PRELIMINARY EXPERIMENTS WITH A DISTRIBUTED, MULTI-MEDIA, PROBLEM SOLVING ENVIRONMENT

Randall B. Smith §†, Tim O'Shea †*, Claire O'Malley *, Eileen Scanlon *, and Josie Taylor *

†Rank-Xerox EuroPARC
61 Regent St.
Cambridge CB2 1AB
U.K.

*The Open University
Walton Hall,
Milton Keynes, MK7 6AA
U.K.

§ Xerox PARC
3333 Coyote Hill Rd.
Palo Alto, CA 94304
U.S.A.

Abstract

We report on studies of pairs of subjects using a system called SharedARK (for "Shared Alternate Reality Kit"). In SharedARK, users at separate workstations interact in real time with the same world of simulated physical objects. In the experiments, two users are in separate rooms with a workstation each, and communicate through a high fidelity, hands-free audio and a camera-monitor device called a "video tunnel" which enables eye contact. For comparison, we have removed the video tunnel for some subjects, and have placed other subject pairs in the same room, giving them only a workstation each. Within SharedARK, subjects are given a "microworld" within which to solve a problem in everyday physics. Subjects are videotaped, and monitored from a remote room. The protocols have been submitted to a preliminary analysis, in which we categorise activities as they relate to use of the interface, task performance and social interaction. We also catalogue eye glances and eye contact in terms of their relation to these activities.

Our primary purpose has been to assess the learnability and usability of this technology, and to identify factors that are important in facilitating collaborative problem solving by directly comparing remote, electronically mediated communication with physical co-presence. We find the system is easily learned and fairly easily used. Subjects normally find the task engaging, and usually work together through largely unarticulated task division to obtain at least a partial understanding of the solution space. Our observations have led us to the hypothesis that this technology can bring subjects in some ways closer together than if they were to engage in a similar task in the real world. We present evidence suggesting that this artificially enhanced proximity may play a positive role in supporting non-interface specific discourse and task division negotiation.

1. Introduction

This research project is motivated by three distinct technological and pedagogical concerns. First our past work on ARK, the Alternate Reality Kit [1], and our experiments with instructional physics simulations implemented in ARK [2], [3] convinced us of the value of interfaces based on physical metaphors. With the design of SharedARK [4], a distributed version of ARK, we were able to experiment with pairs of users who would use SharedARK while communicating from remote locations with an audio and video link.

Second, this line of inquiry is particularly relevant in light of increasing interest in remote work and in distributed education as offered by the various "distance education" institutions. In the U.K., the OU (Open University) now has about 200,000 students studying at home or at work at any time. The sort of technology described in this paper soon could be applied to educational systems like the OU to augment the instructional process. When the OU started 20 years ago, it was not reasonable to assume that students had a telephone or a colour TV. In 1989 over 11,000 students are using home computers as an integral part of their studying, and many of these communicate with other students via computer conferencing and electronic mail. In the near future, with the anticipated arrival of high bandwidth data networks, increasing numbers of individuals will be able to use technology like SharedARK and audio-video links. The computer and video technology described in this paper has already been used to bring users 600 miles apart into the same virtual world, using only two 64 kbps lines (which is within the ISDN standard.)

Our third concern is to understand how to design effective computer-based support for collaborative learning, remote or co-present. This work provides a follow on from studies of both individual and cooperative problem solving with similar tasks not involving the use of computers [5]. Positive effects of peer interaction on learning have been found in several studies [6], [7]. Similar positive results have been found in computer-based synchronous peer interaction [8], [9]. However, most of the systems in these studies were designed for individual users, rather than being designed specifically to support collaborative activity. In these cases it is difficult to see what role the computer is playing in supporting the collaboration. Some studies [8], [11] have involved manipulating the control that students have over input by structuring turn-taking, for example, but this seems a rather impoverished use of technology to support collaboration. Apart from one or two recent studies [12], there is little reported research on peer interaction that provides guidelines or principles for designing systems to support learning via collaborative activity.

The motivating question for the research is "What is different when members of a problem-solving pair are physically separated then reconnected via this type of computer and communications technology?" We have analysed nine pairs of subjects using this system in various settings, to investigate the possible ways in which multimedia technology and interface design affects performance on a collaborative problem solving task.

In our observations, we have varied the configuration of the technology in several ways in hopes of uncovering the role of various components on the working ways of the users. We find that in any configuration, subjects rapidly learn the interface, and usually find the task engaging, working without major difficulty towards a partial solution. There are, we believe, some evidence for effects of the configuration on the non-interface specific interactions of the subjects, and in particular on the way that task division is negotiated. We will report on these effects in section 4.

2. Experimental Infrastructure: SharedARK

SharedARK [4], a distributed, multi-user version of ARK [1], is a virtually unlimited two dimensional plane containing simulated physical objects. An arbitrary number of users at separate workstations interact in real time with the same animated virtual world (see Fig. 1.) In these experiments, two users are in separate rooms with a workstation each, and communicate through a high fidelity, hands-free audio link and with a camera-monitor device called a "video tunnel" which enables eye contact (see Fig. 2.) For comparison, we have removed the video tunnel for some subjects, and have placed other subject pairs in the same room, giving them only a workstation each (see Fig. 3.)

SharedARK provides the users with an essentially boundless two-dimensional plane. A user's rectangular screen enables him or her to see a portion of this plane. Users can "scroll," sliding their rectangular viewpoint over the two dimensional world. Because two users can scroll to different and arbitrary points in the plane, they may not see the same portion of the world: their rectangular views must at least partially overlap if they are to have access to some of the same objects. Thus, users can drag their rectangles into coincidence to work with the same objects, or can move them apart to work on separate portions of the task.

The objects in SharedARK each have an image, a velocity, mass, and the ability to respond to forces.. The users' mouse-operated hands can press buttons, or can grab, carry, and even throw these objects.. Users arrange objects, control devices, sketch diagrams, and make notes on virtual paper in a way strongly analogous to the real world.

In addition to the main rectangular view of their portion of the world, users have a "radar" view of the area surrounding their rectangle. This view is also animated and current, although objects appear in the radar view in reduced detail. When the user scrolls his or her rectangle over the plane, the radar view

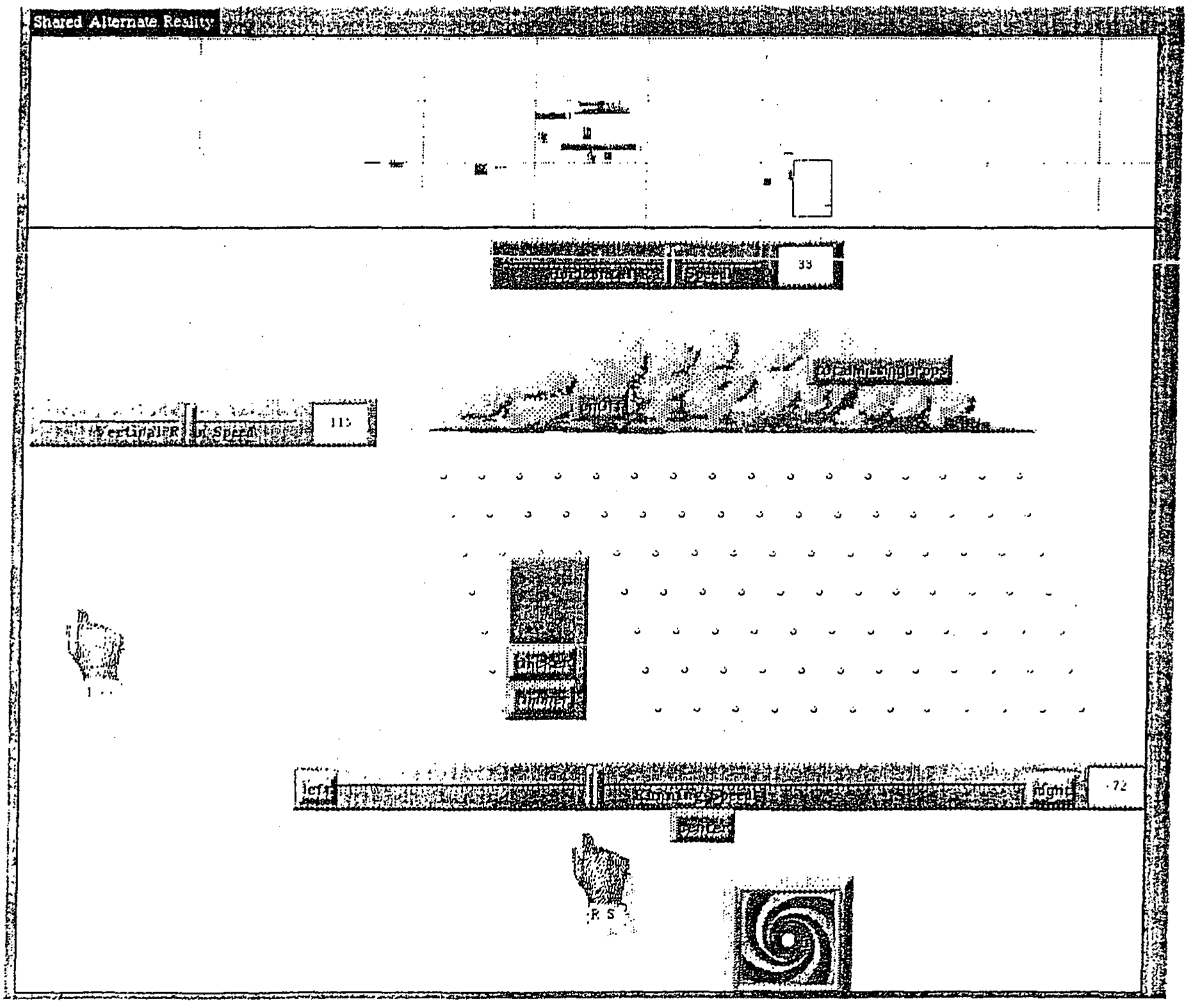

Figure 1. A typical SharedARK screen for the "running in the rain" simulation. Users have a mouse operated hand each. They can use their hand to press buttons, and to carry or throw objects. At the top of the window is a "radar view" in which a greater distance is visible. Users can scroll their view regions apart to visit remote parts of the two dimensional plane. At the bottom of the window is a teleporter, which enables users to jump through vast distances without scrolling.

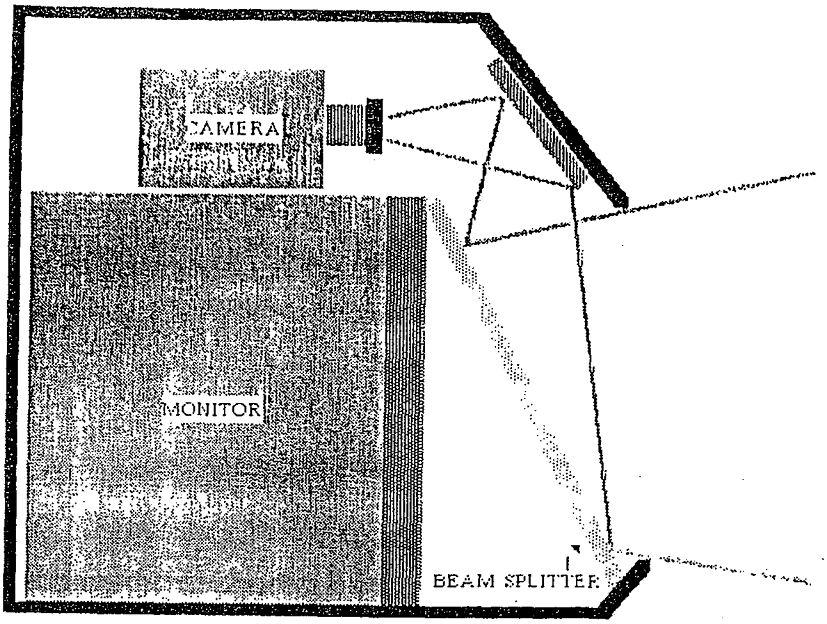

Figure 2. The "Video Tunnel." The effect of the mirror and beam splitter (or "half silvered mirror") is to place the camera's point of view at the center of the monitor. This enables the users to align their line of sight when looking at each other, resulting in a sense of eye contact. The entire arrangement is inside a box which is placed along side the computer screen.

scrolls as well, so that it always remains centered on the user. The edges of one's rectangle, representing one's main view of the world, are visible in the main view of other users. This has been important in informing users of what objects are mutually visible and accessible.

The video tunnel is placed along side the workstation, on the left side for one user, on the right for the other (this enables one's gaze toward the screen to be properly interpreted by one's partner.) Research on interpersonal interaction has demonstrated the importance of eye-contact (for an overview see [13]) and it is for this reason that we prefer the video tunnel over an arrangement with camera and monitor separated. Both the direct manipulation computer world (SharedARK) and the video tunnel are intended to allow users to employ everyday physical and social intuitions.

3. The Experiment

Task

The subjects are first given a eight to ten minute introduction to the interface and task. An experimenter sits at one of the workstations in the place to be occupied by one of the subjects, who looks on. The other subject sits at the appropriate workstation. The experimenter shows the subjects how to use the mouse controlled simulated hand to pick up objects, throw objects, press buttons, and alter slider settings. They are given a brief explanation of how to navigate using the radar view, how to use the "teleporters" for quick movement to "distant" locations, and are shown how they can see from a rectangle that appears around their collaborator's hand what their collaborator is looking at and the extent to which their two screens overlap.

During this introduction, they are told that the object of the activity is for them to jointly agree when it is worth running in the rain. They are shown how to make the runner slimmer or plumper, how to make the runner move, how to adjust the rate and angle of rainfall, and how to switch the rain on and off. Subjects who ask for a more precise problem statement are told that the problem is deliberately underspecified and that part of the task will be to refine the problem so as to make it tractable. The subjects are told that they have "about" an hour. In fact the experiementers continuously monitor the subjects and ask them to summarise their conclusions after they have reached agreement. Some pairs took nearer two hours than one hour and where given a "ten minute warning".

The solution to the "running in the rain" problem has been fully examined by DeAngelis [14]. For the case of vertical rain, it is always best to run as fast as possible, though anything faster than a brisk walk is not very helpful. For the case of rain falling from behind, the solution is more subtle: run with the horizontal component of the rain's velocity unless the ratio of your thickness to height exceeds the ratio of the rain's horizontal to vertical velocity components. (If this ratio is exceeded, then again it is best to run as fast as possible.) None of the subjects were able to fully solve the problem, though they typically did much better than 31 mathematics and physics students who were asked to answer the question given only pencil and paper.

Subjects

We recruited eleven pairs. They were selected on the basis of availability, primarily from the Open University. They were not matched except that where there was an opportunity to use pairs that did not know each other very well this was taken. Potential subjects who were familiar with the rain-running task or with the objectives of our programme of experiments were not selected. Fifteen of the subjects are postgraduate research students, five subjects carry out technical jobs and the other two have a managerial role. All the subjects except one had previously used a mouse. Five pairs were colleagues who knew each

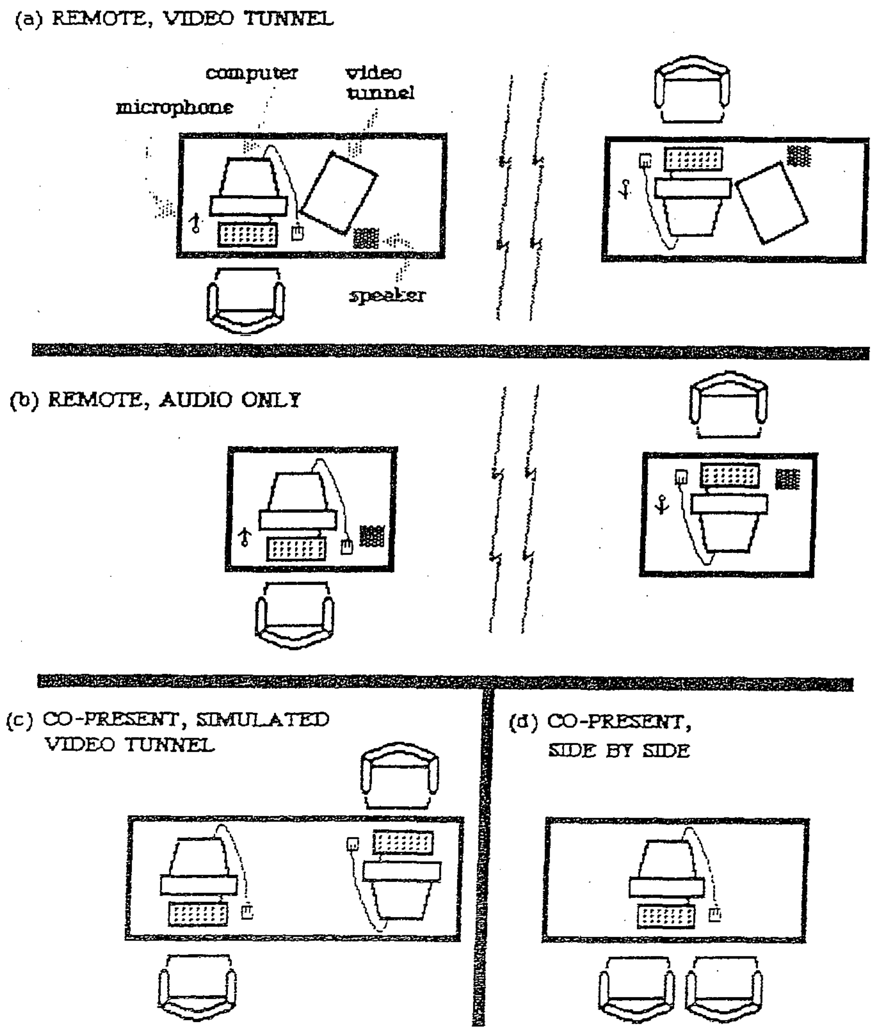

Figure 3. Various settings for this experiment. In the primary setting, (a), users are in separate rooms, but are connected by audio and video links and the common virtual world in the computer. In (b), the video tunnel has been removed, and in (c) and (d) the users are in the same room.

Table 1. Time spent in various activities and associated eye-contact frequency. The nine subject pairs are indexed across the top, and the activity categories are listed along the left.

		Remote + VideoTunnel subject pairs				Co-present subject pairs			Remote + Audio subject pairs	
		1	2	3	4	5*	6	7	8	9
Percentage of time spent in category	1a	25	37	16	6.3	11	9.0	25	21	13
	1b	14	6.2	4.9	18	7.0	15	28	34	20
	1c	--	--	--	--	--	6.8	3.5	6	3.6
	2a	8.2	9.4	4.8	10	1.2	24	1.3	7.0	9.5
	2b	53	47	69	60	69	44	42	27	44
	2c	--	--	--	--	--	0.0	0.0	2.2	9.8
	3a	--	0.0	3.6	1.7	3.3	0.0	0.0	0.3	0.0
	3b	--	0.4	1.2	4.0	8.8	0.5	0.0	1.8	0.5
	3c	--	--	--	--	--	0.0	0.0	--	0.0
Frequency of eye-contacts per category (per min)	1a	4.9	1.9	0.7	1.7	0.3	0.8	2.0		
	1b	2.2	1.8	0.3	3.5	0.8	1.3	3.2		
	1c	--	--	--	--	--	2.0	3.9		
	2a	15.2	13.6	8.9	11.5	5.0	16.8	25.4		
	2b	7.8	5.8	3.0	9.4	1.7	6.9	7.2		
	2c	--	--	--	--	--	0.0	0.0		
	3a	--	0.00	1.8	9.7	3.5	0.0	0.0		
	3b	--	5.9	2.7	6.5	5.5	28.2	0.0		
	3c	--	--	--	--	--	0.0	0.0		

Categories

a Meta-level
b Specific
c Recovery from breakdown /error

1 Interface
2 Task
3 Social

-- indicates no data
* This pair was side-by-side

other well, but would not normally associate out of the work environment, one pair knew each other less well, two pairs were made up of couples who were only introduced just prior to the experiment, and one pair met for the first time via the video tunnel.

Method

Two pairs were treated as pilots, leaving nine pairs used for analysis. Four pairs were recorded in the "primary setting" (Figure 3(a)) with the subjects in separate rooms, communicating over an audio link with video tunnels. Each subject could of course work with partially overlapping SharedARK screens. The data was recorded with four video sources, two from the video tunnels and two from cameras recording the individual SharedARK screens. This data is easy to work with as it appears as a two by two array stored on a single videotape. Each row is readable as a subject and the screen they are working with. Eye-contact is signalled by the two subjects looking simultaneously through the video tunnels at the viewer of the data.

We recorded two co-present pairs working with a "simulated video tunnel" (Figure 3(c).) That is to say, they worked with two Shared-ARK workstations arranged back-to-back on a table with a physical gap (the "simulated" tunnel) through which they could talk and gesture. One co-present pair working with a single workstation was recorded and analysed (Figure 3(d).)

The third contrasting setting was two pairs working in separate rooms with an audio link and SharedARK screens, but no video tunnel (Figure 3(b).) A summary of the data from the nine analysed pairs is given in Table 1.

4. Observations

Analysis

We have conducted a preliminary analysis of the videotapes where we categorised subjects' activities, associated discourse, and eye-contact behaviour according to two sets of factors (see Table 1). The first set of factors describes the type of activity in which subjects were engaged, according to whether it involved the interface (e.g., figuring out how to use the sliders), the task itself (i.e., running in the rain) or social interaction (e.g., telling jokes). These factors are listed in Table 1 as (1), (2), and (3). The second set of factors describes the nature of the activity in the first set according to whether it was concerning meta level activity (e.g., generating hypotheses, discussing problem solving strategies), specific activity (e.g., talk generated in doing the task) or recovery from breakdown (e.g., conversational repair, recovery from interface errors). These factors are listed in the table as (a), (b), and (c). Combination of these factors leads to nine separate categories. Table 1 shows the data according to percentage of time spent and associated frequencies of eye-contact behaviour for each category. It should be kept in mind that some of these categories (in particular, the category of recovery from breakdown) were generated part-way through our analysis, so some of the cells in the table are blank. However, the time spent in such categories is usually quite small, so their inclusion from the beginning is not likely to have had a very large effect. Finally, because this is a report of a pilot study and our sample sizes are small, we have chosen to present the raw data rather than means.

Factors affecting learnability and usability.

The use of the video tunnel was obvious to the subjects. All the subjects who used the tunnels were able to use their hands for explanations in a quite natural way, indicating affect and indicating directionality important in using the interface (e.g.,"up here",) and important in solving the problem (e.g. ,"fast rain at 45 degrees").

A few subjects did not initially use the audio in a natural way (e.g. "can you hear me, over") but quickly learned to speak as if they were in the same room as the other subject. While the personalities and relationship of the two subjects in the pair was important in determining how they worked together, it did not seem to be a limiting factor in the aspects of the technology the pairs eventually learned to use.

All the subjects quickly learned how to use their simulated hands to pick up and move objects and to operate the SharedARK sliders and switches. Obviously prior use of a mouse was helpful but our earlier experiments with ARK [2] have shown that this is not essential . Subjects were sometimes momentarily confused by the movement of the "other" hand when they were moving their own but with one exception this confusion would be quickly solved by stopping their own hand moving or by checking the initials on the cuff of the hand. The exception was one subject who somehow had come to believe that the other subject had effective control of everything. When she made a mistake in operating her hand she attributed the consequence to some action of the other hand. Nearly all the subjects learnt to gesture in a comfortable way with their simulated hands.

Some subjects had a problem following the radar view and keeping track of what the other subject could see on their main SharedARK view. A few subjects also had some difficulty in understanding the operation of the SharedARK teleporters. However, such problems were readily solved with a small amount of instruction. Nearly all the subjects found it hard to make notes of their experiments and data as they followed the task. This is an area which could only be solved by improving the implementation of SharedARK as the drawing and note-taking facilities currently provided are rudimentary, and the somewhat slow animation rate makes sketching with the virtual hand-held markers particularly difficult.

Considering the complexity of the communications hardware and simulation software, it is encouraging that all the pairs mastered its use sufficiently well to tackle a non-trivial problem after eight to ten minutes of instruction. Our hypothesis is that this ease of learn-of-use arises from the physical metaphor associated with the video tunnel (a hole in the wall) and the simpler physical metaphors associated with SharedARK (a pair of hands in a flat plane picking up objects, switching switches, throwing, dropping, etc.) The aspects of SharedARK that use more obscure metaphors (radar, teleporters) are correspondingly harder to learn to use.

How and when is video used?

Video is an expensive communication medium, so we were particularly interested in observing how and when it is used in this kind of setting. We noticed two phenomenon that we did not anticipate: a correlation with non-interface specific activity, and "chaotic glance patterns."

"Chaotic glance patterns" refers to what seemed at first to be an unusual pattern of gaze over time. For example, both subjects may be looking at the computer screen when one subject suddenly looks into the tunnel, then back towards the computer. A few seconds later, the other subject might look into the tunnel, then back towards the computer. This cycle of missed glances might recur irregularly until the two look into the tunnel at the same time. At this point, mutual gaze "locks on," and the subjects proceed to interact through the video medium.

Our first thought was that this was somehow an artifact of the video tunnel. However, we found that subjects in the "simulated tunnel" arrangement exhibited similar gaze patterns. We therefore suspect that this is a general phenomenon arising in any situation in which subjects' attention is split between work surfaces and a readily accessible line of gaze toward collaborating partners. This kind of interaction is reminiscent of the lateral "stutter step" encounter of people avoiding collision as they approach in a hallway. Such chaotic patterns as a function of time can be explained by general theories of cooperating individuals acting with delayed response [15].

Our second observation has to do with how use of the video channel was correlated with activity outside direct manipulation of the interface. Perhaps one of the first things one notices about the subjects' use of the video link is that there are long periods (lasting up to half a minute or so) in which it is not used at all. While the subjects carry out a set of manipulations in order to collect a "data point," for example, their eyes are normally on the computer screen so that they can manipulate objects, adjust controls, or simply watch the progress of the experiment. On the other hand, when subjects talk about what they observe, formulate hypothesis, clarify questions, or suggest strategies, they typically look towards their partner.

One can see from the values in Table 1 that meta level discourse about the task (category 2a) is associated with the highest frequencies of eye contact. We've noticed that subjects sometimes lean right into the video tunnel during meta-level discourse about the task. This is usually when the subject was trying to convince the other about a hypothesis, or when trying to establish eye-contact. Doubt by the other subject was sometimes signalled by brief covert glances from the screen to the video tunnel. When subjects were both looking at the screen while engaged in meta-level talk, one would often glance quickly at the video tunnel to check whether or not their partner agreed.

Meta-level task-related activity also tended to be accompanied by a fairly high degree of non-verbal cues such as gesture and facial expression. Often, subjects in the video tunnel condition seemed to forget that their partner could only see a "head-and-shoulders" view and would start pointing at the screen. Sometimes subjects would realise the error and comment on it, then try to gesture instead with their "virtual hand" in the artificial world. On occasion this sort of mistake would also occur with subjects in the audio-only condition, but seldom with subjects in the co-present condition. Subjects in the video tunnel condition were also observed sometimes gesturing at their partner through the tunnel with one hand, while gesturing with their virtual hand (i.e., the cursor) with the other hand. On these occasions their partner would look at the screen but glance back at the video tunnel every so often to monitor the other's non-verbal cues.

We have also seen times when subjects made use of peripheral awareness of the video tunnel. Subjects would gesture at the video tunnel while both were looking at the screen. On one occasion, when a subject looked behind her, her partner, without looking up from the computer screen, asked "is someone else there with you?" Other times we've seen a subject, who is talking, stay looking at the computer screen but lean his or her body or angle the head towards the video tunnel, almost as a substitute for eye-contact.

With all the non-verbal activity we see associated with the video channel and its correlation with activity outside specific use of the interface, one might expect it to influence the activity of the collaborators.

<u>Does video influence types of activity?</u>

We hypothesised that a video channel (the video tunnel and co-present conditions) might encourage non-interface specific activity. Non-interface specific activity includes everything other than category 1b in Table 1. Category 1b (interface specific) represents direct operation of the interface (pressing buttons, moving objects) and contains dialogue such as "I just turned off the rain," "Okay, I will pick up the runner." A video channel may encourage activity beyond this mechanical level by reducing the distractions of the immediate task within the microworld.

Activity other than interface specific is important. Research in peer interaction suggests that collaboration facilitates learning when students are actively engaged in discussing and exploring alternative hypotheses [8], [10]. This process of critical argument makes it less likely that students will get stuck in "local minima" in solving problems and more likely that they will reflect on their strategies for solving the problem [5], [16].

This hypothesis has some support in the data: according to Table 1, the audio only pairs spent more time in interface-specific activity than did the other conditions. We cannot take these figures too literally, as our sample size does not justify this level of interpretation across pairs. Comparison is further confounded by an informal observation: the nature of the discourse in the video tunnel and co-present conditions was qualitatively different to the discourse in the audio-only condition, in that subjects who could see each other tended to use much more terse and less explicit discourse.

Further support for the hypothesis comes from the observation that meta-level discourse related to the task was accompanied by much higher levels of eye contact than was specific talk about the task or meta-level talk about the interface. In every pair, the highest frequency of eye-contacts was associated with meta-level task related activity (row 2a of Table 1.) These values are usually two or three times higher than those associated with specific task related activity (row 2b), which was normally the second highest value.

To summarize, we feel we have evidence for saying that the addition of a video channel to a remote collaborative technology does influence the users' activity by encouraging interactions about the problem, as opposed to those specific to the operation of the interface. Because discussions about the problem can be important, video may have a positive effect on collaborative learning.

Problem solving strategy and task division

Studies of peer interaction among children [8], [11] have shown that tasks requiring some form of turn-taking result in improved performance. Studies of adults have also found that successful cooperative activity involves task or role differentiation. For example, Miyake [17] found that while one subject led the interaction by engaging in a local task, the other observed from a more global perspective and provided help by criticising and suggesting new ways of approaching the problem. Similar patterns have been found in studies by Garrod and Anderson [18]. However, others have argued that marked role differentiation may have disadvantages for successful peer learning, in that children may learn only part of what is required to complete the task [19], [20]. As Blaye et al. [9] suggest, further analysis of distributing cognitive load by role differentiation needs to concentrate on ways in which software or interface characteristics are associated with particular patterns of role exchange and how such patterns evolve.

In our earlier studies [5] of joint problem solving, we found that role divisions and role exchanges appeared naturally during the interaction, without much need for explicit negotiation. We were interested in the extent to which collaborative technology might affect task or role division. In general, we have found that the SharedARK interface (in particular, the use of overlapping (shared) and non-overlapping work areas) encourages role division and role exchange. Subject pairs in all conditions would alternate their roles about 50 percent of the time.

In terms of how task division was achieved, both the video tunnel and the side-by-side co-present conditions seemed to involve less explicit negotiation than the other two settings. Roughly 50% of task divisions in the video tunnel condition were explicitly negotiated, whereas about 75% of task divisions were explicit in the co-present and audio-only conditions. Our sample sizes are too small to support strong claims. However, we suggest a tentative explanation for the direction of the results. It is not very surprising that subjects in the audio-only condition would need to negotiate task divisions more explicitly: subjects in the other two conditions can make use of non-verbal channels of communication in parallel with verbal means -- a simple nod or moment of eye contact is sufficient to satisfy parties that they are in agreement. What is surprising is that the"simulated video tunnel" pairs did not exhibit this reduced negotiation of task. Again, we are wary of individual differences between pairs, but we do have the impression that subjects in the video tunnel condition sometimes behave as if their partner were seated beside them, with the same view of their computer screen, which may explain why their task division negotiation is more similar to the co-present, side-by-side pair. This is supported to some extent by the finding that subjects in this condition quite often point at the screen, even though their partner can't see this. This hardly ever happens in the co-present "simulated video tunnel" condition, where subjects can see that their partner's workstation is turned the other way round, a constant reminder that the pair may have different views on the virtual world. We are suggesting that a sense of being side-by-side is important for achieving fluid task division.

There were two main types of task division in these sessions. One, which we will call "within-task division", involved dividing the task so that one subject varied the speed of the "runner" through the rain, while the other subject controlled the amount and direction of rain falling. In general, these divisions were largely unarticulated, particularly in the video tunnel condition. The second major type of task division, what we shall call "between-task division", involved dividing the task so that one person ran the "experiment" (i.e., sent the runner through the rain), while the other person moved off to a separate part of the SharedARK world to record the results of the experiment. This type of division was closely related to Miyake's "local/global" divisions, in that the subject who was recording the data tended to adopt an overview of the task as the data came in, and directed the other subject to vary parameters in order to test hypotheses. These divisions were accompanied by more explicit negotiation than within-task divisions and were accompanied by more eye-glances and eye contacts. This makes sense, in that subjects engaged in within-task divisions were in the same virtual workspace, whereas subjects engaged in between-task divisions were in separate areas of the SharedARK world.

There were occasional breakdowns in negotiating both types of task division. In general, subjects in the video tunnel condition seemed to do much better in negotiating successful task divisions than subjects in the other conditions. One exception to this was a pair in the video tunnel condition who tended to get in each other's way, by trying to operate the same control at the same time. This led to some frustration, and one of the subjects tried to defuse the situation with a joke about touching hands. This same subject pair tended to be in conflict about what variables were important to manipulate and on occasions the subjects seemed to be engaged in completely different tasks to each other. The pair only began to get somewhere with the task when one of them suggested explicitly that they divide up the task.

Finally, one pair in the co-present "simulated tunnel" condition seemed to be very successful in negotiating a task division, but their choice of appropriate divisions was rather bizarre. For example, one person decided to be "in charge" of the motion switch, which left the other with most of the work to do in running the experiment and resulted in poor progress in the task.

In conclusion, we hypothesize that the video tunnel condition, which enables gesture and eye-contact while giving a sense of being side-by-side, provides good support for fluid task division.

5. Altering proximity relations

Consider the social distance between a pair of individuals working together in the same room. At times the two are somewhat removed: one may be silently looking out the window while his partner is bent over a table jotting notes. At times, the pair may be close together: they may look each other in the eye, shake hands, or give a pat on the back. We have not sharply defined the concept of social distance, but appeal to an intuitive notion of proximity that varies throughout the working session. In figure 4, we represent the variation of time spent at various distances in a real world setting by the dark curve.

When a technology is introduced as the medium between the two parties, one might expect there to be a sense of greater social distance. The loss of quality in the audio and video channels, the limited visibility of one's collaborator, and the two dimensional representations on the computer and video screen are just a few of the cues that remind a user that his or her partner is not really there.

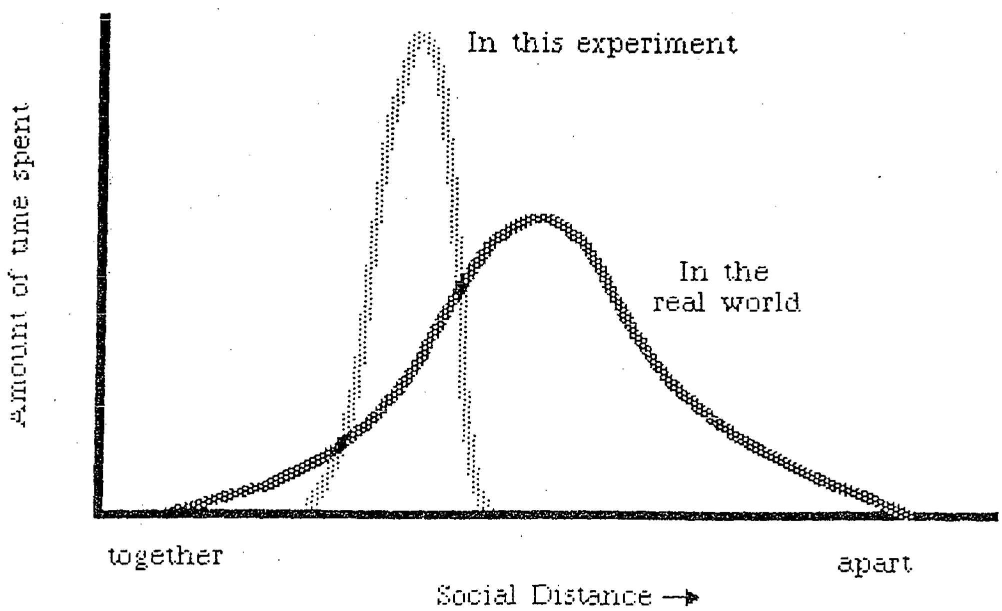

Figure 4. Proximity Relations. In the primary setting involving remote users and the video tunnel, subjects are unable to physically touch, or to increase their separation without breaking off the session: the range of social distances is more limited than that normally available in the real world. Also, users work "side by side" in the computer world, while almost under the direct gaze of their partner. Thus the average social distance may actually be decreased over that of a real-world setting. While enhanced proximity might exaggerate the discomfort of some user pairs, we believe it can facilitate non-interface specific discourse and task division.

However, the SharedARK virtual space (in the "primary setting" with video tunnels) represents a topology that is impossible to achieve in the real world. One's collaborative partner is represented as a disembodied hand unconnected to the face which is seen off to the side. This topology *decreases* social distance in two ways:

- First, as noted in our previous discussion of task division, it is possible to be "side-by-side," facing the same working surface within the computer while being almost face-to-face (through the "video tunnel.") Being side-by-side and being face-to-face both reduce the sense of social distance: in this technology, both are achieved simultaneously. Being side-by-side in the real world puts one's partner ninety degrees off to the side. Note that in Table 1, the frequency of eye-contact for the real world side-by-side pair is much less than for any of those with a video tunnel.
- Second, it is possible to cause one's virtual hand to pass through one's partner's hand. This enables subjects to swap positions rapidly and smoothly, and allows the hands to overlap while working on objects close to each other. Several subjects commented on how disquieting it was to have their virtual hand virtually touched by their collaborator.

Because of these factors, we have a sense that SharedARK's topology may actually reduce the average social distance. The distribution of social distances in the primary setting for this experiment is represented in Figure 4 by a lighter curve, with its peak at smaller values. Proximity is not necessarily an inherently good or bad thing. We have seen subjects who have worked together well, and subjects who were at times obviously uncomfortable working together: a reduced social distance may have amplified these variations.

A collaborative technology can also influence the *range* of social distances available. In this experiment, the range of social distances supported by the technology is limited. For example, it is impossible for the subjects to physically touch, or to enter another's real-world "personal space." (When a subject leans into the video tunnel in an apparent effort to get physically closer to the collaborator, the collaborator is not likely to notice much difference.) At the same time, it is impossible for the subject to step back from his or her position at the workstation without essentially leaving the virtual space completely, due to tight constraints on camera angle and microphone sensitivity. Both the closer social distances and the more removed ones cannot be reached in this technology. Thus the lighter colour curve in Figure 4 is pinched about its central value. There are a number of ways in which SharedARK could be configured to express a greater range of social distances. As an example, the audio and video link, which is under computer control, could switch on and off as subjects come together and move apart on the extended two dimensional space within the computer.

We wish to emphasize that distortion of social proximity relations is not necessarily a bad thing, and may in fact be associated with facilitating higher-levels of discourse and task division as discussed above. Individuals and the task may well dictate what is appropriate use of the collaborative technology in this regard. We suggest that all collaborative tools influence social distance, and this may be an important factor to consider when designing these technologies.

6. Conclusions

Summary

Our study has convinced us that the use of SharedARK, audio links, and video tunnels does not attenuate the problem-solving behaviour of pairs working on the rain problem. The remote video and audio setting is so "natural" that the experimenters (who are madly dashing about at the start and end of the experiments) are often unable to remember within minutes of a conversation with a subject or other experimenter whether they talked "through" the tunnel or were in the same room. Even though physically separated, the subjects in the remote video tunnel setting gesture, talk, and use eye contact in normal ways. Apart from one incident in the "simulated tunnel" setting where a subject leaned across and pointed at a button on the other subject's mouse, we saw nothing done co-present that could not be done in the remote setting.

The subjects quickly learn to use the SharedARK simulations and come to terms with the shared audio-video-computer space in which they find themselves. The shared space created by this technology places subjects into a kind of enhanced proximity in which it is possible to be simultaneously side-by-side and face-to-face. Our findings indicate that being side-by-side can facilitate task division, and that having a visual communication medium encourages interaction outside of interface-specific activity. Both of these kinds of behavior are thought to be important in facilitating collaborative learning.

Further Work

Our observations have lead us to make several hypothesis that could be studied in greater detail. In particular, our observations about non-interface specific activity in general and about task division negotiation in particular could be put to more solid tests with larger populations, or with specialized settings designed for "within subject" variations.

Although this prototype implementation was adequate for our needs, future implementations could benefit from higher animation rates, which we believe would better support drawing and gesture with the virtual hand. Workstations with video windows are becoming available, and it would be natural to integrate the video and computational media. However, if both computer screen and video are to be compressed into the forward direction, care should be taken to insure that direction of gaze of the remote partner can be correctly interpreted, and, ideally, that eye contact can occur.

Interesting directions to consider might include collecting data from settings that supported three remote subjects with three separate SharedARK screens. Also, the use of localised sound could help participants keep peripherally aware of their collaborator's activities and location. Finally, we should point out that the supporting software and environment make it fairly painless to investigate of a range of collaborative interface styles. It would be possible to engineer for different task and social settings, and to test new theories of collaboration as they arise.

There is scope for improvement, particularly by the addition of extra facilities or performance tuning for SharedARK, but this study demonstrates that distributed problem solving can be supported by appropriate technology without attenuation, and that people can enjoy using such technology.

Acknowledgements

We wish to thank the EuroPARC support staff and Bill Buxton for hours of help with audio visual and recording equipment. We also thank Rank-Xerox EuroPARC and the Open University for supporting this work, and our subjects for generously giving of their time.

References

[1] R. B. Smith, The Alternate Reality Kit: an example of the tension between literalism and magic, Proceedings of the CHI + GI87 Conference on Human Factors in Computing Systems (April 5-9, Toronto), New York: ACM, 1987.

[2] Tim O'Shea & R. B. Smith "Learning physics by violating the laws of nature in the Alternate Reality Kit" In preparation.

[3] E. Scanlon & R. B. Smith "A rational reconstruction of a bubble chamber simulation using the Alternate Reality Kit," *Computers and Education*, London: Pergamon Press, 1988.

[4] R. B. Smith, "A prototype futuristic technology for distance education," Proceedings of the NATO Advanced Research Workshop on New Directions in Educational Technology, Cranfield, UK, November 10-13, 1988.

[5] C. O'Malley & E. Scanlon, "Computer-supported collaborative learning: problem solving and distance education," Proceedings of the CAL89 Conference on Computer-Assisted Learning, University of Surrey, 1989.

[6] W. Doise & G. Mugny, *The Social Development of the Intellect*, London: Pergamon Press, 1984.

[7] A-N. Perret-Clermont, *Social Interaction and Cognitive Development in Children*, London: Academic Press, 1980.

[8] P. Light, T. Foot, C. Colbourn & I. McClelland, "Collaborative interactions at the microcomputer keyboard," *Educational Psychology*, 7, pp. 13-21, 1987.

[9] A. Blaye, P. Light, R. Joiner & S. Sheldon, "Joint planning and problem solving on a computer-based task," CHDL Occasional Paper No. 89/5, Centre for Human Development and Learning, The Open University, 1989.

[10] C. Crook, "Computers in the classroom: defining a social context", in J. Rutkowska & C. Crook (Eds.) *Computers, Cognition and Development*, Chichester: Wiley, 1987.

[11] A. Blaye, *Confrontation socio-cognitive et résolution de probléme*, unpublished doctoral thesis, University of Provence, 1988.

[12] J. Singer, S. Behrend & J. Roschelle, "Children's collaborative use of a computer microworld," Proceedings of the CSCW88 Conference on Computer-Supported Cooperative Work (Portland, Oregon), New York: ACM, 1988.

[13] K. R. Scherer & P. Eckman, *Handbook of methds in nonverbal behavior research*, Cambridge, England: Cambridge University Press, 1982.

[14] DeAngelis, "Is it really worth running in the rain?" *European Journal of Physics*, 8, pp. 201-202.

[15] B. A. Huberman and T. Hogg, "The behavior of computational ecologies," in B.A.Huberman, editor, *The ecology of computation*, pp 77 - 115, Amsterdam: North Holland 1988.

[16] C. O'Malley, "Understanding explanation," Proceedings of the Third Alvey Workshop on Explanation (University of Surrey), London: IEE, 1987.

[17] N. Miyake, "Constructive interaction and the iterative process of understanding," *Cognitive Science*, 10, pp. 151-177, 1986.

[18] S. Garrod & A. Anderson, "Saying what you mean in dialogue: a study in conceptual and semantic co-ordination," *Cognition*, 27, 1987.

[19] C. Hoyles & R. Sutherland, *Logo Mathematics in the Classroom*, London: Routledge, 1989.

[20] K. Sheingold, J. Hawkins & C. Char, "'I'm the thinkist, you're the typist': the interaction of technology and the social life of classrooms," Technical Report no. 27, Bank Street College of Education, New York, 1984.