Turn It This Way: Grounding Collaborative Action with Remote Gestures

David Kirk¹, Tom Rodden¹ and Danaë Stanton Fraser²
The Mixed Reality Laboratory
University of Nottingham
Nottingham, UK, NG8 1BB
{dsk, tar}@cs.nott.ac.uk

Department of Psychology
University of Bath
Bath, UK, BA2 7AY
D.StantonFraser@bath.ac.uk

ABSTRACT
Remote gesture systems have been shown to provide a significant enhancement to performance in collaborative physical tasks, an effect ascribed to the ability of remote gestures to help ground deictic references. 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 paper we significantly unpack the effects of remote gesturing on collaborative language, arguing for a more complex role for remote gestures in interaction. We demonstrate how remote gestures influence the structure of collaborative discourse, and how their use can also influence the temporal nature of the grounding process. Through generating a deeper understanding of these effects of remote gesturing on collaborative language we derive implications for the development and deployment of these technologies.

INTRODUCTION
Supporting remote collaborative interactions using a variety of rapidly evolving video-based technologies has been a perennial area of interest for CSCW. One area of particular interest in recent years has been the construction of remote gesturing devices to facilitate interaction in, what have been termed, collaborative physical tasks [14, 18, 20]. Such tasks have been most aptly described by Kraut et al [18] as:

“...the use of remote gestures in such distance collaborations significantly improves performance [8]. Spoken language however, remains the most important communication device, as during these collaborative tasks talk in interaction is the most effective resource through which action can be directed [26]. In these interactions it is primarily through spoken language-use that action is guided, the interaction structured and attention apportioned, and a fundamental result of these activities is the development of inter-subjective awareness between collaborators. As Clark and Brennan argue “all collective actions are built on common ground and its accumulation” [4], and it is the purpose of spoken language in these remote collaborations to help establish this common ground or mutual understanding [3, 4, 5, 6]. This is of particular importance for collaborative physical tasks given their inherently object-focused nature. As Clark 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. There are potentially a variety of ways in which referential information such as this can be grounded, with the use of ‘indicative gestures’ [4], appearing particularly pertinent in collaborative physical tasks. Potentially the use of remote gesture promotes simplified language from the Helpers as instructions are given, and the instructions themselves are more easily interpreted by the Workers making task completion quicker.
Researchers exploring remote gesture tools have suggested that it is the infrequently used but more complex gestures rather than simple deixis which are responsible for the performance enhancement [8]. This suggests that the performance benefits of remote gesture are not derived from more verbose (and therefore time-consuming) referential descriptions being replaced by simple pointing actions, but are in fact derived from a more complex effect of gesture on collaborative language. However, any attempt to understand how remote gesture interacts with language must also take into consideration the temporal nature of the grounding process. Research suggests that the use of task-specific terminology is refined during the course of a task [17]. Referent terms take a certain amount of time to become established. If as suggested above, remote gesturing achieves its performance effects by influencing the course of developing common ground, then remote gestures may be most effective at early stages of the grounding process (wherein, it could be argued, a lack of grounding has the most impact). These issues of exactly how remote gestures influence collaboration in terms of both the temporal course of grounding and the structure of discourse have been given scant regard in prior research.

In this paper we report on a study designed to explore how remote gestures influence discourse during specifically collaborative physical tasks. The study considers the ways in which object-focussed remote interactions are grounded and the temporal aspects of this grounding activity. Our results suggest that the use of remote gesture technologies does indeed influence the structure of language used by the collaborating parties in such tasks. We would suggest that by developing a deeper understanding of the specific impact of gesture technologies on the collaboration process, and specifically by understanding how they influence the achievement of grounded interaction, significant implications can be derived for the design and future deployment of remote gesture technologies.

**PREVIOUS WORK**

**Remote Gesture Technologies**

Experimental studies [24, 34] have indicated that simply linking remote spaces through audio-visual video links (as opposed to audio-only) does not improve collaborative performance, especially not to the levels observed in co-present collaboration [18, 26]. It has been argued that to facilitate performance in object-focussed tasks a representation of gesture between the spaces should be provided [8], this has potentially then been the motivation for the development of remote gesturing systems.

It could be argued that such systems have developed primarily along two separate lines, a division based on the relative perceived role of gesture in the activity to be supported. Taking their lead from design based studies, a class of collaborative sketch and design systems have emerged. These design tools treat the representation of gestures as the goal of communication. Early systems such as Commune [2] supported shared design by representing sketches over a shared digital medium allowing designers to use sketching to remotely gesture around shared digital artefacts. This led to the development of several technologies such as VideoDraw [32], VideoWhiteboard [33] and Clearboard [12]. More recently systems such as the Agora system [22] and VideoArms [31] have extended this arrangement to support multiple party interactions. These systems exploit video projection techniques to support collaboration around the construction of shared representations (such as drawings) within collaborative design activities. One predominant feature of these systems is that they are based on the conception of shared interactive behaviour where all parties share equally in the task and are working in similar roles, usually in some form of shared design activity.

The second broad category of remote gesture system is a class of systems to support collaborative physical tasks. With these tools the representation of gesture is an artefact of communication used to facilitate ongoing object-focussed discourse. These collaborative gesture systems have tended to focus on interactions exhibiting a Worker – Helper dynamic. Systems such as Drawing Over Video Environment (DOVE) [27] allow Helpers to overlay gestural sketches on a video representation of the Workers’ task space, the resulting mixed image being presented back to the Workers via a monitor in their local task space. Other systems such as GestureMan [21] and Wearable Active Camera/Laser (WACL) [29] utilise Helper controlled laser pointers situated directly in the task space, accommodating simple deictic gestures. Research has demonstrated that performance improvements in physical tasks do result from the use of gesture-based systems such as DOVE [8].

**The Interaction of Gesture and Language**

It has been suggested [8, 14] that the performance benefit derived from the use of a remote gesture tool is based on its ability to affect the process of developing, what Clark and colleagues [3, 4, 5, 6] refer to as common ground or grounding. Research evidence suggests that critical to the establishment of this conversational grounding (i.e. establishing mutual knowledge, belief, attitudes and expectations, [3]) is the provision of shared visual access to collaborative task spaces [5, 11, 18, 19]. When provided with this access it has been established that collaborators allocate most of their visual attention to images of the workers’ hands and the shared task artefacts, evidence suggesting that such information is used by the Helper to establish confirmation of understanding from the worker [10]. Clearly gesture is important to establishing common understanding (and has been shown to be especially so with spatially referent language [28]). Clark and Brennan [4] discussed various methods by which a communicative statement can be grounded, highlighting 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. Therefore the provision of remote gesturing should help the grounding process, insofar as it provides increased simple deictic referencing and reduces the numbers of words spoken. Evidence for this has been provided by Fussell et al [8].

However, the study by Fussell et al [8] also demonstrates that it is not simply the deictic qualities of gesture which are of most importance to the improvement of collaborative performance. Results clearly indicated that a less frequently used but more complicated form of gesturing (perhaps such as McNeill’s concrete iconic gestures [23]) provided the performance enhancement. Fussell et al [8] argue that more 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)

(see [3] and [23] for background on these concepts). Other research has demonstrated that complex use of gesture in interaction can have a variety of other uses in collaborative discourse such as helping to marshal turn-taking and to signal understanding [5, 7]. However, some CSCW research in group interaction [1] has tended to marginalise the importance of these aspects of gesturing, even though they are demonstrably some of the most prevalent gestures used [ibid]. Despite this, recent work [16] has shown that remote gesturing increases a sense of remote presence in distributed interactions, and has suggested that this epiphenomenal aspect of gesturing significantly improves collaborative performance (although a link between sense of presence and improved performance was not confirmed). It is not therefore clear as to whether the benefit of gesturing really does stem from an effect on the grounding of spoken language referents or is indeed due to some other effect of gesture on language, such as the ability to convey a sense of presence, or to marshal the structure of the discourse. To establish this one must compare the language used during use of a remote gesture tool to language used during use of standard audio-video connections.

When trying to investigate the grounding process in remote collaborative interaction, it is also obviously important to take account of the temporal nature of grounding. Grounding is an iterative process during any collaborative action [4], and some research has demonstrated continuous improvement in collaborative physical task performance over time [8]. This concept is well illustrated by Krauss and Fussell’s [17] discussion of the ‘evolution of referring expressions’. In this they demonstrated how terminology used to describe an object changes over time to the extent that a term can become significantly different over the lifetime of its use; often being shortened to increase its utility. Clearly therefore, if it is the grounding process that remote gesture tools affect, there is some possibility that they may have greater utility during early stages of the collaboration process. At points where language referents haven’t been securely grounded, remote gesture tools might find their greatest application. And domains where referent terms are already sufficiently grounded might have little use for added gestural support. Such an understanding of the interplay between remote gesture and language has significant potential implications for deriving suitable future applications for remote gesturing technologies. To this end we conducted an experimental analysis of the language used during collaborative physical tasks, to try to ascertain the effects of remote gesturing on discourse.

THE STUDY
Constructing a Remote Gesture Tool

To understand the role of gesture in collaborative tasks we constructed a low tech proto-type remote gesturing system which allowed us to compare gesturing and no gesturing collaboration conditions, and to try to replicate findings found in the DOVE studies [8, 27] for remote gesturing systems which utilize unmediated representation of hands rather than digital sketches. The technology was therefore designed similar to systems constructed by Kirk et al. [14, 15]. This consisted of a closed circuit system of digital video cameras, TV monitor and digital projector, as appropriate. 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 Workers’ desk, focused to capture the entire desk area. This resulting video image of the Workers’ 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 Helpers’ desk (see figure 1), the Helper being sat approximately 60cm from the TV monitor.

![Figure 1 Voice Only Communication](image1.png)

![Figure 2 Voice + Projected Hands](image2.png)
To allow the Helpers 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 Workers’ desk, projecting a video image (approx. 40x53cm) of the Helpers’ hands and anything else on/over their desk space onto the centre of the Workers’ desk (see figure 2 for illustration). The Helpers’ 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 Workers’ desk by viewing video feedback from the Workers’ desk presented on their TV monitor.

The system was constructed such that both participants would be in the same room during the study, 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. This enabled us to retain full audio in all conditions without having to use any audio communications technology. Participants were allowed to speak to one another at all times during the study.

Whilst some might consider that it would make most sense to allow Helpers to align their gestural actions with a projection of the Workers’ task space onto their own desk space, we made the decision (similar to Kirk et al. [14, 15]) to have Helpers align their gestures with the Workers’ task artefacts by viewing their hands on the live video feed. Arguments supporting such an approach are presented in ([14]) but we reiterate that this decision is based on the argument that it is of benefit to those creating gestures for presentation to others, to be aware of how those gestures will be perceived. So when gestures are to be presented in a two-dimensional format (i.e. video-capture and projection), it is of benefit to the Helpers if they can construct their gestures in a similar two-dimensional environment.

Participants
A total of 48 participants took part in the study, volunteering in pairs. There were 24 single sex pairs in total, comprised of 26 males and 22 females. Participants’ ages ranged from 18-26 (mean 20.83, St. Dev. 1.59), and they were mostly undergraduate / postgraduate students. Participants were each paid a small amount of money for taking part in the study. All participants had normal or corrected to normal vision. One pair was excluded from subsequent analysis as they failed to reach required levels of task completion.

A Collaborative Task
To motivate collaborative action, pairs of participants were engaged in a Lego assembly task, using both remote gesturing and non-remote gesturing communication methods. Lego was chosen as it represents a generic object-focused task and is comparable to the tasks used in previous work (e.g. [5, 8]). The nature of Lego construction ensures that during the accomplishment of any given stage of a model a variety of actions are required, including search and select, rotate and align, attach and detach and pattern matching to ensure the final model follows the plan.

After being randomly assigned roles (Helper or Worker), participants were given training in how the remote gesture 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 (during the relevant trial) the gesturing system could be used.

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 – the manual was to the side of the Helpers’ desk and not visible to the Workers). The Helpers were instructed that they were to talk to their Worker and explain how to construct the Lego model working sequentially through the stages described in the manual and additionally in the appropriate trial using gestures to help provide guidance – if they so wished.

As the nature of the Lego pieces precluded any guessing of how they should be put together, and the Workers had no visual guide of what the end model should look like, the Workers relied completely on instruction from the Helpers. Pairs were given 10mins to complete as much of the model as they could (no pairs ever managing to complete a model within 10mins). 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.

RESULTS
Our initial desire had been to understand whether remote gesture systems using simple presentation of hands rather than other methods could produce benefits to working collaboration. To this end we compared basic performance (measuring the time to complete the first three stages of a model) for collaborating pairs in both voice only and voice and gesture communication conditions. The results of this comparison can be seen in Table 1 below.

<table>
<thead>
<tr>
<th></th>
<th>First trial</th>
<th>Second trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice Only</td>
<td>227 (87.7)</td>
<td>151.73 (75.49)</td>
</tr>
<tr>
<td>Voice + Gesture</td>
<td>155.27 (28.16)</td>
<td>164.08 (41.51)</td>
</tr>
</tbody>
</table>

Table 1. Mean times (in seconds) for performance of first three stages of model, by trial and communication condition (St. Dev. in brackets)

Noticing an apparent difference in effect of gesture by trial we performed a 2x2 mixed design ANOVA. The main effects of both gesture condition and trial were found to be
non-significant (F (1,45) = 2.78, p= n.s.) and (F (1,45) = 3.42, p= n.s.) respectively. There was however a significant interaction effect of gesture condition by trial order (F (1,45) = 4.8, p= 0.03). This result suggests that allowing remote gesturing improved performance but only in the first trial. Using two-tailed independent measures t-tests for post-hoc comparison revealed that performance with remote gesturing was significantly faster than with voice only communication in the first trial only (t (22) = 2.24, p= 0.04). The data potentially suggests that using remote gesturing in an early trial has lead to performance equal to later more ‘grounded’ collaboration. Potentially therefore the use of remote gesturing has helped to ameliorate the process of achieving conversational grounding. To test the veracity of this assumption we decided to investigate the language being used by collaborators, attempting to find evidence of the grounding process in action.

The first stage of this process was to understand some of the basic characteristics of the language used during collaboration. To better manage our data we took a random sample of 12 of our pairs (using their conversation from both trials) and after creating transcripts from the video records of their experimental trials, calculated the total number of words used in the first 5 minutes of each interaction. A two-way independent-measures t-test showed that, as would be expected given the nature of the task, Helpers speak significantly more than Workers (mean of 574 vs. 191.6 words) (t (46) = 14.45, p≤ 0.001). We also noticed that amongst the Workers total word use in the first trial was increased when the Helpers used voice only communication (mean of 237.5 words) as opposed to voice and gesture communication (mean of 177 words). 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 finding of longer average completion times for voice only communication in the first trial. To explore this further we began to analyse the content of the interactions. A first point of analysis was the use of questions during collaboration. Our analysis revealed that Workers were far more likely to ask questions than Helpers t (46) = 10.33, p< 0.001), and generally questions were more likely to be asked in the first trial rather than the second (t (22) = 2.39, p≤ 0.03). Equally the total number of words used between Worker questions showed significant increase over the course of the two trials (t (22) = -2.25, p≤ 0.03), this suggests that the Workers were asking less questions, less frequently as practice with the task increased. Whilst firm conclusions about the effects of gesturing on number of questions asked are hard to draw, all 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 gesture restricted. Suggesting 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 below¹ (taken from the transcripts of interactions) we can understand the nature of the differences behind this desire to adopt questioning behaviour. Whereas in Excerpt 1 the Helper is very directive (the Worker not needing to respond verbally) using 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.

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

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

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

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

H Yep that’ll do= ((fingers move from finish point of last gesture to off the table))

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

H erm (.) and (.) that should be (.) just hold it up a bit (.) erm (.) 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]

H just flip it round

W you mean just could of turned it? hahah

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

W right.

At this point it was felt that a more in-depth analysis of speech content was needed. To control for differences in the Lego model stages that might be discussed by any given pair during a specific time sample (some stages requiring significantly more instruction), the analysis was refined to focus on performance in specifically two stages of one Lego model only. To bolster the data additional pairs of participants were included creating a sample of 23 transcripts. For these more detailed transcripts a Conversation Analysis style notation to mark speech acts was applied. We began the analysis by considering the use of deictic referencing within conversation.

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 deixic references staying largely the same between trials. This suggests that early components of

---

¹ Guide to notation in excerpts – H refers to Helper, W to Worker. Pauses in speech are marked with ( ). Parts of text accompanied by a gesture are underlined. 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. -
Workers’ 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

A final point of interest considering the use of deixis was the finding that Helpers are significantly more likely to use proximal deictic references when allowed to use remote gesture tools as opposed to when relying on voice only communication (t (21) = -2.23, p≥ 0.04). If we consider the observations of correlations between deictic linguistic features and sense of presence reported in Kramer et al [16] 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 support. This type of shared working space, as opposed to providing external support.

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 ( ) 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 yep
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.

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 our sample we turned to analyse the extent of overlapped speech within the collaborative discourses. The prevalence of overlapped conversational exchanges varied by trial and communication condition and can be seen in Table 2 below.

<table>
<thead>
<tr>
<th></th>
<th>First trial</th>
<th>Second trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice Only</td>
<td>38.19% (19.42)</td>
<td>27.39% (10.59)</td>
</tr>
<tr>
<td>Voice + Gesture</td>
<td>9.97% (8.94)</td>
<td>24.91% (17.28)</td>
</tr>
</tbody>
</table>

Table 2. Percentage of overlapped turns, by trial and communication condition (St. Dev. in brackets)

To understand and further test the interaction of gesture use and trial order we analysed the overlap data in a 2x2 independent measures ANOVA. We 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. We 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 we did find 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 examples of pieces that might fit the bill. 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  [with] five holes on the top and three going down=
W  =you 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 huh huh yeah
W  yeah
H  erm and now that goes over the other side of 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 Helpers’ discourse where there is hesitation there is a tendency for the Workers to feel obliged to start to attempt to force a change of turn. In excerpt 7 below however, we can see an example 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 erm 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 the one James yeah=
W  =right

In this example the Helper’s initial turn is littered with 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 our raw data had demonstrated that amongst our sample there was significant difference between individual collaborating pairs and the extent to which they adopted use of the remote gesturing tool, we chose to perform a correlational analysis of the data. 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), we 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, we 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 we 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 correlational analysis therefore appear to suggest that the use of a remote gesture tool improves performance and reduces the probability of disfluent speech and overlaps during discourse. With further evidence that increases in overlaps degrade collaboration we have perhaps come to some understanding of how gesture interacts with language to improve performance.

In summary therefore, we 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 Workers in the interaction need to speak. Whilst questioning behaviour from the Workers 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 Workers are forced to point to various alternative pieces for the Helpers – so as to establish their common points of reference. When gesture is used, the Helpers seem more engaged directly in the task space, exhibiting increased use of proximal deixis. In turn we have seen that the use of remote gesturing is associated with a reduction in the occurrence of speech overlaps; this suggests that remote gesture smoothes interaction and facilitates clear turn-taking. This smoother more structured form of interaction allows for better performance.

**DISCUSSION**

In this paper we have explored the use of a remote gesturing technology to support collaborative physical tasks. In this section we wish to reflect on the lessons from our study and the implications for the development and deployment of these technologies.

**Achieving grounded interaction**

Fussell et al [8] 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. Our results 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.

Our work however, also demonstrated 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 Helpers, with more
effort in establishing mutual referents being shifted to the Workers, who consequently had to increase the amount of words they used during interaction. We also noted evidence 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, one of our most important observations must be the effects of remote gesture tool use on overlaps in speech. In early trials 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.

We would not wish to challenge previous arguments that more complex forms of gesture are used to replace difficult to interpret complex referential descriptions with representational gestures, as such an argument is clearly valid (see [15] for discussion of the ability of hands to represent complex information). We merely wish to add to this our evidence which demonstrates an additional role for remote gestures in structuring discourse during remote collaboration. Through our analysis of overlapped and interrupted speech we can see a common association between presenting remote gestures, reducing overlaps (therefore smoothing turn-taking) and improving performance. Previous research in video-mediated communication (VMC) has suggested that using video-links between spaces increases the occurrence of overlapped speech [25]. And these increases have been presumed to be a positive sign of increased speech fluency [30] therefore supporting the adoption of VMC, but were never really linked to an objective measure of task performance. These previous studies were also based on negotiation and discussion style tasks. We have shown that alternatively in an object-focussed task such increases in the occurrence of overlaps are actually associated with poorer task performance. But we have also demonstrated that this can be potentially alleviated through the use of additional remote gesturing links between the spaces.

The results support the work of Kramer et al [16], finally demonstrating the link between increased remote presence and improved performance. It could be argued that the increased presence and the visual feedback of action and intention within the shared visual space [11] which is derived from the use of a remote gesture tool actually supports the indication (in the absence of face-views) of who is taking a turn and when, it is this feature which reduces overlapped speech.

Further, we would argue 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 [4] and back-channeling speeds up the process of grounding terms [3]. It is also possible that remote gesturing is influencing the collaboration process in this way.

Integral to these arguments is a consideration of the costs of grounding [4]. 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. This is exemplified in discussions of the translational overheads of working without remote gesture [14, 15]. 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 ameliorate delay, speaker change and repair costs of grounding. The addition of remote gestures to speech, alleviates the likelihood for interruption when utterances are being formed or modified or when a speaker wishes to retain the floor, therefore reducing the number of failed attempts at turn-taking, requiring that significantly less time be expended on costly sentence repair phases.

Our study has also observed 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 (a hardly surprising effect). 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 [17] and the words and effort required to refer to shared artefacts are reduced [4, 17]. Given enough time (and therefore practice) all remote collaboration will become grounded. Communication restricted to an audio-only connection, does not become impossible, achieving grounded interaction merely takes longer to achieve.

Considering that in a collaborative physical task, where one person is directing the activities, the key hindrance to performance is this lack of common understanding, once grounding has been achieved performance becomes optimised. If gestures are primarily helping to ground speech (which is what we argue our results demonstrate) then once common terms, referents and practices have been established, the presence of remote gestures will cease to have such a discernible impact on performance. This explains why performance in the second trial was not further improved by the presence of remote gestures.

**Implications for Remote Gesture Tools**

The evidence of this work has hopefully highlighted the role of remote gesture technologies as tools for supporting the achievement of grounded interaction in object-focussed tasks (much as other work [34] has suggested the benefits to grounding of simple VMC links in negotiation based tasks). Stated frankly, these results in themselves offer little in the
way of implications for design of remote gesture technologies. Granted we have offered further support for arguments presented elsewhere [22] that designing gesture tools using specifically unmediated representations of hands can offer positive benefits for collaborative performance (similarly to or as opposed to other gesture representations). And we have presented some evidence that suggests that the addition of gestural capacity to standard VMC connections can improve object-focussed collaborations.

But we feel that the main argument of this paper does not necessarily support a concern for the design of the technologies per se, other work already offering a much more thorough discussion of how best to design these technologies e.g. [8, 14]. What this work does offer however is a set of implications for the future deployment of these technologies, an issue that has been given scant consideration in previous work.

Understanding how best to deploy a technology is critical to its successful adoption and continued use. Remote gesture technologies are often built to support working practices that will emerge but don’t currently exist. This can make it hard to foresee limitations in the technology that will limit its adoption in a specific working scenario. Locking in a technology design to support a working task that is ultimately unsuitable can be a costly mistake. To help this our results provide 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 collaborative task they have in mind.

Given our observations of the role of remote gesturing in expediting the grounding process, only impacting on performance in early ungrounded interactions and the observed importance of smooth turn-taking in object-focussed interactions we have identified three key indicators which help to identify the potential applicability of remote gesturing for any given task. They focus on determining the levels of prior ground in a collaborative scenario and the importance of time to the interaction

- **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 unencountered 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 indicators suggests a number of cooperative arrangements as ideal applications 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 sensitise technology designers to the factors which influence how their remote gesture technologies will be applicable to different future collaboration scenarios.

**CONCLUSIONS**

In this paper we have discussed the effects of remote gestures on language used during collaborative physical tasks. We have highlighted previous research which has suggested that the benefit of remote gesturing is in its ability to support the grounding process. We have presented evidence concerning the influence of gesture on the structure of collaborative action, and have also drawn the first reference to the significant interaction between gesture and the temporal course of grounding, highlighting the significant implications that these factors have for the future deployment of such technologies.

**ACKNOWLEDGMENTS**

Thanks to the EPSRC Equator project for funding (www.equator.ac.uk)

**REFERENCES**


