T. Koschmann, D. Suthers, & T.W. Chan (Eds.), *Computer Supported Collaborative Learning 2005: The Next 10 Years!*
Mahwah, NJ: Lawrence Erlbaum Associates.

The Effects of Remote Gesturing on Distance Instruction

David S. Kirk
Computer Science & IT
University of Nottingham
dsk@cs.nott.ac.uk

Danaë Stanton Fraser
Department of Psychology
University of Bath
D.StantonFraser@bath.ac.uk

Abstract. In this paper, we describe an experimental research study, investigating the impact of varying communication media on the quality of learning. Our study investigates remote instruction using an object assembly task. The between-subjects independent-measures study compared instruction via audio only, with instruction via a remote gesturing system. Measures included assembly speed and assembly accuracy and were recorded during instruction and post-instruction at 10min and 24hr intervals. Perceived Instructor presence and other interpersonal variables were assessed via questionnaire. Results showed that remote gesturing during instruction led to significantly faster self-assembly 24hrs post instruction ($t\ (13) = 1.73$, $p \leq 0.05$). Whilst the use of gesture reportedly reduces communicative rapport, we conclude that gesture-based remote instruction improves the overall efficiency of remote collaboration.

Keywords: expert, novice, remote gesturing, remote instruction, collaborative physical tasks

INTRODUCTION

Learning is often characterized in terms of the relationship between Instructor and Learner with the Instructor either passing on knowledge or creating an environment for the Learner that is rich for self-discovery (e.g. Vygotsky's zone of proximal development, Vygotsky, 1978; Bransford et al., 2000). During instruction the Instructor must be able to define the limits of understanding of the Learner, they must successfully pass on knowledge and they must be able to competently assess that the Learner has understood (Tharpe & Gallimore, 1988). All of this is an interactive process based on communication feedback loops. Indeed, Garfinkel (1967) and Sacks (1992) stress the social construction of meaning during dyadic interactions. Similarly, research within the CSCL community has highlighted the importance of dyadic communication for remote interactions, where the Instructor may not be co-present with the Learner. As Stahl (2002, 2004) indicates, CSCL communication takes place primarily though discourse with communication breakdowns being resolved through the process of the discourse. The work of Garfinkel (1967) and Sacks (1992) also stressed the ways in which interactive factors other than speech (such as non-verbal behaviour) also help to construct meaning.

One factor in particular, that is important for the extraction of meaning from an interaction is the expression of gesture (McNeill, 1992; Clark, 1996; Kendon, 1996). Indeed, research has demonstrated that the adequate expression of gesture can be critical for establishing conversational grounding (Fussell et al., 2004), especially in those elements of discourse, which have a strong spatial reference (Rauscher et al., 1996). Inherently without gesture representation in a remote interaction much of the discourse becomes an attempt to secure conversational grounding (Kraut et al., 1996), e.g. the talk becomes 'about the talk'.

In this paper we are particularly interested in the ways in which gesture can be supported in remote learning environments and examining the learning effects of remote gesturing techniques. Our interest is in tasks that have been characterized as 'remote help giving' (Tolmie et al., 2004) where one of the collaborators has the task knowledge and one of the collaborators manipulates the task artefacts. Kraut et al. (2003) state that such tasks:

> "…fall 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 about the task." (p.16)

In these situations there is a clear asymmetry between the roles and requirements of the collaborators, and the task clearly resembles a learning or instruction experience. Typical examples of such tasks include remote expert medical assistance, supporting remotely located junior surgical teams or paramedics in the field, or situations in manufacturing, e.g. machine repair or plant maintenance incorporating expert guidance (see Fussell et al., 2004). Whilst an ideal instructional situation might involve co-locating instructor and learner, practical constraints may interfere, such as pressures on time or budget. Remote instruction may overcome such practical constraints.

However, the reduced availability of embodied behavior in remote instruction may seriously degrade the experience of the learner.

Research effort is therefore being expended in the design of technologies to support such remote instruction situations, with an emphasis being placed on the remote representation of non-verbal behaviour, most prominently gesturing (Kato et al., 1997). There are a variety of ways in which this can be achieved, different approaches including human proxy robots (GestureMan; Kuzuoka et al., 2000), direct video-based representations of hands (Agora; Kuzuoka et al., 1999) and video-based sketching (DOVE, Ou et al., 2003). However, when collaborators are not side-by-side they have different perspectives on the task depending on the medium of communication between the remote sites. As a result, they may approach the task with differing levels or types of knowledge. This mismatch of perspectives has been referred to as 'Fractured Ecologies' (Luff et al., 2003) and creates observable problems in collaboration. Each of the systems mentioned above displays this issue in varying degrees.

This paper describes initial experiments in overcoming this fracture in the ecologies of instruction by providing technical arrangements that provide remote gesturing support. We have developed a system with which to explore how a closer alignment between remote ecologies increases the presence of the remote collaborator in the task space. The aim is to understand whether such an increased alignment will give a more useful representation of non-verbal behaviour from instructor to learner. This paper begins by motivating the use of aligned gesture in providing mixed ecologies for remote instruction. We then discuss existing technologies for gesture support. We proceed by describing our system and experiments that investigate the use of remote gesturing. Finally, we discuss how the findings of our experiment support the use of aligned remote gesturing in conducting instruction.

The Emergence of Remote Gesture Technologies

Remote gesture systems emerged from early media space research where experimental studies (Ochsman & Chapanis, 1974; Daly-Jones et al., 1998; Kraut et al., 2003) indicated that merely linking spaces through audio-visual video links does not improve performance to the levels observed between side-by-side collaborators. The importance of gestures in face-to-face collaboration was stressed by Tang (1991) with later studies by Bekker observing that many hand activities in physical workspaces were gestures to express ideas (Bekker et al., 1995). These studies suggested that support for remote gesturing could improve cooperation beyond the capabilities of simple video links and motivated research into a number of remote gesture systems.

Two broad classes of gesture system have emerged. ***linked gesture systems*** directly represent remote gestures within the local environment while ***mediated gesture systems*** use an artificial representation of remote gestures. Linked gesture systems have emerged from efforts to study remote collaborative design work using video connections (Tang, 1991) and led to the development of several technologies such as VideoDraw (Tang & Minneman 1990), VideoWhiteboard (Tang & Minneman, 1990) and Clearboard (Ishii & Kobayashi, 1992). These systems exploit video projection techniques to support collaboration around the construction of shared 2-D artefacts such as drawings. Mediated gesture systems are more diverse. Early systems such as Commune (Bly & Minneman, 1990) used sketching to remotely gesture around shared digital artefacts and a range of systems have emerged that use a visible embodiment such as a telepointer to convey gestures (Gutwin & Penner, 2002). More recently mediated gesture systems have focused on how gestures may be manifest in the real world and support the physical manipulation of 3D objects.

Systems such as Drawing Over Video Environment (DOVE) (Ou et al., 2003) allow an Instructor's remote gestures to be fed to a local Worker. Gestural sketches are overlaid on a video representation of the working area presented via a monitor in the local task space. The work of Kuzuoka et al.. in the development of GestureCam, GestureCar and GestureMan (Kuzuoka et al., 2000) has focused on directly embedding remote gestures into a working environment through the use of a laser pointer. However, the laser pointer obviously has a lower bandwidth for the expression of gestural information than the direct presentation of hand gestures or sketches.

Realizing remote gesture systems has not been without its difficulties. A particular concern has been the extent to which 'Fractured Ecologies' (Luff et al., 2003) have emerged where the remote and the local ecologies are too distinct, creating a barrier to understanding and conversational grounding. This is most prominent in the mediated gesture systems concerned with collaborative physical tasks. For example, within the GestureMan system local workers could not assess the situational awareness of the remote instructors as they were not aware of what the experts could see (Luff et al., 2003). While in the DOVE system (Ou et al., 2003) the local worker needs to extrapolate from the overlaid sketched information from the remote helper presented on a separate video monitor to their own local ecology.

Gestures in Instruction

As researchers have developed these various technologies to support remote gesturing it has become necessary to find ways of isolating improvements in the quality of interaction. One common methodology used to demonstrate the success of the technology has been to provide evidence for immediate performance benefits. Experiments are constructed which demonstrate whether a particular remote gesturing device improves performance speed in a standardized collaborative physical task (e.g. Fussell et al., 2004). However, the use of such metrics circumvents the inspection of particular applications. The dynamics of situations such as those where an instructor guides a learner through some physical process in the hope of successfully imparting knowledge, require further investigation. Specifically, experimental approaches to understanding remote gesturing systems have failed to consider the impact of such devices on learning. By focusing solely on the immediate task performance benefits rather than any assessment of longer-term knowledge development the research literature rarely discusses whether the newly developed remote gesturing techniques actually provide benefits for remote learning, which cannot be replicated by current methods of remote help giving (such as telephones or videoconferencing).

We would argue that successful learning-oriented interaction depends on the access for both instructor and learner. The use of a system to provide a remote worker access to an instructor's non-verbal behaviour (such as gesture) should improve the quality of learning that is achieved during the interaction. Remote gesture should facilitate conversational grounding (Fussell et al., 2004) meaning that less time in a time-limited interaction is given over to 'talking about the talk' and more time can be spent discussing salient learning features. Whilst there might be something to be gained from extending discussion during a learning interaction, there are often clear economic constraints for this class of remote instruction situations, which necessitate that learning should be expedited. We would anticipate that the facilitation of gesture, which normally occurs as either a component of utterances in alternation with speech or in conjunction with speech (Kendon, 1996) should improve understanding in collaborative physical tasks, especially given that discourse must relate to spatial concepts (Rauscher et al., 1996). Equally, in situations where a learner attempts to perform the task at a later time on his or her own, they might be able to recognize hand shapes and gestures that they are performing, which would prompt instruction recognition. This hypothesis is reinforced in experiences with the use of previous remote gesturing systems (Kirk et al., 2004) which have shown evidence that users will map their hand movements onto the hand movements of instructors demonstrating physical manipulations of task artefacts or indicating locations of interest.

Nonetheless, there exists a counter-argument that might indicate that providing a representation of gesture for remote instruction could impair learning. If one were to consider the 'Agentic' personality role described by Milgram (1974) or indeed theories of automatic processing within work on attention (Shiffrin & Schneider, 1977) it could be argued that with increased physical presence during remote instruction and less interactive discourse, learners might simply perform actions as they are instructed without considering in depth the nature of the task they are performing.

Our technological arrangement

We wish to explore gesturing in remote help giving situations where the technologies seek to minimize the differences between the ecologies of the local Worker and the remote Helper. To effectively embed remote gestures in the local ecology and provide a rich representation of hand gestures we exploit direct video-projection. Figure 1 illustrates the general technological arrangement.

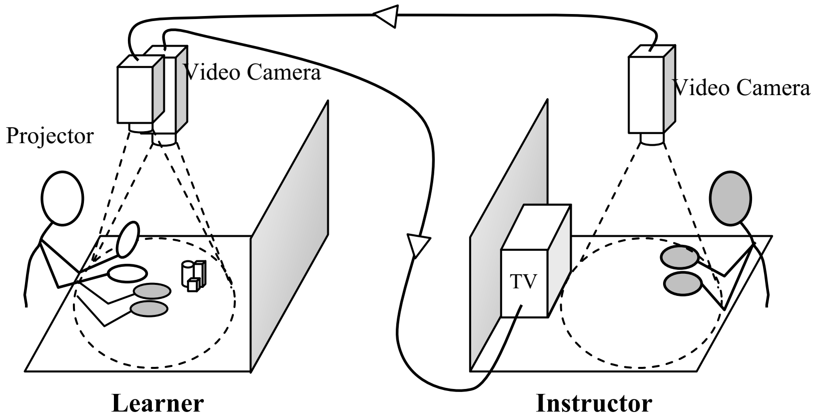

Figure 1. Schematic of Gesture Projection System

A video camera was used to capture images of one collaborator's hands (***the Instructor***); these gestures were then projected onto the desk of the other collaborator (***the Learner***), who had the task artefacts on their desk. These remote gestures were therefore captured and posited directly into a remote ecology, creating a mixed reality surface at the level of the task space. The resulting images played out at the mixed reality surface were also captured by a second video camera and passed back to a TV monitor situated on the desk of the Instructor. This allowed the Instructor to see artefacts in the task space, to see the Learner's progress in assembly and to see their associated gestures and also to guide their own gestures in relation to the shared artefacts. This arrangement exploits two key features to help align the remote and local ecologies:

- *The gestural output from the remote situations is directly embedded in the local environmen*t. Remote gestures are directly projected into the local space. This arrangement extends the approach suggested by systems such as DOVE (Ou et al., 2003) where remote gestures are made available on a separate display.
- *The gestures are un-mediated.* We directly project gestures captured form video camera allowing us to preserve the richness of expression of the remote user's gestures and reduce the costs of interpretation.

The asymmetric nature of the Learner-Instructor dynamic is also reflected in the physical arrangement of the technology. Essentially, our aim here is to encourage the remote Instructor to share the same ecological arrangement as the local Learner. In order to do this we made two design choices in assembling our technologies:

- *The remote Instructor shares the same orientation to the task space as the local Learner* with their gestures projected on top of the local Learner's rather than arranged face to face.
- *The remote Instructor views their gestures on the remote work surface* alongside the artefacts and the gesture of the remote user rather than projecting the work surface into the Instructor's environment.

This arrangement is in contrast to the use of video projections within Agora (Kuzuoka et al., 1999) and VideoArms (Tang et al., 2004), which adopt a face-to-face (or side-by-side) orientation for remote and local participants and more symmetric projections that reflect the more equal collaborative arrangement they seek to support.

STUDYING THE TECHNOLOGY

We have developed the technological arrangement described in the previous section in order to assess its value in supporting remote interactions for collaborative physical tasks involving a strong instructional emphasis. Rather than studying performance effects, therefore, we developed a method of understanding the role of instruction itself in such scenarios. Given the paucity of literature available on learning effects in remote instruction, we chose to study post-instruction performance by asking learners to complete a task on their own after being instructed. Testing post-instruction effects should eliminate the possibility that learners are blindly following instructions without retaining task knowledge in their own right.

Design

The study was conducted using a between-subjects independent-measures design. We employed one independent variable, communication condition, which consisted of two levels, voice-only and voice-plus-gesture. One participant 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 as a distraction task during the 10-minute interval after the instruction period.

Equipment

The gesture projection apparatus (see figure 1 for schematic, figure 2 for illustration of system in use) consisted of two bespoke wooden frames, positioned on a standard non-adjustable working desk. Frame 1 held a digital video camera attached to a boundary microphone and an LCD projector. Frame 2 held a digital video camera only, and incorporated a 14" Television. A Lego™ kit (model no. 8441) was used for the assembly task. Video recordings of the experiment were taken from the video camera on Frame 1 (so as to cover in-depth the mixed reality surface) and an additional video camera was used to give a contextual perspective that recorded participant's behaviour during the post-instruction learning assessment.

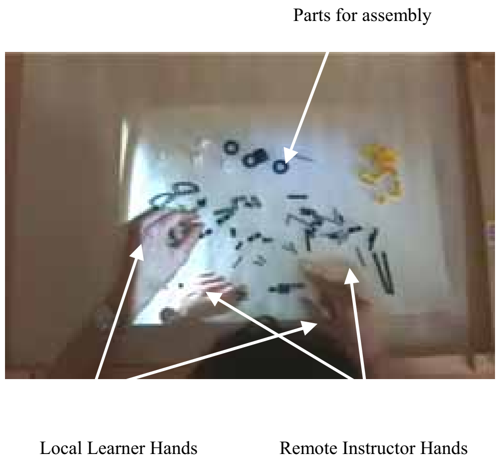

Figure 2. Gesture Projection System in use

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. This participant was not told the hypotheses of the study.

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.

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 2nd self-assembly was also calculated. Responses to the questionnaire items were also analysed.

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.

RESULTS

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

The Time Taken to complete the assembly can be seen in Figure 3 and the pattern of mistakes over the experimental phases is shown in Figure 4.

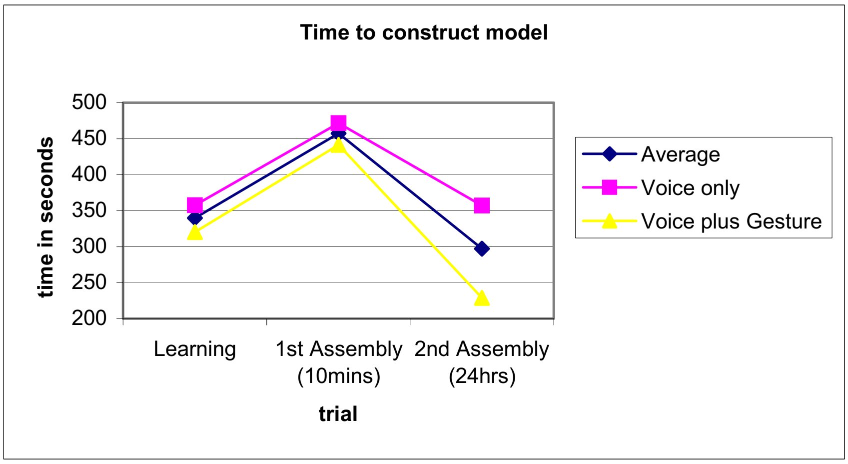

Figure 3. Time to complete model in each of three phases

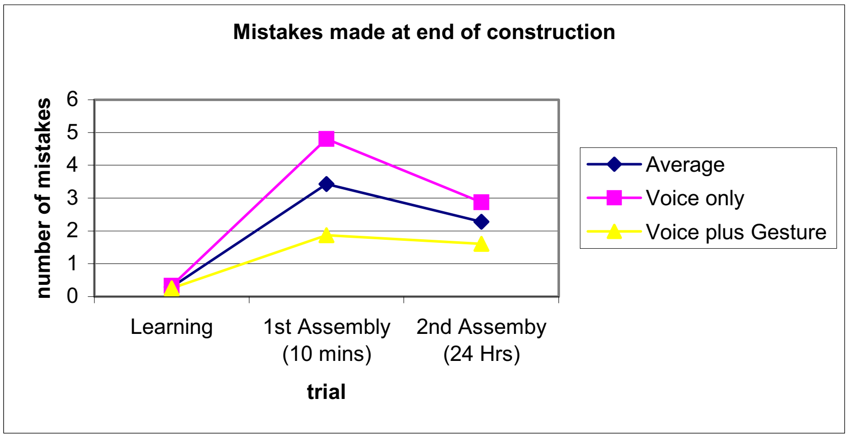

Figure 4. The numbers of mistakes made in each experimental phase

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 ($t(13)=1.73$, $p\leq0.05$). Intriguingly, as demonstrated in Figure 3, 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 2).

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

Table 2. 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 ($t\ (13) =1.83$, $p\leq0.045$). 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. We conclude that remote gesturing during instruction has improved task learning.

Improved performance with a poorer perception of involvement

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.

Two statements (highlighted in figure 6) were found to significantly differ by instruction communication group. Those participants who had experienced instruction utilizing remote gesture actually rated the instructor as slightly less likeable ($t\ (16) =-2.08$, $p\leq0.05$) and simultaneously were actually more likely to agree with the statement "I felt like I just did what I was told to do" ($t\ (16) =2.65$, $p\leq 0.02$), which demonstrates a perceived lack of involvement with the task. 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 rapport emerging during the instruction.

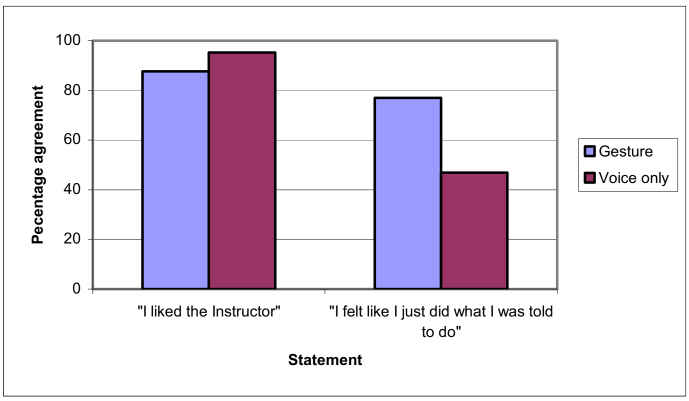

Figure 6. Responses to two statements by Instruction communication group

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 we have shown 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.

DISCUSSION

The aim of this study was to investigate the impact of using a remote gesturing device on the quality of learning achieved during remote instruction. In line with this aim the study has demonstrated that the use of such a device during instruction in a physical task leads to significantly improved speeds of self-performance of the task 24 hours post-instruction. Intriguingly, however, the study has also demonstrated that the relationship between the instructor and the learner is affected by the use of the technology, slightly impairing the ability of the instructor to develop a rapport with the learner. 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 results 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. We would argue 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. We would suggest that our two conditions reflect different levels of translational overhead.

The overhead of "translating" representations

In our voice only case, the instructor can see items in the task space but not point. This means that then they need to translate their visuo-spatial instructions into a verbal code which must be transmitted to the learner 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 instructor and the learner will render the process of decoding talk and then resituating visuo-spatial information within the learner's ecology much harder.

Alternatively, a particularly close alignment of remote and local ecologies such as that used in our experiment provides direct visuo-spatial reference intact. The instructor can make gestural references, which are aligned with the learner's visual perspective on 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 reduces the effort required 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).

Improved effects over time

Our results found no significant difference in times taken for initial instruction between the voice-only and voice-plus-gesture groups. There is a possibility that the similar times for instruction are derived from different types of interaction. It may be that in the gesture condition more time was spent on salient features of the task and less time was spent 'talking about the talk'. Nonetheless, analysis of our data by studying the composition of the talk used in the two conditions would be required to substantiate this claim, and such claims have already been made with regard to the impact of gestural information during instruction (Clark & Krych 2004). Relying on the questionnaire data, results suggest that in the remote gesture condition learners felt more directed and less involved in the task. Perhaps the continual resolution of difficulties in talk in the voice-only condition allows greater immediate reflection on the necessary features of conducting the task. However, the answer to this problem probably lies in a consideration of the nature of recall and recognition memories (Baddeley, 1990). It is possible that the improved performance after 24 hours for those in the voice-plus-gesture condition derives from the ability of the task to trigger memories of the physical and embodied demonstration of task performance available with the gesture instruction. Despite Kendon's (1996) comments that gestures are largely unconscious and most gesturers would be hard pressed to recall exact gestures that they had used in prior moments, there is evidence that gestures do implicitly convey information (see Kendon, 1994, for a review), enriching the learning environment. When the learners have been instructed with the aid of remote gesturing it could be argued that they are receiving visual cueing of their actions as they manipulate the model. This contextual cueing should promote recognition memory (Chun & Jiang, 1998) of the instructions. Certainly we might appropriate distributed cognition to support this idea, given that performance could be enhanced if the cognitive processing of an instruction is performed inherently by its representation.

We have not collected data that might be used to assess the differences in level of understanding of the task between the two groups, so no conclusions can be drawn as to whether those instructed via voice-only better understood the task. However, given the simplicity of the task in this situation, there is very limited capacity for developing a deep understanding and indeed this factor would vary with tasks of an increased complexity. This raises the issue of whether a technology should be designed to facilitate the making of mistakes for learning. Such a complex domain requires many task-dependent metrics to understand how the technology supports the learning involved.

CONCLUSIONS AND FUTURE WORK

In this paper we have explored the use of remote gesturing technologies to support the situated learning involved in remote help giving. We have shown that the use of gesture for remote instruction significantly improves subsequent task performance in the performance of physical tasks over audio-only methods. We have also provided evidence that, whilst performance is improved, learners may actually have poorer perception of the instructor, regarding them as more impersonal. This can lead to a perception of less involvement in the instructed task.

There are limitations to the scope of this study. Firstly we have demonstrated only a simple assembly task, and such results need to be compared with instruction in more complex physical tasks. Equally only one instructor was used and as such gesturing behavior itself was idiosyncratic. Further work is therefore required in understanding the capacity for various instructors to adequately use a remote gesture tool. One final limitation that is of importance is that learners were made aware that they would have to perform the task on their own post-instruction. This may have influenced how well information was retained and the results could vary if subjects were not aware of a later need for the knowledge. This is an especially important point to consider given questionnaire results that indicate participants felt more directed and therefore less engaged in the gesture instruction condition. Such an effect might produce poorer performance in informal *ad hoc* learning situations.

Conversely, we have also provided an indication that tools and technologies for remote instruction may prove beneficial given adequate consideration of the alignments of local and remote ecologies. Systems designers may benefit from our study in understanding how remote instruction systems may be optimized for instruction, but such work requires further results to fully understand the relationships between remote instruction, technological arrangement and learning benefits. Finally, we plan to analyse and consider the basic structure of the remote gesturing apparatus, i.e. the representations of gesture used (unmediated views of the instructor's hands versus video-sketching) and the location of the gestural output relative to the task space (embedded, as in this experiment versus externalized with a video window). These analyses will emphasise the features of our study that we have demonstrated to be of importance for supporting remote instruction. The impact on learning of mixed ecologies both during and after remote instruction must be considered.

ACKNOWLEDGMENTS

We would like to thank Lizzy Gregory for assistance with data collection, and suggestions for analysis, Prof. Tom Rodden for discussing the work and the Equator IRC (EPSRC GR/N15986/01) for funding this work.

REFERENCES

Baddeley, A (1990) *Human Memory: Theory and Practice*. Psychology Press

Bekker , M. M., Olson, J. S., & Olson, G.M. (1995) Analysis of gestures in face to face design teams provides guidance for how to use groupware in design. *In Proc. DIS 1995*, ACM, 157-166

Bly, S.A., & Minneman, S.L. (1990) Commune: a shared drawing surface. *In Proc. Conf. Off. Info. Sys. 1990*, 184-192

Bransford, J. D., Brown, A . L., & Cocking, R. R. (Eds.) (2000) *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press

Chun, M. M. & Jiang, Y. (1998) Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology* **36**, 28–71

Clark, H. H. (1996) *Using Language*. Cambridge, UK: Cambridge University Press

Clark, H. H. & Krych, M. A. (2004) Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, **50** (1), 62-81.

Daly-Jones, O., Monk, A., & Watts, L. (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

Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E. & Kramer, A. D. I. (2004) Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction*. **19** 273-309

Garfinkel, H. (1967) *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall

Gutwin, C., & Penner R. (2002) Visual Information and Collaboration: Improving interpretation of remote gesture with telepointer traces. *In Proc. CSCW 2002*, ACM, 49-57

Hutchins, E (1995) *Cognition in the Wild*. Cambridge, MA: MIT Press

Ishii, H., & Kobayashi, M. (1992) Clearboard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. *In Proc. CHI 1992*. ACM, 525-535

Kato, H.,Yamazaki, K., Suzuki, H., Kuzuoka, H., Miki, H., & Yamazaki, A. (1997) Designing a video-mediated collaboration system based on a body metaphor. *In Proc. CSCL'97*, Kluwer, 142-149

Kendon, A. (1994) Do Gestures Communicate? A Review. *Research on Language and Social Interaction* 27 (3), 175-200.

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

Kirk, D. S., Rodden, T. and Stanton Fraser, D. (2004) Supporting embodied gestures for remote help giving using a mixed reality surface. *Equator Technical Report*. www.equator.ac.uk/xaraya

Kraut, R. E., Fussel, S. R., & Siegel, J. (2003) Visual Information as a Conversational Resource in Collaborative Physical Tasks. *Human-Computer Interaction* **18**, 13-49

Kraut, R. E., Miller, M. D. & Siegel, J. (1996) Collaboration in performance of physical tasks: Effects on outcomes and communication. *In Proc. Of CSCW '96*, ACM, 31-40

Kuzuoka, H., Oyama, S., Suzuki, K., Yamazaki, K., & Mitsuishi, M. (2000) GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions. *In Proc. CSCW'00*, ACM, 155-162

Kuzuoka, H., Yamashita, J., Yamazaki, K., & Yamazaki, A. (1999) Agora: A Remote Collaboration System that Enables Mutual Monitoring. *In Proc. CHI 1999*. ACM, 190-191

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003) Fractured Ecologies: Creating Environments for Collaboration. *Human-Computer Interaction*, **18**, 51–84

McNeill, D. (1992) *Hand and Mind. What gestures reveal about thought*. Chicago: University of Chicago Press

Milgram, S. (1974) *Obedience to Authority*. New York: Harper & Row

Ochsman, R.B. & 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

Ou, J., Fussell, S., Chen, X., Setlock, L.D., & Yang, J. (2003) Gestural Communication over Video Stream: Supporting Multimodal Interaction for Remote Collaborative Physical Tasks. *In Proc. ICMI '03*. ACM, 242-249

Rauscher, F. H., Krauss, R. M. & Chen, Y. (1996) Gesture, Speech and Lexical Access: The Role of Lexical Movements in Speech Production. *Psychological Science*. **7**, (4), 226-231

Sacks, H. (1992) *Lectures on conversation*. Oxford, UK: Blackwell

Shiffrin, R. M. & Schneider, W. (1977) Controlled and automatic human information processing II. Perceptual learning , automatic attending and a general theory. *Psychological Review* **88**, 127-90

Stahl, G. (2002) Contributions to a Theoretical Framework for CSCL. *Proc. of CSCL 2002*, Boulder, Colorado, 62-71

Stahl, G., Herrmann, T. & Carell, A. (2004) Kommunikationskonzepte, [Concepts of Communication in CSCL] In J. Haake, G. Schwabe & M. Wessner (Eds.), *CSCL-Kompendium*, Oldenburg, Frankfurt , Germany

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, J.C. (1991) Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*. **34**, 143-160

Tang, J.C., & Minneman, S.L. (1990) VideoDraw: a video interface for collaborative drawing. *In Proc. CHI 1990*. ACM, 313-320

Tang, J.C., & Minneman, S.L. (1991) VideoWhiteboard: video shadows to support remote collaboration. *In Proc. CHI 1991*. ACM, 315-322

Tharpe, R. G. & Gallimore, R. (1988) *Rousing minds to life*. Cambridge, MA. Cambridge University Press

Tolmie, P., Grasso, A., O'Neill, J. & Castellani, S. (2004) Supporting Remote Problem-Solving with Ubiquitous Computing: Research Policies and Objectives. Workshop paper for 'Giving Help at a Distance' workshop at the *Sixth International Conference on Ubiquitous Computing*, Nottingham, September, 2004

Vygotsky, L. (1978) *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press