.13 Iterative Development of a Complex Drawn Structure

In figure 6.13 the various stages of one Helper's attempts to describe a complex figure can be seen. As parts of the assemblage were found and added, the Helper built on their existing

drawn model in the hope that it would express the multitude of complex spatial relationships that were developing among the parts in a more economical or at least clearer fashion than could be expressed verbally. This sequential ordering of pieces to be assembled directly mirrored the instructions that the Helper possessed and was trying to convey. By drawing the items out it allowed the Helper to implicitly describe the relative orientations of several pieces without a) entering into lengthy description of already accomplished orientations and connections or b) gesture toward the assemblage as the Worker held it and was trying to attach pieces. For the Helper, the ability to provide a drawn image allowed them to represent information that they were gleaning from an instruction manual, keeping the visual information in its original orientation. By doing this the effort of translating how spatial relationships were constructed relative to the orientation of the assemblage as it was being held was shifted onto the Worker. They must translate the diagram such that it was relative to what they could see of the model. When hand gestures are used the Helper was required to make this translation themselves as they moved their hands to represent additional parts on the model – moving the hands to align with the model as it was held, or insisting that the Worker move the model to align with the Helper's desired view. Clearly for the Helper, it is much easier to directly replicate the information they already have (which is essentially a sketch of how the parts fit together) rather than negotiate a complex manoeuvre with the Worker.

### 6.4.4 Presenting alpha-numerics

In some instances (although such uses were very infrequent) alpha-numeric characters were incorporated into the sketched environment. Such sketching (using the term somewhat loosely to describe what is essentially writing) was used primarily to annotate more descriptive sketches as they were being produced. Figure 6.14 below demonstrates such an occurrence.
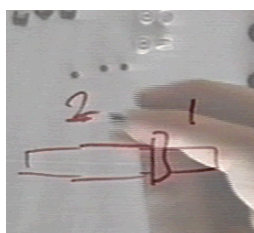


**H**: It's like basically twice th- (.) this end here (.) is twice as long as that one

**W**: yeah ok got one

Figure 6.14 Use of alpha-numerics to annotate sketches

In most instances the use of such symbols was entirely redundant and only sought to express information already presented relatively clearly through verbal means and was not used in any significant capacity by the Workers in making their attempts to understand the instructions of the Helper.

### 6.4.5 Delineating areas

A final further use of sketching that was observed, and one that was also somewhat infrequently used was the use of sketched lines to demarcate specific areas of the working surface to apportion regions to specific task actions. Such a use is demonstrated below in figure 6.15:



**H**: So here (.) to construct the things

Figure 6.15 Sketching to delineate areas

In this instance the Helper wanted to ensure that an area of workspace was held clear of any obstruction from intruding assembly parts which were otherwise being spread across the entire surface. By using the presence of a visible line, used primarily as it was being projected onto the Worker's desk, and therefore holding some affordance as an object definitely residing within the workspace, the Helper was able to reserve space for herself to create sketches and annotate the model as it was being assembled.

### 6.4.6 Problems encountered

Despite the richness of gesturing that was available to those using a hands and sketch based system, which presented both unmediated views of the Helpers' hands and the ability to provide sketched gestures, a number of difficulties were encountered. In the rest of this section some of these difficulties are highlighted and discussed.

Some Helpers decided not to use the pen at all and their gesturing was performed with hands only, but for those that did opt to use the pen this often meant that it severely impaired their use of hand based gesturing. It appeared that rather than some hybrid of hand based gesturing and sketching being performed, Helpers would opt for one method only. The reason behind this may best be understood if the nature of pen use and the strength of hand gesturing are taken into consideration. Hand gesturing (in this given working context) was of significant benefit because the Helpers had two hands with which to gesture. Admittedly a significant proportion of the gesturing required only one hand – pointing after all is not a complex activity, but for those more complex gestures which perhaps had more impact on the collaboration, the ability to use two hands to model relative spatial orientations was a positive benefit. As soon as a Helper picks up a pen however, their dominant hand is now no longer

available for gesturing. With the pen in hand, the unitary pointing gestures could still be performed, but any gesture that comfortably needed two hands required the Helper to put down the pen and particularly focus on how to construct the gesture with the hands. Such an approach significantly breaks the fluidity that facilitates smooth interaction, and forces the Helper to really think about their gesturing, whereas in other instances such gestures may be produced quite spontaneously. The presence of a drawing tool in the hand therefore significantly affecting how the Helpers devised a course of action for expressing gestural information, if a pen was in the hand then the gestures produced were most likely sketch-based.

In certain instances, once gestures become sketch-based (and ignoring the simple sketches used to highlight objects, sections 6.4.1. and 6.4.2) an additional problem of interpretation is encountered. As discussed previously (section 6.3.5) hand-based gestures representing complex shapes are often interpreted with a certain degree of latitude. Workers do not expect Helpers to be able to accurately portray a shape in question with their hands and therefore approximate from the gesture presented. With a drawn image however, Workers are more likely to accept the representation given as entirely accurate. Two examples of this are given below in figure 6.16.

**H**: A black object and it's shaped like that and it's got another side and is sort of shaped like that

(notice how the Worker is picking up a piece with dominant curves in the shape – this is completely wrong – but similar to the diagram)

**H**: I can't think of anything in black that big.

(The Worker is clearly looking for a large object – similar to the drawing - but the target object is actually very small)

Figure 6.16 Misinterpretations of sketches

In the first example (figure 6.16) the Helper's drawing is literally interpreted such that the Worker moves pieces in an entirely inappropriate fashion. In the second example (figure 6.16) the image drawn is taken to be of scale in relation to other previously drawn items and therefore as the Worker searches for the specific item in question all of the pieces they find are far too large and they ignore any items (including the actual target) which they perceive to be too small in relation to the drawing. This misunderstanding is not immediately obvious to the

Helper, who is unaware that the Worker has an entirely wrong mental model of how the shape they are searching for looks.

An additional problem encountered when sketching becomes heavily used is the ability to clutter-up the shared working space with old no longer required sketches. An example of how this can manifest itself is shown below in figure 6.17, which demonstrates the positive enhancement of the work space that is created by removing the old unused sketches.



Before                                    After

Figure 6.17 Cluttered over-sketched screen

In the example of 6.17 the Helper expressly states "I can't see anything anymore" as she then picks up a cloth to wipe her sketching surface. The actual process of wiping is very quick although it is not as quick as would be engendered by an auto-wipe system as utilised in systems such as DOVE (Ou et al 2003). The system used herein, of manual wiping, means that work surfaces can become excessively cluttered and time must be spent making explicit movements to clear the space. The time required for wiping however, in the hands and sketch based system reported here, was extremely small as the wiping facility was so simple. The reliance of some Helper's on the ability to return to gestures previously created also suggests that the presence of an auto-wipe system might run counter to the sketching practices of some of the Helpers. For them, such a function might prove exceptionally annoying as complex layered information is either removed whilst it is still required, or in fact may not ever be possible to generate, depending on the time between auto-wipes. Fussell et al (2004) argued for the use of an auto-wipe, demonstrating some slim evidence that it improved performance, but this may well be impacted by the personal preferences and working style of individuals. More research is needed to fully understand this separate issue.

Perhaps adding to this problem of cluttered sketching surfaces was the presence of highly unnecessary sketches which served to convey very little actual useful information. Some instances were observed, see figure 6.18 below, where sketched gestural information was presented which in itself was highly redundant only serving to directly replicate what was being said, and which could be expressed with much more concise means.

A                              B                              C
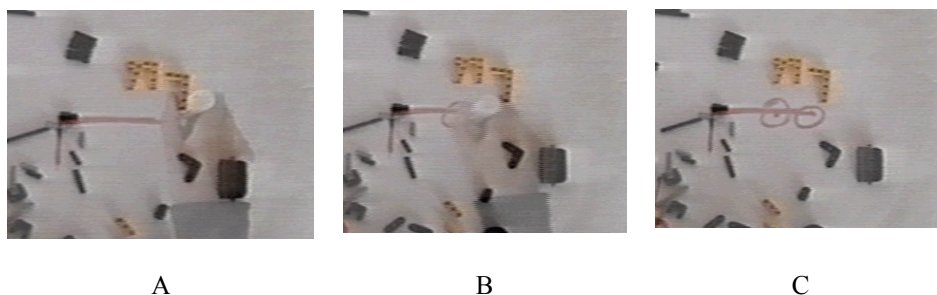
Figure 6.18 Redundant sketched information

In the example given above the Helper starts by saying "you've got your yellow bit" whilst drawing a brief image of an L shape (figure 6.18 – A). They then begin to add holes to the L shape by saying "then you've got one hole" (figure 6.18 – B) "two holes" (figure 6.18 – C) before finally stating "you put it in the second hole". The sketched information is largely redundant expressing only what is directly being said, furthering the information in no way. The Worker would presumably have been able to understand the instruction without the sketch and with the presentation of only the final statement. There is a potential therefore that once the use of the pen has been adopted for the presentation of complex information, reliance upon and overuse of the tool can become present. This was not an issue that affected many Helpers especially as for some the use of the pen for complex drawing was only adopted when, as discussed previously, all other means had been exhausted. But none-the-less there is clear ability for a pen based gesturing system to be able to provide extraneous, and as discussed above, potentially misdirecting information. Such a conclusion must however be balanced with a consideration that potentially a much richer level of description can be provided through sketch-based means than could be expressed through hand-based means.

**6.5 Functions of Sketch Only Gesturing**

The following section considers the nature of gesture use in a remote gesture system that employed only sketch-based gestural means removing the ability to present views of the Helpers' hands. Obviously a large proportion of the gestures used were witnessed in observations of the hand and sketch system discussed above (as the gestures used with that system were largely sketch-based, ignoring a role for the hands). Therefore the section starts by briefly reiterating the forms of gesture witnessed in sketch only gesturing and goes on to discuss some of the aspects of gesturing which were uniquely observed with only this system and some of the problems that this system generated.

**6.5.1 Observed forms of digital sketch**

The various forms of digital sketch that were observed are illustrated below in figure 6.19.
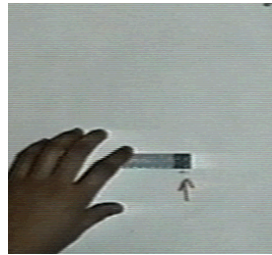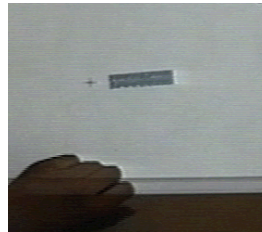


| | | |
|:---:|:---:|:---:|
| Arrows | Cursor pointing | Circle |
| Dot | Draw shape – on model | Draw shape – on surface |
| Shading | Underline | |
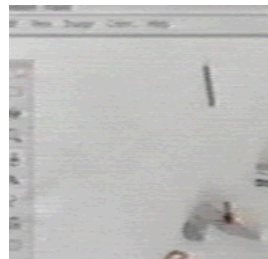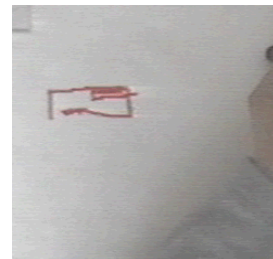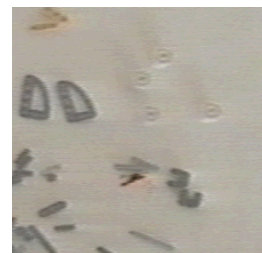
Figure 6.19 Forms of digital sketch

From observation of the videos it became apparent that with digital sketching there was significantly less likelihood that Helpers would engage in creating drawn sketches. They were much more likely to annotate the actual assemblage rather than create a drawn model and annotate that. The gesturing was therefore reduced to more indicative forms. This is not say that sketching was entirely absent as clearly some shapes were drawn, but the prevalence of more complex sketches showing the interrelationship of multiple parts to be assembled, was significantly reduced. Why this was is not immediately clear, but it may be due to the unfamiliarity with the drawing tools used. Obviously in the hands and sketch system the participants were using a pen to sketch – a highly familiar tool, but with the sketch only system the tool was switched to be a Wacom Tablet, a device with which the participants had little

experience. The process of drawing with such a tool is somewhat different to pen use, especially as in normal pen use the resultant pen etchings are witnessed at the site of action, whereas with a tablet the pen lines are represented only on screen, such a property may influence the use of the tool. In those few instances where drawings of a more complex nature were constructed it was very much as a last desperate attempt to describe some relative orientations of pieces that the Helper had had great difficulty in articulating such that the Worker could understand. Further examples of this, seen with digital sketching, are show below in figure 6.20.
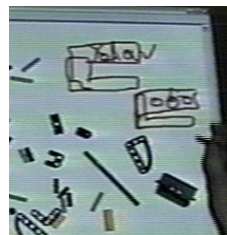


**H**: Then after that (.) err: this: (.) I still cannot see it is not very clear so I'm just trying to draw it (.) then this is the yellow one (.) so basically  I mean that you stick the tube in the second hole (.) of the yellow bit

**W**: Second hole as in this one?
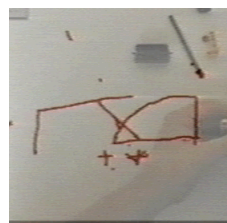
**H**: erm: it's sort of sticking out the front I mean I assume it rotates round (.) so s- s- so with this where am I? Ah here with this picture here its sort of coming out like that

Figure 6.20 Digital Sketch Drawings

Observation of the use of the simpler drawings in the digital sketch system did however provide another interesting insight into the nature of sketched gesturing. For at least one collaborating pair unusual shapes could be drawn to aid search. These resultant sketches were then left within the work space and at later stages when similar pieces were required again the Helper merely pointed to the existing sketches. Figure 6.21 below illustrates this process.
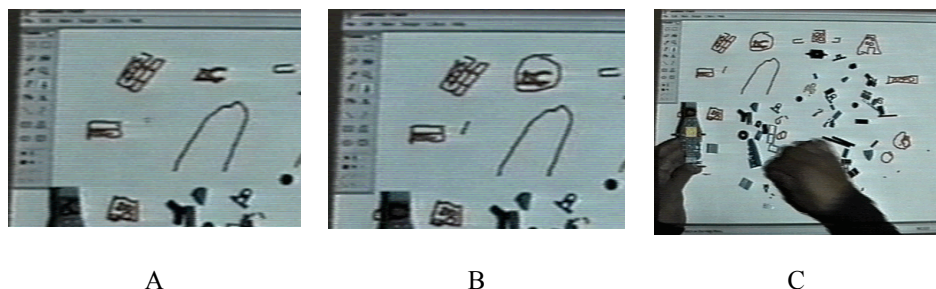


A                                    B                                    C

Figure 6.21 Referring to a catalogue of sketched items

In the example the Helper is articulating which piece is the next required item. Realising that the image is already sketched, rather than searching the work space for the desired item or describing it further to the Worker they begin a pointing movement toward their previous sketch (figure 6.21 – A). The Helper then highlights the image of the piece to be searched by circling it (figure 6.21 – B). The Worker then begins searching for it, eventually finding the correct piece, matching it to the drawn image, rather than a verbal description (figure 6.21 – C). This example nicely illustrates the way in which drawn sketches become objects of use within the task space. Unlike hand based gestures with their fleeting presence, a sketch can become an object for manipulation that exists in the shared space alongside the other task artefacts such as the assembly parts. As assembly parts are asymmetrically manipulated by the Worker the sketches are asymmetrically manipulated by the Helper. Shared visual and gestural access to these items however allows the participants to mutually interact with one another's efforts within the shared space.

A final function of remote sketches that is of particular interest to highlight is their ability to convey information implicitly that is not otherwise accurately expressed in speech. An example of this is presented below:



**H**: Now we wanna puts some things on top they're kinda like the shark's fin things we used before (.) but they're kinda blue (.) yeah those one's there yeah

Figure 6.22 Sketches express more than words

In the example above the Helper refers to an object for selection. The reference she uses however, is a term ("shark's fin bits") which she has never used before. If presented in isolation such a term could well be misinterpreted as there were a variety of items within the workspace which could fill the somewhat vague description, even taking into account the colour qualifying statement she gave. The drawing however, that she concurrently produced, looked very little like her verbal description of a shark's fin (especially considering the common orientation that one would expect to see a shark's fin presented). The sketch of the desired item was however a very accurate representation of the structure of the piece she needed, and as such as he saw it being drawn the Worker was made immediately aware of which piece (that they had previously used) to which she was referring. This exemplifies a very powerful use of sketching which is that in certain instances the strength of a visual representation will be such that it will overcome the conflicting ambiguities presented through verbal description.

### 6.5.2 Problems encountered with digital sketching

Having demonstrated the many uses of digital sketching as a form of gesturing it is worth briefly articulating a problem that this approach generated.

The key problem stemmed from the use of a cursor to perform some of the gestural functions such as pointing. In some instances, such as the example given below in figure 6.23, the Worker was unaware of what the Helper was trying to point at because of the small size of the cursor and their inability to locate it within the work space.
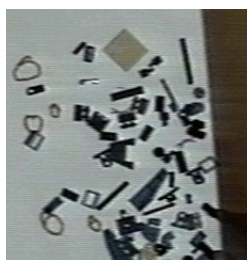


**H**: Is it this one?

**W**: Where's the thing? What are you pointing at?

**H**: err:

Figure 6.23 Difficulties finding cursor

The construction of adequate cursors to support digital sketching as a form of remote gesture is a significant issue. Losing the location of the cursor causes problems. If it is removed from the screen when no gesturing occurs it makes it harder to find again when gesturing starts, and some of the epiphenomenal benefits of gesturing discussed in section 6.3.7, such as presenting a continued sense of the Helper's presence within the task space will be impaired. The alternative however to have a pervasive cursor which is always present confers its own set of problems as discussed in Luff et al (2003). A constantly visible cursor can make it harder to differentiate when cursors are being moved and when they are being used to gesture. The ability to create permanent traces with sketched lines does however make significant headway into resolving this problem.

### 6.6 The Nature of Remote Gesture – Some Conclusions

Having presented a wealth of description about the nature and practice of remote gesturing during collaborative physical tasks, using a variety of mediums to represent those remote gestures, this section attempts to summarise these earlier findings. The section then goes on to attempt to draw conclusions on the basis of this evidence as to why gesturing with hands has been observed to lead to better levels of performance than other means (see chapter 5), drawing specific reference to the construction of sketched objects, the satisficing nature of hand gestures and the development of mixed ecologies.

### 6.6.1 A taxonomy of remote gestures and gestural use

The simplest way to summarise the practices of gesture use that have been observed in the context of a collaborative physical task (namely a fine detailed assembly task) is to present an overall taxonomic view of the broad functions of these gestures. As stated previously the desire in observing the gestures was not to quantify them and assign them to taxonomic classifications adopted from the psychology literature, as this would have done little to explain how the gestures were being used in context. Whilst other work (e.g. Fussell et al 2004) has already categorised the broad functions of sketched gestures, the observations presented within this chapter extend the body of understanding of these sketched gestures and incorporate hand based gestures as a direct comparison.

| *Sketch / Gesture* | *Function* | *Description* |
|---|---|---|
| Pen-tip pointing and Movement<br><br>Cursor Pointing and Movement | Identification and Orientation | Deixis and some non-permanent figurative actions showing relative movement and manipulation |
| Circle<br><br>Dot<br><br>Underline (solid line or dashed)<br><br>Shading | Identification | Deixis and highlighting objects for perception. Focussing attention on items for selection, places on models and places on existing sketches |
| Arrows | Orientation and some Identification | Some simple highlighting (see above), but largely figurative actions demonstrating suggested movements and relative orientations of pieces |
| Draw Shape<br><br>Multiple Shapes | Identification and Orientation | When single shapes are presented in clear space they are being used to identify for search, when single shapes are drawn onto assembly parts the use is for both or separately identifying pieces for search and for showing relative orientation of application. Multiple pieces drawn together is to display relative orientations of pieces for assembly |
| Alpha-numerics | Extraneous information | Providing additional discourse supporting information |
| Delineating areas | Task-space management | To mark out areas of task space for specific activities |

Table 6.1 Functions of Sketching

The above table, Table 6.1 highlights the various sketch based gestures, describing their function and characteristics so that common patterns can be determined. Table 6.2 alternatively demonstrates a taxonomy of functions of the observed forms of *hand* gesture.

| Hand Gesture | Function | Description |
|---|---|---|
| Flashing Hands | Identification | Establishing reciprocity of perspectives |
| Wavering Hands | Identification | Indicates search for and location of items and coordinates selection of correct pieces |
| Negating Hands | Identification | Clarifies instructions and focuses attention on correct pieces |
| Drawing Hands | Identification and some Orientation | Further describes shapes of search items and in some instances is used to display relative orientations of pieces on assemblage |
| Mimicking Hands | Orientation of objects | Hands represent pieces to be assembled and show relative orientation for movement and connection |
| Inhabited Hands | Orientation of Worker | Hands represent Worker's hands showing relative orientation of hand movements to assembled pieces |
| Parked Hands | Task-space management | Marshals turn-taking behaviours within the task-space |

Table 6.2 Functions of Hand-based Gesturing

Comparison of the two tables demonstrates that a variety of sketch and gesture types serve similar purposes. The primary uses of both forms of gesturing appear to be the identification of objects in the task space currently being referred to, and the demonstration of relative orientation of the assembly pieces, at various stages of assembly. Some aspects of gesturing behaviour are however, uniquely afforded by the medium of their production. For example with sketching there is the ability to provide extraneous information such as alpha-numeric characters and the ability to mark off or delineate areas of the task space for specific functions. For those using hand based gestures there is comparatively, the ability to engage in a unique form of embodied interaction, where the hands can be used to directly model not just pieces for assembly but also the Worker's hands, to show relative manipulations that are required.

The analysis of observed gestures points to two key differences between sketched-based gesture systems and hand-based gesture systems, namely the construction of sketched objects and the ability to employ high level embodied gestures. These issues will be considered in the following sections.

### 6.6.2 The construction of sketched objects

A key feature of sketch-based remote gesture systems is their ability to create for common workspace consumption an additional class of collaborative object (additional to the task artefacts already existing such as the pieces to be assembled and the other forms of gesture being used). When sketch-based systems are used Helpers have the ability to create free-standing drawings. These new artefacts created during the process of interaction remain visible (in the absence of auto-wipe features) in the work-space for as long as required, and become an object in their own right, which is referenced and annotated. When complex drawings are constructed the Helper would invariably cease annotating the model that was being assembled

and would focus their attention on annotating the diagram that they were constructing. As discussed above, these diagrams are of particular use in certain contexts and can express relationships between items that would be hard to express verbally. However, the analysis has demonstrated that there are a variety of problems that are associated with the use of these sketches. The most prominent of these problems are the time required to construct the sketches, and the literal interpretation that recipients seem to give them. Quite separately an additional problem perhaps stems from the very nature of complex sketched diagrams being constructed as separate objects within the task space. In their absence all focus is attached to the items to be assembled and the collaborators refine their comments and actions to comments and annotations specifically regarding the piece in question. With the added entity of a sketched object which is being iteratively developed and referred to, then necessarily, attention must be split between the actual model and the drawn simulacrum. Such a process would be unlikely to occur in a face-to-face consultation between Worker and Expert (Helper) and as such may serve to fracture the interaction. Especially given the considerations highlighted previously which suggested that Helper's are likely to draw figurative diagrams by effectively copying the instructions that they have ensuring that the diagrams stay true to the orientations of the pieces that the Helper would wish to see rather than the orientations that the Worker might currently being holding the assemblage in. Therefore the effort of translating the spatial relationships demonstrated in the sketch into workable instructions relative to the orientations of pieces as they are currently held and viewed by the Worker is shifted from the Helper to the Worker. The Helper effectively absolves responsibility, saying 'this is what it should look like', and the Worker is left to figure this out. In the absence of a sketched object the Helper is more likely to suit their gestural instructions to the orientation that the assemblage is currently held in or else they will guide the Worker through manoeuvring the assemblage into a more useful orientation for the Helper. The asymmetric access to the drawn object (i.e. only the Helper can create and manipulate it) therefore becomes a problem, which could in turn affect collaborative processes.

### 6.6.3 The strengths of hand-based gesturing

Research evidence has suggested that gesturing with unmediated representations of hands during collaborative physical tasks, may lead to better performance than when gesturing is achieved through other means. Some of the problems associated with the use of sketched objects have been discussed above, as a partial explanation of this effect, but clearly for a full understanding the comparative strengths of the hand gesturing approach should be considered.

The strengths of unmediated views of hands start with the fact that there are two hands available (usually) for gesturing). Whilst it is acknowledged that the majority of gesturing only really requires one point source, either an indexical finger or a pen tip or a cursor, for the performing of pointing gestures, at the point at which a complex gesture demonstrating relative

orientations is required, the ability to use two hands is a significant advantage. People are used to manipulating an object with one hand and pointing at it or gesturing toward it with another. However, pen use reduces the number of gestural information points available down to one. Coupled with this is the fact that hands are in themselves highly complex features. They have multiple degrees of freedom of movement, with individual digits available, each offering separate gestural ability, multiple fingers can be used to point at multiple objects or just as easily be re-shaped and combined to model the interrelationships of a complex assembly piece. As discussed previously, hand gestures are less likely to be interpreted in an overly literal sense (compared to sketches) as such latitude is given to their interpretation. Although this potentially poses a comparative problem in that even the best hand gestures may well at times be insufficient to model a particular complex shape which may well be much more easily interpreted from a quick sketch. Hands do however have the advantage of being able to be re-used quickly and formed into multiple representations at varying levels of abstraction with relative ease. It is easy for a hand to represent a piece of Lego in one second and then be switched to represent the Worker's hand in the next. Sketches don't have this facility, and have to be re-drawn, demonstrating a reduced fluidity of multiple gesturing and sometimes leaving a temporal residue of un-required old sketches littering the work-space.

Consideration of the fluid nature of gesturing leads also to the understanding that hand-based gestures are also performed in an animated fashion. They effectively animate instructions. A sketch must convey the relative spatial orientations of pieces by showing the resultant end images of how combined pieces should look, unless an extremely complex gesture is sketched with multiple stages of interaction and functional descriptions of movement added with the use of arrows. Such an endeavour is a complex process, hands being much quicker as a way to show a real-time animation of an interaction in progress.

It was believed initially that the ability to refer to established sketched objects, which had been developed in the cultural-historical context of the collaborative task, would be a particularly powerful tool, but this was not the case. It may be that having such objects added to the space which can only be manipulated by the Helper and are therefore asymmetric elements of a shared space, creates fractures within the interaction. Hand gestures alternately, whilst not having the ability to develop highly complex objects, may satisfy the key requirements of the communication element of the task in that they can model complex spatial relationships enough to aid understanding of description.

The construction of a mixed ecology tries to develop a working environment in which collaborators share equally, and which is as close to the presumed optimal standard of face-to-face communication as possible. In a face-to-face situation a Helper would be unlikely to draw a complicated diagram and continually refer and gesture toward that rather than the pieces to be interacted with. When available however, the desire to sketch rather than explain can be quite powerful, but sketches can be easily misinterpreted. By using unmediated representations

of hands (as suggested by designing from a mixed ecologies perspective) collaborators are forced to focus their gestures more on the model as it is assembled, they do not have the ability to create a separate plane of information that the Worker must shift attention towards and collaborators are more likely therefore to stay task-object focussed. Gesturing with hands will also allow Helpers to perform complex embodied gestures conveying detailed implicit information regarding movements which are as discussed difficult to represent in a static sketch. Where results have demonstrated the superior ability of hand gestures to convey information over sketches in collaborative physical tasks, it is potentially due to these considerations.

## 6.7 Chapter Summary

Previous chapters have stated that gesturing with hands improved performance and that using hands in particular was best, but the analysis contained therein didn't explain why. To increase knowledge of how to develop remote gesture technologies appropriately, this 'why' must be fully understood. Knowledge of why remote gesture improves interaction helps to distil what the key aspects of interaction are, and understanding these key aspects suggests which features future technologies should be focused on supporting.

The work of this chapter therefore detailed the common practices and cycle of activities of remote gestural use during collaborative physical tasks. Within an identified common structure, or cycle of activities, including search, select and direction of manipulation gestures, observation was made of what gestures were used in each different gesture format and importantly how they were used and the impacts that this had on collaborative performance.

Through the development of an understanding of the praxiological character of remote gesture, insight has been derived which has articulated the way in which representations of gesture must be commonly understood between collaborators, they must be fast and fluid and ideally leave only as much temporal residue as is absolutely necessary (to keep the workspace clear). Gestures also directly benefit from occurring in three-dimensions and in being presented in a format such that the appropriate temporal course of an action can be determined. The analysis also suggested the benefits of gestures being presented in a format which avoids over literal interpretation, to allow for inconsistencies in understanding and to implicitly prompt the gesturer to keep trying to relay their intent until they are satisfied that the recipient has understood. This final observation in particular, has helped to further understanding of how interaction might become fractured through processes of misdirection. Put together these observations further suggest why the unmediated representation of hands as a remote gesture format should be considered an integral aspect of a mixed ecology communication device.

## Chapter 7 – How Gesture Interacts with Language

### 7.1 Introduction

Thus far the research presented within this thesis has demonstrated that remote gesture systems provide a significant enhancement to performance in collaborative physical tasks. It has been demonstrated that the structure of a remote gesturing system will have an impact on both the usability of the system and the performance benefit to be gained from its use. In particular an argument has been put forward that remote gesturing systems should be designed from a mixed ecologies perspective, which takes as its central tenets the notion that the reconstructed communicative environment should be as similar as possible to a naturally occurring face-to-face interaction. One element of this was the argument that unmediated views of hands should be the primary medium through which gesturing activity is facilitated, and the last chapter (6) has aptly demonstrated the complexity of gesturing behaviour that can be produced through such means. At various points in the thesis, arguments have been mooted in attempts to explain exactly how it is that this form of remote gesture has such an impact on performance. The arguments frequently centring on the notion that remote gestures in some way help to 'ground' spoken deictic references, reducing the translational overheads incurred when attempting to understand verbal instructions meant for a visuo-spatial medium. However, the argument that this effect works by replacing complex referential descriptions with simple pointing behaviours has been drawn into question by recent research. In this chapter the effects of remote gesturing on collaborative language are significantly unpacked, in an attempt to further understand the complex role for remote gestures in interaction. The research presented demonstrates how remote gestures, rather than merely acting as replacements for referential descriptions, actually significantly influence the *structure* of collaborative discourse, and consequently the temporal nature of the grounding process. Through generating a deeper understanding of these effects of remote gesturing on collaborative language a set of significant implications for the design, development and deployment of these technologies is generated.

### 7.2 Understanding Common Grounding and Remote Gesture

Previous research (e.g. Ochsman and Chapanis, 1974) suggests that the use of spoken language is the primary tool for achieving successful interactions in collaborative physical tasks. In these interactions it is primarily through spoken language-use that action is guided, interactions are structured and attention apportioned, and a fundamental result of these activities is the development of inter-subjective awareness between collaborators. As Clarke and Brennan argue *"all collective actions are built on common ground and its accumulation"* (1991, p.127), and it is the purpose of language in these remote collaborations to help establish this common

ground or mutual understanding (Clark, 1996, Clark and Brennan, 1991, Clark and Wilkes-Gibbs, 1986, Clark and Krych, 2004). This is of particular importance for collaborative physical tasks given their inherently object-focused nature. As Clarke and Brennan point out

> "*Many conversations focus on objects and their identities; when they do, it becomes crucial to identify the objects quickly and securely. Conversations like these arise, for example, when an expert is teaching a novice how to build things, and the two of them refer again and again to pieces of the construction*." (p.136)

It is imperative that collaborators possess common ground knowledge of mutual referents and mutual understanding within this class of assembly tasks. When a Helper directs a Worker to pick up a piece the Worker must understand which piece is being referred to, for the interaction to be considered successful. Perhaps then it can be assumed that remote gestures influence collaborative performance by, in turn, influencing the language that is used during interaction.

Such an argument relies heavily on the notion as discussed above that for interactions to be successful they must become adequately grounded. Research evidence suggests that critical to the establishment of this conversational grounding is the provision of shared visual access to collaborative task spaces (Clark and Krych, 2004, Gergle et al 2004a, Kraut et al 2002, Kraut et al 2003). When provided with this access it has been established that collaborators allocate most of their visual attention to images of the worker's hands and the shared task artefacts, evidence suggesting that such information is used by the Helper to establish confirmation of understanding from the worker (Fussell et al 2003). Clearly gesture is important when establishing understanding (and has been shown to be especially so with the understanding of spatially referent language (Rauscher et al 1996)).

At several previous points within this thesis, an argument has been put forward, suggesting that remote gestures have the influence they do because of their power to reduce the 'translational overheads' encountered when decoding complex instructions. The idea behind this is that complex referential descriptions are merely replaced by easier to understand figurative gestures. The gestures being presented visually managing to convey complex spatial information without this having to be coded into a verbal medium by the signaller and then decoded by the receiver and made relevant to their working context. Previous research, such as the studies of the DOVE system (Fussell et al 2004) has also suggested such a relationship between remote gesture and language. Fussell et al (2004) agreeing that complex forms of representational gesture,

> "*may facilitate conversational grounding in collaborative physical tasks by allowing speakers to communicate multiple pieces of information simultaneously.*" (p.280)

This ties in nicely with Clark and Brennan (1991) who, in their discussion of the various methods by which a communicative statement can be grounded, draw particular reference to the role of indicative gestures in grounding *deictic* references. They argue that according to the principle of 'Least Collaborative Effort' a deictic reference when accompanied by an appropriate gesture is particularly easy to interpret and therefore preferable to more complex sentence constructions for most collaborating partners.

A possible additional element that may contribute to the reduction in translational overheads is the 'evolution of referring expressions' (Krauss and Fussell, 1991). In their exploration of this concept Krauss and Fussell (1991) demonstrate that the explicit descriptions used to describe individual items being used / referred to during collaboration become names, the names of these items then going through a process of normalization until (if the term is used frequently) a shortened form is commonly accepted. This shortened form may bear very little resemblance to the original phrasing or indeed may be relatively indecipherable to any outsiders who have not been made aware of the development of the term. Remote gesture may influence this process by allowing items to be identified without the initial period of lengthy description – they are merely pointed at, a shortened name being used in these early stages, therefore the process of refining the description and name, taking significantly less time. This may in turn impact on the observable performance results.

In chapter 4 a basic finding was presented that demonstrated that remote gesturing improves collaborative performance. This finding is represented in figure 4.1 (repeated below as figure 7.1 for reference). Statistical analysis of the performance data from this study showed that the impact of remote gesturing was most significant during the first trial of use. Use of a remote gesture system in later trials (when participants had become much more practiced at this class of task) had not conferred any observable performance benefit. A key feature of this finding was that performance levels in the first trial, when using remote gesturing, were indistinguishable from later performance. The significant differences between gesturing and non-gesturing conditions stemming from participants in the voice only communication condition demonstrating significantly poorer performance in the first trial.
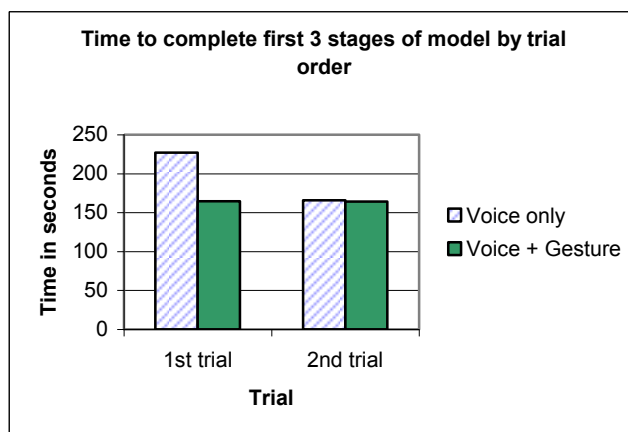
**Figure 7.1**

The data appears to suggest therefore that performance in a collaborative task will improve over time, a reasonable assumption demonstrated in similar studies (e.g. Fussell et al 2004). However, the data also suggests that using remote gesturing in an early trial has somehow lead to performance equal to later more practiced collaboration. The fact that remote gesture improves distance collaboration is now confirmed, but the process through which this phenomenon occurs has not been established, and why this effect should be so prevalent in early stages of interaction rather than later stages of interaction is also not understood.

One could seek to explain such findings with a consideration of the above discussion of the principles of grounding. It can be seen that for those participants who started collaboration with a voice only communication method, there was a significant impedance to performance. Their ability to ground deictic references was limited and they had to rely on more complex verbal descriptions to identify objects. It could be argued that this is what lead to the performance difference between gesturing and non-gesturing conditions in the first trial. Clearly if grounding of terms is a process which is inevitable (which surely it must be if a task is achievable) then it is simply the case that the cost of achieving grounded interaction is higher without gesture present and the price that must therefore be paid is time to achieve that grounding.

The model being presented is therefore that the process of grounding can be expedited through the use of indicative gestures (Clark and Brennan 1991) to replace complex referential descriptions and the use of such gestures to convey multiple layers of complex spatial information without the need to recode this information into a non Visio-spatial medium. This role of replacing verbose information with speedy hand gestures is however, less clear cut than it may at first seem. The study by Fussell et al (2004) clearly demonstrated that it was not the replacement of the referential descriptions by simple pointing behaviours that was responsible for the performance benefits of the gesturing technology. The performance benefits, they surmised, were derived from the infrequent use of more elaborate and complex forms of

gesture. Exactly what these uses of gesture were, or how they interacted with the collaborative discourse being engaged in, were issues not sufficiently discussed. Given that there is a large body of work within the social science literature that has considered the issue of how gesture interacts with discourse (e.g. Argyle, 1988, Bull, 2002, Kendon, 1994, McNeill, 1992), and a quantity of this has specific comment to make about how gesture can influence the structure of interaction (Duncan, 1972, Duncan and Fiske, 1977, Duncan and Fiske, 1985), it is perhaps a possibility that the uses of remote gesture may be more complex than has thus far been conceived.

To formulate any understanding of exactly how remote gestures are influencing collaboration one must look to the impact on the discourse as it is being engaged in during collaboration. This being the main conduit for collaborative action it must be through this medium that the influence of gesture is best observed. To this end, an experimental analysis of the language used during collaborative physical tasks was conducted, comparing linguistic performance in the gesturing and non-gesturing collaborative pairs from the first experiment (chapter 4). This re-analysis of the earlier experiment, focussing on the recordings of language use, attempted to ascertain the effects of remote gesturing on collaborative language and to test the veracity of the assumptions presented above about the nature of these effects of remote gesturing on object-focussed task-related conversation.

### 7.3 Study Methodology

### 7.3.1 Experimental design

For the language and gesture study the data was generated from transcripts of collaborative discourse during one of the experiments presented and discussed in chapter 4 (specifically, Remote Gesture vs. Voice Only Communication, section 4.2). The experimental design is therefore identical to that described in section 4.2.1.1.

### 7.3.2 Participants

Participants are the same cohort as presented in section 4.2.1.2.

### 7.3.3 Equipment

Equipment used was as reported in section 4.2.1.3 (and as detailed in depth in section 3.5.1).

### 7.3.4 Materials

Materials present during the study were the same as reported in section 4.2.1.4, however, the questionnaires etc. detailed, were not of relevance to this analysis.

### 7.3.5 Procedure

As reported in section 4.2.1.5.

### 7.3.6 Analysing the language

To analyse the language used by the participants as they performed their tasks, video recordings were taken from the camera placed above the Worker's desk (with an attached boundary mic situated between the two desks). From these video recordings transcriptions were created. The transcriptions formed two large samples. The first sample consisted of transcriptions of the first five minutes of interaction for 12 of the pairs of participants in both of their two trials (and therefore concerning collaboration for two different models). This sample was used to generate data about the average numbers of words and questions used by participants during their tasks. Having accomplished this broader spectrum analysis a more refined analysis was conducted. The interactions of 23 pairs were transcribed (including the 12 already transcribed, and excluding one pair from the analysis as they failed to reach the required level of task completion). This analysis focused specifically on conversation and action during the completion of stages 4 and 5 of one model only (therefore having data for each pair from one trial only). The transcriptions performed for this sample were significantly more detailed, following a conversation analysis (CA) methodology (see Antaki, 2005, tutorial on CA). This allowed a deeper level of analysis of the data, providing information on the use of verbal deixis during speech (specifically for proximal deixis, uses of the terms here, this, and these and for distal deixis, there, them, that, those, they), the extent of overlapping of speech and the coordination of physical hand gesture use with language. Data of this nature was aggregated and subjected to statistical analyses. Various other features of language structure were noted for description and to aid reading of transcripts, but were not directly analysed, these included pauses (recorded but left un-timed), increases in speed of speech, decreases in volume of speech and cut-off's during words.

### 7.3.7 Problems encountered

There were no significant problems encountered during the linguistic analysis of the data, other than as stated above that one pair of participants had to be excluded from the data analysis as they had failed to reach the required stage of the model that would have enabled their interaction to be analysed. Such a decision may seem unusual to those more regularly involved with conversational analytic work, as this disfluent pair who had significant technical difficulties would perhaps be of some interest to such researchers. However, maintaining consistency of the samples compared is a pre-requisite of effective *experimental* analysis, the methodology chosen for investigation of the *quantified* elements of language that were being investigated.

### 7.3.8 Statistical analysis

Statistical analysis was by t-test comparison between mean scores for various measures calculated for each of the two experimental conditions (i.e. the remote gesture enabled group and the voice only communication group), where appropriate post-hoc comparisons were also made, using t-tests, to compare differences between the first and second trials.

### 7.4 Results

### 7.4.1 Main findings

The first stage of the analysis was to understand some of the basic characteristics of the language used during collaboration sessions. From the sample of 12 pairs over two trials various measures of language use were recorded (taken from the first 5mins of interaction in each trial) including, total numbers of words used, total number of exchanges (turns) made, efficiency (in number of words per turn) of language used and relative proportion of total words spoken by each participant. All measures were aggregated and where appropriate split-out by specific task role. The separation by task role was seen to be appropriate as a two-way independent-measure t-test showed that, as would be expected given the nature of the task, Helpers speak significantly more than Workers (t (46) = 14.45, p≤ 0.001). It seemed appropriate therefore to treat them as very different samples for many of the measures. The basic results from the analysis for these various measures of language use can be seen below in table 7.1. All results are split-out by communication condition (e.g. voice only or voice and gesture communication). All variables were statistically compared using two-way repeated-measures t-tests to assess the differences between the communication conditions, table 7.1 includes the significance scores of these various t-tests.

| Measure | Total | Voice Only | Voice and Gesture | T-test Significance |
|---|---|---|---|---|
| Total Words Spoken | 765.67 | 765.33 | 766.00 | 0.49 |
| Number of Exchanges | 94.75 | 98.42 | 91.08 | 0.15 |
| Total Words (Helper) | 574.00 | 560.83 | 587.17 | 0.59 |
| Total Words (Worker) | 191.58 | 204.50 | 178.67 | 0.27 |
| Efficiency (Helper) | 12.51 | 11.66 | 13.37 | 0.15 |
| Efficiency (Worker) | 4.09 | 4.24 | 3.94 | 0.26 |
| Helper's proportion of total words | 74.70% | 72.69% | 76.72% | 0.20 |
| Worker's proportion of total words | 25.29% | 27.31% | 23.26% | 0.20 |

Table 7.1 Average numbers of various elements of language use during 1st 5mins of interaction (from 2 trials), split by gesture communication conditions.

As can be seen from table 7.1 above there were no statistically reliable differences observed between the communication conditions, for any of the measured variables. However, considering that it had been confirmed from the performance analysis that the effects of gesturing were most evident during the first trial only, the data collected was split-out by trial order and reanalysed. The results of this reanalysis can be viewed below in table 7.2 and 7.3 (table 7.3 is an addendum to table 7.2, showing the results of statistical comparison between figures in 7.2 in the first trial and second trial columns).

It would appear that for the majority of the measures taken there was again no significant difference between the gesture conditions, however this analysis revealed that this was so regardless of the trial order. Certain elements of discourse remained consistent despite changes to communication condition and largely, although there were what appeared to be changes of language use over the two trials, language seemed to remain consistent for the measures calculated over the course of the two trials. There was however, some significant reduction in the total words used between the two trials for those Workers using voice only communication.

| Measure | First Trial | | | Second Trial | | |
|---|---|---|---|---|---|---|
| | *Voice Only* | *Voice and Gesture* | *T-test Significance* | *Voice Only* | *Voice and Gesture* | *T-test Significance* |
| Total Words Spoken | 755.00 | 758.67 | 0.96 | 775.67 | 773.33 | 0.97 |
| Number of Exchanges | 102.17 | 96.67 | 0.55 | 94.67 | 85.50 | 0.47 |
| Total Words (Helper) | 517.50 | 581.67 | 0.38 | 604.17 | 592.67 | 0.87 |
| Total Words (Worker) | 237.50 | 177.00 | **0.03** | 171.50 | 180.33 | 0.81 |
| Efficiency (Helper) | 10.19 | 11.91 | 0.19 | 13.12 | 14.82 | 0.56 |
| Efficiency (Worker) | 4.76 | 3.76 | 0.10 | 3.73 | 4.13 | 0.58 |
| Helper's proportion of total words | 68.31% | 76.12% | **0.05** | 77.06% | 77.32% | 0.96 |
| Worker's proportion of total words | 31.69% | 23.88% | **0.05** | 22.94% | 22.64% | 0.95 |

Table 7.2 Average numbers of various elements of language use during 1st 5mins of interaction (from 2 trials), split by gesture communication condition and trial (Significant differences between gesture conditions shown in bold)

| Measure | Significance of First trial to Second Trial Changes | | |
|---|---|---|---|
| | *Total* | *Voice Only* | *Voice and Gesture* |
| Total Words Spoken | 0.72 | 0.75 | 0.85 |
| Number of Exchanges | 0.22 | 0.45 | 0.36 |
| Total Words (Helper) | 0.32 | 0.26 | 0.87 |
| Total Words (Worker) | 0.18 | **0.04** | 0.92 |
| Efficiency (Helper) | 0.07 | 0.19 | 0.23 |
| Efficiency (Worker) | 0.48 | 0.17 | 0.53 |
| Helper's proportion of total words | 0.12 | 0.07 | 0.77 |
| Worker's proportion of total words | 0.11 | 0.07 | 0.76 |

Table 7.3 T-test significances (two-way independent-measures T-tests) **comparing first and second trials** for various measures of language use, split by gesture communication condition (statistically significant scores shown in bold)

Looking in detail at the significant scores it became apparent that the relative proportion of the total words spoken by both Helper and Worker appeared to be influenced by the use of the remote gesture tool, but only in the first trial (for further illustration see table 7.2 and figure 7.2).
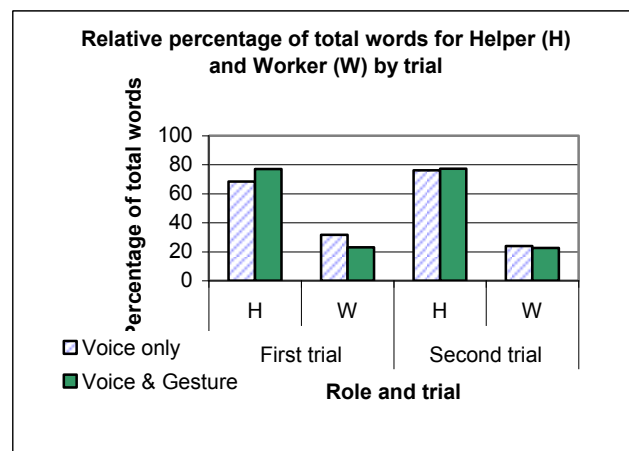


Figure 7.2

Two-way independent-measures t-tests revealed that that there was a significant difference (t (10) = -2.20, p≤ 0.05), between voice only and voice and gesture conditions in the first trial for both Helper and Worker percentages of total words used. By the second trial however the relative percentage of Helper and Worker words had become consistent between the communication conditions (again see figure 7.2). The biggest impact of remote gesturing appeared to be on the Workers total word use in the first trial. Again a two-way independent-measures t-test confirmed that there was a significant difference (t (10) = 2.48, p≤ 0.03) between the gesturing conditions. This increase in Worker words in the first trial mirrors the performance results from chapter 4, discussed above, that found longer average completion times for voice only communication in early stages of collaboration. Having gained evidence therefore that the differences in performance time, attributed to the use of remote gesture, could in fact be based on some property of the language being used and how gesture interacts with this, and potentially an alteration of the content of the language being used, the investigation was focused more on the specific content of the speech during interaction.

The first point of the content analysis was to consider the use of questions during collaboration. Similar to the prior analysis a variety of measures were calculated concerning the use of questions. Starting with a basic count of the number of questions used in total, and then split out by Helper and Worker roles and trial order. This analysis progressed to investigate such questioning/language behaviour as the number of Worker Words and Total words per Worker question and the number of Total Words per Total Questions and finally the proportions of Total, Helper and Worker exchanges containing a question statement. In several instances the number of Helper questions was not further analysed because the prevalence of Helper questions was so low compared to the occurrence of Worker questions. Indeed statistical comparison using a two-way independent-measure T-test demonstrates that Helpers are significantly less likely to ask questions during interaction than Workers (t (46) = 10.33, p< 0.001). The results for these various analyses are shown below in tables 7.4, 7.5 and 7.6.

| Measure | Total | Voice Only | Voice and Gesture | T-test Significance |
|---|---|---|---|---|
| Total Questions | 27.08 | 28.42 | 25.75 | 0.32 |
| Helper Questions | 5.58 | 5.42 | 5.75 | 0.84 |
| Worker Questions | 21.50 | 23.58 | 19.42 | 0.12 |
| Worker Words per Worker Questions | 9.45 | 9.27 | 9.63 | 0.80 |
| Total Words per Worker Questions | 33.61 | 30.51 | 36.72 | 0.09 |
| Total Words per Total Questions | 29.91 | 27.76 | 32.05 | 0.23 |
| Total Turns containing a question | 28.73 | 29.27 | 28.19 | 0.60 |
| Helper's Turns containing a question | 12.22 | 11.92 | 12.52 | 0.88 |
| Worker's Turns containing a question | 45.47 | 46.78 | 44.15 | 0.52 |

Table 7.4 Average numbers of Questions and Questions per various Word counts asked during 1st 5mins of interaction (from 2 trials), split by gesture communication conditions

| Measure | First Trial | | | Second Trial | | |
|---|---|---|---|---|---|---|
| | *Voice Only* | *Voice and Gesture* | *T-test Significance* | *Voice Only* | *Voice and Gesture* | *T-test Significance* |
| Total Questions | 30.67 | 29.17 | 0.51 | 26.17 | 22.33 | 0.39 |
| Helper Questions | 5.33 | 6.83 | 0.53 | 5.50 | 4.67 | 0.73 |
| Worker Questions | 25.33 | 22.33 | 0.27 | 20.67 | 17.67 | 0.51 |
| Worker Words per Worker Questions | 9.49 | 8.41 | 0.48 | 9.04 | 10.85 | 0.48 |
| Total Words per Worker Questions | 30.44 | 35.36 | 0.35 | 43.25 | 49.21 | 0.60 |
| Total Words per Total Questions | 24.96 | 26.03 | 0.68 | 30.57 | 38.07 | 0.20 |
| Total Turns containing a question | 30.41% | 30.38% | 0.99 | 28.14% | 26.01% | 0.53 |
| Helper's Turns containing a question | 10.53% | 14.46% | 0.43 | 13.42% | 10.81% | 0.68 |
| Worker's Turns containing a question | 50.29% | 46.30% | 0.43 | 42.85% | 41.21% | 0.80 |

Table 7.5 Average numbers of Questions asked during 1st 5mins of interaction (from 2 trials), split by gesture communication condition and trial

| Measure | Significance of First trial to Second trial changes | | |
|---|---|---|---|
| | *Total* | *Voice Only* | *Voice and Gesture* |
| Total Questions | **0.03** | 0.13 | 0.12 |
| Helper Questions | 0.54 | 0.95 | 0.31 |
| Worker Questions | 0.08 | 0.20 | 0.25 |
| Worker Words per Worker Questions | 0.49 | 0.79 | 0.32 |
| Total Words per Worker Questions | **0.03** | 0.20 | 0.11 |
| Total Words per Total Questions | **0.01** | 0.14 | **0.03** |
| Total Turns containing a question | 0.10 | 0.48 | 0.10 |
| Helper's Turns containing a question | 0.92 | 0.68 | 0.38 |
| Worker's Turns containing a question | 0.11 | 0.20 | 0.39 |

Table 7.6 The statistical significance (two-way independent-measures T-tests) **comparing first and second trials** for average numbers of questions asked and proportions of questions per various measures of words (also split by gesture communication condition, statistically significant scores shown in bold)

This analysis revealed little statistically reliable difference between the two conditions. There was however, more evidence of a change in gesturing behaviour through the progression of the two trials. The analysis of the differences between the first and second trials revealed that the total number of questions asked decreased from the first trial to the second (t (22) = 2.39, p≤ 0.03), accompanied by significant increases in the number of words used per questions asked (t (22) = -2.88, p≤ 0.01). In particular the total number of words used between Worker questions showed significant increase (t (22) = -2.25, p≤ 0.03); this suggests that the Worker's were asking less questions, less frequently as practice with the task increased. This lowering in the frequency of questions asked over time was particularly significant for those collaborating using remote gesturing (t (10) = -2.51, p≤ 0.03), remote gesturing in a later trial being marked by a reduction in the need to ask questions. Whilst no firm conclusions about the effects of gesture on number of questions asked could be derived from the analysis most trends in the

data suggested that more questions were asked by the collaborators when they were communicating in the voice only condition and had their access to instructional gesture restricted. This suggested that part of the component of increased speech for Workers in early trials and voice only conditions was based on the need to formulate more questions.

By comparing excerpts 1 and 2 below[11] (taken from the transcripts of interactions) the nature of the differences behind this desire to adopt questioning behaviour can be understood.

**Excerpt 1 – Pair T2 – Voice & Gesture –Trial 2**

H      and place the short end erm (.) in on the sticky thing erm <u>other way round</u> (.) ((*index finger pointed circles hand in vertical plane, then moves hand towards object extends thumb and performs a rotate motion*))

H      Yeah. (.) err <u>rotate it that way</u> ((*uses thumb and forefinger on desk to trace desired angle of rotation, with thumb as the axis point*))

H      <u>the the yellow bit</u> (.) ((*uses two hands index fingers touching desk one high one low tracing movement in opposite directions till fingers are level*))

H      <u>Yep</u> that'll do= ((*fingers move from finish point of last gesture to off the table*))

**Excerpt 2 – Pair T15 – Voice only –Trial 2**

H      er:m (.) and (.) that should be: (.) just hold it up a bit (.) er:m (.) ok >that should actually <take it off again and put it on the other way round in the same hole

W      in the same [hole]=

H      [yeah]

W      =but the other way round?

H      just flip it round

W      you mean just could of turned it? hahah

---

[11] Guide to notation in excerpts – Excerpts use a simplified version of conversation analytic notation conventions as used in Fraser (2000). H refers to Helper, W to Worker. Pauses in speech are marked with (.). Parts of text accompanied by a gesture are <u>underlined</u>. Descriptions of the gestural action are given in brackets (( )) at the end of turns. Overlaps are marked with [ ]. Rapid speech is marked with > <. Changes of turn with no discernible gap are marked with =. Words cut short are noted with a dash e.g. -

H       err: yeah and now just swivel it round a bit ok keep going keep going stop there

W       right.

Whereas in Excerpt 1 the Helper is very directive (the Worker not needing to respond verbally) using figurative gestures to clarify difficult to describe concepts such as relative angle of rotation, in Excerpt 2 the Worker is forced to question the instructions, manipulating the pieces first and then waiting for or requesting clarification that the action is correct.

At this point it was felt that a more refined analysis of data was needed so, as discussed previously, a new data sample was constructed which allowed a more focused analysis of the language to be performed. Taking the sample of 23 transcripts (all based on interactions with exactly the same stages of the same Lego model) an analysis was conducted into the use of deictic referencing within the collaborative object-focused discourse. Various measures were calculated including, the number of turns used that included a deictic phrase (specifically for proximal deixis, uses of the terms here, this, and these and for distal deixis, there, them, that, those, they). The percentage of turns containing such a phrase was also calculated, as was the total number of deictic phrases used (independent of turns taken) and specifically the numbers of both proximal and distal deictic references made. These calculations were made for both participants combined and also calculated separately for Helpers and Workers. The results were also split out by communication condition and trial order. The average scores for each of these various sub-groups can be seen below in table 7.7.

Table 7.7 Average numbers of various measures of deictic referencing split by communication condition and trial order

|  |  | Voice Only | | | Voice and Gesture | | | Trial Order Totals | |
|---|---|---|---|---|---|---|---|---|---|
|  |  | 1st Trail | 2nd Trial | Total | 1st Trail | 2nd Trial | Total | 1st Trail | 2nd Trial |
| Total | Turns with a deictic phrase | 12.00 | 9.67 | 10.83 | 17.20 | 8.17 | 12.27 | 14.36 | 8.91 |
|  | Percentage of turns | 30.24% | 31.53% | 30.89% | 40.02% | 35.15% | 37.37% | 34.68% | 34.04% |
| Helper | Turns with a deictic phrase | 6.17 | 6.17 | 6.17 | 10.00 | 4.83 | 7.18 | 7.91 | 5.45 |
|  | Percentage of turns | 29.91% | 41.06% | 35.49% | 46.27% | 39.59% | 42.63% | 37.35% | 40.96% |
|  | Total deictic references | 8.33 | 8.33 | 8.33 | 13.80 | 7.83 | 10.55 | 10.82 | 8.27 |
|  | Proximal | 0.00 | 0.00 | 0.00 | 1.40 | 0.83 | 1.09 | 0.64 | 0.45 |
|  | Distal | 8.33 | 8.33 | 8.33 | 12.40 | 7.00 | 9.45 | 10.18 | 7.82 |
| Worker | Turns with a deictic phrase | 5.83 | 3.50 | 4.67 | 7.20 | 3.33 | 5.09 | 6.45 | 3.45 |
|  | Percentage of turns | 30.64% | 20.67% | 25.65% | 33.45% | 27.26% | 30.07% | 31.91% | 24.54% |
|  | Total deictic references | 6.17 | 4.00 | 5.08 | 8.20 | 3.67 | 5.73 | 7.09 | 3.91 |
|  | Proximal | 2.83 | 0.67 | 1.75 | 2.00 | 0.33 | 1.09 | 2.45 | 0.45 |
|  | Distal | 3.33 | 3.33 | 3.33 | 6.20 | 3.33 | 4.64 | 4.64 | 3.45 |

These various sub-groups were then compared statistically using two-way independent-measures t-tests to ascertain whether there were any significant differences between the communication conditions. The results of these t-test comparisons can be seen in table 7.8 below.

| | | Voice Only vs. Voice and Gesture T-test Comparisons | | | Trial Order (1st trial vs. 2nd trial) T-test Comparisons |
| --- | --- | --- | --- | --- | --- |
| | | 1st Trial Only | 2nd Trial Only | Both trials combined | |
| **Total** | Turns with a deictic phrase | 0.35 | 0.68 | 0.67 | 0.09 |
| | Percentage of turns | 0.33 | 0.74 | 0.36 | 0.85 |
| **Helper** | Turns with a deictic phrase | 0.20 | 0.48 | 0.56 | 0.16 |
| | Percentage of turns | 0.15 | 0.89 | 0.34 | 0.69 |
| | Total deictic references | 0.16 | 0.83 | 0.32 | 0.21 |
| | Proximal | 0.15 | 0.16 | **0.04** | 0.69 |
| | Distal | 0.26 | 0.58 | 0.60 | 0.23 |
| **Worker** | Turns with a deictic phrase | 0.61 | 0.93 | 0.80 | **0.05** |
| | Percentage of turns | 0.78 | 0.63 | 0.60 | 0.34 |
| | Total deictic references | 0.52 | 0.87 | 0.73 | 0.07 |
| | Proximal | 0.69 | 0.21 | 0.53 | **0.05** |
| | Distal | 0.55 | 1.00 | 0.37 | 0.37 |

Table 7.8 T-test significance scores for deixis use data comparisons of gesture conditions and trial order (significant scores, to the p≤0.05 level, are shown in bold)

The results indicated that Workers were more likely to include a deictic phrase in a turn in the first trial as opposed to the second ($t(21) = 2.03$, $p= 0.05$), and in particular they were more likely to use a proximal deixis reference (such as here, this, these) in the first trial ($t(21) = 2.03$, $p= 0.05$), rates of usage of distal deictic references staying largely the same between trials. This suggests that early components of Worker's conversation are more likely to include explicit deictic reference to items near them, potentially considering the results above, questioning whether specific items are those to which the Helper is referring. This type of interaction is exemplified below in Excerpt 3, wherein the Worker has to repetitively refer to different bits of the piece in question until the right area is located.

**Excerpt 3 – Pair T12 – Voice only – Trial 1**

H      =err the third hole away from the corner (.) from the shortest end (.) if you
          [s-]

W    [tha-] that hole?

H    err away from the corner (.) the other end

W    >there<?=

H    =>no no< (.) >other side< (.) >other one<

W    there?=

H    =the short end

W    short end [here? one]=

H    [short en-]

W    =or two

H    >no just go< (.) get the other part of the L

W    the other part what this part?

H    yeah

An additional point of interest considering the use of deixis was the finding that Helpers are significantly more likely to use proximally deictic references when using remote gesture tools as opposed to when relying on voice only communication (t (21) = -2.23, p≥ 0.04). Considering the observations of correlations between deictic linguistic features and sense of presence reported in Kramer et al (2006) this would suggest that Helpers, when allowed to use remote gesture, feel more like they are actually part of a shared working space, as opposed to providing external guidance to actions in somebody else's working space. A comparison of Excerpts 4 and 5 below demonstrates this point. In excerpt 4 the Helper (line 1) explicitly changes their sentence construction to adopt a proximal deictic reference e.g. 'this piece here'. This is combined with an explicit gestural action to direct the Worker's attention. The gestural action is a requisite function, enabling the proximal deictic reference to be adequately grounded for the Worker to understand its' reference point. In excerpt 5 such an approach to grounding is not possible, a proximal deictic reference could not be supported by a gestural action, so the work of grounding is shifted to more linguistic effort. References such as 'this piece here' are dropped in favour of explicit verbal description e.g. 'the yellow L shape'.

**Excerpt 4 – Pair T22 – Voice & Gesture – Trial 2**

H    right then you need a:: (.) the: this piece here (.) yep (.) ah: then you need to put the: the shorter edge the end (.) this end (.) yeah that needs to clip through on the black piece ((*right index finger pointed forward to yellow L*)) ((*right index and thumb form C and bounce vertically*)) ((*right index points at end hole of yellow L*))

**Excerpt 5 – Pair T7 – Voice only –Trial 2**

    H      ok then the yellow L shape

    W      yeah

    H      erm should go onto that (.) the piece you just put into the corner of the L

    W      which [piece]?

    H      [it should] (.) the err sorry the bottom of the smaller side of the yellow shape the bottom half (.) and it should go over the other bit as if it's like a crane.

    W      right.

Whilst the use, therefore, of proximal deixis may necessarily be constrained by the availability of gestural support, when such language is used, it demonstrates that the Helpers do indeed accept that they are working in a shared space with the Workers, rather than providing external support.

Having looked at the use of deictic referencing within the sample the analysis was turned to focus on the extent of overlapped speech within the collaborative discourses. The percentage of overlapped turns (i.e. those exchanges of turn which are marked by overlapped or 'interrupted' speech) are shown below in figure 7.3
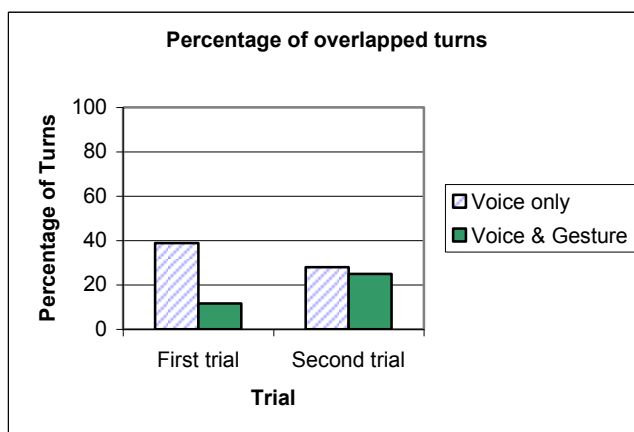


**Figure 7.3**

As can be seen in Figure 5, voice only communication appeared to lead to increased overlapping of speech during speaker exchanges. This appeared to be most prevalent in the

first trial. To understand and further test the interaction of gesture use and trial order the overlap data was analysed in a 2x2 independent measures ANOVA. This found a significant main effect of communication condition ($F_{(1,45)}$ = 5.51, p≤0.03), suggesting that the use of remote gesture is significantly associated with a reduced occurrence of overlapped turns. It failed however, to find a main effect of trial order ($F_{(1,45)}$ = 0.02, p=n.s.) suggesting that the prevalence of overlaps did not alter over time per se. But there was a significant interaction effect ($F_{(1,45)}$ = 4.26, p≤0.05). This suggests that *after* an early critical period, the use of a remote gesture tool is unlikely to have an effect on the overlapping of speech. However, in *early* periods of use not using a remote gesturing tool leads to increased levels of disfluent overlapped speech. Clearly as can be witnessed in excerpt 6 below (overlaps marked by [ ] symbols) when speech becomes disfluent and overlaps occur communication can become quite effortful. Excerpt 6 starts with the Helper having to describe a piece to pick up ('a yellow bit'), this description is clearly inadequate so the Worker understands that more must be done to ground the message, they begin to offer example pieces that might satisfy the search criteria. Realising this is about to happen the Helper then tries to force a change of turn by overlapping speech, extending the word 'with' until the Worker ceases talking. At this point the Helper can then finish their description work. Note that even after the more exact description the Worker seeks clarification that they have the right piece (not needed when gesturing is available cf. excerpt 4). The real problems occur seconds later however, when the pieces selected must now be connected. The Helper needs to draw reference to a specific area of one Lego piece (namely the bottom hole), without the ability to point to this area again the Helper describes it verbally. Attempts to do this however are hampered by the Worker, who keeps interjecting and offering alternatives. In this instance the Helper must keep repeating themselves, as the Worker's extra words are not actually helping to ground the Helper's meaning.

**Excerpt 6 – Pair T16 – Voice only –Trial 1**

  H  err right ok a yellow bit

  W  erm  >[yeah we've got]<

  H  [wi:th] five holes on the top and three going down=

  W  =yep this one?

  H  yeah you wanna flip it over (.) no so that (.) no put yeah but put it down
     flat with (.) yeah but the other way round huhuh yeah

  W  yeah

  H  erm and now that goes over the other side of that that you've just added on

  W  oh ok so that what so [it goes]

  H  [so the] bottom hole of that pick up the yellow thing

W        yeah (.) [that bit]

H        [the bott]

W        [under my finger]

H        [no no the bottom hole] the bottom hole go down no down from your left finger

W        [it's]

H        [yeah] the bottom one of there

W        yeah

It would appear that at those points in a Helper's discourse where there is hesitation there is a tendency for the Worker to feel obliged to start to attempt to force a change of turn. In excerpt 7 below however, an example can be seen of how gesturing can be used to prevent this from happening.

**Excerpt 7 - Pair T4 – V&G –Trial 1**

H        ok now we're gonna need the bl- the other black L shape thing it's er:m it's the one with one bit on the end (.) ((*right hand index out circles over the pieces and withdraws*)) ((*a second gesture with same hand circles some items on the lower right hand side of the desk*))

W        this one here?

H        that's that's the one James yeah=

W        =right

In this example the Helper's initial turn contains disfluencies such as cut-off words and continuation terms like 'erm'. But as the use of erm is covered by an accompanying gesture no interruption is seen. Whilst the end of the Helper's first spoken turn is marked by silence, the Worker does not immediately take over the turn as they wait for the Helper's gesture to be completed first. This should be compared with the Worker's fifth turn in excerpt 6. In this the obvious pause after 'yeah' ensures that as the Helper attempts to continue their turn with 'that bit' the Worker, in the absence of any gestural evidence to the contrary, marks the pause as a potential point of interjection to take over the turn. Clearly there is evidence therefore of gesture enabling smoother turn-taking.

As the raw data had demonstrated that amongst the sample there was significant difference between individual collaborating pairs and the extent to which they adopted use of the remote gesturing tool, a correlation analysis of the data was performed. Correlating the percentage of Helper turns that contained a physical gesture component with a basic measure of performance

(time taken to complete first three stages of model), it was observed that there was a negative relationship between task performance time and the total number of gestures used ($r_s$ = -0.79, p≤0.01). Extending this analysis to factor in the percentage of turns which were overlapped, it was also observed that an increase in the percentage of turns including a physical gesture from the Helper was associated with a decrease in the number of turns including an overlap ($r_s$ = -0.68, p≤0.03). And finally it was noted that increases in number of overlaps in a discourse is associated with increases in performance time ($r_s$ = 0.69, p≤0.02), more overlaps therefore being associated with slower performance. The results of the correlation analysis therefore appear to suggest that the use of a remote gesture tool improves collaborative performance and reduces the probability of disfluent speech and overlaps during discourse. With further evidence that increases in overlaps degrade collaboration this bolsters understanding of how gesture interacts with language to improve performance.

### 7.4.2 Results summary

In summary therefore, the results have demonstrated that the performance benefits of remote gesture tools appear to be strongest during early stages of an interaction. During these early stages if a remote gesture tool is used it has the potential to reduce the amount the Worker in the interaction needs to speak. Whilst questioning behaviour from the Worker is slightly lessened by gesturing it stays fairly consistent over time, however, it is likely to be combined with deictic referencing in early voice only interactions as Worker's are forced to point to various alternative pieces for the Helper – so as to establish their common points of reference. When remote gesture is used, the Helper seems more engaged directly in the task space, exhibiting increased use of proximal deixis. In turn it has been shown that the use of remote gesturing reduces the occurrence of overlapped speech, therefore demonstrating that remote gesture helps to smooth interaction and facilitate clear turn-taking during collaboration. This smoother more structured form of interaction allows for better performance. Clearly therefore the effects of remote gesture are not simply a benefit gained from the replacement of verbose referential descriptions being replaced by simple deictic references accompanied by pointing gestures. The role of gesture in collaborative language is much more complex and a large proportion of the benefits to be gained from using remote gesture tools are to be found in the way in which gestures can be used to regulate and structure interaction at early points of confusion in a task.

### 7.5 Discussion

### 7.5.1 Achieving grounded interactions

The aim of this chapter's investigation of language use during collaborative action was to attempt to ascertain the effects of remote gesturing on collaborative language and to test the

veracity of assumptions about the *nature* of these effects of remote gesturing on object-focussed interaction. It was felt that to further support the notion that such communication devices should be designed from a mixed ecologies perspective, a deeper understanding was required of exactly how remote gesture improves performance. The detailed results presented above, expand understanding of the role of gesture in remote collaborations. Combined with existing research in the area, a complex role for gesture in interaction is illuminated, which has implications for the design, development and deployment of remote gesture technologies. Discussion of these issues is presented below.

Fussell et al (2004) had already demonstrated the performance benefits of using remote gesture tools in collaborative physical tasks; showing that higher rates of remote gesture use were correlated with faster task performance, and that the use of a gesture tool leads to higher rates of proximal deixis use amongst Helpers. The results of the language analysis confirmed these findings demonstrating that they remained true when the format of remote gesture was altered from a digital sketch to an unmediated representation of hands.

The study did however, also demonstrate effects of *receiving* remote gestures on *Worker* language. These effects suggested that where remote gesturing was not used the interaction was less directed by the Helper, with more effort in establishing mutual referents being shifted to the Worker, who consequently had to increase the amount of words they used during interaction. Evidence was also noted that this increase in Worker words was related to an increase in the need to formulate questions early on in interactions. In addition to these findings, an unexpected but potentially influential result was the observation that use of remote gesture tools significantly impacts on the presentation of overlaps in natural discourse, remote gesturing significantly effecting smooth turn-taking. Equally of interest, it was in early trials that there was an increased likelihood of overlapped exchanges between Helper and Worker unless remote gesturing was used; if remote gesturing was used this lead to a significant *reduction* in the amount of overlapped turns, but the presence of this effect was only observed in the first tested trial.

At various points in the thesis an argument has been put forward concerning the nature of the influence of gesture on collaboration. Put simply the argument states that complex (but easy to produce) representational gestures are used to replace difficult to interpret, complex referential verbal descriptions. Such a position has been suggested by other researchers (e.g. Kraut et al 2003, Fussell et al 2000, Karsenty, 1999) and it would be indefensible to argue that this is not true to some extent. The extensive discussion presented in chapter 6 clearly demonstrating the complex layering of information that can be produced by easily formed concrete iconic gestures. The data presented here in chapter 7 should be interpreted as adding a further layer of explanation to the role of gesture in remote collaborations. Clearly the turn-taking 'parked hands' style gestures referred to in the previous chapter have a significant impact on the structure of the interaction, even though they are not directly object-focussed actions, the

research evidence demonstrates a clear role for remote gestures in structuring discourse during collaborative action. The results of the overlaps analysis specifically demonstrating that remote gesturing has a role to play in facilitating smooth turn-taking and supporting inter-subjective awareness. Where previously a correlation had been demonstrated between remote gesture use and task performance (Fussell et al 2004), this has been extended by demonstrating correlations between remote gesture use and speech overlaps, and speech overlaps and task performance. This suggests a firm link between how an interaction is structured and how successful it will be. Clearly the finding that gesture can be used to facilitate smooth turn taking has been demonstrated previously in other social science fields (see Duncan, 1972, Duncan and Fiske, 1977, Duncan and Fiske, 1985). In most of this research however, the focus has been on dyadic, face-to-face interactions during conversational communication. Little focus being given to task-oriented object-focussed interactions. And certainly no one has considered if such effects of gesturing would remain true when gestures are disassociated from the signaller's body and artificially projected into another's working space. This study provides evidence that this is indeed the case.

Further, it could be argued that if interaction is smoothed and overlaps reduced and the Helper has the ability to present gestures of their own, the ability to provide back-channels and therefore demonstrate mutual understanding is enhanced. It has been suggested that the ability to provide gestural information acts as a back-channeling device (Clark, 1996, Clark and Brennan, 1991) and back-channeling speeds up the process of grounding terms (Clark and Brennan, 1991, Clark, 1996). It is also possible then that remote gesturing is influencing the collaboration process in this way.

Integral to these arguments is a consideration of the costs of grounding (see Clark and Brennan, 1991 for discussion of this concept). As Clark and Brennan discuss there are a variety of different grounding costs in communication which can be more or less prevalent depending on the communication media adopted. Traditional views of the effects of remote gesturing, that saw its benefit purely in terms of the replacement of referential descriptions with observable gesture, focused attention on how remote gesturing reduces the production, reception and understanding costs of grounding. However, given this study's observations that remote gesturing has a significant influence on the structure of collaborative discourse it is apparent that remote gesture use should also reduce delay, speaker change and repair costs of grounding. The presence of remote gestures, alleviating the likelihood for early interruption when utterances are being formed or modified, therefore reducing the number of failed attempts at turn-taking requiring that significantly less time be expended on costly sentence repair phases.

The language analysis has also illustrated that for all significant results from the basic performance effects to the overlap analysis, the benefits of remote gesturing are inherently tied to the time course of the grounding process and are affected by experience with the study

tasks. As participants became more experienced with the tasks, performance improved (as would be expected). It would seem that what is happening in these tasks is performance is becoming grounded. As a task progresses the collaborators are establishing and adding to a shared communicative environment (Krauss and Fussell, 1991) and the words and effort required to refer to shared artefacts are reduced (Clark and Brennan, 1991, Krauss and Fussell, 1991). Comparing early trials with later trials demonstrated clear differences in behaviour, as evidenced by the changing pattern of questions asked across the trials. What the result demonstrate however, is that through the use of a remote gesture tool the effects of an early lack of grounding can be ameliorated. When collaborators used the gesture technology in early trials there was a significant improvement in performance. In fact performance was at levels observed in later stages, where it could be argued that grounding had been successfully achieved. The benefits of remote gesturing disappear however, by the later trials. Given enough time all distance collaborations will become grounded. If communication is restricted to an audio connection only, it does not become impossible. The process of achieving grounded interaction (and therefore optimal performance) merely takes longer to achieve. It would appear therefore that in a collaborative physical task a remote gesture tool will have its most influence during early un-grounded stages of interaction. This has significant implications for the deployment of such technologies which will be discussed further in the next chapter.

### 7.5.2 Implications for mixed ecologies

A key feature and aim of this chapter was to use an analysis of the language being used during collaboration to determine support for a mixed ecologies approach to the design of remote gesture technologies. The results of the study have clearly demonstrated that where possible collaborating partners will (probably unconsciously) adopt features of remote gestural activity as indicators of complex behavioural strategies. The evidence gleaned from studies of face-to-face interaction has demonstrated that gesturing behaviour can act as a turn-taking device and a floor-holding device, and as a body of interactive behaviours represents a whole host of subtle and possibly epiphenomenal interaction structuring devices. That such behaviours would be continued when gesturing was promoted purely as a device for conveying remote spatial information, and when gestures were literally divorced from the signaller's body, is somewhat un-expected. However, this would confer support for any approach which seeks to design communications technologies from a mixed ecologies perspective. The ultimate aim of a mixed ecologies approach being the construction of a shared communicative environment in which the key salient features of face-to-face interaction are captured and made available. With the use of naturalistic forms of gesture (i.e. unmediated hand representation and embedding directly into the working environment) a variety of these natural interaction structuring

properties of gesturing can be preserved for use, and it has been demonstrated by the results of this study that in doing so collaborative performance can be enhanced.

**7.6 Chapter Summary**

Chapter 7 significantly unpacked the effects of remote gesture on collaborative language used during interaction in physical tasks, in an attempt to further understand the complex role for remote gestures in interaction. The chapter presented previous research which suggested that the benefit of remote gesturing is in its ability to support the grounding process, increasing mutual understanding between collaborators. Investigating the language used during earlier experiments, a more complex role for gesture was highlighted, centred on observations of the changing nature of the content of discourse when remote gesture devices were used to aid collaboration. In particular evidence was noted of the changing pattern of question usage and the changing pattern of speech overlaps, in the presence of remote gesture.

Remote gesturing was demonstrated to have a specific influence on the structure of collaboration, facilitating smooth turn-taking behaviours, and therefore working on more levels than had previously been considered. The evidence presented also demonstrated a significant influence of gesture on early trials, when interaction was considered to be ungrounded. An effect which diminished over time, suggesting that the real benefits of remote gesture technologies are time and context dependent, and are influenced by the extent of prior grounding of terms and interaction between collaborating parties. The nature of this effect can be used to derive key indicators to help guide future deployment of remote gesture technologies. The results were also discussed as an extension to the argument for the design of remote gesture technologies from a mixed ecologies perspective, arguing that the use of a technology designed from such a perspective had enabled the adoption of naturalistic turn-taking behaviours, which had been demonstrably responsible for improvements in collaborative performance.

## Chapter 8 – Conclusions

### 8.1 Introduction

This thesis has explored the use of remote gesturing technologies in the support of collaborative physical tasks. It has explored how and why remote gesturing works to improve remote communications and has investigated the process of fracturing of interaction and the causes behind this, articulating and testing a hypothesis of a mixed ecologies approach to designing communications infrastructure for object-focussed tasks. The results of the thesis have significant implications for the design, deployment and development of future remote gesturing tools.

The rest of this chapter proceeds by firstly re-visiting the original research motivations and by restating the main research questions and hypothesis of the thesis. After this the nature of mixed ecologies is reflected upon, providing answers, based on the evidence from previous chapters, for these main research questions. After this, the chapter discusses the implications of these research findings for the design and deployment of remote gesturing technologies. The chapter concludes by reflecting on the future development of such tools, discussing a program of future work which would extend the work of the thesis by addressing general issues raised by the research and specific issues which are raised by the system design guidelines presented.

### 8.2 Re-stating the problem

In Chapter 1 (p. 2-3) a scenario requiring remote collaboration was illustrated. In this scenario a paramedic arrives at the scene of an accident and realises that to save the life of a casualty they must perform a procedure which they are not comfortable performing without additional guidance from surgically trained staff. It was illustrated in the scenario that in this situation, in which two people might collaborate over some physical task, in which one might be imparting expert knowledge and directing action whilst another manipulates artefacts in the task-space, there is a need to develop strong communication links between the distributed working spaces. It was pointed out that current best practice would probably see the establishment of a video-link between the person at the scene and the remote expert. This form of video link between spaces can however, be significantly flawed, and can fail to achieve the levels of natural collaboration that might occur in side-by-side interactions.

This thesis then has been concerned with the study of video-mediated communication and in particular adds to the body of work seeking to explore how video-mediated communication systems can be improved upon to support distributed interactions in specifically *collaborative physical tasks* (i.e. tasks of remote collaboration that are inherently object-focussed in nature). This thesis has explored ideas of how to develop technologies that *do* support such forms of

interaction. It has specifically studied the design and potential implementation of extensions to video-mediated communication systems which allow for the remote representation of non-verbal behaviours and artefact-focused actions *in addition* to providing visual access between spaces.

Previous research (discussed in detail in chapter 2, pp. 12-67) has however demonstrated that there are a variety of competing ways in which such remote gesturing devices can be constructed and implemented. These different approaches have been critiqued, as in many ways by trying to make interaction at a distance more like face-to-face interaction they have inadvertently fractured the process of interaction that occurs between collaborators, in many respects therefore failing to achieve their potential as communication tools. On the basis of the scenario mentioned above and this reported literature on perceived limitations with remote gesture tools the fundamental question driving the research of this thesis became phrased as - how can technologies be built to improve remote collaborations for physical tasks, that don't fracture ecologies between remote spaces, but make the interactions as close to the perceived-to-be-optimum standard of face-to-face communication as possible?

In addressing these issues a research hypothesis was proposed and evaluated. Previous research had argued that the presence of dichotomous ecologies in such working collaborations is inevitable, and the role of communication tools is to mediate between the two locations. Based on Moles' notions of communication a mixed ecology approach to communication device design was proposed, which assumes that rather than linking and mediating between spaces the technology should seek to construct a unified environment in which both parties are *effectively* co-present. It was hypothesised that the use of such a communication device would optimise performance in object-focussed interactions as the mixed ecology supports communication by using technology to give collaborating partners access to the most salient and relevant features of communicative action that are utilised in face-to-face interaction, reportedly mutual and reciprocal awareness of commonly understood, yet richly complex object-focussed actions and mutual and reciprocal awareness of task-space perspectives. It was proposed that a mixed ecology therefore has more ability to successfully relay contextually embedded physical representations which have been shown to be of importance to collaboration in shared ecologies (i.e. co-present interactions).

To begin to address the fundamental research question and to facilitate an exploration of the benefits of the notion of designing from a mixed ecologies perspective it was felt that there was a need to investigate and understand how gesturing is actually used in remote collaboration. Further to this it was felt pertinent to explore specifically how the process of the fracturing of interaction is actually caused or mediated by the introduction of remote gesturing technologies. To firmly establish how to build these potentially useful remote gesturing systems, and to understand how their design impacts on their use and potential deployment, three main research questions were investigated:

- How and why does a representation of gesture improve remote communications (and consequently performance) in collaborative physical tasks?

- What creates a fractured ecology, how does interaction breakdown and how can a remote gesture simulacrum overcome this problem?

- What does an understanding of answers to the above questions mean for the design, deployment and development of such technologies?

In the following sections answers to these questions are reflected upon and the progress made in developing remote gesture tools to adequately support collaborative physical tasks is evaluated.

## 8.3 Reflecting on Mixed Ecologies

In this section of reflection and evaluation the first two key research questions are considered, and the progress made in the thesis towards answering them is evaluated in the context of understanding what this might mean for the mixed ecologies hypothesis proposed. The two questions were "How and why does a representation of gesture improve remote communication?" and "What creates a fractured ecology?", each question is considered in turn.

### 8.3.1 The how and why of remote gesturing

So turning first to a consideration of how and why representations of remote gesture improve communication in collaborative physical tasks, there was a wealth of evidence presented in the thesis over the four core chapters of empirical work. Starting in Chapter 4 (pp. 85-111) two experiments were presented. The first experiment (pp. 87-102) presented some basic effects of remote gesture tool use, focusing on both physical performance effects and cognitive measures. The study demonstrated how a representation of remote gesture during distributed collaboration can increase performance speed (in an exemplar form of collaborative physical task). Equally there were observed benefits to *cognitive* aspects of performance, with self-reported measures of cognitive effort being reduced amongst remote experts when their remote gesturing was enabled. The second experiment of chapter 4 (pp. 102-109) progressed even further the notions of performance enhancement during remote gesture tool use, by demonstrating distinct learning benefits during remote instruction (in an assembly task) via remote gesturing means. Together these studies demonstrated that remote gesturing improves performance speed, reduces cognitive load and can improve retention of instructed actions.

In Chapter 5 empirical work began to provide evidence that performance effects can be maximised by using technologies designed from a mixed ecologies perspective. In the thesis' third experiment (pp. 113-121) the orientation of presented remote gestures within a task-space was experimentally varied, illustrating some preference for aligned orientations, which kept

task-space views and relative gesturing angles consistent between remote collaborators. There was discussion on the basis of the experimental evidence of how certain orientations (such as 45° angle collaborations) can impede the production of more complicated forms of gesturing activity, effectively hampering the production of preferred gestural actions. In the fourth experiment (also presented in Chapter 5, pp. 121-135) system configuration variables of gesture representation format and gesture location (embedded within or external to the specific site of task-artefact action) were also tested. The results of this study demonstrated that a representation of gesture which was based on an unmediated view of video footage of the hands was of significant benefit to performance and outstripped the performance that could be achieved with sketch-based methods of gesturing. The second element of that study, which focussed on gesture location, was less conclusive. Evidence did not find a consistent performance difference between embedding gestures within a work-space or presenting them written over an external video feed. There was however, some user preference for the embedded method as this reduced the need to split the Worker's focus of attention between a site for action and a site for instruction. The failing to find a difference between the two locations was in part explained by epiphenomenal aspects of video view, which confounded the study, but demonstrated the importance during collaboration of developing mutual awareness of relative task-space and gesture-space perspectives.

Having seen some of the ways in which remote gesture improves performance in a communication based task, and having seen some evidence that this could be enhanced by designing the remote gesturing device along mixed ecology principles, Chapters 6 and 7 began to address *why* these representations of gesture and this mixed ecologies approach improved collaboration, showing how remote gesturing works as a process. In Chapter 6 a taxonomy of gestural use was derived (pp. 170-171) and the life-cycle of common gesture-based interactions during collaboration were articulated (pp. 144-169). This was done in conjunction with a specific comparison of sketch-based and hand-based gesturing. The results of the study demonstrated that hands are particularly versatile, allowing rapid and fluid changes between gestural actions, and facilitating gestural activity at a variety of conceptual levels (from pointing to embodied gesturing) and were principally satisficing tools. Alternatively, sketch-based gesturing suffered from problems of over-literal interpretation, excess abstraction, cluttering of limited visual resources and in some instances involved the creation of separate entities (drawn objects) within the work-space which fractured interaction by diverting attention and confusing perspectives. An argument was also posited that that the use of sketching might increase cognitive load amongst the workers, shifting relative responsibility for task progress during the task from the remote experts and their description activities, onto the workers and their efforts to interpret instruction.

In Chapter 7 (pp. 175-203) the empirical work of the thesis took the research on remote gesturing down to the level of considering in detail how remote gesture interacts with language. And for the first time provides some elements of evidence which show how

gesturing is not just used as a tool to support language directly, but also shows how the provision of remote gesturing actually re-structures interaction. This analysis was achieved by transcribing and comparing speech excerpts from uses of the remote gesture system in the experiments of Chapter 4. The results of the analysis (using a technique piloted by Kramer et al 2006) demonstrated that the use of a remote gesture tool significantly increases the remote Expert's self-expressed presence in the Worker's task space (the remote or in-the-field site). As the Helpers increased their gesturing behaviour this was seen to correlate with a decrease in the number of interruptions and overlaps between the collaborating pair, and this increase in gesturing was also seen to be correlated with a decrease in performance times. Importantly increases in the number of overlaps in collaborative speech were also positively correlated with poorer performance times. This demonstrated that the use of a remote gesturing tool was apparently leading to more structured interactions with smoother turn-taking and this was correlated with improvements in collaborative performance, both physical and cognitive. Some of the importance of these findings stems from a consideration of the prior video-mediated communication literature. In many studies there has been an express belief that more fluid and therefore *better* interaction was signalled by increased interruptions (Sellen, 1992, 1995, O'Connaill and Whittaker, 1997). However, the results of Chapter 7, as stated, clearly demonstrated that increased overlap in speech patterns was associated (in collaborative physical tasks, at least) with poorer task performance. And having a more structured and smoother turn-taking practice was associated with improved performance, with remote gesturing observably being significantly associated with increases in the adoption of such turn-taking patterns.

Not just having implications for the role of 'fluid' speech and interruptions as a marker of productive discourse, these findings challenge some pre-conceptions about how gesture works. A large amount of the previous theorising (as discussed in chapter 2) demonstrated a pre-occupation (at least in the development of DOVE) with a cognitive processing view of the benefits of remote gesturing. From this perspective a distributed cognition reading of remote gestural interaction sees the role of gestural action as one of keeping visuo-spatial information in a visual medium such that it can be transmitted and decoded with the minimum amount of translation effort possible. In this scenario, in the absence of the capacity to gesture all visuo-spatially relevant instructions must first be translated into a verbal medium, then transmitted and then decoded before they can be made relevant again to the visual task-artefacts embedded in the working space. The work of Chapter 7's linguistic analysis has then demonstrated some benefits from understanding that the role of gesture in remote collaboration is not just a process of transmitting object-focussed information between spaces but also to transmit a sense of physical presence between spaces. And the use of this physical presence then becomes a structuring tool in establishing the smooth flow of communication between spaces. This finding is perhaps then also corroborated by observations made in experiment 2 (p. 109), that the use of gesture during instruction affected the personal perception of the collaborators.

Those learners who were instructed via voice *and* gestural means were actually less accepting of their instructor, they felt less positively inclined toward them and felt more directed and less involved in the task, doing more what they were told than *discussing* what to do next. It is interesting that this might be as a consequence of increased presence in the space, and underlines how presence can dictate the actual flow of the interaction.

### 8.3.2 What creates fractured ecologies?

So understanding more now about how gesture is actually used and seeing that a large part of it is also to do with having a presence within the working space, it is pertinent to reflect on the causes of fractured ecologies, discussing what the thesis has told us about how interaction can become fractured. To begin this discussion it is pertinent to briefly re-iterate what the term fractured ecology actually means. It is generally considered that interaction becomes fractured when during communication one party's perception of their colleague's *intention* becomes divorced from their perception of their colleague's *action*. This can logically happen in one of two ways. One collaborator might be able to perceive the actions of the other but not be aware of the intentions of those actions, for example they might be able to see some remote gestures but not be able to understand what those gestures are meant to mean. Likewise, a collaborator might perceive of the intention of the other but not be aware of the action, for example, where remote gesturing is not present and the remote helper, referring to a mutually visible object, suggests that the worker should 'turn it <u>this</u> way' whilst simultaneously gesturing and forgetting that their colleague cannot see the relevant action. In this instance the worker is aware of the intention that they should manipulate a task-artefact but the actions have been stripped of their contextual relevance.

Building on these concepts of a divorce between intention and action, studies which have observed the nature of co-present interactions and the situationally-embedded context-dependent nature of communicative action, have suggested two principle features of co-present object-focussed interaction which are essential for successful communication. These two features are 'Mutual and reciprocal awareness of commonly understood yet richly complex object-focussed actions' and 'Mutual and reciprocal awareness of task-space perspectives.' The thesis proposition was that to alleviate the problems of fractured ecologies remote gesture tools should be designed from a mixed ecologies perspective. This mixed ecologies perspective saw the communication tools being designed in ways which it was felt were more in tune with these two critical components of shared ecologies. Through developing and evaluating prototype remote gesturing tools the thesis has found evidence for ways in which these two features of successful collaboration can be supported by certain aspects of remote gesturing tools. In turn by performing these analyses the nature of these two components of co-present interaction have been further broken down into constituent parts further developing understanding of both the nature of fractured ecologies and therefore the necessary nature of a

mixed ecology as a derivative solution. Each of these two key components of a shared ecology is considered in turn in the following sections.

*8.3.2.1 Commonly understood yet richly complex object-focussed actions*

The first factor of interest is the development of commonly understood but complex object-focussed gestures. This was investigated directly in chapters 5 and 6. In attempting to understand how this feature should be achieved three key technical components were observed. These were the format of representation of gesture, the location in which it was displayed and the relative orientation to each collaborator and the task artefacts in which it was presented. The findings of experiment 4 (in chapter 5, pp.125-135) demonstrated that an unmediated representation of the hands as the gesturing medium had the most efficacy in terms of facilitating task communication. The reasons behind views of hands being more useful than more abstract representations such as sketches were discussed at length in chapter 6 (pp. 171-174). The analysis demonstrated the ways in which hand gestures were commonly understood, fluidly and rapidly constructed and linked together into chains of communicative action, and also highlighted the ways in which they could seamlessly be used at a variety of conceptual levels, giving them much increased versatility. Additionally the role of sketched forms of gesturing was critiqued for the ways in which it had the capacity to cause fractures by disguising the intentions of the sketcher, because of a tendency toward over-literal interpretation of sketched gestures and diverting attention away from the site of action, as at times it would not have been clear whether a Helper's instructions were being said relative to the task artefacts of the sketched representations.

Further to this notion of the desire to understand the gestures in context (as this facilitates understanding) was the technical consideration of where to physically locate the remote gestural output. This aspect was directly evaluated again in experiment 4 (chapter 5, pp. 121-135). In this work although no firm differences in performance could be generated by varying gesture output location between embedded and externalised views, there was evidence that users preferred and encountered less problems with projected gestures (i.e. those embedded in the workspace). By projecting gestures into a task space there is less chance that a gesture will be missed, as the site of gestural action is the site of actual artefact manipulation. It is also more likely that a remote gesture can effectively be aligned with not just artefacts within the space but also *actions* within the space, something that an externalised view might find harder to initiate and which was observably a common element of more complex forms of gesturing that were enabled by the use of the hands as a gesturing medium.

Extending this concept of a desire to facilitate complex gestures the third issue of interest moves on from considering grossly where the gestures are presented to specifically where within the task space they are presented (i.e. the relative virtual angle of orientation between collaborators and artefacts within the space). This was investigated in particular in experiment 3 (Chapter 5, pp. 113-121). The study found user *preferences* for adopting common

orientations, which essentially demonstrated the ways in which interaction can be harmed by forcing collaboration to occur when there is a discrepancy between the perspectives on and actions towards the task-artefacts. A fundamental property of establishing a common orientation to task artefacts was the way in which it facilitated the remote gesturer in producing a particularly high level of complex 'embodied' gesturing (discussed in Chapter 6, pp. 149-152 and p. 173). With the use of the 'Mimicking' and 'Inhabited' hands forms of gesturing, a remote gesturer could literally use their hands to recreate a visual exemplar of a desired manipulation of an assembly piece. In doing this there was a benefit to grounding instructions that were given. Obviously the ability of the remote gesturer to use their hands to directly mimic the hands of the worker to show a desired movement is greatly enhanced if both the worker and remote gesturer share the same orientation to the task artefacts. And therefore in most instances users expressed a desire to *not* have gesture representations constructed at a 45° angle, as this restricted their capacity for constructing just these forms of naturally occurring high level gestural actions.

### 8.3.2.2 Mutual and reciprocal awareness of task-space perspectives.

The final component of a remote gesture tool that should be considered is the development of mutual and reciprocal task-space perspectives. This is essentially the most important of the components of a co-present communication ecology and is a fundamental part in regulating the fracturing of interaction in remote communications. As a feature of a remote gesture tool it underpins all of the other components and is therefore a specific element of both of the most important features of shared ecologies i.e. commonly understood object-focussed actions and common task-space perspectives, and as such its considered arrangement is the most critical component in ensuring that interaction does not become fractured during remote collaboration.

To begin to understand this component of remote gesture systems it is pertinent to briefly unpack what this term means. Firstly, mutual awareness, this element implies that both collaborating parties should have visual access to the task space (this is the most fundamental aspect of the interaction and is the primary concern of all video-mediated communication devices). The second element, reciprocal awareness, implies that both collaborators are aware that one another can see the task space, but what is more important, 'mutual and reciprocal', implies that both parties are simultaneously aware of the extent of their collaborator's view of the task space and that each party knows this of the other. The final element of the term 'relative task-space perspectives', refers however, to more than simple views of the task-space. It is also a reference to a more conceptual notion of the disposition towards the task and current levels of attention and other elements of presence within the task-space that would be otherwise conferred by mutual co-presence.

As stated previously, mutual and reciprocal awareness underpins all of the proposed components of a mixed ecology. To establish mutual and reciprocal awareness a collaborating pair requires commonly understood gesturing mediums. One cannot be made aware of

another's intentions, or have an adequate belief that another is correctly interpreting your actions, unless those actions are coherent with commonly used forms of gestural action, are presented at orientations which might enhance intelligibility and which are embedded within the working environment such that they can be contextually relevant and validly decoded. In such circumstances assumptions can be made of the ability of others to perform accurate interpretation of meaning, as in co-present interactions these functions are available and are the means by which gestural action is grounded.

Many studies have previously demonstrated the necessity of ensuring that all participants in a shared task are mutually aware of what each other can see of the task space (and consequently interpret the actions performed within the space as contextually relevant). The studies presented herein are no exception. Indeed experiment 4 (Chapter 5, pp. 121-135) specifically demonstrated evidence to show how collaborators were keen to exploit the ability to coordinate their actions in light of being able to see explicitly how those actions would be perceived on the other side of the interaction. During the experiment it became evident that in the video-window system it was much easier for the Worker to be made explicitly aware of exactly what the Helper's view of the task space was like. In the projection condition the Worker had to assume much more of what the Helper's perspective on the task space was, whilst they knew the limits of what the Helper could see, they could not see for themselves how this looked to the Helper. This understanding of relative perspectives was observed to be of significant use to the Workers on a number of occasions when they wanted to show something in detail to the Helper. Knowing explicitly what their actions looked like to the Helper helped the Workers to form their actions accordingly at their own site.

This notion of understanding how one's actions will be interpreted is important when the aspects of adequate communication environments, as espoused by Moles (1975, 1966, see chapter 1, p. 1), are taken into consideration. Moles argued that for effective communication one must make more complex the space time surrounding the point of reception, by creating a micro-replica of the complexity created at the origin of transmission, but to do this, Moles suggested that one must use the items of knowledge that they have in common. It is to this then that the notion of mutual and reciprocal awareness lends itself. In establishing awareness of both one's actions and the potential perceptual interpretation of these actions a gesturer can ensure that they are effectively taking regard of the items of knowledge that they share in common. This facilitates the development of phenomenal coherence between the site of remote gesturing and the local task-space. As this is a fundamental aspect of the mixed ecologies approach to the design of remote gesture tools and has been encompassed in the basic system design that was tested throughout the thesis this factor has been repeatedly demonstrated to be beneficial when performing collaborative physical tasks.

The work of this thesis has then verified the mixed ecologies approach to remote gesture tool design, demonstrating its efficacy and expanding understanding of how interaction can become

fractured. Whereas previous conceptual theorising on the role of communications media might have posited the role of the technology as one of mediating between disparate spaces, this work has suggested that in collaborative physical tasks this is not the most beneficial approach. Studies of shared visual spaces in apparent interaction tasks (see Kraut et al 2002, Gergle et al 2004, Gergle et al 2006) have emphasised the apparent importance of mutual awareness. This research has demonstrated however that for object-focussed tasks, which possess a requirement for interaction with three-dimensional artefacts, a mixed ecology approach which seeks to render mutual *and reciprocal* awareness is of vital importance to fully embed and construe the necessary remote gesturing behaviours.

### 8.4 Implications for the Design and Deployment of Remote Gesture Tools

The third research theme of the thesis was to explore what an understanding of remote gesturing and mixed / fractured ecologies meant for the design, deployment and development of remote gesture technologies. Leaving the development of the technologies to one side momentarily (to be reprised in the 'Program for Future Work' of the following section) this section details the implications of the above findings and discussions for the actual *use* of remote gesture tools. So what have the results of the thesis informed us about how remote gesture tools should be constructed and engaged?

### 8.4.1 Design

With currently available technologies there are a variety of possibilities for constructing remote gesture tools. The designs that have been thus far explored or are logically possible, tend to differ through three principle features, namely, how the gestures are displayed, how the gestures are generated (often inherently tied to the method for display) and the extent to which the system is mobile. The potential current design choices are mapped out below in figure 8.1.
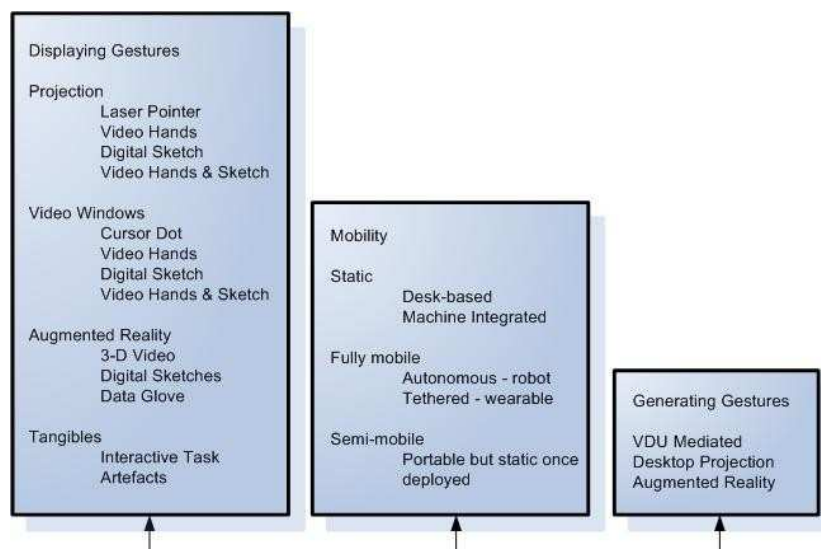
Figure 8.1 Possible system design alternatives for remote gesture tools

The individual pros and cons of each of these system design choices and examples of use in any current systems are detailed in full in appendix 8.1.

The work of chapters 4 and 5 has however, provided some specific recommendations for features of remote gesture tools, which enable a considered choice to be made when considering some of the options presented above. These have variously been discussed as key features of a mixed ecology and elements to which design should adhere to prevent the fracturing of interaction. Taken from section 5.4.4 (pp.138-9) and discussed further in section 8.3.2, the study of mixed ecologies suggests that remote gesture tools for collaborative physical tasks should provide:

- *Hand-based gestures*

- *Shared orientations to task artefacts*

- *Projection of gestures*

- *Mutual awareness of relative task-space perspectives*

If these features are built into a remote gesturing tool then it is likely that they will limit the potential for interaction becoming fractured. In the following, the requirements for establishing each of these systems properties are considered.

To establish hand-based gestures there are a variety of technical possibilities but the most likely of these is to use direct video capture of a remote collaborator's hands. Increasing bandwidth capacity of communications means that direct video feeds between sites are

becoming increasingly feasible and the hard-wired closed-circuit links of the low-tech prototype utilised in this research could be easily replicated with mobile wireless technologies. The use of video facilitates the production of naturalistic forms of gesturing behaviour, or at least an approximation of these that is readily adapted to. This is not to say that other forms of gesture representation are unusable, but merely to suggest that hand-based representation is optimal for the forms of task tested in this work. Other more abstract gesture representations might find application but the key features of hand-based gesture such as the fluidity, the bilateralness and the multipoint gesturing should all be incorporated. Likewise to ensure the fluidity it is perhaps pertinent to generate gestures by literally recording natural hand-based gestures, so if this is already being captured then it seems unwise to mediate the representation (such as has been attempted in the Mixed Presence Groupware of Tang, Boyle and Greenberg, 2004, presumably for the sake of bandwidth), in essence this is reducing the cost of transmitting and processing data at the expense of communicative information bandwidth.

The development of shared orientations to task artefacts principally requires attention to be paid to the location of video feeds at the site of the collaborator located with the task artefacts. In providing a remote gesture tool it is a given that their must be a video link between the spaces. A legacy of research within the video-mediated communication community has demonstrated the importance of providing shared visual access to the task space (all reviewed extensively in chapter 2). And this body of work has demonstrated that the talking heads model of video link is inappropriate for object-focussed tasks, the video data is much more relevant and useful if it is focussed on the actual objects of interest. But in doing this the video data could be captured from a variety of angles. The data of this thesis however demonstrates that a shared orientation and common task perspective helps to prevent fractures in interaction. To maintain a shared orientation it is beneficial therefore to ensure that the video device capturing context views of a workspace is as closely aligned to the view held by the person local to the task space as possible. A primary objective however, is to achieve this without obscuring the view of that person. This can be achieved with either static cameras held above a work surface or utilising 'over-the-shoulder' style views, or can be effected by having cameras which are attached to the body of the person in the local task-space. Choosing between the two approaches is largely dependent on the context of the technology deployment and will be determined by whether it must be mobile, semi-mobile or can be static. A key feature that should be borne in mind however is that video-devices tethered to a local worker must be constructed such that they maintain a static view of the space and are not constantly re-oriented by movements of the local collaborator.

The use of projected gestures largely requires that the person in the local space is equipped with projection technology. Again the format of this might be determined by whether the whole device is intended for mobile operations. With the use of projection technologies there are a variety of ergonomic considerations. The technologies themselves are currently still relatively power intensive which either requires access to permanent power supplies or the

provision of large battery packs. Equally the technologies themselves tend to overheat, which combined with the power considerations means that their use in a mobile context (or even for that matter an outdoor context) might be somewhat problematic. In static contexts obviously these considerations are less relevant. An additional point of concern is the ability to adequately project images (especially video data) in external environments with alternating and often less than optimal lighting conditions. In many instances this can lead to projected images being too faint to be effective. To deal with some of these considerations there are principally two possible solutions. The first is to utilise projection technologies which are designed with a specifically miniaturised, and intended for mobile use, form factor. Such devices whilst not currently commercially available, are under development and the next few years will see a large expansion in the development of the mini-projector market, making mobile remote gesture tools more feasible from this perspective. An alternative strategy and the second possibility is to utilise an entirely different form of projection technology, augmented reality. In this case the actual projection is based on the delivery of digital information to a personal viewing device such as a see-through heads-up display, which posits digital information artificially into the three-dimensional world. This approach which is already designed for mobile applications alleviates a variety of the ergonomic difficulties that projection technologies confer. However, successfully transmitting video images of hands is not compatible with this approach. Research would therefore be needed to resolve how best to construct remote gestures (i.e. their representations) in an augmented reality context.

The final system property, and the one that is perhaps the hardest to successfully achieve is the development of mutual and reciprocal awareness of task space perspectives. With video windows as the location of remote gesturing this is relatively simply achieved, but projection technologies largely suffer in this respect as they are inadvertently much more asymmetric in their connections between spaces. It is not currently clear how one might make the limits of another's view expressed through a projection technology. Augmented reality however, again might offer some solutions, as it might enable the construction of virtual avatars representing remote collaborators which could be anchored to static video cameras, and from which virtual representations of gesturing arms could emanate. In this way then, it might take the innate adult human capacity to infer relative view when co-present by observing apparent relative bodily orientation and re-situate this skill in a distributed working context.

### 8.4.2 Deployment (situating the technology)

When considering the design options discussed above and talking through how the system design recommendations might be implemented it becomes apparent that much of the final decisions about how to construct the technology are actually inherently tied to considerations of when and where the technology might be used. Clearly whilst some might consider that all technology should be developed in a situated context, so as to be able to determine from the

beginning exactly how it will be integrated into the socio-technical systems of the organisation in which it will be used, this is not always possible. With remote gesture technologies in particular, a new class of communicative device is being constructed and being constructed to support working practices that might emerge but don't currently exist. Of those contexts which have been proposed as viable scenarios for application (such as the scenario detailed in the introduction) there are often ethical or practical barriers to the use and evaluation of experimental technologies. Consequently the technology has been developed 'in-the-lab' and there has been little opportunity to consider how it would be most effectively deployed in the real world.

The results of this thesis and in particular the results of chapter 7 can, however, be used to derive some guidance for both remote gesture tool designers wishing to better understand the possible applications of their technologies and for those who are interested in understanding whether a remote gesturing tool is suitable to support a specific collaborative task they have in mind. The results presented in chapter 7 would strongly suggest that remote gesture tools are effectively a technology to support the achievement of grounded interaction in collaborative physical tasks. Understanding the technology from this perspective allows focus to be re-drawn on the prospective applications of the technology helping to structure some guidelines for the effective deployment of these technologies.

For that class of design-oriented gesture-based technology (e.g. Clearboard, Ishii and Kobayashi, 1992, VideoWhiteboard, Tang and Minneman, 1991) gesture has an expressive role in collaboration. To this extent, and given the fact that as discussed previously the gesture is itself the object of communication, the benefits from the use of remote gesturing should persist over a significant timescale, balancing the cost of introducing the technology. However, for those technologies designed to support a class of 'collaborative physical tasks' gesture is an artefact of communication. It is there to support communication as and when necessary, and the research of this chapter has demonstrated that it is most necessary, or at least has the most significant benefits, at early stages of ungrounded interaction. This would suggest that the advantage of introducing this technology for familiar and often repeated assembly and physical manipulation tasks needs careful consideration. These tasks would include routine repair and maintenance where the field engineer has an established rapport with the helper. In these situations the heavily grounded use of terms and common understanding of task practices means that orientation to the problem space cannot be expedited. The collaborators already have a highly developed sense of mutual awareness and understanding.

The issue is then, how those applications where the costs of grounding either persist for long periods of time or at least a significant proportion of the lifetime of cooperative engagement might be identified. It was felt that three key factors exist which can be examined in any given possible context for deployment of a remote gesturing technology to be used as determinants of the applicability of such a device. These three concerns are presented below:

- *The level of experience of the participants involved*. Have they performed this kind of task before (do they have a good task knowledge)? Is there a significant and possibly affecting disparity in knowledge between the collaborators (is this an expert – novice interaction and will it matter)?

- *The novelty of the task*. Is the task new to the parties involved? Do they have experience of working together on this form of task? Is it a familiar task but presented in a previously un-encountered environment/situation (and will this affect task performance)?

- *The urgency of the task*. Is the task time-critical requiring significant action to take place under time pressure? Would the additional time required to achieve grounding through other means have critical implications?

A consideration of these three factors suggests a number of cooperative arrangements as ideal situations for the application of this form of technology.

- *Non-routine physical manipulations* where the nature of the task and the settings vary considerably and each cooperative interaction requires significant effort to ground the interaction. This sort of activity would include remote diagnosis of problems (e.g. medical, mechanical) where the context of the remote setting is unknown and needs to be understood and interpreted in order to guide the work.

- *Regular changes in the participants* where the remote Worker or the Helper have not had the opportunity to build a world known in common or have to reestablish this frequently. This might occur even for routine repair and assembly task where the remote worker is new to the task at hand. Consider for example replacing a trained field engineer with a consumer who is guided through the repair by an expert.

- *Rapid cooperative diagnosis settings* where rapid coordination is required in order to decide the best possible action. This would include settings such as remote medical diagnosis and intervention. In these settings the ability to rapidly orient to a task is extremely critical.

These criteria therefore provide guidelines for the possible applications of remote gesture technologies. It is not necessarily an exhaustive list but it will hopefully sensitize technology

designers to the factors which influence how their remote gesture technologies will be applicable to different future collaboration scenarios.

If consideration is given to the above guidelines for determining potential applications of remote gesturing tools it becomes evident that there is a potentially significant argument for these technologies becoming mobile devices, or at the very least partially mobile, in that they could be transported easily and set-up quickly. Equally however, many of the potential applications might also suggest ad hoc connections and casual use by non-expert users which might imply that such technologies would benefit most from being developed in light-weight form factors that are integrated into existing technical devices. For example, given the applications that might be suggested by the above there is significant potential for exploring how remote gesturing devices might be incorporated within existing mobile communications infrastructure. Given the increasing miniaturisation of projection technologies, and the acknowledged fact that mobile handset manufacturers are considering ways in which to incorporate projection technologies in their devices (see figure 8.2 below) then there is significant potential for the development of more lightweight remote gesture technologies.



Figure 8.2 Compal Projector Phone exhibited at 3GSM 2006 Barcelona

A development of lightweight remote gesture technologies would have further implications for the development of remote gesture tools. The guidelines for deployment given above and the guidelines for design taken together provide important directions for the minimum technical specifications for these technologies to be useful.

## 8.5 A Program of Future Work

Considering the general research findings as presented above and the implications that have been drawn from these it is appropriate to begin to consider what the implications of all of these are for the development of future generations of remote gesture tools. This thesis has largely been an exercise in developing an understanding of the principles behind remote gesture technologies and as such has only required the use of low-tech prototypes. For the technologies to actually be implemented however there needs to be a significant phase of technical development. Likewise, as the technologies are deployed the findings of this thesis have implications for the ways in which the deployed technologies might best be evaluated and

in turn there are further implications derived from the thesis as issues of general interest which are raised as potential future research directions. As such, there are two broad themes of future work, one a technical program and the other more concerned with the experiential aspects of using and understanding remote gesture tools and working within mixed ecologies. Each of these is addressed separately although clearly they would logically overlap in many respects.

### 8.5.1 A technical program

A logical progression for the work would be to find a suitable context of application (perhaps utilising the guidelines presented above). From this the development of an actual full implementation of a gesture technology could be initiated, which would facilitate a process of *in situ* development. Clearly the work thus far has been limited to some extent by the fact that it has been developed in-the-lab and there are potentially numerous environmental factors which might influence system usability. These issues would be addressed by actually deploying and examining the technology as it is used in support of a task with increased ecological validity.

Considering the spheres of potential deployment and the discussions presented above, two common themes can be derived for the development of remote gesture technologies. These are the development of augmented reality remote gesturing tools and the development of mobile phone based remote gesture tools. Each of these is considered in turn.

The development of augmented reality systems would address many of the current limitations with projection technologies. Projection of gesture is one of the key criteria of a remote gesture system presented above (in section 8.4.1). However environmental constraints such fluctuating light sources, often bright daylight conditions and uneven projection surfaces can render current projection technologies virtually useless when taken outside. As such the use of augmented reality allows remote gestures to be posited into a remote environment without relying on optimal visual conditions. However, if such a system were to be developed techniques would need to be explored to adequately support the other suggested key components of remote gesturing systems. Clearly video representations of the hands would be difficult as a video capture is a two-dimensional view and augmented reality presumes three dimensions. The augmented reality system could present a two-dimensional view flattened over the working surface in the three dimensional context or some three-dimensional representation of the remote gesturer's arms would need to be presented. This has implications for both the production of the gestures and the capture of the gestures. Likewise developing shared orientations and relative task-space perspectives might suggest the use of video cameras in an augmented reality system which are physically tethered to the person in the local space, but in doing this techniques must be explored for keeping the remote gesturer's view consistent (prior research showing that head-mounted cameras which cannot be controlled by the remote viewer significantly fracture interaction). Taken together the production of an

augmented reality gesturing system poses several key technical challenges which would need to be overcome.

The alternative major trajectory for remote gesture tool development is the investigation of more lightweight mobile phone based gesture tools. Technologies which combine video capture enabled mobile phone technology and miniaturised projection technology (as discussed above) have great potential for supporting remote gesturing in a variety of environments. For example, in high ambient temperatures or when ease of movement is a high priority, bulky augmented reality equipment (which normally must be carried), would be inappropriate. Equally if the technology was only for use in an *ad hoc* fashion and was not required as a dedicated device then such a lightweight form factor might be a significant benefit. Obviously research is needed to explore the necessary bandwidth requirements for supporting two-way simultaneous video capture and transmission and projection at a significant fidelity as to be useful. Also problems of video feedback loops of projecting and capturing at the same point source would need to be resolved dictating a need for the development of phone based camera / projector management software. Obviously a strength of such approaches is the ability to utilise common forms of hand based gesturing by using direct projections of unmediated video of hands, however this would then lead to the problems of projection mentioned above. Projection technologies are naturally limited by fluctuations in local lighting conditions and by projecting onto uneven surfaces which can distort the images. This would suggest then that if a mobile phone based technology was explored a program of research would be needed to explore how to optimise the strength of projections to cope with fluctuating light conditions and software based techniques would need to be developed to explore how to format projected images to cope with uneven surfaces. An additional problem that might need to be addressed which is possibly resolved in augmented reality systems is the development of mutual and reciprocal awareness. With mobile phone based technologies it may be difficult to alert collaborators to the extent of relative view, although exploring linking the view enabled to the extent of the projected area (and therefore illuminated area) might be a possible way of achieving this. This too would need to be evaluated through further research.

### 8.5.2 An experiential program

A desire to develop and explore remote gesture technologies *in situ* also has implications for how one might evaluate the technology. The work of this thesis has demonstrated, in most cases, base performance effects, but has also demonstrated that remote gesturing can influence performance in a variety of ways, several of which are more cognitively derived. In many instances where benefits were not shown in performance time, it was discussed that there was still a sense in which some technologies tested were preferable to others and seemed easier to use. Better, more objective measures of cognitive workload, might then be developed for assessing such performance effects. The subjective measures used in this thesis were at times

felt to be less rigorous than would have been hoped owing to the often large between-subjects differences in self reports. As such more objective measures would be valuable. Equally, as the technologies actually become deployed and tested in-the-field other considerations of usability must also become paramount. Knowing that there is a base utility to the technology other factors such as comfort during use become important requirements that must be addressed. The mobile solutions discussed above such as augmented reality systems would require significant evaluation from a physical ergonomics perspective before realistic adoption of the technology can be expected.

A large portion of the research of the thesis has been given over to developing an understanding of exactly how remote gesture tools work and the role of developing mixed ecologies of communication, and future work should extend these findings and explore the implications of this further. In the process of this research it was demonstrated that a key function of the remote gesturing technologies was their ability to instil a sense of presence of a remote collaborator into a local task space. Understanding new measures for reporting and experimenting with sensations of remote presence should also therefore be explored. The existing measure suggested by studies such as Kramer et al (2006) are heavily reliant on circumstantial evidence based on linguistic terms used. In these instances only measures of the remote expert's sense of presence within the space remote to them can be measured. The relative experience of receiving a sense of remote presence cannot be measured by these techniques. But study evidence from chapter 4 (experiment 2) in particular, demonstrated that there is a significant impact of remote presence on inter-subjective perceptions. Continuing research with new methods for evaluating sense of presence would enable researchers to explore why increased remote presence may have lead to the development of less favourable perceptions of colleagues. This is potentially an important issue if such technologies are to be used in situations in which collaborators are unlikely to know each other very well, as sources of tension in technology use can become significant problems in technology adoption. Generally, understanding the socio-emotional impact of working with remote colleagues who, whilst distributed geographically, are actually given significant remote presence within your personal working space, is an avenue of research that this thesis has raised as an issue of some potential interest and is fundamental to the adoption of mixed ecologies forms of video-mediated communications technologies.

This thesis has shown that human communication is remarkably versatile and will always persevere, even in distributed interactions. The work then has sought not to create communication where it was not possible but to optimise it where it was needed. Continuing this tradition the development of remote gesturing tools must now pass from a phase of trying to understand how remote gestures work and how mixed ecologies can limit the fracturing of interaction to a more global evaluation of how to optimise these technologies for becoming situated within and alongside current and developing working practices.

# References

Adler, A. and Henderson, A. (1994) A room of our own: Experiences from a direct office-share. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'94)*. Reading, MA: Addison-Wesley. pp. 138-144

Anderson, A. H., Mullin, J., Newlands, A., Doherty-Sneddon, G. and Flemming, A. M. (1994) Video-Mediated Communication in CSCW: Effects on Communication and Collaboration. In *Proceedings of Workshop on VMC* at CSCW'1994. Chapel Hill, NC.

Anderson, A.H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A. M. and Van der Velden, J. (1997) The impact of VMC on Collaborative Problem Solving: An Analysis of Task Performance, Communicative Process, and User Satisfaction. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp.133-155

Angiolillo, J. S., Blanchard, H. E., Israelski, E. W. and Mané, A. (1997) Technology Constraints of Video-Mediated Communication. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 51-73

Antaki, C. (2005) *An Introduction to Conversation Analysis*. http://www-staff.lboro.ac.uk/~ssca1/sitemenu.htm

Argyle, Michael, (1988). *Bodily Communication*. 2nd Ed. London: Routledge.

Baecker, R. M. (1993) *Readings in Groupware and Computer Supported Cooperation Work: Software to Facilitate Human-Human Collaboration*. Morgan Kaufmann Publishers

Bannon, L. (1991). From human factors to human actors: The role of psychology and human-computer interaction studies in system design. In Greenbaum and Kyng (Eds.) *Design at work: Cooperative design of computer systems* Hillsdale, NJ: Lawrence Erlbaum. pp. 25-44

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

Bellotti, V. and Dourish, P. (1997) Rant and RAVE: Experimental and Experiential Accounts of a Media Space. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey, pp. 245-272

Bellotti, V. and Sellen, A. (1993) Design for privacy in ubiquitous computing environments. *Proceedings of the Third European Conference on Computer-Supported Cooperative Work (ECSCW'93)*. Milan, Italy, pp. 77-92

Benford, S., Greenhalgh, C., Reynard, G., Brown, C. and Koleva, B. (1998) Understanding and Constructing Shared Spaces with Mixed Reality Boundaries, *ACM Transaction on Computer-Human Interaction* (ToCHI), 5 (3), ACM Press, pp. 185-223

Billinghurst, M. and H. Kato (2002). "Collaborative Augmented Reality." *Communications of the ACM* 45(7): 64-70

Bimber, O. and Raskar, R. (2005) *Spatial Augmented Reality: A Modern Approach to Augmented Reality*. A K Peters

Bly, S. A. (1988) A Use of Drawing Surfaces in Different Collaborative Settings. In *Proceedings of the Conference on Computer-Supported Cooperative Work*. Portland: Oregon pp. 250-256

Bly, S., A., S. Harrison, R., and Irwin, S. (1993). Media Spaces: Bringing People Together in a Video, Audio and Computing Environment. *Communications of the ACM* **36** (1): 27-47

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

Brave, S., Ishii, H. Dahley, A. (1998) Tangible Interfaces for Remote Collaboration and Communication. In *Proceedings of CSCW 1998*, Seattle, Washington, USA. ACM. pp. 169-178.

Bull, P. (2002) *Communication Under the Microscope. The Theory and Practice of Microanalysis*. Sussex, UK: Routledge

Button, G., Coulter J., Lee, J. and Sharrock, W. (1995) *Computers, Minds and Conduct*. Polity Press: UK

Buxton, W. (1997) Living in Augmented Reality: Ubiquitous Media and Reactive Environments. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp.363-384

Byers, J. C., Bittner, A. C. and Hill, S. G. (1989). Traditional and raw Task Load Index correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety*, London: Taylor and Francis. pp. 481–485

Chapanis, A. (1975). Interactive human communication. *Scientific American* 232 (3): 36-42.

Chapanis, A., Ochsman, R. B., Parrish, R. N., and Weeks, G. D. (1972). Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors*, 14(6), 487-509.

Clark, H. H. (1996) *Using Language*. Cambridge, UK: Cambridge University Press

Clark, H. H., and Brennan, S.E (1991). Grounding in Communication. In L.B. Resnick, R.M. Levine, and S.D. Teasley (Eds.). *Perspectives on socially shared cognition*, (1991). pp. 127-149. Washington, DC: APA.

Clark, H. H., and Krych, M. A. (2004) Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50* (1) 62-81.

Clark, H. H., and Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition*, 22 (1) 1-39.

Clarke, D. (2004) "Structured judgment methods" – The best of both worlds? In Z. Todd, B. Nerlich, S. McKeown and D. D. Clarke (Eds.) *Mixing Methods in Psychology. The integration of qualitative and quantitative methods in theory and practice.* Psychology Press. pp. 81-100

Cohen, K. M. (1982) Speaker interaction: Video teleconferences versus face-to-face meetings. In L. A. Parker and C. H. Olgren (Eds.) *Teleconferencing and electronic communications: Applications, technologies and human factors.* Madison, WI, University of Wisconsin Extension, Center for Interactive Programs. pp. 189-199

Crabtree, A. (2003) Designing Collaborative Systems: A Practical Guide to Ethnography. London, UK: Springer-Verlag

Daft, R. and Lengel, R. (1984) Information Richness: A new approach to managerial behaviour and organizational design. In B. Straw and L. Cummings (Eds.) *Research in organizational behaviour*. Greenwich, CT: JAI Press. pp. 191-223

Daly-Jones, O., Monk, A. F., and Watts, L. A. (1998). Some advantages of video conferencing over high quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies*. **49**, 21-58

Doherty-Sneddon, G., Anderson, A.H., O'Malley, C., Langton, S., Garrod, S., and Bruce ,V. (1997). Face-to-face and video mediated communication: a comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, 3, 1-21

Dourish, P. (1993) Culture and control in a media space. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work (ECSCW'93)*. Milan, Italy. pp. 125-137

Dourish, P. and Bellotti, V. (1992) Awareness and coordination in shared workspaces. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW'92)*. Toronto, Canada. pp. 107-114

Dourish, P. and Bly, S. (1992) Portholes: Supporting Awareness in a Distributed Work Group. In *Proceedings of the Human Factors Conference on Computing Systems CHI 1992,* ACM Press (1992), pp. 541-547

Dourish, P., Adler, A., Bellotti, V. and Henderson, A. (1996) Your place or mine? Learning from long-term use of video communication. *Computer-Supported Cooperative Work.* **5** (1), 36-62

Duncan, S. (1972) Some Signals and Rules for Taking Speaking Turns in Conversations. *Journal of Personality and Social Psychology*. **23** 283-292

Duncan, S. and Fiske, D.W. (1977) *Face-to-Face Interaction: Research Methods and Theory*. Hillsdale, NJ: Erlbaum

Duncan, S. and Fiske, D.W. (1985) *Interaction Structure and Strategy*. New York: Cambridge University Press

Ekman, P. and Friesen, W. V. (1969) The repertoire of nonverbal behaviour: categories, origins, usage and coding. *Semiotica* **1** 49-98

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors Special Issue: Situation Awareness*, *37*, 32-64.

Fairclough, S. H. (1991). *Adapting the TLX to assess driver mental workload* (DRIVE I V1017 BERTIE, Deliverable 71). Loughborough, UK: HUSAT Research Institute

Finn, K. E. (1997) Introduction: An Overview of Video-Mediated Communication Literature. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 3-21

Finn, K. E., Sellen, A. J. and Wilbur S. B. (1997) *Video-Mediated Communication*. LEA: New Jersey

Fish, R., Kraut R. E. and Chalfonte, B. L. (1990) The VideoWindow system in informal communications. *In the Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW 1990)*. New York: ACM Press. pp. 1-11

Fish, R., Kraut, R. E., Root, R. and Rice, R. (1993) Video as a technology for informal communication. *Communications of the ACM*. **36** (1), 48-61

Flor, N. V. (1998) Side-by-side collaboration: A case study. *International Journal of Human-Computer Studies*. **49** 201-222

Fraser, M. (2000) *Working with Objects in Collaborative Environments*. Unpublished PhD thesis, Department of Computer Science, University of Nottingham. UK

Fussell, S, R., Setlock, L. D. and Kraut, R, E. (2003). Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks. In *Proc. of the Human Factors in Computing Systems conference* CHI 2003. April 5-10, Ft. Lauderdale, Florida, USA. ACM. pp. 513-520

Fussell, S. R., Kraut, R. E., and Siegel, J. (2000) Coordination of communication: Effects of shared visual context on collaborative work. *Proceedings of CSCW 2000*, ACM Press. pp. 21-30.

Fussell, S. R., Setlock, L. D., Parker, E. and Yang, J. (2003) Assessing the Value of a Cursor Pointing Device for Remote Collaboration on Physical Tasks. In *Proceedings of CHI 2003 Extended Abstracts*. NY: ACM Press. pp. 788-789

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E. and Kramer, A. D. I. (2004) Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction*. **19** 273-309

Fussell, S.R., Setlock, L. D., and Parker, E. M. (2003) Where do helpers look? Gaze targets during collaborative physical tasks. In *Proceedings of CHI 2003 (Extended Abstracts)*, ACM Press. pp. 768-769

Gale, S. (1989) Adding audio and video to an office environment. *In Proceedings of the 1st European Conference on CSCW (ECSCW, 1989)*. London. pp. 121-130

Garfinkel, H. (1967) *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall

Gaver, W. (1992) The Affordances of Media Spaces for Collaboration, In *Proceedings of CSCW'92*. pp. 17-24

Gaver, W., Sellen, A., Heath, C. and Luff, P. (1993). One is not enough: Mulitple Views in a Media Space. *In Proceedings of INTERCHI'93* (24-29 April 1993). ACM, New York. pp. 335-341

Gergle, D., Kraut, R. E., and Fussell, S. R. (2006). The impact of delayed visual feedback on collaborative performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM Press, New York, NY, pp. 1303-1312

Gergle, D., Kraut, R.E., and Fussell, S.R. (2004a). Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, 23 (4), 491-517

Gergle, D., Kraut, R.E., and Fussell, S.R. (2004b). Action as language in a shared visual space. In *Proceedings of ACM Conference on Computer Supported Cooperative Work* (CSCW 04), NY: ACM Press. pp. 487-496

Gergle, D., Millen, D. R., Kraut, R. E., and Fussell, S. R. 2004. Persistence matters: making the most of chat in tightly-coupled work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM Press, New York, NY, pp. 431-438

Glucksberg, S., Krauss, R. M., and Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology*, 3, 333-342

Goodwin, C. (1981) *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

Goodwin, C. (1994) Professional vision. *American Anthropologist.* 96 (3), 606-633

Goodwin, C. (1995). Seeing in Depth, *Social Studies of Science*. 25: 2, 237-274

Greenberg, S., Cutwin, C. and Roseman, M. (1996) Semantic telepointers for groupware. *Proceedings of OzCHI'96 Australian Conference on Computer-Human Interaction*, pp. 54-61

Grudin, J. (1988) Why CSCW Applications fail: Problems in the design and Evaluation of organizational interfaces. In *Proceedings of Conference on Computer-Supported Cooperative Work 1988*. ACM. pp. 85-93

Gutwin, C., and Penner R. (2002) Visual Information and Collaboration: Improving interpretation of remote gesture with telepointer traces. *In Proceedings of Conference on Computer-Supported Cooperative Work (CSCW) 2002*, New Orleans: ACM, pp. 49-57

Harrison, S., Bly, S., Anderson, S. and Minneman, S. (1997) The Media Space. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 273-300

Hayne, S., Pendergast, M. and Greenberg, S. (1993) Gesturing Through Cursors: Implementing Multiple Pointers in Group Support Systems. *Computer Science Technical Report 1992-490-28*. Department of Computer Science, University of Calgary

Heath, C. (1986) *Body movement and speech in medical interaction*. Cambridge University Press, Cambridge.

Heath, C. (1997) The Analysis of Activities in Face to Face Interaction Using Video. David Silverman (ed.) *Qualitative Sociology*, London: Sage, 1997, pp. 183-200

Heath, C. and Luff, P. (1991) 'Disembodied conduct: communication through video in multi-media office environment', *In Proceedings of Conference on Human Factors in Computing Systems CHI '91*, New Orleans: ACM Press. pp. 99-103

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

Heath, C. C. and Luff P. K. (1996) Convergent activities: collaborative work and multimedia technology in London Underground Line Control Rooms. In D. Middleton and Y. Engestrom (Eds.), *Cognition and Communication at Work: Distributed Cognition in the Workplace*. Cambridge University Press. pp. 96-130

Heath, C., Luff, P and Sellen, A. (1997) Reconfiguring Media Space: Supporting Collaborative Work. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 323-347

Heath, C., Luff, P., Kuzuoka, H., Yamazaki, K., and Oyama, S. (2001). Creating Coherent Environments for Collaboration. In W. Prinz, M. Jarke, Y. Rogers, K. Scmidt and V. Wulf (Eds.) *Proceedings of the Seventh European Conference on Computer-Supported Cooperative Work*. (16-20 September 2001, Bonn, Germany). Kluwer Academic Publishers: Netherlands. pp 119-138

Heiser, J. and Tversky, B. (2006) Arrows in Comprehending and Producing Mechanical Diagrams. *Cognitive Science,* **30** 3, 581-592

Heritage, J. (1997) Conversational Analysis and Institutional Talk. David Silverman (ed.), *Qualitative Sociology*, London:Sage, 1997, pp. 161-182

Hinds, P. J. and Kiesler, S. (2002) *Distributed Work*. MIT Press

Hughes, J., King, V., Rodden, T. and Andersen, H. (1994) Moving Out from the Control Room: Ethnography in System Design. In *Proceedings of Conference on Computer-Supported Cooperative Work '94*. ACM Press. pp. 429-439

Hutchins, E. (1991) The Social Organization of Distributed Cognition. In In L.B. Resnick, R.M. Levine, and S.D. Teasley (Eds.). *Perspectives on socially shared cognition*. Washington, DC: APA. pp. 283-307

Hutchins, E. (1995) *Cognition in the Wild*. Cambridge, MA: MIT Press

Hutchins, E. and Palen, L. (1997) Constructing Meaning from Space, Gesture and Speech. In L. B. Resnick, R. Saljo, C. Pontecorvo and B. Burge (Eds.) *Discourse, Tools and Reasoning. Essays on Situated Cognition.* NATO ASI series. Series F: Computer and System Sciences v.160, Springer. pp.23-40

Hutchins, E. and T. Klausen (1996). Distributed cognition in an airline cockpit. In D. Middleton and Y. Engestrom (Eds.), *Cognition and Communication at Work: Distributed Cognition in the Workplace*. Cambridge University Press. pp. 15-34

Isaacs, E. A. and Tang, J. C. (1993) What video can and can't do for collaboration: A case study. *In Proceedings of ACM Multimedia 93*. New York: ACM Press. pp. 199-206

Isaacs, E. A. and Tang, J. C. (1994) What video can and can't do for collaboration: A case study. *Multimedia Systems*. 2, 63-73

Isaacs, E. A. and Tang, J. C. (1997) Studying Video-Based Collaboration in Context: From Small Workgroups to Large Organizations. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 173-197

Isaacs, E. A., Whittaker, S., Frohlich, D. and O'Conaill, B. (1997) Informal Communication Reexamined: New Functions for Video in Supporting Opportunistic Encounters. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 459-485

Ishii, H (1990) TeamWorkStation: Towards a Seamless Shared Workspace. In *Proceedings of Conference on Computer-Supported Cooperative Work 1990.* ACM pp. 13-26

Ishii, H. and Arita, K. (1991) ClearFace: Translucent multiuser interface for TeamWorkStation. In *Proceedings of European Conference on Computer-Supported Cooperative Work (ECSCW 1991)* Amsterdam: Netherlands. pp. 163-174

Ishii, H., and Kobayashi, M. (1992) Clearboard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. *In Proceedings of Conference on Human Factors in Computing Systems CHI 1992*. Monterey: ACM Press, pp. 525-535

Ishii, H., and Miyake, N. (1991) Toward an open shared workspace: Computer and video fusion approach of TeamWorkStation. *Communications of the ACM*. 34 (12) 37-50

Ishii, H., Kobayashi, M. and Arita, K. (1994) Iterative Design of Seamless Collaboration Media. *Communications of the ACM*. 37(8) 83-97

Ishii, H., Kobayashi, M. and Grudin, J. (1993) Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments. *ACM Transactions on Information Systems*. 11 (4) 349-375

Ishii, T., Hirose, M., Kuzuoka, H., Takahara, T. and Myoi, T. (1990) Collaboration System for Manufacturing System in the 21st Century. In *Proceedings of the International Conference on Manufacturing Systems and Environment*, pp. 295–300

Karsenty, L. (1999) Cooperative work and shared visual context: An empirical study of comprehension problems and in side-by-side and remote help dialogues. *Human-Computer Interaction*, 14 (3) 283-315

Kato, H.,Yamazaki, K., Suzuki, H., Kuzuoka, H., Miki, H., and Yamazaki, A. (1997) Designing a video-mediated collaboration system based on a body metaphor. *In Proceedings of Conference on Computer Supported Co-operative Learning CSCL'97*, Kluwer, pp. 142-149

Kendon, A. (1994) *Gesture and Understanding in Social Interaction*. Lawrence Erlbaum Associates

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

Kirsh, D. (2001). The Context of Work. *Human-Computer Interaction*. 16, 305-322.

Koleva, B. N., Schnädelbach, H. M., Benford, S. D. and Greenhalgh, C. M. (2000) Traversable interfaces between real and virtual worlds, in *Proc. ACM Conference on Human Factors in Computing Systems* (CHI 2000), Hague, Netherlands, ACM Press, pp. 233-240

Koleva, B., Taylor, I., Benford, S., Fraser, M., Greenhalgh, G., Schnädelbach, H., vom Lehn, D., Heath, C., Row-Farr, J. and Adams, M. (2001) Orchestrating a Mixed Reality Performance, in *Proc. ACM Conference on Human Factors in Computing Systems* (CHI 2001), 31 March – 5 April, ACM Press, pp. 38-45

Kramer, A. D., Oh, L. M., and Fussell, S. R. (2006). Using linguistic features to measure presence in computer-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM Press, New York, NY, pp. 913-916

Krauss, R. M. and Fussell, S. R. (1990) Mutual Knowledge and Communicative Effectiveness. In J. Galegher, R. E. Kraut and C. Egido (Eds.) *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 111-147

Krauss, R. M. and Fussell, S. R. (1991) Constructing Shared Communicative Environments. In L.B. Resnick, R.M. Levine, and S.D. Teasley (Eds.). *Perspectives on socially shared cognition*. pp. 127-149. Washington, DC: APA

Kraut, R. E., Egido C. and Galegher, J. (1990) Patterns of Contact and Communication in Scientific Research Collaborations. In J. Galegher, R. E. Kraut and C. Egido (Eds.) *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 149-172

Kraut, R. E., Fussel, S. R., and Siegel, J. (2003) Visual Information as a Conversational Resource in Collaborative Physical Tasks. *Human-Computer Interaction* 18 (1) 13-49

Kraut, R. E., Gergle, D., and Fussell, S. R. (2002) The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of CSCW 2002,* ACM Press. pp. 31-40

Kraut, R. E., Miller, M. D. and Siegel, J. (1996). Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication. In *Proceedings of CSCW'96*. ACM Press, Cambridge, MA. pp. 57-66

Kraut, R.E., Gergle, D., and Fussell, S.R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW 2002)*, ACM Press. pp. 31-40

Kurata, T., Sakata, N., Kourogi, M., Kuzuoka, H. and Billinghurst, M. (2004a) Remote collaboration using a shoulder-worn active camera/laser. In *Proceedings of International Symposium on Wearable Computers 2003*. IEEE Press. pp. 62-69

Kurata, T., Sakata, N., Kourogi, M., Kuzuoka, H. and Billinghurst, M. (2004b) The Advantages and Limitations of a Wearable Active Camera/Laser in Remote Collaboration. In *Extended Abstracts of ACM Conference on Computer Supported Cooperative Work* (CSCW 04) NY: ACM Press.

Kuzuoka, H. (1992) Spatial workspace collaboration: A Sharedview video support system for remote collaboration capability. *Proceedings of CHI'92* ACM Press. pp. 533-540

Kuzuoka, H. and Shoji, H. (1994) Findings from Observational Syudies of Spatial Workspace Collaboration. *Electronics and Communications in Japan*. 77 (8) 58-68

Kuzuoka, H., Ishimoda, G., Mishimura, Y., Suzuki, R. and Kondo, K. (1995) Can the GestureCam be a Surrogate? In *Proceedings of the Fourth European Conference on CSCW*. Stockholm, Sweden. pp. 181-196

Kuzuoka, H., Kosaka, J., Oyama, S. and Yamazaki, K. (2003) GestureMan PS: Effect of a Head and a Pointing Stick on Robot Mediated Communication, In *Proceedings of HCII2003* Volume 3 (Human-Centered Computing), pp. 1416-1420

Kuzuoka, H., Kosaka, J., Yamazaki, K., Suga, S., Yamazaki, A., Luff, P. and Heath, C. (2004a) 'Mediating dual ecologies', *In Proc of CSCW '04*, Chicago: ACM Press. pp. 477-486

Kuzuoka, H., Kosuge, T., and Tanaka, K. (1994) GestureCam: A video communication system for sympathetic remote collaboration. *Proceedings of CSCW 1994* ACM Press pp. 35-43

Kuzuoka, H., Oyama, S.,Yamazaki, K., Suzuki, K. and Mitsuishi, M. (2000). GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions. *Proceedings of CSCW'2000*. ACM, Philadelphia, PA. pp. 155-162

Kuzuoka, H., Yamashita, J., Yamazaki, K., and Yamazaki, A. (1999) Agora: A Remote Collaboration System that Enables Mutual Monitoring. *In Proceedings of Conference on Human Factors in Computing Systems CHI 1999*. Pittsburgh: ACM, pp. 190-191

Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, S., and Heath, C. (2004b) 'Dual Ecologies of Robot as Communication Media: Thoughts on Coordinating Orientations and Projectability. *In Proceedings of CHI '04*, Vienna, Austria: ACM Press. pp. 183-190

Latour, B. (1992). Where Are the Missing Masses? The sociology of a few mundane artefacts. In Bijker and Law (Eds.) *Shaping Technology/Building Society*. MIT Press. pp. 225-258

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., and Oyama, S. (2003) Fractured ecologies: creating environments for collaboration, Special Issue of the *HCI Journal*: 'Talking About Things: Mediated Conversation about Objects', 18, (1 and 2), 51-84

Luff, P., Heath, C., Kuzuoka, H., Yamazaki, K. and Yamashita, J. (2006) Handling Documents and Discriminating Objects in Hybrid Spaces. In *Proceedings of CHI 2006*. Montreal, Quebec: ACM Press. pp. 561-570

Mantei, M., Baecker, R., Sellen, A., Buxton, W., Milligan, T. and Wellman, B. (1991) Experiences in the use of a media space. *In Proceedings of the Human Factors Conference on Computing Systems CHI 1991,* ACM Press, pp. 203-208

McNeill, D. (1992) *Hand and Mind. What gestures reveal about thought*. Chicago: University of Chicago Press

Minneman, S.L. and Bly, S.A. (1991) Managing a trois: a study of a multi-user drawing tool in distributed design work. *In Proceedings of Conference on Human Factors in Computing Systems (CHI) 1991*, New Orleans: ACM Press. pp. 217-224

Moles, A. (1966) *Information Theory and Esthetic Perception*. (Trans. Joel, E. Cohen). University of Illinois Press

Moles, A. (1975) Le mur de la communication. *Actes du XV<sup>e</sup> Congrès de la ASPLF*. Vol. II. Referenced in A. Mattelart and M. Mattelart (Trans. G. Taponier and J. A. Cohen); *Theories of Communication: a short introduction*. (1998) London: SAGE (p.49)

Monk, A., McCarthy, J., Watts, L. and and Daly-Jones, O. (1996) Measures of process. In M. MacLeod and D. Murray (Eds.), *Evaluation for CSCW*. Berlin: Springer-Verlag. pp. 125-139

Moore, G. (1997) Sharing Faces, Places, and Spaces: The Ontario Telepresence Project Field Studies. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 301-321

Muter, Paul (1996) *Interface design and optimization of reading of continuous text*, in van Oostendorp, H. and de Mul, Ajaak, Eds. *Cognitive Aspects of Electronic Text Processing*. Ablex, Norwood, N.J. pp. 161-180

Nardi, B. A., Kuchinsky, A., Whittaker, S., Leichner, R. and Schwarz, H. (1997) Video-as-Data: Technical and Social Aspects of a Collaborative Multimedia Application. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 487-517

Nardi, B. A., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S., and Sclabassi, R. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. In *Proceedings of the Conference on Human Factors in Computing Systems CHI '93* New York: ACM Press. pp. 327-334

O'Conaill, B. and Whittaker, S. (1997) Characterizing, Predicting and Measuring Video-Mediated Communication: A Conversational Approach. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 107-131

O'Conaill, B., Whittaker, S. and Wilbur, S. (1993) Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction*. 8, 389-428

Ochsman, R.B. and Chapanis, A. (1974) The effects of 10 communication modes on the behaviour of teams during co-operative problem-solving. *International Journal of Man-Machine Studies*. 6, 579-619

Olson, G. M., Olson, J. S., Carter, M. and Storrøsten, M. (1992) Small group design meetings: An analysis of collaboration. *Human Computer Interaction*. 9, 427-472

Olson, J. S., Olson, G. M., and Meader, D. K. (1995) What mix of video and audio is useful for small groups doing remote real-time design work? In *Proceedings of Conference on Human Factors in Computing Systems (CHI) 1995*. pp. 362-368

Olson, J. S., Olson, G. M., and Meader, D. K. (1997) Face-to-Face Group Work Compared to Remote Group Work With and Without Video. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 157-172

Olson, J. S., Olson, G. M., Storrøsten, M. and Carter, M. (1993) Groupwork close up: A comparison of the group design process with and without a simple group editor. *ACM Transactions on Information Systems.* 11, 321-348

Ou, J., Chen, X., Fussell, S. R. and Yang, J. (2003b) DOVE: Drawing over Video Environment. *Demonstration paper for Multimedia 2003* (MM'03) conference. November 2-8, Berkeley, California, USA. ACM. pp. 100-101

Ou, J., Fussell, S. R., Chen, X., Setlock, L. D., and Yang, J. (2003a) Gestural communication over video stream: Supporting multimodal interaction for remote collaborative physical tasks. In *Proceedings of ICMI 2003*, Vancouver: ACM Press, pp. 242-249

Ou, J., Min, L., Yang, J. and Fussell, S. R. (2005) Effects of Task Properties, Partner Actions and Message Content on Eye Gaze Patterns in a Collaborative Task. In *Proceedings of CHI 2005*. ACM Press, pp. 231 – 240

Ou, J., Oh, L. M., Fussell, S. R., Blum, T., and Yang, J. 2005. Analyzing and predicting focus of attention in remote collaborative tasks. In *Proceedings of the 7th International Conference on Multimodal interfaces* (Toronto, Italy, October 04 - 06, 2005). ICMI '05. ACM Press, New York, NY, pp. 116-123

Ou, J., Oh, L., Yang, J., and Fussell, S. R. 2005. Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '05* (Portland, Oregon, USA, April 02 - 07, 2005). ACM Press, New York, NY, pp. 231-240

Rauscher, F. H., Krauss, R. M. and Chen, Y. (1996) Gesture, Speech and Lexical Access: The Role of Lexical Movements in Speech Production. *Psychological Science*. 7 (4) 226-231

Robertson, T. (1997a) Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions. In *Proceedings of the Fifth European Conference on Computer-Supported Cooperative Work*, Lancaster, UK, September 7-11, 1997, Kluwer Academic Publishers, pp. 205-220

Robertson, T. (1997b) *Designing Over Distance: A study of cooperative work, embodied cognition and technology to enable remote collaboration*. Unpublished PhD Thesis, University of Technology, Sydney, Australia

Robson, C. (2002) Real World Research, 2. ed. Oxford: Blackwell

Root, R. W. (1988) Design of a multi-media vehicle for social browsing. *In the Conference Proceedings of CSCW 1988*. New York: ACM Press. pp. 25-38

Rutter, R. and Robinson, R. (1981) An experimental analysis of teaching by telephone. In G. Stephenson and J. Davies (Eds.) *Progress in applied social psychology*. London: Wiley. pp.143-178

Rutter, R. R., Stephenson, G. M. and Dewey, M. E. (1981) Visual communication and the content and style of conversation. *British Journal of Social Psychology.* 20, 41-52.

Sacks, H. (1992) *Lectures on conversation*. Oxford, UK: Blackwell

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

Sakata, N., Kurata, T., Kato, T., Kourogi, M. and Kuzuoka, H. (2003) WACL: Supporting Telecommunications Using Wearable Active Camera with Laser Pointer. *Proceedings of International Symposium on Wearable Computers 2003*. IEEE Press. pp. 53-56

Schmalstieg, D., Fuhrmann, A., Szalavri, Z. and Gervautz, M. (1996) Studierstube: An environment for collaboration in augmented reality. In *Proceedings of the Collaborative Virtual Environments (CVE'96) Workshop*. (Nottingham, UK, Sept. 19-20)

Schnädelbach, H, Penn, A., Steadman, P., Benford, S., Koleva, B., Rodden, T. Moving Office: Inhabiting a Dynamic Building. In *Proceedings of CSCW 2006*, Banff, Canada. pp. 313-322

Scott, S. (2005). *Territoriality in Collaborative Tabletop Workspaces*. Unpublished PhD thesis. Department of Computer Science. University of Calgary: Canada

Sellen, A. (1992) Speech patterns in video-mediated conversations. *In Proceedings of the Human Factors Conference on Computing Systems CHI 1992*. ACM Press, pp. 49-59

Sellen, A. (1995) Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction*. 10 (4), 401-444

Sellen, A. (1997) Assessing Video-Mediated Conduct: A Discussion of Different Analytic Approaches. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 95-106

Sellen, A. and Harper, R. (1997) Video in Support of Organizational Talk. In K. E. Finn, A. J. Sellen, and S. B. Wilbur (Eds.) *Video-Mediated Communication*. LEA: New Jersey. pp. 225-243

Shapiro, D. (1994) The Limits of Ethnography: Combining Social Sciences for CSCW. In *Proceedings of Conference on Computer-Supported Cooperative Work '94*. ACM Press. pp. 417-428

Short, J., Williams, E. and Christie, B. (1976) *The Social Psychology of Telecommunications.* Chichester: Wiley

Smith, R., O'Shea, T., O'Malley, C., Scanlon, E. and Taylor, J. (1991) Preliminary experiments with a distributed, multimedia, problem-solving environment. In J. Bowers and S. Benford (Eds.) *Studeis in computer-supported co-operative work: Theory, practice and design*. Amsterdam: Elsevier. pp. 31-48

Suchman, L. (1996). Constituting shared workspaces. In D. Middleton and Y. Engestrom (Eds.), *Cognition and Communication at Work: Distributed Cognition in the Workplace*. Cambridge University Press. pp. 35-60

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communications*. Cambridge, UK: Cambridge University Press.

Tang, A., Boyle, M. and Greenberg, S. (2004) Display and Presence Disparity in Mixed Presence Groupware. *Proceedings of Australasian user Interface*. ACM Press. pp. 73-82

Tang, A., Neustaedter, C., Greenberg S. (2004) Embodiments and VideoArms in mixed presence Groupware. *Technical Report 2004-741-06*, Dept of Computer Science, University of Calgary.

Tang, A., Owen, C., Biocca, F. and Mou, W. (2002). Experimental Evaluation of Augmented Reality in Object Assembly Task. *Proceedings of the International Symposium on Mixed and Augmented Reality* (ISMAR 02). IEEE. Sept 30 – Oct 01. Darmstadt, Germany. pp. 265-267

Tang, A., Owen, C., Biocca, F., and Mou, W. (2003) Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '03* (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003). ACM Press, New York, NY, pp. 73-80

Tang, J. C. and Isaacs, E. A. (1993) Why do users like video? Studies of multimedia-supported collaboration. *Computer-Supported Collaborative Work: An International Journal.* 1 (9), 163-196

Tang, J. C. and Rua, M. (1994) Montage: Providing teleproximity for distributed groups. *In Proceedings of the Human Factors Conference on Computing Systems CHI 1994,* ACM Press, pp. 37-43

Tang, J. C., Isaacs, E. A. and Rua, M. (1994) Supporting distributed groups with a montage of lightweight interactions. *In the Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW 1994).* New York: ACM Press. pp. 23-34

Tang, J.C. (1989) *Listing, Drawing and gesturing in Design: A Study of the Use of Shared Workspaces by Design Teams.* Unpublished PhD thesis. Department of Mechanical Engineering. Stanford University: California

Tang, J.C. (1991) Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies.* 34, 143-160

Tang, J.C. and Leifer, L. J. (1988) A framework for understanding the workspace activity of design teams. In *Proceedings of the Conference on CSCW.* Portland, Oregon, pp. 244-249

Tang, J.C. and Minneman, S.L. (1990) VideoDraw: a video interface for collaborative drawing. *In Proceedings of the Conference on Human Factors in Computing Systems. CHI 1990.* Seattle: ACM, pp. 313-320

Tang, J.C. and Minneman, S.L. (1991a) VideoDraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems* 9 (2) 170-184

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

Tatar, D, G., Foster, G., and Bobrow, D. G. (1991). Design for conversation: lessons from Cognoter. *International Journal of Man-Machine Studies.* 34, 185-209.

Tversky, B., Zacks, J., Lee, P. U. and Heiser J. (2000) Lines, Blobs, Crosses, and Arrows: Diagrammatic Communication with Schematic Figures. In M. Anderson, P. Cheng, and V. Haarslev (eds.) Theory and Application of Diagrams. Berlin: Springer pp. 221-230

Weiser, M. (1991) The Computer for the Twenty-First Century. *Scientific American.* 265, (3) pp. 94-10

Wellner, P. (1993). Interacting with Paper on the DigitalDesk. *Communications of the ACM*. 36 (7) 87-96

Whittaker, S. and O'Conaill, B. (1997) The role of vision in face-to-face and mediated communication. In K. Finn, A. Sellen , and S. Wilbur (Eds.) *Video-mediated communication*. Mahwah, NJ: Lawrence Erlbaum Associates. pp. 23-49

Williams, E. (1977) Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*. 84 (5), 963-976

Yamashita, J., Kuzuoka, H. and Yamazaki, K. (1999) Agora: Supporting Multi-participant Telecollaboration, In *Proceedings of HCI International HCII '99*, Vol. 2, pp. 543-547

Yamazaki, K., Yamazaki, A., Kuzuoka, H., Oyama, S., Kato, H., Suzuki, H. and Miki, H. (1999). GestureLaser and GestureLaser Car Development of an Embodied Space to Support Remote Instruction. In S. Bodker, M. Kyng, and K.Scmidt (Eds.) *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work CSCW'99* (12-16 September 1999, Copenhagen, Denmark). Kluwer Academic Publishers. Netherlands. pp. 239-258

# Appendices

**Appendix 3.1 Prompt image available during use of the sketch only gesture system** (not shown actual size)



To delete a sketch, select this icon (by pressing with pen) and then outline the area for deletion., then click in the area (using the button on the pen)and select cut from the menu

To draw make sure this icon is selected and then press the pen against your desk surface
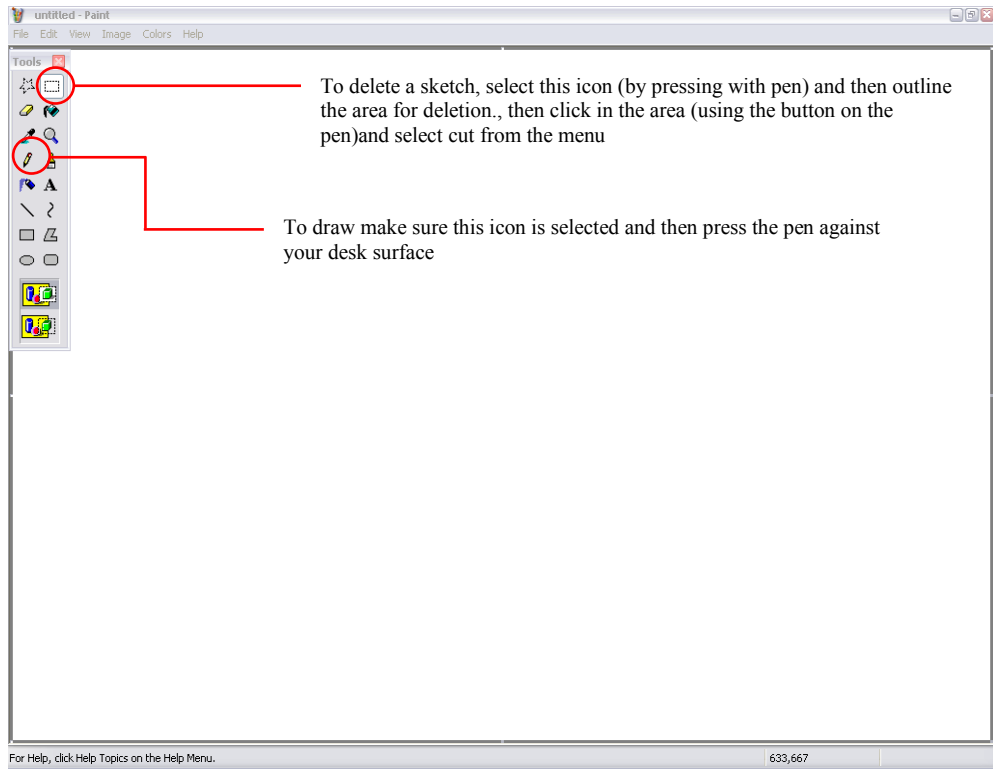
**Appendix 4.1 Lego models used and sample instructions**

Lego kit 8441 as Forklift



Lego kit 8441 as Car



Example of a Lego manual's instructions (first 4 stages of Car model – 8441)
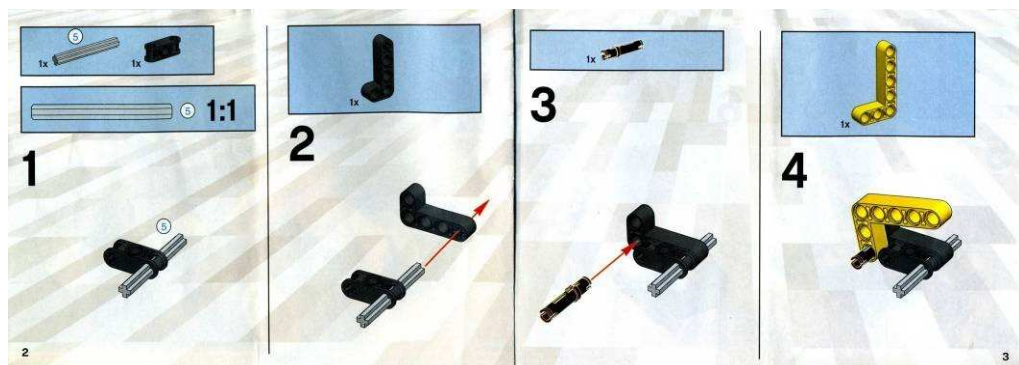
**Appendix 4.2 NASA TLX questionnaire for subjective assessment of mental workloads**

Subjective rating subscales

Participant_____                                          Date  _____

Trial  1  or  2   (please circle)                    Worker   or   Helper   (please circle)

### NASA TLX - MENTAL WORKLOAD

(Please mark along the line from Low to High)

| | Low | High |
|---|---|---|
| Mental Demand | ├─────────────────────────────┤ | |
| | Low | High |
| Physical Demand | ├─────────────────────────────┤ | |
| | Low | High |
| Temporal Demand | ├─────────────────────────────┤ | |
| | Low | High |
| Effort | ├─────────────────────────────┤ | |
| | Good | Poor |
| Performance | ├─────────────────────────────┤ | |
| | Low | High |
| Frustration Level | ├─────────────────────────────┤ | |

Subscale paired-comparisons form

Participant_____                                          Date _____

**For the tasks that you have just done which measures contributed most to the <u>workload</u> of the task?**

(Please circle the more important measure in each pair)

| | | | |
|---|---|---|---|
| 1) | Mental Demand | Or | Physical Demand |
| 2) | Temporal Demand | Or | Mental Demand |
| 3) | Temporal Demand | Or | Physical Demand |
| 4) | Physical Demand | Or | Effort |
| 5) | Performance | Or | Physical Demand |
| 6) | Physical Demand | Or | Frustration Level |
| 7) | Effort | Or | Temporal Demand |
| 8) | Mental Demand | Or | Effort |
| 9) | Performance | Or | Mental Demand |
| 10) | Mental Demand | Or | Frustration Level |
| 11) | Temporal Demand | Or | Performance |
| 12) | Frustration Level | Or | Temporal Demand |
| 13) | Effort | Or | Performance |
| 14) | Frustration Level | Or | Effort |
| 15) | Performance | Or | Frustration Level |

Definition of scales

| RATING SCALE DEFINITIONS | | |
|---|---|---|
| Title | Endpoints | Descriptions |
| MENTAL DEMAND | Low/High | How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? |
| PHYSICAL DEMAND | Low/High | How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious? |
| TEMPORAL DEMAND | Low/High | How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic? |
| EFFORT | Low/High | How hard did you have to work (mentally and physically) to accomplish your level of performance? |
| PERFORMANCE | Good/Poor | How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals? |
| FRUSTRATION LEVEL | Low/High | How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task? |

**Appendix 4.3 Evaluation questionnaire for Experiment 1**

**Remote Gesturing in Collaborative Physical Tasks**
**Evaluation Questionnaire**

1) Have you met your collaborator in the experiment before? (please circle)

Yes     No

2) In the trial where gesture projection was used were you the Helper or Worker? (please circle)

Helper    Worker

3) Did you enjoy using the gesture projection? (please circle)

Yes     No     Don't Know

4) Do you think it made your communication easier? (please circle)

Yes (go to 4a)    No (go to 4b)     Don't Know (go to 5)

4a) If yes why? (go to 5)
_____
_____
_____
_____

4b) If no why? (go to 5)
_____
_____
_____
_____

5) Did you encounter any difficulties using the gesture projection?
_____
_____
_____
_____

6) If you had to do a similar physical task to the one you have just done would you prefer to have use of such a gesture projection system?
_____
_____
_____
_____

7) Date of Birth_____

8) Gender (please circle)

Male    Female

9) Are you a student? (please circle)

Yes (go to 9a)    No

9a)

Current Course_____          Year_____

**Thank you for taking part!**

**Appendix 4.4 Participant information sheet experiment 1**

**Remote Gesturing in Collaborative Physical Tasks**
**Participant Information Sheet**

You will be working in a pair for this experiment. Prior to starting the experimental trials you will be given the chance to try-out the remote gesturing technology so that you understand what it does and how it works.

The experiment has two trials, each lasting 10 minutes, after each trial you will be required to fill out some questionnaires.

During each trial you are asked to collaboratively assemble a Lego kit. You will be randomly assigned to either the 'Helper' or 'Worker' roles, and you get to swap role for each trial. The 'Helper' is the person who has the instructions for the Lego kit, they tell the 'Worker' what to do. The 'Worker' is the only person allowed to touch the Lego pieces, but they are not allowed to look at the instructions or know what it is that they are building.

For each trial the 'Helper' and the 'Worker' sit at separate desks, the 'Helper' is able to see what the 'Worker' is doing because of a video link between the desks. In one of the trials the set up remains exactly like this, with you working collaboratively by speech alone.

However, in the other trial the gesture projection technique is used. In this trial, if the 'Helper' sticks their hands out in front of them, this is picked up by a video camera, which is linked to a projector. Their hand gestures are then projected over the surface of the 'Worker's' desk, thus allowing the 'Helper' to point things out to the 'Worker'.

As mentioned above, in each trial you have 10mins, in this time you must complete as much of the Lego model as you can.

After each trial you must complete a mental workload questionnaire (the NASA TLX), and after the second trial you must fill out the third and final part of the mental workload questionnaire and a final evaluation questionnaire as well.

**Appendix 4.5 Consent Form**

**Participant consent form**

I, (insert full name in capitals) --------------------------------------------------------
Consent to take part in an experiment testing the performance effects of using a desktop-based mixed reality gesturing system. An explanation of the nature and purpose of the procedures has been given to me by:

(Investigator's name here) --------------------------------------------------------------

I also confirm that I am over 16 years of age and that I understand that I may withdraw from the experiment and that I am under no obligation to give reasons for withdrawal. I undertake to obey the laboratory regulations or the instructions of the experimenter regarding safety, subject only to my right to withdraw as declared above.

I understand that any information about myself that I have given will be treated as confidential by the experimenter, I will not be personally identifiable from any of the data stored about me and all experimental data will be kept in accordance with the Data Protection Act.

Video Images (Please circle as appropriate)

I also consent for video data containing my image to be stored and used for academic purposes as necessary in:

Presentations:        Yes / No

Research Videos:    Yes / No

Websites:              Yes / No


Name (in full) _____

Signature _____

Signature of investigator _____

Date _____

**Appendix 4.6 Practice model (3 – parts) experiment 1**

- 254 -

Model's diagrammatic instructions
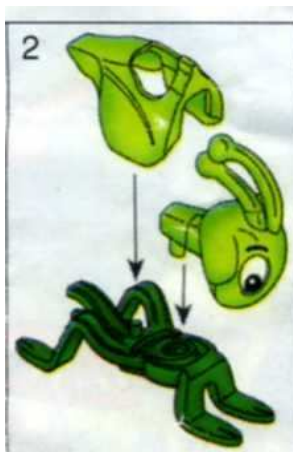


Model in pieces



Model assembled

**Appendix 4.7 Evaluation questionnaire for Experiment 2**

The University of **Nottingham**

UNIVERSITY OF **BATH**

**Mixed Reality Learning Experience Evaluation Questionnaire**

Please read the following 12 statements. Considering each in turn decide how much you agree or disagree with the statement. Mark along the line according to how much you agree or disagree. All responses will be kept strictly confidential, and you will not be personally identifiable from this questionnaire.

1. I found the Instructor's instructions clear to understand [please mark along the line]

Disagree |_____| Agree

2. I feel like I can remember how to assemble the model [please mark along the line]

Disagree |_____| Agree

3. I trusted the Instructor [please mark along the line]

Disagree |_____| Agree

4. I found communicating this way difficult [please mark along the line]

Disagree |_____| Agree

5. I liked the Instructor [please mark along the line]

Disagree |_____| Agree

6. The Instructor didn't think I understood what they meant [please mark along the line]

Disagree |_____| Agree

7. The Instructor noticed when I didn't understand an instruction [please mark along the line]

Disagree |_____| Agree

8. The Instructor understood me [please mark along the line]

Disagree |_____| Agree

9. I feel I did well with the task [please mark along the line]

Disagree |_____| Agree

10. I felt like I just did what I was told to do [please mark along the line]

Disagree |_____| Agree

11. I asked lots of questions [please mark along the line]

Disagree |_____| Agree

12. I was very aware of the presence of the Instructor [please mark along the line]

Disagree |_____| Agree

End of questions. Thank you for taking part!

## Appendix 4.8 Participant information sheet experiment 2

Dave Kirk – University of Nottingham

### The Impact of Mixed Reality Gesturing Systems on Remote Learning
### Participant Information Sheet

In this study you will initially be working with the assistance of an Instructor. During the first part of the study you will be remotely instructed in how to assemble a multi-part object (a Lego kit). This will be done via one of two methods, depending on the group to which you are randomly assigned.

One group of participants experiences the instructions from the 'Instructor' with the aid of projected gestures, the other group experiences the instructions in audio only – without the aid of remote gesturing (see Figure below for illustration of how the remote gesturing is achieved).

Instruction in object assembly lasts for up to 10mins.

After which you will fill out an evaluation questionnaire and then complete a distraction task for the remaining time, up to a total of 10mins.

You are then given a further 10mins to try and complete again as much of the object assembly task as you can, but this time on your own without assistance.

This attempt at self-assembly is then repeated roughly 24hrs later, you need to try and return to do the experiment at roughly the same time that you did it the first time, although arrangements can be made to do it later or earlier if this is not possible. On the second day the session only lasts for 10mins. All attempts at self-assembly will be video-recorded as will all instruction (using recordings from the video cameras integral to the technological set-up).
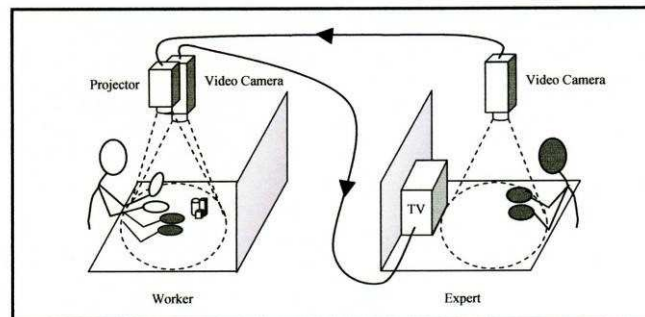


Figure 1. Schematic of gesture projection system

**Appendix 5.1 Evaluation questionnaire for Experiment 3**

Dave Kirk

**Mixed Reality Assembly Evaluation Questionnaire**
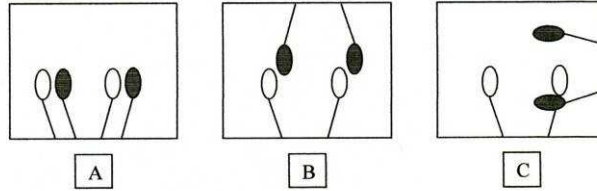


Please answer the questions below by referring (where appropriate) to the diagrams above.

1a. Which orientation was easiest to use? [please circle your choice]

     A     B     C

1b. Why? [please write a comment below]

_____
_____
_____
_____
_____

2a. Which orientation created the most confusion? [please circle your choice]

     A     B     C

2b. Why? [please write a comment below]

_____
_____
_____
_____
_____

3a. If you did the task again, which orientation would you prefer to use? [please circle your choice]

     A     B     C

3b. Why? [please write a comment below]

_____
_____
_____
_____
_____

[Can you please also provide the following information]

4. In this experiment were you the Helper or Worker? [please circle your choice]

     Worker          Helper

5. Gender_____

6. Age_____

7. Current course or job_____

End of questions. Thank you for taking part!

**Appendix 5.2 Participant information sheet experiment 3**

Orientations to Remote Tasks

**Lego study – Orientations to Remote Tasks**
**Participant Information Sheet**

You will be working in a pair for this experiment. Prior to starting the experimental trials you will be given the chance to try-out the remote gesturing technology so that you understand what it does and how it works (see the schematic below for an illustration of how it works). Communication during the experiment is by normal speech and by using the remote gesture device.

The experiment has three trials, each lasting 10 minutes, during each trial you will have a different Lego model to assemble.

For the experiment one of you will be the helper and one of you will be the worker. The helper is the person who has the instructions for the Lego kit, they tell the worker what to do. The worker is the only person allowed to touch the Lego pieces, but they are not allowed to look at the instructions or know what it is that they are building.

For each trial the helper and the worker sit at their separate desks, the helper is able to see what the worker is doing because of the video link between the desks, the helper can stick their hands out in front of them, this is picked up by a video camera, which is linked to a projector. Their hand gestures are then projected over the surface of the workers' desk, thus allowing the helper to point things out to the worker.

As mentioned above, in each trial you have 10mins, in this time you must complete as much of the Lego model as you can. What changes between each trial is the angle at which the remote helper's hands are projected onto the worker's desk.

After all of the trials each member of the pair must complete a brief evaluation questionnaire. You will be video-recorded during the experiment.
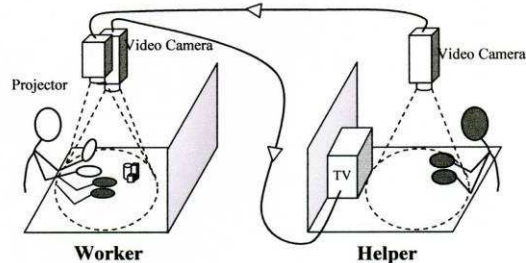
**Appendix 5.3 Transcript of evaluative comments experiment 3 evaluation questionnaire**

**Participant: 1a**
**Role: Helper**
Questions:
1b – it allowed me to be in his place when directing him with instruction, so he could easily copy my hand movements (like I was making it myself)
2b – Coming in from the side was awkward, I found it harder and longer to adjust to the view and guide him. I had to keep on asking him to show me what he had done so far by getting him to turn the model round
3b –  It allows me to be in his place, so enable me to work the problem out – as if I was there and ten tell him how and what to do in the next step

**Participant:1b**
**Role: Worker**
Questions:
1b – B was easiest to use as there was little overlap of our hands and one person could point to the object without obstructing the view of the other person due to hands being in the way. It also seemed to be the most natural way of doing things if we had actually been working together
2b – A caused the most confusion because hands were overlapping each other and blocking the other person's view and I found that sometimes we mistook our own hands for each other's, as it is unnatural to be working in that orientation if we had been sitting at the same desk and building the object together.
3b – I would prefer to use b again as it seemed the most natural and effective way to do things almost as if you were actually sitting opposite someone and working together to build the object. This created the best environment for me to complete the task.

**Participant: 2a**
**Role: Helper**
Questions:
1b –It (overlapped) made the instructors feel as though he/she was doing the actual work, made it easier to point to the pieces
2b – (Lateral) Was hard to see which way up the pieces were
3b – (Overlapped) Was the easiest to use

**Participant:2b**
**Role: Worker**
Questions:
1b – (f-2-f) because it was like like sharing a table with someone. A (overlapped) was weird because it felt like I had four hands
2b – with A and B I hadn't have to change the object sideways each time. With C (lateral) I had to turn it sideways and it confused even them
3b – Because it felt like the person was right in front of me showing me what to do. Instead of being on top of me

**Participant: 3a**
**Role: Helper**
Questions:
1b – (f-2-f) It was clearer to point out objects, not as easy to guide the usage but clarity is less frustrating
2b – It (lateral) was also a lot harder to explain how objects fit together when approaching from the side. It also doesn't help when pointing out objects, its harder to guide the worker
3b – I just found it (f-2-f) the simplest to use alongside speech. There was no confusion or overlaying of hands

**Participant: 3b**
**Role: Worker**
Questions:
1b – (f-2-f) Helpers hands could be more easily distinguished as they were not overlapping my own. Could tell which pieces they were indicating to clearly
2b – (Overlapped) While working on the pieces it was made difficult by not being able to see which pieces were indicated to, and also meant harder to assemble having to move my hands out of the way
3b – (f-2-f) less difficulty in understanding the gestures

**Participant: 4a**
**Role: Worker**
Questions:
1b – (Lateral) Hands were not interfering with his display – maybe easier to find a particular piece from different angles
2b – (f-2-f) Had to "flip" pieces so it matched his instructions – this didn't give me the best view
3b – (Lateral) Easiest to use, didn't cause too much confusion and easy orientation for both of us

**Participant: 4b**
**Role: Helper**
Questions:
1b – It's (overlapped) easiest to "be the other person" spatially – as you're seeing things exactly the way they would

2b – (f-2-f) Seeing from opposing viewpoints was harder to grasp. We seemed to encounter more problems with orientation.
3b – (Overlapped) It's much easier to perform a task as if you're the other person

**Participant: 5a**
**Role: Helper**
Questions:
1b – (overlapped) I can see what his hands are doing from my view
2b – (f-2-f) Because left and right was different for me. When I wanted to show him certain pieces
3b – (overlapped) same reason as 1b

**Participant: 5b**
**Role: Worker**
Questions:
1b – (overlapped) I could see where her hand was moving therefore it was easier
2b – (f-2-f) could not help a lot as hands were opposite
3b – (overlapped) it's really easy therefore faster

**Participant: 6a**
**Role: Helper**
Questions:
1b – Would be the same if you were actually facing someone
2b – That wouldn't happen in real life, overlapping made it confusing
3b – (f-2-f) because it was easiest – less confusing

**Participant: 6b**
**Role: Worker**
Questions:
1b – (f-2-f) Probably because J was actually sat opposite me in reality
2b – (overlapped) hands got mixed up. However A (lateral) was almost as confusing
3b – (f-2-f) it was easiest to use

**Participant: 7a**
**Role: Helper**
Questions:
1b – built more pieces because worker and helper are using the same orientation
2b – (f-2-f) total inverse way
3b – (overlapped) easy to cooperate

**Participant: 7b**
**Role: Worker**
Questions:
1b – Because the helper is at the same orientation with me, its more easy for me to understand his instruction
2b – (lateral) the gesture projected on the paper is in different orientation. Difficult for me to understand
3b – (overlapped) easier for me to understand the gesture, easier for me to know where the helper point to

**Participant: 8a**
**Role: Worker**
Questions:
1b – I actually noticed the hands pointing in this orientation(f-2-f), whereas in others I wouldn't notice them come onto the projection
2b – (overlapped) couldn't really see the hands
3b – (f-2-f) easiest to see the other persons hands

**Participant: 8b**
**Role: Helper**
Questions:
1b – (Lateral) was quite like a normal way to be sitting when doing a task like this
2b – Strange for both sets of hands to come from same direction. Not like anything that would normally happen
3b – (lateral) seemed to be the easiest way to point and show things accurately

**Participant: 9a**
**Role: Worker**
Questions:
1b – Instructions were outlined clearer since it was as if both participants were facing one another – the other two orientations resulted in overlapping of hands which created confusion
2b – (overlapped) Both the participants hands were overlapping so confusion created, resulted in instructions not being outlined as clear
3b – (f-2-f) clearer instructions

**Participant: 9b**
**Role: Helper**
Questions:
1b – (f-2-f) This is how would probably give instructions to someone in real life

2b – I think because this in reality is impossible (i.e. it is not possible to sit in this position)
3b – (f-2-f) It was easier to explain how to put the pieces together in this orientation

**Participant: 10a**
**Role: Worker**
Questions:
1b – equal
2b – equal
3b – easier communication as 'left', 'right', 'top' and 'bottom' terms would have the same meaning for each person

**Participant: 10b**
**Role: Helper**
Questions:
1b – (f-2-f) I could visualise easier and point easier to the piece required. And it was more comfortable to imagine this setting
2b – (overlapped) As it isn't the most common position to be sitting in! it was hard to imagine sitting in that position. However, it was easier to say to the worker to get the piece pointing towards you as I was looking at it from their angle
3b – (f-2-f) due to reasons stated above

**Participant: 11a**
**Role: Helper**
Questions:
1b – (f-2-f) It was the closest to the orientation we were already at and the most common orientation. Voices were coming from the correct area…but the colour on the TV made it more difficult
2b – (overlapped) because it's an unusual seating configuration for joint work
3b – (f-2-f) It was the easiest

**Participant: 11b**
**Role: Worker**
Questions:
1b – (f-2-f) The helper could point to pieces without getting in the way
2b – (lateral) If worker was working on pieces it was hard to distinguish which pieces the helper was pointing at
3b – (f-2-f) It's the most natural, as if someone was sitting opposite you

**Participant: 12a**
**Role: Worker**
Questions:
1b – (overlapped) More similar to "Helper's" hands being my own – able to direct me as if she were doing the task herself; same orientations etc.
2b – (f-2-f) was almost as if everything was reversed; got confused with orientations etc.
3b – (overlapped) Clearest had fewest problems and seemed to perform best at the task using this orientation (completed most of the model)

**Participant: 12b**
**Role: Helper**
Questions:
1b – (overlapped) Because when pointing to the pieces and where they should go we were both looking at the model / Lego from the same perspective. Also the easiest model by far!
2b – (lateral) Because it was harder to see the model as you had to keep changing the way it was facing in order to communicate exactly where each piece should go.
3b – (overlapped) Because it felt far more simple to communicate effectively where pieces should be put

**Participant: 13a**
**Role: Worker**
Questions:
1b – In A our hands got on top of each other. In B it was confusing which is left and right. With C (lateral) are hands didn't clash which made it simpler
2b – (f-2-f) because it was upside down
3b – (lateral) as discussed above

**Participant: 13b**
**Role: Helper**
Questions:
1b – C (overlapped) was difficult as we got in each others way B (lateral) was good as I had a similar reach to Al. A (f-2-f) meant I had to stretch all the way across the board
2b – (f-2-f) difficult to describe the correct orientation
3b – (lateral) I found it easiest to use

**Participant: 14a**
**Role: Worker**
Questions:
1b – Same direction as mine
2b – I don't actually think any orientation has caused me confusion. They are more or less the same.
3b – (overlapped) easiest

**Participant: 14b**
**Role: Helper**
Questions:
1b – felt more face-to-face
2b – (lateral) harder to reason 'left' and 'right' hands i.e. to demonstrate
3b – (f-2-f) it felt more comfortable

**Participant: 15a**
**Role: Helper**
Questions:
1b – Orientations were the same – 'the block nearest to you' or 'on your left' were more interpretable
2b – (f-2-f) Both axes were in effect reversed, wasting a few seconds confusion
3b – (overlapped) Easiest to use and allowed both people to feel like they were experiencing the same thing

**Participant: 15b**
**Role: Worker**
Questions:
1b – Because the Helper was sitting in the same orientation the builder was and both had the same angle and point of view as each other
2b – (lateral) The angle of view was not the same. It was hard to understand which way the pieces were supposed to fit
3b – Both people have the same point of view and angle

**Participant: 16a**
**Role: Worker**
Questions:
1b – Our ands were in the same orientation and directions
2b – (lateral) It was hard to see when his hands were pointing from the sides
3b – (overlapped) it was the easiest

**Participant: 16b**
**Role: Helper**
Questions:
1b – (overlapped) Because it was as if I was the person assembling the model but couldn't touch it
2b – (lateral) It was just harder to give directions when it came to explaining which parts were to face where
3b – (overlapped) Was probably the most efficient

**Participant: 17a**
**Role: Worker**
Questions:
1b – (lateral) They were all about the same, but if (?) with the hands on the right it was like my own right hand. It also did not create too much shadow
2b – (f-2-f) Because it came from an opposite direction, and was distracting
3b – (lateral) for the reasons in 1b

**Participant: 17b**
**Role: Helper**
Questions:
1b – (overlapped) It is easier to point and find pieces. And maybe cos the set is easier, easier to distinguish the pieces
2b – (lateral) It was hard to find the pieces
3b – (overlapped) It is easier to find the pieces and point to them

**Participant: 18a**
**Role: Worker**
Questions:
1b – (lateral or f-2-f) B (overlapped): obstructed view of the other person's hands
2b – (overlapped) Obstructed view of other person's hands i.e. couldn't always see their hands
3b – (lateral or f-2-f) So that you have a clear view of the other person's hands

**Participant: 18b**
**Role: Helper**
Questions:
1b – (f-2-f) The hands overlapped the least. This preference is only very slight
2b – Not sure
3b – (f-2-f) The hands did not overlap so much

**Appendix 5.4 Evaluation questionnaire for end of trial 1 Experiment 4**

(NB, questionnaire contents remained the same, with minor alteration to title for all conditions of the study e.g. unmediated hands, hands and sketch, sketch only)

Dave Kirk, MRL, Exp_4,5,6 Spring/Summer '05

**Remote Gesturing Questionnaire -**
Projection Vs Video Window (unmediated hands)

First trial

1) Were you the Worker or the Helper? (please tick the box)

Worker (assembling the pieces)
Helper (giving the instructions)

2) How easy or hard was the task to complete? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Very
Hard                    Neither                              Very
                                                           Easy

3) How did you find communicating in this way? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Very
Easy                    Neither                              Very
                                                           Difficult

4) How do you feel you did in the task? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Very
Badly                   Neither                              Very
                                                           Well

5) Did you understand what your partner was saying? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Yes –                   Half of                              No –
always                  the                                  never
                        time

6) What problems did you encounter (if any)? (please write, continue overleaf if necessary)

_____
_____
_____
_____
_____
_____
_____
_____
_____

alpha

**Appendix 5.5 Evaluation questionnaire for end of trial 2 Experiment 4**

(NB, questionnaire contents remained the same for all conditions of the study)

Dave Kirk, MRL, Exp_4,5,6 Spring/Summer '05

<u>Second trial</u>

7) Were you the Worker or the Helper? (please tick the box)

Worker (assembling the pieces)
Helper (giving the instructions)

8) How easy or hard was the task to complete? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Very
Hard         Neither                Very
Easy

9) How did you find communicating in this way? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Very
Easy         Neither                Very
Difficult

10) How do you feel you did in the task? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Very
Badly        Neither                Very
Well

11) Did you understand what your partner was saying? (please circle a number)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Yes –
always        Half of
the
time             No –
never

12) What problems did you encounter (if any)? (please write, continue overleaf if necessary)

_____
_____
_____
_____
_____
_____
_____
_____

<u>and finally</u>

13) Which did your prefer, gestures projected or presented on a TV? (please circle)

Projected         Neither / Don't know         Presented on TV

alpha

**Appendix 5.6 Participant information sheet experiment 4 (Hands only)**

Projected vs. external viewed Hands

**Lego study – Remote gesturing (with hands)**
**Participant Information Sheet**

You will be working in a pair for this experiment. Prior to starting the experimental trials you will be given the chance to try-out the remote gesturing technology so that you understand what it does and how it works (see the schematic below for an illustration of how it works). Communication during the experiment is by normal speech and by using the remote gesture device.

The experiment has two trials, each lasting 10 minutes, during each trial you will have a different Lego model to assemble.

For the experiment one of you will be the helper and one of you will be the worker. The helper is the person who has the instructions for the Lego kit, they tell the worker what to do. The worker is the only person allowed to touch the Lego pieces, but they are not allowed to look at the instructions or know what it is that they are building.

For each trial the helper and the worker sit at their separate desks, the helper is able to see what the worker is doing because of the video link between the desks, the helper can stick their hands out in front of them, this is picked up by a video camera, which is linked to either a projector or another TV.

As mentioned above, in each trial you have 10mins, in this time you must complete as much of the Lego model as you can. What changes between each trial is that in one trial the image of the Helper's hands is projected over the surface of the workers' desk, in the other trial the worker can see the Helper's hands on a TV monitor on their desk.

After all of the trials each member of the pair must complete a brief evaluation questionnaire. You will be video-recorded during the experiment.
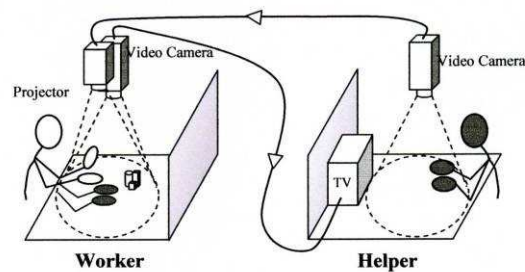
Projected Hands



Externally represented hands
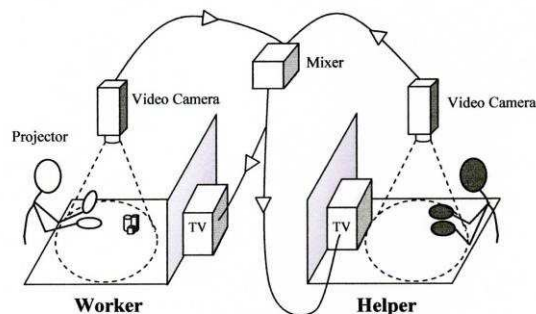
**Appendix 5.7 Participant information sheet experiment 4 (Hands & Sketches)**

Projected vs. external viewed Hands & Sketches

**Lego study – Remote gesturing (with hands & sketching)**
**Participant Information Sheet**

You will be working in a pair for this experiment. Before starting the experiment you will be given the chance to try-out the remote gesturing technology so that you understand what it does and how it works (see the images below for an illustration of how it works). Communication during the experiment is by normal speech and by using the remote gesture device.

The experiment has two trials, each lasting 10 minutes, during each trial you will have a different Lego model to assemble.

One of you will be the helper and one of you will be the worker. The helper is the person who has the instructions for the Lego kit, they tell the worker what to do. The worker is the only person allowed to touch the Lego pieces, but they are not allowed to look at the instructions.

For each trial the helper and the worker sit at their separate desks, the helper is able to see what the worker is doing because of the video link between the desks, the helper can stick their hands out in front of them to make gestures and also (using the marker pen provided) write/draw on the whiteboard on the desk in front of them, these actions are picked up by a video camera, which is linked to either a projector or another TV.

As mentioned above, in each trial you have 10mins, in this time you must complete as much of the Lego model as you can. What changes between each trial is that in one trial the image of the Helper's hands and sketches is projected over the surface of the workers' desk, in the other trial the worker can see the Helper's hands and sketches on a TV monitor on their desk.

After all of the trials each member of the pair must complete a brief evaluation questionnaire. You will be video-recorded during the experiment.

Projected Hands & Sketch
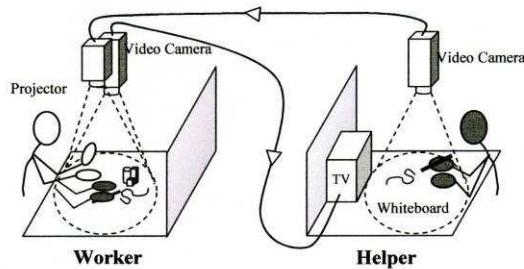


Externally represented hands & sketch
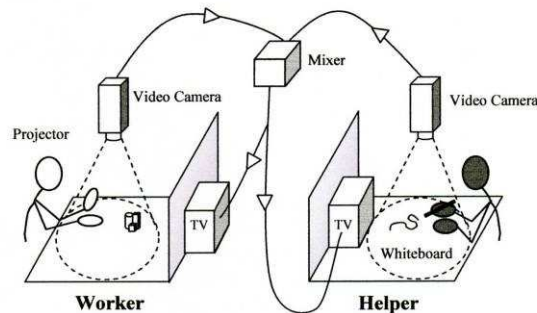
## Appendix 5.8 Participant information sheet experiment 4 (Sketching only)

Projected vs. external viewed Sketches

**Lego study – Remote gesturing (sketching only)**
**Participant Information Sheet**

You will be working in a pair for this experiment. Prior to starting the two experiment trials you will be given the chance to try-out the remote gesturing technology so that you understand what it does and how it works (see the schematic below for an illustration of how it works). Communication during the experiment is by normal speech and by using the remote gesture device.

The experiment has two trials, each lasting 10 minutes, during each trial you will have a different Lego model to assemble.

For the experiment one of you will be the helper and one of you will be the worker. The helper is the person who has the instructions for the Lego kit, they tell the worker what to do. The worker is the only person allowed to touch the Lego pieces, but they are not allowed to look at the instructions or know what it is that they are building.

For each trial the helper and the worker sit at their separate desks, the helper is able to see what the worker is doing because of the video link between the desks, the helper (using the Wacom tablet and pen provided) can write/draw on the drawing tablet in front of them, these actions can be seen on the TV in front of them and are either projected onto the worker's desk or shown to them via a TV of their own.

As mentioned above, in each trial you have 10mins, in this time you must complete as much of the Lego model as you can. What changes between each trial is that in one trial the image of the Helper's sketches is projected over the surface of the workers' desk, in the other trial the worker can see the Helper's sketches on a TV monitor on their desk.

After all of the trials each member of the pair must complete a brief evaluation questionnaire. You will be video-recorded during the experiment.

Projected Sketching
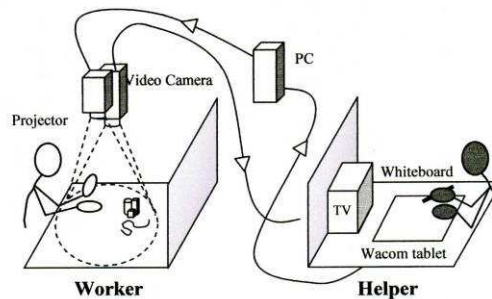


Externally represented Sketching
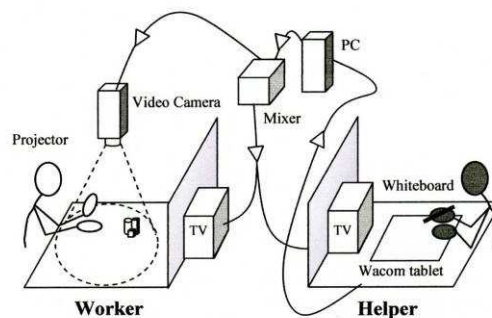
**Appendix 8.1 Pros and cons of system design choices and examples of use in current systems**



Figure 8.1 Possible system design alternatives for remote gesture tools

A) **Choices for Displaying Remote Gestures**

**A1) Display Type:** *Projection*

**Example:** *The remote gestures are generated and then directly projected into/onto the working task-space of the remote Worker.*

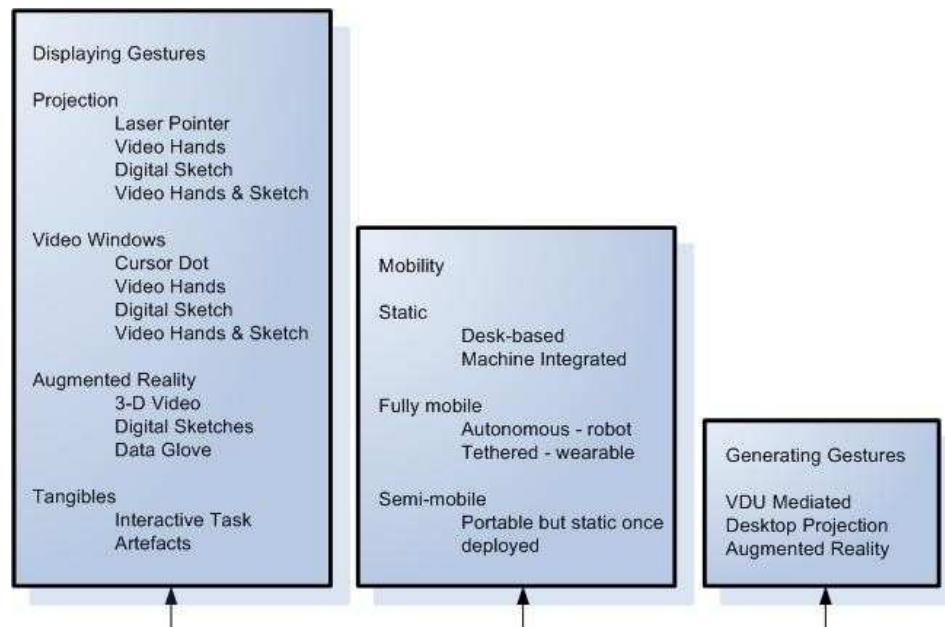**General Benefits and Limitations:** Enables the gestures to be linked directly to the task artefacts at the site of manipulation, keeping views of gesture and action synchronous. Some mild support for a user preference for this approach (in comparison with video windows) has been observed. Relatively easy to construct / prototype such a system. Supports a wide variety of different gesture representations. Is better suited to indoor environments with controlled conditions. Will encounter difficulties if used in bright natural daylight. Projection surface must be flat and plain coloured to enable gestures to be easily viewed. A projection system that does not deform the image when it is projected onto an uneven surface would be of extreme benefit but such systems are currently at an early stage of development (see Bimber and Raskar 2005 for overview). Ever-present projection can clutter or obscure items in the task-space. Projection equipment can often be temperamental and produces significant heat output. Size of digital video projectors is currently still relatively large. Needs further work to ensure that collaborators have mutual task-space awareness.

**Possible Gesture Representations:**

**A1.1) Laser Pointer**

**Comparable Existing System:** GestureMan  (Kuzuoka et al 2000)

**Pros and Cons:** Facilitates deixis but otherwise is a very poor medium for expressing gestural communication. Complex surfaces of projection area may obscure such a small projected image making coordination of gestural action and Worker attention difficult.

**A1.2) Video Hands**

**Comparable Existing System:** Agora (Luff et al 2006) *

**Pros and Cons:** Proven to improve collaborative performance, but potentially difficult to integrate with digital content. Gestures presented are highly natural and usually therefore easy to interpret. The gestures can however be obscured by the hand actions of the Worker, although systematic practices for negotiating access to the space are developed. Results of experimentation demonstrated that more formalised turn-taking such as this, can improve performance in collaborative physical tasks. Multiple levels of gesturing behaviour can be achieved. Whilst obviously not 3-Dimensional (given the 2-Dimensional capture and presentation) there is in essence a 2 ½ - Dimensional image presented allowing some instruction in 3-Dimensions.

**A1.3) Digital Sketch**

**Comparable Existing System:** N/A*

**Pros and Cons:** Whilst this offers less use of embodied gesturing than unmediated representations of hands, highly complex gestural objects can be created. These gestural objects can incorporate multiple levels of information but are generally restricted to only one point source of information being used at any given moment, can therefore be difficult to represent dynamic information. Sketched information is also resolutely 2-Dimensional and providing sketched gestures to represent actions in 3-Dimensional planes is extremely difficult. Sketched information can be directly annotated onto artefacts in the physical environment, but moving the artefacts renders the extant sketches useless.

**A1.4) Video Hands and Sketch**

**Comparable Existing System:** N/A*

**Pros and Cons:** Although it was believed that such an approach would dramatically increase the utility of both digital sketch only and unmediated hand only approaches it became clear that in most instances the use of the pen limited the use of two-handed gesturing. Whilst a hand held pen with a fine tip can be used for fine-grained pointing actions, more complex gestural actions are often inhibited. Sketched objects are particularly easy to create and as such too much time may be spent constructing elaborate sketched objects, so despite the site of gestural action being tied to the location of artefact manipulation in many instances and

additional area of focus was inserted into the task-space as a new 'sketching' zone was created at a remove from the artefact manipulation space.

**A2) Display Type:** *Video Window*

**Example:** *The remote Worker has their task-space and an additional external VDU which runs a live video feed of the task space which is annotated by having the Expert's gestures overlaid.*

**General Benefits and Limitations:** Keeps actual task-space uncluttered. Relatively simple to construct / prototype. Can be usable in a variety of conditions, less likely than projection systems to be affected by fluctuations in lighting levels or produce such heat output. Easily supports a variety of gesture representations. Requires a Worker to have access to a suitable VDU, which might limit mobility (although using laptop or tablet PCs might minimise this problem). Divorces the site of gesture representation from site of artefact manipulation, causing a potential fracture in interaction which potentially requires significant additional work to overcome (both *cognitive* work for the Worker in extrapolating gestures oriented to the video feed view of the task-space to their own perspective and *collaborative* work through the verbal channel as increased back-channelling to confirm understanding may be required). Does however provide good implicit feedback to the Worker of the bounds and limitations of the Expert's view of the task-space, which can be essential for interpreting instructions.

**Possible Gesture Representations:**

**A2.1) Cursor Dot**

**Comparable Existing System:** Prototyped in Fussell et al 2004

**Pros and Cons:** Demonstrably shown to have low utility, whilst this probably has more presence in the gesturing space than its projected counterpart such a low bandwidth communication method for expressing gestural content has particular difficulty in supporting interactions in a 3-Dimensional task such as a collaborative physical task.

**A2.2) Video Hands**

**Comparable Existing System:** N/A*

**Pros and Cons:** This approach essentially shares all of the benefits and problems of its projected counterpart.

**A2.3) Digital Sketch**

**Comparable Existing System:** Drawing Over Video Environment - DOVE (Ou et al 2003a) *

**Pros and Cons:** With this system the problems encountered in having a separate video window (namely the potential fracturing of the interaction) are compounded by the abstract nature of the gestural representation and the natural difficulties of such a representation to

adequately represent 3-Dimensional information. Information sketched over a view of the task space that is not consistent to one's own view might be particularly hard to reconcile, requiring further cognitive effort to extrapolate from the gestural input to the actual working task-space.

### A2.4) Video Hands and Sketch

**Comparable Existing System:** N/A*

**Pros and Cons:** Again this approach is not qualitatively different in terms of costs and benefits from its projected counterpart.

### A3) Display Type: *Augmented Reality*

**Example:** *The Worker wears a head mounted display system incorporating 'see-through-lenses' so that they can see their task-space but additional digital (possibly video) information can be inserted into their natural view.*

**General Benefits and Limitations:** Avoids many of the disadvantages of both the projection and video window systems and incorporates many of the advantages of the projection systems such as keeping site of gestural action and artefact manipulation consistent. Can be used in highly mobile contexts and will be significantly less affected by environmental constraints such as fluctuating lighting conditions. Is however significantly more demanding technically to establish, as gestural information must be securely anchored into a 3-Dimensional context (which may require careful deployment of fixed markers in the remote task-space). Once established however this approach would provide a high level of fidelity for gesturing in 3-Dimensions, much more so than either projection or video window systems (which is clearly of use in most collaborative physical tasks, given their inherent physicality). Facilitates several different forms of gestural representation and can easily incorporate the inclusion of further digital information / resources.

**Possible Gesture Representations:**

### A3.1) 3D Video

**Comparable Existing System:** N/A

**Pros and Cons:** Some form of 3-Dimensional video view would obviously facilitate the use of naturally occurring forms of gestural activity. The capture of such video information would obviously require a much more significant investment in technology at the Expert's end of the communication device (presumably involving video capture of the Expert's arms from multiple perspectives in a blue screen environment, which are then mapped onto existing 3-D models of arms). Bandwidth required for transmission of such data would presumably also be quite extensive necessitating particularly established communication infrastructure.

**A3.2) Digital Sketches**

**Comparable Existing System:** N/A

**Pros and Cons:** Whilst digital sketches would be a lower bandwidth way of communicating it is apparent that it would potentially be difficult to present such sketches in a 3-Dimensional environment. The most realistic approach would allow 2-D sketches to be transmitted which would be held statically in a plane that the Worker could see relatively easily. Obviously however, if they were to move, their view of these 2-D sketches might become significantly obscured. For example if the Worker moved to face the task-space from an alternate 90° degree angle their flat sketches would essentially become invisible unless additional 3-D markers were added to notify the location of the sketched diagram (obviously sketches themselves once made, could not be moved, as this would mean that they lose their relevance to the objects that they were appended to, as per the Expert's view of the task space).

**A3.3) Data Glove / Virtual Hands**

**Comparable Existing System:** N/A

**Pros and Cons:** Perhaps the simplest method for representing remote gestures in an augmented reality environment is to provide the remote expert with a pair of data gloves. Whilst clearly not as high fidelity as an unmediated representation of a hand it would confer many of the benefits of such an approach. And could comfortably be combined with some sketching or other digital facilities. It would clearly require less computational effort than the video arm capture, and a variety of existing hardware (data glove) devices could be utilised, rather than requiring the establishment of bespoke technologies.

**A4) Display Type:** *Tangible Bits*

**Example:** *There is no direct representation of gesturing per se but there would be task artefacts that can be remotely manipulated in some fashion by the Expert to help guide / focus the attention of the remote Worker. Working systems incorporating this approach are likely to be machine repair scenarios and would work by having bits of a machine that is being repaired illuminate themselves as the Expert wishes to draw attention to them.*

**Pros and Cons:** A key problem with this approach is that it does not offer fine-grained gesture support. There is a relative paucity of information content of the conceptual gestures, they are merely a form of remote deixis, as such more elaborate orientational and artefact manipulation gestures would not be supported. Studies have demonstrated that such limited gesture support is often insufficient to improve performance in collaborative tasks (e.g. Fussell et al 2004). The technology also only works in highly situated contexts (i.e. it would not be a generic device that could be deployed in various contexts, each application is therefore specifically tailored. There are also clear issues around the concept of granularity. A decision would need to be made about how small an individual constituent item could be before it is deemed inappropriate to have functionality added to it such that it can be manipulated (made to

illuminate or vibrate perhaps). Despite this however such a system might be easy to install and would be very useful for supporting interactions where items are difficult to describe without expert knowledge. It will also be easy to incorporate the gestural action with other digital information resources as it is tied to the site of action which will inevitably be a networked machine, therefore if screen space is incorporated gestural actions can be used in conjunction with other collaborative instructional resources such as animations.

**Possible Gesture Representations:**

**A4.1) Interactive Task Artefacts**

**Comparable Existing System:** N/A

**Pros and Cons:** Artefacts could be remotely manipulated possibly in different modalities (e.g. illumination, auditory or tactile signals) but no actual alternative gestural representations available (see Pros and Cons (*Tangible Bits*))

**B) Choices for Mobility of Remote Gesture Tools**

**B1) System Mobility:** *Static*

**Example:** *The most common set-up for a static installation would be a desktop system for regular interaction. Maybe the system would include a desk with the apparatus of communication attached to it (i.e. projectors above or VDUs next to it), and the physical task artefacts that were to be examined would be of the category that they could be brought to the desk for discussion and sharing. Alternatively if one considers the example of tangible bits in table 9.1, it is easy to imagine that a static system could be integrally woven into the fabric of an existing machine, for example car bonnets that have camera and projection facilities attached to their underside such that the bonnet can be lifted and the device automatically provides both visual and gestural access to the engine for a remote expert.*

**Pros and Cons (*Static*)** The benefits of a static system are that they are generally easier to construct. Mobility brings with it a variety of problems about how to keep task perspectives consistent and more importantly how to keep gestural actions securely anchored to their target artefacts, a static system is not worried by these concerns as the perspectives on the task space do not change and the projection of gestural action and the capture of visual access remain tethered. Obviously the problem with this approach is that each system that is built must be largely custom made for a specific form of interaction such as the machine repair discussed above using tangible bits as the gesturing medium. Desktop systems could be built that integrate into office environments, but given the nature of collaborative physical tasks and the guidelines for application summarised previously, it seems unlikely that there are many applications in such environments that would be suitable for remote gesture support. This

would mean that the investment in space and financing for establishing such installations is unlikely to be made.

**B2) System Mobility:** *Fully mobile*

**Example:** *Fully mobile systems largely occur in one of two differing classes of device. The system can be fully autonomous in the workspace, such as a human proxy robot (like the GestureMan, Kuzuoka et al 2003) which is controlled and used as a virtual embodiment by the remote expert, or the system can be tethered to the Worker as a set of wearable devices which provide visual and gestural access to the remote working space for the Expert (as per Wearable Active Camera/Laser, Sakata et al 2004).*

**Pros and Cons (***Fully Mobile***)** The fully mobile systems again would require a large investment both financial and technological. GestureMan for example utilises a human proxy robot, the technology for which would not make necessarily make it a commercially viable product. There a re also many issues still remaining about the value of using human proxy robots in interactions when other simpler technology may be just as effective. In the situations depicted as sensible targets for deployment of such technologies the emergency nature of many of the working environments might seriously preclude the use of a cumbersome robot device. Those fully mobile systems which rely on being anchored to the body might offer a more realistic alternative, although as previously discussed it is unlikely that projection systems could be used in this way owing to the difficulties of projecting stably from a moving person. And equally systems such as DOVE which might at first appear as though they could be made fully mobile by moving the video window to an HMD would encounter significant difficulties. They could not be fully mobile as the camera view would have to be static, otherwise the gestural actions would lose their links to their target artefacts, and an HMD would still be divorcing the gestural action from the site of artefact manipulation. To avoid these problems the only sensible fully mobile system must surely utilise Augmented Reality gesturing. This would alleviate many of the problems highlighted above, but would require a much more significant investment in technology at the site of the remote Expert. It would however, enable gestures to be firmly posited into the Worker's task space in a relatively unobtrusive non-distracting fashion, that would enable the maximum mobility for the system.

**B3) System Mobility:** *Semi-mobile*

**Example:** *A semi-mobile mobile system would retain many of the features of the static system, but would be constructed of materials such that it could be moved and deployed with relative ease. Unlike fully mobile systems it would not be able to either move independently or require being attached to the Worker. Systems that incorporate closely aligned small digital projectors and video cameras, tethered to PC support with wireless internet links, that are posable such that they can be oriented to different working surfaces would be of particular use in a variety of contexts.*

**Pros and Cons (*Semi-mobile*)** A semi-mobile system would adequately support many of the potential application scenarios previously discussed. To utilise projection technology however, it would require significant power resources (large battery packs) and the difficulties of projecting into various lighting conditions and over deformed surfaces would either have to be ignored and worked around, or technology would need to be improved to directly alleviate these difficulties. If this could be achieved (ignoring the complexities of the advanced projection abilities) the technology involved would remain relatively simple. With a consistently held video capture and gesture projection the advantages of the Static system could be utilised and complex Augmented Reality techniques would not be required to ensure that gestures remain securely anchored to target artefacts. Whereas an Augmented reality system in a fully mobile device could of course provide a rich level of 3-D gesturing the semi-mobile system would likely be restricted to the 2 ½ -D representation of gestures provided by video projection. This level of detail has however been demonstrably shown to be of sufficient fidelity to improve collaborative action at various points throughout this thesis.


## C) Choices for Gesture Generation Environments (Technologies)


### C1) Gesture Generation Environment: *VDU mediated*

**Example:** *As in the low-tech prototype presented herein, gestures are created by the Expert as they watch what is being fed to them through a live video feed. They can see their own hands in relation to the task artefacts only by viewing the separate video feed, and guide their gestural actions accordingly (they can essentially see their own hands in their own task space and a separate video representation of their hands in conjunction with the video representation of the remote workspace).*

**Pros and Cons (*VDU mediated*)** The benefits of this approach have been variously discussed at different points within the thesis. To briefly reiterate the main benefit of such an approach is that it is relatively simple to construct yet supports a fundamental aspect of the mixed ecologies approach in that it supports mutual awareness of artefact focussed actions. In most scenarios given in table 9.1, in particular the projection and video window conditions it is clear that the resultant gestural representations will be delivered in a 2-D format. It is of observable benefit to the signaller when constructing their gestures to be aware of exactly how their gestures will appear to the receiver. Possession of this ability allows the signaller to refine their actions such that they are more suited to the medium of presentation. For example, issues which may arise because of the lack of 3-Dimensionality of the resultant output may be immediately obvious to the Expert as they view how their hands appear on the VDU monitor. This does however require that the Expert guides their actions whilst viewing a separate monitor, this is obviously not particularly natural behaviour, which may incur some performance deficit. It is however, worth noting that complex keyhole surgery is now

conducted using exactly this process of guiding one's actions in relation to feedback via a video monitor.

**C2) Gesture Generation Environment:** *Desktop projection/capture*

**Example:** *The task-space of the Worker is video captured and projected directly onto the Expert's desktop (or wall surface as appropriate), and the Expert guides their gestural actions in relation to the 2-D projection in their environment (the expert can essentially see their own 3-D, real world hands (or other gesturing device) on top of an actual 2-D projection of the remote task-space).*

**Pros and Cons (*Desktop projection*)** Whilst this system would produce relatively naturally occurring gestures similar to that which would be found in a normal face-to-face desktop interaction there are certain problems with the approach. In relation to the point concerning VDU mediated gesture generation above there is a potential fracture of the interaction that would be caused because of the distinct difference between the gesture input modality and the gesture output modality. Gesturing over a 2-D representation of a task space with 3-D movements which will then be translated into 2-D gestures projected over 3-D objects ensures that as a signaller it will be very difficult to accurately interpret how your gestures will be perceived. Minimising the numbers of these translations of dimensionality is clearly preferable. Some problems may also occur with such a system whereby the Expert's hands or other devices used to gesture will obscure the projection of the images, although this problem is easily resolved by using rear projection surfaces for gesturing over. The technical demands of this approach are increased however as capturing data (such as video feeds of gestural action) from above a projection of the remote task-space will lead to the inevitable problems of video feedback loops. This issue however, is not insurmountable, and has previously been resolved using spliced video feeds (e.g. Agora, Luff et al 2006).

**C3) Gesture Generation Environment:** *Augmented Reality*

**Example:** *The Expert wears a head-mounted display. In this display they see a live video image of the remote task-space (occupying the largest area possible). The Expert does not have direct visual access to their unmediated gestural actions. They can only see their arms (or potentially their virtual reality hands in the case of a data glove installation) as they would be perceived by the remote Worker.*

**Pros and Cons (*Augmented Reality*)** The immediate drawback of generating gestures in an augmented 3-D world application is that one would not necessarily be allowed to see one's own arms, which might cause difficulties. The context view that the Expert looks at would need to be provided via fully enclosed HMD. Gestural information from the arms would then be captured and this could be attached to the video feed, so the Expert would see the resultant composite image. Depending on how this is done there is potential for problems with delays in visual feedback of the arm movements, but faster networked connections would reduce these problems. Such a system does confer the large benefit that the Expert would largely be able to

view the gestures as they are being made exactly as they will be perceived by the Worker receiving them. And, depending on how the video capture is achieved at the remote site participants could be relatively sure of achieving mutual awareness of general task perspectives. Such a system would however require that the Expert be encumbered with significant items of technology and their movement be somewhat restricted, which might limit their productivity and obviously negates the seamless movement between working spaces (i.e. the personal and the public) that some authors have argued for (Ishii, 1994). However given the likely applications for the technology and given that it is largely based on a model of the provision of expert support in emergency situations it is perhaps not unreasonable to expect the total engagement in the task of the expert, to the exclusion of other tasks. And such an approach is consistent with the mixed ecologies approach of constructing new shared equally engaged in working spaces as opposed to mere links between disparate ecologies.