A Collaborative Guidance Case Study

Duncan Stevenson[1], Jane Li, Jocelyn Smith[1] and Matthew Hutchins

CSIRO ICT Centre, GPO Box 664, Canberra 2601 ACT, Australia
[1]Also Department of Computer Science, Australian National University, Canberra

Duncan.Stevenson@csiro.au

Abstract

This paper reports on a collaborative guidance case study, which investigated the use of remote pointing and drawing technologies in a system designed for spatially focussed collaborative tasks. Four guidance technologies were available to the participants – pointing and drawing over video of the remote site and pointing and drawing into the remote workplace itself. The experimental task was designed to mimic the actions observed in an actual application setting. The purpose of the study was to see how the participants would use the technology and how they would collaborate with each other during the performance of the task..

Specifically, the experiment looked at how the participants selected from the choice of guidance technology, and changed their selection, as the task progressed. It looked at how they used the technology and how they created working, 3-dimensional, shared frames of reference for the task. Finally it explored the way the system supported emerging collaborative behaviour between each pair of participants.

The paper concludes that the participants were able to make reasoned choices about their selection of guidance technology, and that they evolved effective guidance strategies as the task progressed. They adapted their understanding of each other's frame of reference with respect to the task by focusing on reference objects created during the task. Finally, the paper concludes that the experimental system did indeed foster emerging collaborative behaviour between the participants.

Keywords: Collaborative guidance, tele-collaboration, remote pointing, remote drawing.

1 Introduction

The case study which is reported in this paper looked at the way in which the participants used pointing and line-drawing gestures to guide a remotely located partner to complete a specific task in a three-dimensional setting. The study also looked at how they collaborated during the process of completing the task. During each run of the study, pairs of participants completed self-paced training in two technologies for delivering remote pointing and line-drawing gestures, then they were free to use their own choice of these to complete the task. The participants had a real-time face-to-face video and audio link, in addition to their shared view of the task workspace, and their collaboration included use of hand, face and body gestures, spatial references and task-specific dialogue.

The motivation for this study was to investigate the remote guidance component of a tele-health system prior to deploying that system in a clinical pilot study. In the tele-health scenario we have a surgeon at a major hospital conducting a post-operative paediatric consultation with the patient and a clinic assistant, both located remotely at, say, a regional medical centre. The role of the guidance technology is to allow the surgeon to guide the assistant in the very three-dimensional task of examining the patient and, in particular, the healing surgical site.

This study explored the way the guidance technologies, and their broader video and audio links, could support the task of directing an activity, and also how they could support collaborative activity between the surgeon and the assistant. To place this work in context, Figures 1 and 2 show the surgeon's office and patient's clinic for the tele-health system. Figures 3 & 4 show the experimental system for this study as a sub-system of the tele-health system. Figures 1 and 2 also show the two-way face-to-face video and audio links between the surgeon and the remote participants, including the clinic assistant.

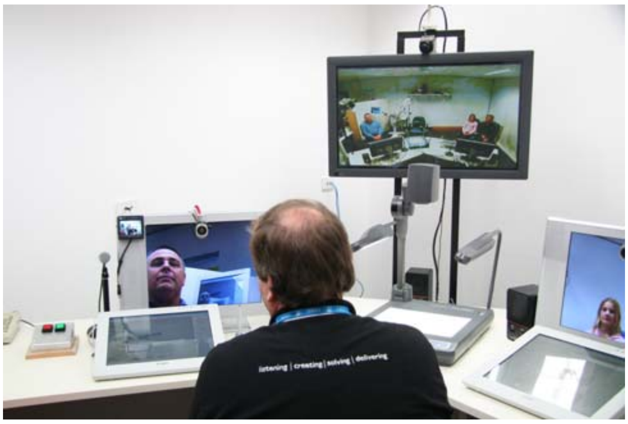

Figure 1: The surgeon's end of the tele-health system

The complete tele-health system uses three fixed video channels to connect the surgeon's office and patient's clinic. As seen in Figure 1, the left-hand video connects the surgeon to the examination area of the clinic room, the central plasma video display gives an overview of the whole clinic room and the right-hand video connects the surgeon to where the patient and family would sit during discussions with the surgeon. Two tablet displays are visible on the surgeon's desk. Figure 2 shows the matching video channels in the patient's clinic, and an overhead camera and laser device mounted on a pole. During this study only the tablet displays, the overhead camera and laser device, and the examination face-to-face video displays were used.

Figure 2: View of patient's clinic

The study task was designed to mimic the types of actions that we had noted during observations of outpatient clinics at the hospital and in discussions with our clinical partners. The types of actions, however, were framed in such a way that the task could be carried out by non-medically-trained participants. This allowed us to design and conduct the study in a relatively self-contained manner, and to get some general insight into the performance of the guidance component of the larger system. The alternative, designing a health-related task and having health professionals as experimental participants, would have been much more expensive to develop and would have consumed scarce time and effort on the part of our clinical partners. We fully expect that the results of this study will influence the way that we incorporate this guidance technology into the larger tele-health system.

Within this study we refer to the person mimicking the role of the surgeon as the "instructor" and the person mimicking the role of the clinic assistant as the "actor". In transcripts of their dialogue we refer to them as "I" and "A".

The study explored the following aspects of collaborative guidance between the pairs of participants:

- Their choice of pointing and drawing options and the reasons behind these choices
- The way the instructor guided the actor to complete the components of the task
- The way the instructor and actor dealt with the mismatch between their 3D spatial frames of reference
- The emergence of collaborative behaviour between the instructor and actor

The guidance technologies available to the instructor were controlled by a tablet display and electronic pen (Figure 3). A window on the display showed live video of the remote actor and workspace, taken from a video camera mounted on a pole above the actor. By moving the pen just over the surface of the tablet the instructor could control a cursor that functioned as a pointer for the actor, and by drawing on the surface of the tablet the instructor could create transient drawings with which to guide the actor.

Figure 3: Instructor using the tablet display and electronic pen to guide the actor

This interface controlled two different guidance technologies. One pointing and drawing technology, referred to here as "on-video", placed the cursor or drawn lines over the live video as seen on the tablet display. The actor had an identical tablet display mounted in front of the workspace and, by looking up from the workspace, could see, in real time, the cursor and drawing gestures of the instructor.

Figure 4: Experimental configuration for the actor

The second technology, referred to here as "in-workspace", used a custom-built laser device mounted with the overhead camera. The instructor's gestures with the electronic pen over or on the tablet display controlled the laser device to project cursor graphics or real-time drawing into the actor's workspace. The green lines projected by the laser device into the workspace can be seen in Figures 4, 5 and 7.

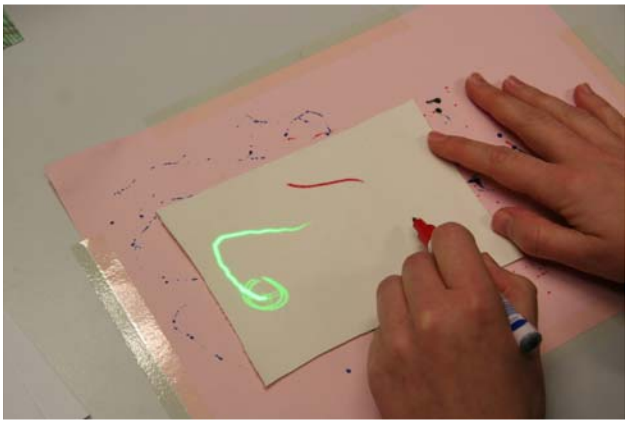

Figure 5: The green laser-projected line will guide the actor to draw an irregular shape

The two technologies each had different advantages. The "on-video" technology provided a precise 2-D interface to the remote guidance, and was interpreted by the actor on a 2-D display at the remote site. The cursor and line drawing were exactly where the instructor had placed them. The actor, however, needed to look up from the 3-D workspace to see the 2-D display to interpret the instructions.

The "in-workplace" technology was slightly less precise and required the instructor to be aware of the 3-D spatial relationships of the objects in the workspace. The displayed cursor and lines were, however, visible to the actor in the workspace, requiring no shift of visual attention or matching of the 2-D display to the 3-D workspace. Figures 3 and 4 show the instructor and actor using the in-workplace line-drawing guidance technology.

2 Related Prior Work

Several different technologies have been used to support remote guidance. One technology is to draw over live video of the task space, displaying the output on a screen in the remote workspace (Ou et al. 2003; Fussell et al. 2004) or on a head-mounted display (Bauer et al. 1999; Kraut et al. 2003).

A second technology is to project guidance gestures into the actual workspace. The GestureCam system, developed by Kuzuoka and colleagues, combines a laser pointer and a camera on a moveable actuator controllable from a remote site (Kuzuoka et al. 1994; Kuzuoka et al. 1995). This work was extended to include placing the camera and laser pointer on mobile devices (Yamazaki et al. 1999; Kuzuoka et al. 2000). Kurata and colleagues describe a wearable system, called the Wearable Active Camera/Laser (Kurata et al. 2004).

A third technology is to project live video of the guider's (instructor's) hands into the remote workspace (Kirk et al. 2005; Kirk et al. 2005; Kirk and Fraser 2006). The hands are then available for the full richness of natural gestures.

Several of these papers consider these technologies from the point of view of comparative evaluation. They identify potentially equivalent generic ways of addressing a given task then conduct controlled experiments. For example, one study projected video into the remote workspace and used three conditions: instructors hands, hands and digital sketching, digital sketching only (Kirk and Fraser 2006). Various performance metrics are then evaluated. These include time to completion and accuracy of completion of the task (Fussell et al. 2004) and time taken and mistakes made (Bauer et al. 1999; Kirk et al. 2005).

Our paper takes a different approach. We take the view that each of the technologies has its strengths and weaknesses, and is appropriate for different situations. For example, the on-video drawing can be more precise (it is a 2D user interface to a 2D display) but there is a cost for the "worker" in looking up from the task space to view it ["worker" in Kirk's terminology, "actor" in ours] (Kirk et al. 2005). In-workplace drawing overcomes this cost, but at the expense of precision of placing the drawn sketch (it uses a 2D user interface to a 3D workspace display with a 3D camera/laser offset disparity). In the relatively complex task that we are using we expected that there would be a role for both on-video pointing and drawing and in-workplace pointing and drawing. These roles would depend on the particular stage of the task and on how the participants chose to interact. We therefore provided all four of these options and observed how the participants chose to use them as the experimental task unfolded. For the "in-workplace" options the method for translating the 2D user interface into the 3D workspace is discussed by Palmer and colleagues (Palmer et al. 2007).

We used a pair of experimental subjects in each run of the experiment. This use of pairs of participants has been followed in several previous studies (Bauer et al. 1999; Fussell et al. 2004; Kirk and Fraser 2006). Other studies used a small number of experiment confederates or trained participants as instructors and experimental participants as actors. By not having an experiment confederate as the instructor we avoided any preconceptions of how the task should be conducted.

Several previous studies have placed an emphasis on fast and accurate completion of the task. In one study the participants were offered an additional $25 "by being the fastest and most accurate pair to complete the task" (Fussell et al. 2004). In our study we placed much more emphasis on the collaboration between the instructor and actor over the task. This mimics the corresponding tele-health situation of surgeon and remote clinic assistant collaborating to examine the patient, where the process of examination is important. In our experimental task, therefore, both the training and the task phases were self-paced, and it was left to the two participants to agree that they had reached a satisfactory conclusion to the task.

In designing their experiment, Fussell and colleagues implemented a feature that automatically erased any lines that had been drawn after a certain display time (ibid.). They based this on the idea that the underlying video was transient and that this would free their participants from having to manually erase old parts of a drawing. We followed their idea and implemented an automatic fade for our drawing, both on-video and in-workplace. After some fine-tuning we settled on a 5-second time-to-fade.

Kraut and colleagues emphasise the importance of both participants having a shared view of the task space (Kraut et al. 2003). In our experiment we extend this concept in the following ways; both participants have similar views of the actual workspace – the actor can see exactly what is in front of him or her and the instructor can see the workspace via the overhead camera. In addition to this, the instructor can see the actor himself or herself sitting at the workspace and the actor can see the instructor's view, which is replicated on the swivel-mounted tablet display. The instructor and actor can, therefore, see the task space and can also see (and understand) what each other can see.

The work reported here is a sequel to an earlier guidance interface study (Smith et al. 2007) which looked at different user interface concepts for controlling on-video and in-workplace guidance technology. In that study only the instructor was an experimental subject and the purpose was to explore three different user interfaces. One conclusion from this study was that, for tasks that require a detailed sequence of spatial instructions, the collaboration between the instructor and the actor was very important. This collaboration included both voice and gestures, and interface features that interrupted it interfered with their ability to complete the task.

The present study differs from our earlier study in the following ways. The user-interface for this study incorporates the design results from the previous study. This study uses experimental subjects for both instructor and actor. It has an extended, self-paced training phase that is closer to the study task than previously. Lastly, it focuses on the evolving choice of guidance technology and on the evolving collaboration between the two experimental subjects.

3 Collaborative Task Used in This Study

3.1 Task Description

The basic task required the instructor to guide the actor in drawing the shape, shown in Figure 6, onto a piece of fabric, cutting out the shape and then pinning it onto an article of clothing (a child's dress) which was fitted over a foam rubber core to give it its proper 3-dimensional shape. The instructor was able to point or draw over a video of the actor's workspace, and also to point and draw with a laser projector into the actor's workspace. The instructor and actor also had a face-to-face video and audio link so that they could conduct a dialogue about the progress of the task. Because the shape was relatively amorphous it could not be drawn from high-level verbal instructions. As one instructor said to their partner "It's an odd looking patch, [laughs] that's the best way to describe it." The instructor needed to use a moving pointer or line-segment drawing to guide the actor in drawing the shape.

The instructor then guided the actor to place the shape in the correct position and orientation on the dress. Because the dress itself had no clearly describable structure and relatively few landmarks the instructor had to use the guidance technology to point and gesture the instructions for placing the shape. Two things contributed to the complexity of the placement component of the task: the mis-match of the actor's physical view of the workspace and the instructor's video view, and the haphazard way that the actor picked up the dress and presented it to the instructor's view.

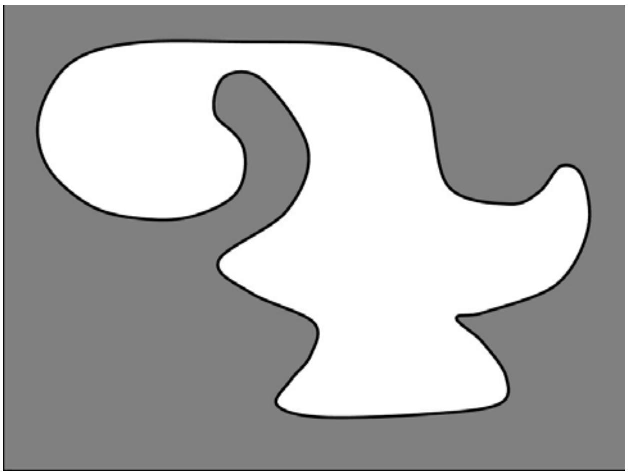

Figure 6: The shape that the instructor guides the actor to draw

Figure 7: Placing a fabric patch on the dress. The green laser cursor (below her left thumb) shows the position of the patch directly in the workspace

3.2 Rationale for This Task

There were several objectives for the experimental task. Firstly, the task would involve objects and actions both on a flat surface and in the three-dimensional workspace which were sufficiently amorphous that high-level instructions or prior knowledge were not appropriate. This was in contrast to the assemble tasks (Lego, bicycle, etc.) used in many guidance studies (Fussell et al. 2004; Kirk et al. 2005; Kirk and Fraser 2006) where common words match the component shapes, the components themselves can join in only a limited number of ways and many people have extensive childhood experience with the objects.

Secondly, the actions involved in the task mimic the actions observed by participants in outpatient clinics at the children's hospital. These include pointing to the surgical site, indicating by gesture a region of the body and using directional motion to indicate mobility of a joint or limb. The movements take place in a workspace roughly the size of a small child and have an amorphous, line-sketching character rather than precise circles, rectangles and straight lines.

Thirdly, the purpose of the task was to entice the participants to use the technology, to think about how they represented guidance intent and to engage in collaboration with their experiment partner. Success in the task was a matter for the satisfaction of the participants themselves, rather than being measured in terms of success rates, times to completion or errors made.

4 Method

4.1 Experimental Protocol

The basic task was embedded in a larger experimental protocol:

- Information about the experiment was given to the participants the day before
- A brief initial interview was conducted to collect demographic data and sign the participant consent form
- The audio-visual equipment and tablet display interfaces were explained, and the two participants were introduced using the face-to-face video link
- The four guidance options (pointing and drawing, either on-video or in-workspace) were demonstrated
- The participants undertook self-paced training using a sequence of computer-screen instructions. This training consisted of 4 tasks each using one of the guidance technologies to guide the actor to draw relatively simple shapes onto paper, thus anticipating in a simpler form the actual task.
- They then completed the task using whatever combination of guidance techniques they chose, again following a sequence of computer-screen instructions. The task consisted of selecting the correct fabric, drawing the shape, cutting it out and placing it on the dress.
- Each participant completed an exit interview

The study took place between remotely located rooms in two different CSIRO laboratories, one in Sydney and the other in Canberra. An Internet connection between the two rooms carried the real-time video and audio channels for the face-to-face communication system, and supported the remote display of video and instructional data on the tablet displays.

4.2 Data Collection

The full experimental sessions were recorded on videotape at both sites. Each video record contains a view of the actions of the participant at the particular site as well as the audio of the whole interaction between the sites. Each participant was accompanied by an experimenter for the whole time, and the experimenter took notes of their observations. The exit interviews were also recorded so that the participants' full responses were available for analysis.

4.3 Hardware Configuration

The hardware used for this study was a subset of the full tele-health system. The instructor used the left-hand face-to-face video subsystem as shown in Figure 1, and in detail in Figure 3, consisting of camera, microphone, small fold-back video monitor, LCD screen and audio speaker. The actor used the corresponding subsystem located in the area of the patient's clinic used for patient examinations, shown in Figure 2 and in detail in Figure 4. The other face-to-face video subsystem and the room overview subsystem were switched off during this study.

The instructors used a Wacom tablet and electronic pen to control all four guidance options, as shown in Figure 3. The results of these guidance options appeared for the actors either on the arm-mounted Wacom tablet shown in Figure 4, or directly projected into the actors' workspace from the pole-mounted laser projection system. The pole mounting is visible on the right-hand corner of the curved desk in Figure 2.

A video camera mounted on the pole, together with the laser projection system, provided the video view of the actor and his/her workspace. This was the video component on the Wacom tablet which the instructors used to reference their guidance actions. This video, together with any superimposed pointing or drawing, was shown to the actors on their arm-mounted Wacom tablet. The actors, therefore, could see exactly the view of their workspace that their instructors were seeing. When the in-workspace laser was used the actors could see the pointing/drawing directly in their workspace and, if they looked up at the tablet, could see the view of the laser that the instructors had.

5 Results

5.1 Participants

Six pairs of participants took part in this study. They were recruited from within CSIRO and from the university populations nearby. The participants lived in different cities from their partners for this experiment and did not know each other.

5.2 Choice of Pointing and Drawing Options

In the fabric selection subtask the instructor directed the actor to choose one of six differently patterned pieces of fabric. Four instructors used purely verbal instructions, one used on-video pointing and one used in-workplace pointing together with spoken instructions. An example:

I: OK, so we need this piece of fabric. Look on your screen, we need that piece of fabric [indicates with on-screen pointer]

The instructors chose a range of drawing options to guide the actor to draw the shape. One used the in-workspace drawing mode, with the actor drawing directly onto the fabric. Another used the on-video mode to illustrate the shape, then the in-workspace mode to guide the actor to draw directly onto fabric. The four other pairs negotiated to draw draft shapes onto paper (they used spare paper from the training phase), transfer the shape to fabric then edit the shape. The instructors used a combination of all four modes (point/draw on-video/in-workspace) to guide this process. The final editing was mostly done using the more precise on-video drawing.

All six instructors used pointers to guide the actor to place the patch on the dress – three used on-video and three used in-workplace. All used a mixture of verbal direction and pointing to orient and shift the patch towards its correct location.

Decisions on these choices were based on the participants' perceptions of the tools, as these two dialogue fragments show:

I: Which tool did you prefer?

A: The laser that draws is pretty cool, but the red one's probably better

I: That's what I was thinking as well

I: Maybe if I use the red one [on-video] to show you, because it seems to stay on the screen longer, then I'll use the green one [in-workspace] when we go to do the real drawing

5.3 Guidance Strategies

All six instructors used the basic strategy of drawing the shape as a sequence of curved line segments and asking the actor to draw each segment, with spoken fragments like "in", "out", "around" and "here" as they drew the segments. All six prefaced this drawing sequence with a short discussion with the actor about the nature of the shape and how they would go about drawing it. Four pairs decided to co-opt the spare paper from the training tasks to draw draft versions of the shape.

A: Do you want to practise first, on a bit of paper?

I: Yeah, that would be good.

When an initial shape was in place, either on the fabric or on paper, the instructor framed the guidance with reference to the draft, which they could both see and to which they could point and gesture. To edit one feature, three instructors described it as "like a shark fin", and their actors immediately understood what they meant.

I: And up here has gotta go just a little bit farther out so it's a big bulbous sort of thing and then here

A: So like out here

I: Yup, [pause] and just a bit farther out there

To place the patch in the correct location and orientation on the dress the instructors all asked the actors to place the dress (on its foam core) onto the work surface and then to orient it so that they could see it clearly. They then used the pointing tools (either on-screen or in-workplace) to locate the patch followed by spoken instructions to rotate and shift the patch to its correct position.

5.4 Frames of Reference

The instructor's and actor's views of the workspace were approximately at 90 degrees to each other. In spite of this, all six instructors used words like up/down, top/bottom and across in their verbal directions. In one case the actor pointed out the mis-match between points of view but the instructor later reverted to his earlier usage.

At a smaller scale, the instructors used the centre of the shape as the reference point, with phrases "curve outwards" and "curve inwards". Two instructors explicitly asked their actors to rotate the fabric patch "so the horizontal side is the longest". This showed awareness of the actor's orientation and served to align the instructor's view of the fabric with the displayed shape on the instructor's private view of the experimental instructions.

All six pairs modified the drawn shape to better match the required shape. This editing was done using instructions and actor responses made with respect to the draft shape. Both could point directly into the space and could see the editing instructions and actor responses. These formed a tightly coupled dialogue.

I: You kind of see that little thing that juts out again, [gestures with in-workspace tool] just copy that shape, that's exactly what we need

I: This here and this here are meant to be like sort of aligned [points with in-workspace tool]

A: Here and here? [points with felt-tip pen]

In placing the patch at its correct location on the dress the instructors used spoken instructions such as "towards the hem", "it's in the middle, below the pocket" and "a centimetre or two towards the seam". The dress itself provided the frame of reference for the instructions.

5.5 Collaborative Behaviour

All six pairs of participants exhibited collaborative behaviour throughout the experiment. The experiment started with introductions using the face-to-face video link, fulfilling the social meeting process. During the experiment they spoke to each other often, looking up from the task to make eye contact.

The training phase was self-paced by the instructors, who read the instructions to their partner as they progressed. They shared the humour of the unusual shapes and verbally celebrated success at the end of drawing each shape. While drawing each shape they exchanged verbal confirmation of each curve segment.

All six pairs edited together the first version of the shape that they drew for the main task.

I: This area there, a bit closer

A: Closer? You mean further away from this side? [Points]

I: Yes

During this editing process the actor sometimes took the lead.

A: So from here [points with felt-tip pen]

I: Yeah

A: Can you direct me from here?

I: OK from there …

The pairs were observed to negotiate the initial approach to the task and repairs when difficulties were encountered.

I: [Selects on video tool] How about if I draw it and you just sort of like have a look and then we'll, we'll start again

A: Yeah

I: [Under breathe] Na actually yeah it's probably

A: Could you actual give a demonstration using the green maybe?

I: I was just going to say you can see my red one on your screen can't you?

A: Yeah

I: Alright but if I [pause] if um, maybe with the demonstration if I just show you with the red one cause the red one seems to stay the lines seem to stay on a little bit longer

A: Yep

I: So if I use the red one and you just have a look and then I'll use the green one when we go to do the real drawing, ok?

A: Yeah

5.6 Exit Interview

An exit questionnaire was given to both instructor and actor during the exit interview. It covered the choice of the guidance tool, the guidance strategies, the participants' awareness of each other's view of the workspace and strategies for repairing task-based misunderstandings.

a. Instructors: ***What influenced your choice of tool?***

Four said that they chose the in-workspace (laser) tool because they felt that it would be easier for their partner to follow. One pair had technical problems with the laser tool and agreed to use the on-screen tool. The last pair had zoomed the overhead camera, which created calibration problems in its 3-dimensional behaviour so they chose the on-screen marker to give the instructor more precision in giving instructions.

I: I guess it depended on what was easy for either party and we kind of just discussed which one was better

b. Instructors and actors: ***Did you change your strategy during the experiment?***

Five instructors said that they did not change their strategy during the experiment. The sixth regarded the change from on-screen drawing of the overall shape to in-workplace editing of the shape as a change in strategy.

Four actors replied that they had changed strategies, referring to the editing process. One said that their pair had developed a strategy and stuck with it and the last said that there had been no change in strategy.

c. ***I understood my partner's orientation to the workspace***.

On a 5-point Likert scale (strongly disagree to strongly agree) the actors responded with either agree(3) or strongly agree(3). The instructors' responses were agree(4) and strongly agree(2).

d. Instructor: ***How did you know when your partner was confused?***
Actor: ***How did you let your partner know when you were confused?***
Both: ***How did you and your partner deal with this confusion?***

This set of questions explored their sense of working together. Three instructors did not think their partner became confused. Two others heard their partner express confusion verbally and the sixth observed confusion because of the mis-match of their respective viewpoints. They responded to the confusion by re-drawing the curve segments.

Five actors said something to their instructors when confused and the sixth actor said there was no confusion. Four pairs dealt with confusion with verbal responses and re-drawing the shape, and the fifth pair resolved it verbally.

5.7 Training and Task Times

Both the training in the use of the four guidance modes and the actual study task were self-paced, with the instructor working through a sequence of instruction pages shown on the tablet display. The mean training time (in minutes:seconds rounded to nearest 5 seconds) was 9:15 (range 6:15 to 12:30). The mean task time was 8:15 (range 5:20 to 14:10).

This variation in training and task times reflects the different approaches that each pair of participants took. Some pairs were content to complete a single pass at drawing a shape, especially in the training phase, whereas others wanted to refine the shape until the instructor was happy with it. One pair completed three paper drafts of a training shape before moving on. During the actual task this variation was also obvious. One pair drew the shape in essentially a single drawing pass, whereas another pair took elaborate steps to draft and refine the drawing of the shape.

6 Discussion

6.1 Pointing and Drawing

All six pairs made their choice of pointing and drawing tool (on-video or in-workplace) based on their perception of the differing attributes of the tool. In particular, they were able to see the trade-off between the precision of the on-video tool against the immediacy of the in-workplace tool, and were able to make agreed decisions about when to use which tool. This was particularly so for the pairs who drew an initial draft on paper, where the situation changed as they moved from initial drawing to modifying the completed draft shape and finally drawing onto fabric.

Four of the six pairs drew various forms of draft shape first and then they used editing instructions to bring the shape closer to the desired shape. This had the effect of converting the task from low-level following of line segments to a higher-level shape editing action accompanied by verbal shape descriptions and confirmation dialogue. It also removed the temporal urgency created by the 5-second fade of the drawn lines.

This understanding and ability, or willingness, to switch between different tools and modes suggests that providing multiple technologies to support remote physical guidance is worthwhile. Just as people can decide whether to verbalise, gesture or both they also seem able to decide between the four options offered in this experiment.

6.2 Guidance Strategies

The guidance strategies employed by the participants centred on creating a mutually visible artefact, initially a curve segment and later a full draft version of the shape, then conducting a discussion with words, drawings and directional gestures to converge on the required shape. This draft version of the shape corresponds to the "publicly available artefact" that Robertson refers to when she emphasises that the objects in a remotely collaborative environment need to be available to all of the participants (Robertson 2002).

Where the participants were able to recognise a describable shape component they used this higher level description to assist in editing the shape. Three instructors referred to part of the shape as "like a shark fin", which was easily recognised by their actors.

Guidance studies frequently use one or two trained instructors and multiple actors. In this study we frequently observed that the guidance strategy was often collaboratively established, both initially and as the task progressed. The actual guidance strategies would seem to depend on the initial training experience and on the personalities of the pair of participants, so we would expect quite a range of strategies to appear. This contrasts with other researchers' experiments, where a small number of previously trained instructors worked with a larger number of experimental participants, and where the strategies would be heavily influenced by the experience of the instructors.

6.3 Frames of Reference

Heath, Luff and colleagues discuss the importance, for participants in tele-collaborative environments such as this one, of being able to "determine the location, orientation and frame of reference of others" (Heath et al. 2001). In this experiment we see this importance played out in mixed ways. All six actors responded that they understood their instructor's orientation to the workspace, even though each instructor used directional words that were consistent with their own [instructor's] orientation but not consistent with their actor's orientation to the workspace. All six actors simply ignored the spoken words and followed the drawing gestures. Only one actor tried, unsuccessfully, to correct the instructor about this point.

We discussed this with a participant in our earlier guidance experiment, who had taken the role of "instructor". He noted that, as a trained teacher, he used terminology like "to your left" when talking to his partner. In this experiment we did not observe the instructors using their partner's body as a frame of reference.

Heath, Luff and colleagues (ibid) also refer to "stable reference objects" in the collaborative environment and the participants made frequent use of these. The primary reference object was the rectangle of fabric or paper on which the shape was to be drawn, together with the tip of the actor's pen and the display of the instructor's cursor and line. When a draft of the shape had been drawn it became both a stable reference object and a local frame of reference. The two participants conducted discussions with respect to the draft shape, and edits to it were made with reference to the inside and outside of the shape. It might be interesting to ask what they would have done if the shape's outline had not been topologically closed.

During the patch placement component of the task both participants demonstrated a comfortable grasp of the three-dimensional space around the foam-core dress and of its spatial orientation. The instructors pointed directly to the required location and used spatially appropriate words to rotate and adjust the patch.

6.4 Collaborative Behaviour

Collaborative behaviour emerged very early in this experiment, with the participants conducting the self-paced training task. This behaviour continued through the experimental task itself, finally reaching an agreed conclusion to the task. The behaviour was exhibited at the small scale with guidance, response and query gestures and words in the drawing of the curve segments, and at the larger scale with face-to-face video contact to discuss issues and make decisions.

The participants were also able to take their partner's situation into account when making suggestions about which guidance tool to use. This suggests that the training tasks had given them enough experience with the tools and enough observation of their partner's responses to effectively consider their partner's situation.

7 Conclusion

The study reported in this paper has shown the following:

- The participants are able to chose from the pointing and drawing options supported by the two guidance technologies based on the precision of each technology and the ease-of-use for their partner
- Guidance strategies when using these technologies can evolve from the base level of sequential drawing of curve segments to an effective higher level of

interaction centred on shared availability of the drawn shapes and the interaction tools

- The participants are aware of each other's frame of reference at the level of working with relative references to shared stable reference objects, and this level of working can override potential confusion about each other's spatial orientation towards the workspace.
- This technology (remote guidance technology in a shared media spaces setting) can support emerging collaborative behaviour between participants.

The participants in this study actively used and chose between the four available guidance options, and in doing so were able to successfully collaborate to complete the task. The implication for use of remote guidance technology in actual applications, therefore, is that providing a range of options is a design decision that should be seriously considered.

8 Future Work

The purpose of this study was to examine, in a relatively isolated situation, the way that these guidance technologies might be used in a relatively complex 3D task. Although the immediately intended target user group, from the tele-health project's point of view, consists of hospital-based clinical staff, the study was conducted with members of the general university and CSIRO population precisely because the hospital-based target group were not available for such a time-consuming exploratory study. This continues to be the case, and there are no plans to repeat such a laboratory study with clinical participants.

We do, however, intend to take the overall tele-health system into a hospital setting and to conduct pilot and clinical trials. In this context, the guidance technology will be a part of the overall tele-health task. While we will not have control over how the clinicians use particular aspects of the technology we will be able to observe how they use the technology in their clinical practice.

At the time of revising this paper for publication we have just completed a four-week pilot study at the Royal Children's Hospital, Melbourne, Australia, in which we used the tele-health system to conduct 45 surgical outpatient consultations. Each of these involved detailed examination of the patient by the clinic assistant under direction and guidance from the remotely located surgeon. This direction and guidance involved a wide range of delivery modes ranging from pointing and drawing gestures through to spoken high-level clinical instructions, all conducted in the context of close collaboration between surgeon, clinic assistant and patient. The data is currently being analysed and results from this data will be submitted for publication in the near future.

9 Acknowledgements

This work was supported in Australia by the Australian Government through the Advanced Networks Program (ANP) of the Department of Communications, Information Technology and the Arts and by the CSIRO ICT Centre. The authors acknowledge Dr Andrew Greensmith and Dr Wellington Davis at Royal Children's Hospital, Melbourne, and Dr John Lambert, intensive care specialist at Orange Base Hospital. They also acknowledge the support of their CSIRO colleagues and thank the volunteers who took part in the experiment. The experiment has been approved by the CSIRO Human Research Ethics Committee – reference number 06/41.

10 References

Bauer, M., G. Kortuem and Z. Segall (1999). Where Are You Pointing At? A Study of Remote Collaboration. *Proc Third IEEE International Symposium on Wearable Computers IEEE*.

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

Heath, C., P. Luff, H. Kuzuoka, K. Yamazaki and S. Oyama (2001). Creating Coherent Environments for Collaboration. *Proc. Seventh European Conference on Computer-Supported Cooperative Work.* Bonn, Germany, 119-138, Kluwer Academic Publishers

Kirk, D., A. Crabtree and T. Rodden (2005). Ways of the Hands. *Proc. Ninth European Conference on Computer-Supported Cooperative Work.* Paris, France, 1-22, Springer

Kirk, D. and D. S. Fraser (2006). Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. *Proc. CHI 2006*. Montreal, 1191-1200, ACM.

Kirk, D., D. S. Fraser and T. Rodden (2005). The Effects of Remote Gesturing on Distance Instruction. *Proc. Computer Supported Collaborative Learning*. Taipei, Taiwan, 301-310, Lawrence Erlbaum Associates

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

Kurata, T., N. Sakata, M. Kourogi, H. Kuzuoka and M. Billinghurst (2004). Remote Collaboration using a Shoulder-Worn Active Camera/Laser. *Proc. Eighth International Symposium on Wearable Computers* ICSW'04, IEEE.

Kuzuoka, H., G. Ishimoda, Y. Nishimura, R. Suzuki and K. Kondo (1995). Can the GestureCam be a Surrogate? *Proc. Fourth European Conference on Computer-Supported Cooperative Work*, Stockholm, Sweden.

Kuzuoka, H., T. Kosuge and M. Tanaka (1994). GestureCam: A Video Communication System for Sympathetic Remote Collaboration. *Proc. CSCW'94*. Chapel Hill, NC, 35-43, ACM

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

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

Palmer, D., M. Adcock, J. Smith, M. Hutchins, C. Gunn, D. Stevenson and K. Taylor (2007). Annotating with Light for Remote Guidance. *To appear in Proc. OZCHI 2007*. Adelaide.

Robertson, T. (2002). The Public Availability of Actions and Artefacts. *CSCW* **11**: 299-316.

Smith, J., D. Stevenson and M. Hutchins (2007). Exploring Interface Design for Remote Guidance. *CSIRO ICT Centre Internal Report 07/372*.

Yamazaki, K., A. Yamazaki, H. Kuzuoka, S. Oyama, H. Kato, H. Suzuki and H. Miki (1999). GestureLaser and GestureLaser Car - Development of an Embodied Space to Support Remote Instruction. *Proc. Sixth European Conference on Computer-Supported Cooperative Work*. Copenhagen, Denmark, 239-258, Kluwer Academic Publishers