Shared Workspaces: How do they work and when are they useful?

Steve Whittaker — Hewlett Packard Labs, Bristol

Erik Geelhoed — Hewlett Packard Labs, Bristol

Elizabeth Robinson — Psychology Department, Birmingham University

Abstract

We investigated the effect on synchronous communication of adding a Shared Workspace to audio, for 3 tasks possessing key representative features of workplace activity. We examined the content and effectiveness of remote audio communication between pairs of participants, who worked with and without the addition of the Workspace. For an undemanding task requiring the joint production of brief textual summaries, we found no benefits associated with adding the Workspace. For a more demanding text editing task, the Workspace initially hampered performance but, with task practice, participants performed more efficiently than with audio alone. When the task was graphical design, the Workspace was associated with greater communication efficiency and also changed the nature of communication. The Workspace permits the straightforward expression of spatial relations and locations, gesturing, and the monitoring and co-ordination of activity by direct visual inspection. The results suggest that, for demanding text-based tasks, or for complex graphical tasks, there are overall benefits in adding a visual channel in the form of a Workspace. These benefits occur despite the costs involved in attempting to coordinate activity with this unfamiliar form of communication. Our findings provide evidence for early claims about putative Workspace benefits. We also interpret these results in the context of a theory of mediated communication.

Running head

Shared Workspaces evaluation.

Introduction

Recent changes in organisations and work practices mean that it is increasingly important to provide effective methods for supporting remote collaborative work. Three trends are distancing coworkers in both space and time. The first is telecommuting, with a number of corporations experimenting with employees working from home (Harkness, 1977; Kelleher and Cross, 1985; Kraut, 1989). There are also increases in people working away from the office, eg. from customer sites or from hotel rooms while travelling (Sproull and Kiesler, 1991). A final trend is the globalisation of manufacturing, with designers, suppliers and manufacturers being increasingly located in different countries and time zones (Johansen, 1984; Sproull and Kiesler, 1991). Given the importance of interpersonal communications in the workplace (Kraut, Egido and Galegher, 1988), these trends point towards the need for technology to support high quality remote interpersonal communication.

A current pervasive technology for supporting remote interpersonal communications is the telephone. It is cheap and easy to connect remote collaborators by telephone, but for some purposes this may be inadequate or inefficient. A major aim of this study is to identify those situations where audio alone is insufficient and to test the impact of new communications technologies in those situations. One problem with the intuitive application of technology, however, is that mediated communication differs in psychologically important ways from face-to-face communication, and it is not always clear a priori when adding media to audio will provide communicative benefit (Noll, 1992; Rutter and Robinson, 1981; Williams, 1977). We therefore need to understand the principles of mediated communication if we are to meet the new set of remote communication requirements effectively. One way of advancing our understanding is to identify the sets of tasks for which particular mediated communication systems are most or least effective. By comparing the requirements of these tasks with the communication properties afforded by certain combinations of media we should gain insights into mediated communication principles. In addition, by comparing new technology with existing technology, we will also have clear ideas about where the new technology can be most effectively applied. A final objective of the study is to produce guidelines for the designers of new communications technology, with the aim of minimising the opportunities for communication failure arising from the properties of new systems.

We investigate these issues for Shared Workspaces, in the form of an electronic whiteboard on which remote collaborators can draw or write (Bly, 1988; Minneman and Bly, 1991; Stefik, Foster, Bobrow, Kahn,

Lanning and Suchman, 1988; Tang, 1991, Whittaker, Brennan and Clark, 1991). Despite extensive interest in Shared Workspaces and various claims for their utility, no recent research has identified the set of tasks for which Workspaces combined with audio are more efficient than audio alone, nor a theory of why such advantages might occur. In addition, there is no clear set of applications to which Workspaces have yet been commercially applied. On the other hand, there is a large body of early research showing that adding visual media to audio does not lead to improvements in communication efficiency for a variety of office tasks such as information transmission and collaborative problem-solving (Chapanis, Ochsman, Parrish, and Weeks, 1972; Chapanis, 1975; Reid, 1977, Williams, 1977). Initial research into Workspaces has focussed on their technological feasibility, design and on documenting the ways in which they are used (Bly, 1988, Minneman and Bly, 1991, Tang, 1991, Whittaker et al, 1991). Little work so far has evaluated the weaknesses of existing phone technology, in identifying key tasks where adding a Workspaces is maximally useful, and giving theoretical explanations for these findings. This study addresses these three issues, while attempting to extend and replicate that initial work.

Some early research on Shared Workspaces began with the Workspace and investigated the impact on communication of adding other media to it (Gale, 1989; Whittaker et al., 1991). While this approach might enable designers to improve the Workspace system by incorporating additional media links, it does not help us to specify the circumstances in which Shared Workspaces offer advantages over existing communications technology, such as the more pervasive and cheaper telephone. Other work has attempted to characterise the ways in which Workspaces are used in design tasks (Minneman and Bly, 1991; Tang, 1991). Again however that work did not draw contrasts with existing audio technology.

Work by Bly (Bly, 1988) did compare styles of interaction between graphic designers in (i) face-to-face settings; (ii) working with an audio link only, and (iii) working with an audio and video workspace link. However the object of this work was not to make comparisons between the efficacy of communication, but rather "to understand the use of drawing spaces for design" in the three settings. In contrast, we aim to identify conditions in which a Shared Workspace improves communication via an audio link, and to explain why such effects might occur. We also hope to provide quantitative tests for the observations reported in early Workspace studies and also to extend their analyses to a broader range of tasks. For these reasons we employ a different method from the early Workspace studies: we compare the advantages of adding

the Workspace to audio alone, for a series of measures for 3 different textual and graphical tasks that are representative of workplace activity. Evaluation across multiple tasks is important given the task-specific nature of communication and cognition (Reid, 1977; Suchman, 1985).

Clark and Brennan (1991) provide a cost/benefit analysis of the process of communication via different media, and this is the basis for the theoretical framework we use here. They point out that a crucial aspect of communication is the establishing of "common ground" in the form of shared understanding and agreement between speaker and listener. They also describe how the techniques available for this differ according to the medium of communication. For example, nonvocal clues to listeners' understanding such as headnods and gaze are not available to people talking by telephone, and timing of feedback responses such as backchannels is difficult to achieve with typed input or lagged speech (O'Conaill, Whittaker and Wilbur, in press; Oviatt and Cohen, 1989; Rutter and Robinson, 1981; Whittaker and O'Conaill, 1993). Hence communication media differ in the costs and benefits they impose on participants attempting to establish common ground. We apply this framework to understand how the process of achieving shared understanding and agreement differs for communication using phone and Workspace, compared with phone alone.

What are the communicative costs and benefits of using the Shared Workspace? The Workspace allows communication by drawing, and the use of gesture to identify an item or location (Bly, 1988; Minneman and Bly, 1991; Tang, 1991), and for certain purposes, these benefits might enhance communication effectiveness compared with communication by audio channel alone. Whittaker et al. (1991) list further possible benefits of Workspaces over audio communication. They point out that Workspaces allow: concurrent rather than just sequential input from partners; the easy depiction of spatial relations; and, most critically, a permanent record of activity which contrasts with the ephemeral information generated by audio alone. In their study, the fact that the participants could record information using the Workspace served as a focus for group activity and the recorded information also served as the target for gesture. On the other hand, they identified costs associated with the Workspace. These include: accidental deletions or over-writing of the partner's work; the need to organise the physical use of space, e.g. who writes at what location; and co-ordination of scrolling and deletions.

How might these costs and benefits weigh against each other in different potential uses of the Workspace? We made predictions for 3 different types of task possessing different critical task parameters and tested these

for a series of measures. For all tasks, pairs of participants worked using an audio link, with and without a shared electronic Workspace. We studied the effect on communication performance and obtained subjective reactions to the addition of the Workspace. We also analysed the content of the audio interactions with and without Workspace. This was to identify whether the presence of the Workspace changed the style of spoken communication, and also to allow us to explain observed differences in performance and subjective measures. The pairs were part of the same work group and were therefore experienced collaborators. This method enabled us to directly compare the existing technology for synchronous collaboration, namely the phone, with our new technology. By using experimental tasks we were able to make observations for multiple users carrying out representative collaborative tasks, and hence test specific theoretical predictions, by manipulating key task parameters. These results should complement and extend early work documenting the nature of Workspace use (Bly, 1988; Minneman and Bly, 1988; Tang, 1991; Whittaker et al., 1991).

We evaluated Workspaces for three different tasks, possessing key properties relevant to testing the theory. Given the importance of text and documents in Workplace communication activity (Short, Williams and Christie, 1976), we thought it was critical to test predictions for textual material, especially given that early studies focussed on design and drawing tasks. There were therefore two types of text-based task (Studies 1 and 2), and a graphic design task (Study 3). In the first, undemanding, text-based task, pairs had to agree on a summary slide of a specified part of their group work objectives. This was a task which they had all carried out in the past and hence was a realistic workplace activity. The second, more demanding, text-based task involved joint editing of a long document, again the kind of activity that collaborators are likely to engage in as part of their everyday work (Kraut et al., 1988). In the final graphical task participants had to agree on the placing of differently shaped pieces of 2D "furniture" into a given "floorplan". Although this task was less realistic than the text-based tasks, it had the key properties of a collaborative design task.

We predicted that for text-based editing where there are many ideas, several pages of text, and complex syntax changes, it is difficult or impossible to remember detailed points discussed orally. Here Workspaces should provide benefits over audio alone, because the shared text allows participants to easily identify key regions of text and record detailed changes directly and visibly. In contrast, for the less demanding text-based summary task, a shared permanent representation may not be as critical because people can easily describe and remember the points they are discussing. Here the unfamiliarity of using a Workspace in addition to an

audio channel may also impose more costs than benefits. For the third, graphic design task, we predicted that the Workspace would be particularly helpful because of the need for shared understanding and memory for spatial arrangements, and because the oral description of components and locations is laborious. Our overall expectations were therefore that for the undemanding text task, the shared visual permanent record provided by the Workspace would offer fewer benefits in contrast to the demanding text and design tasks. In all tasks we had several outcome measures to test overall communication efficiency with and without Workspace.

One problem with a cost/benefit framework of Clark and Brennan (1991) is that it is difficult to define a way of quantifying costs and benefits a priori. At worst the framework may allow no more than a post hoc explanation of when and why a particular form of mediated communication turns out to be more effective for a particular task. In the studies which follow, we therefore attempted to quantify costs and benefits by comparing the content of the conversation with the different communication media.

In our analysis of conversation content we focus on three critical aspects of communication that must be supported in any work-related conversation. We suggest that each participant has to be able to: (i) Track the other's focus of attention, e.g. detect whether or not the collaborator is working, and if so on what part of the task; (ii) Retain the context, e.g. remember what has already been done and know what still has to be done; and (iii) Achieve joint reference, e.g. be able to offer information to, or receive new information from the partner, and be able to relay or receive developments of that information. For each of these three measures, the audio and audio plus Workspace communication media offered different costs and benefits according to the task. These different aspects of communication have been shown to be important in other theoretical work on discourse (Clark and Brennan, 1991).

Achieving joint reference forms the heart of collaborative problem-solving. In the slide task, for example, this consists of writing ideas on the Shared Workspace or saying them for the other person to hear, then suggesting orders of priority for these ideas. In the text-editing task, this consists of identifying the paragraphs that have to be edited, and suggesting alterations. In the furniture task, it consists of describing pieces and suggesting where to put them. Other research has also suggested the utility of the Workspace for reference to concepts and relations that are difficult to verbalise (Bly, 1988, Minneman and Bly, 1991, Whittaker et al., 1991). The Workspace should thus especially facilitate reference in the furniture task because of its

inherently spatial character.

The other two aspects of work-related conversation, tracking the other's attention and retaining the context, can sometimes be achieved at little cost to the participants when they use the Workspace. For example, when each participant can see the other's inputs, it is easy for them to track when and where the other is working. This can occur in the Workspace without deliberate effort to communicate this information. In contrast, in audio only communication, when the other's inputs cannot be seen, conversational turns have to be taken up to convey this type of information. Other research indicates the utility of visible inputs for signalling attention and co-ordinating activity without the need for verbal interchange (Nardi, Schwarz, Kuchinsky, Leichner and Whittaker, 1993; Whittaker et al., 1991). Disruptions of communication structure and overwriting have also been shown to result from the failure to provide this type of information with typed synchronous communications (McCarthy, Miles, Monk, Harrison, Dix and Wright, 1991; Tatar, Foster and Bobrow, 1991).

As far as retaining context is concerned, the Shared Workspace offers the obvious advantage of providing a permanent record of what has been achieved so far, and to some extent, of what has still to be done. In the slide task, for example, collaborators can see the list of suggested ideas while putting them in order. In the text-editing task, they can see both what has been edited and what remains to be tackled. In the furniture task, both participants can see the plan and what pieces have already been inserted. Other research has discussed the value of shared context for design tasks (Bly, 1988).

We also collected subjective measures by having users rate their satisfaction with what they had produced, how well they had collaborated with their partners, and how well-suited the system was to the task. These subjective measures are important, because whether or not people choose to use a particular communication system may depend as much on perceived, as on objective effectiveness (Short et al., 1976). For example, if people feel that the Workspace helps them to reach agreement with their partner, they may choose to use it even if objective measures show no advantage over audio alone. We present the results of these subjective measures for each of the studies independently. However, there are obvious problems with subjective ratings given by participants who were employees of the company which developed the new technology being evaluated. In order to avoid bias, we also present comparisons between questionnaire ratings made in the three studies. Any bias should be equivalent in all three studies, showing itself as an overall preference for having

the Workspace, but over and above this we should be able to see specific differences according to task. We also collected spontaneous user remarks during use, noted user errors and elicited user comments after each study. We used these to derive design requirements for Workspaces and suggest ways in which our systems could be improved. These design requirements are reported in the Conclusions.

A final problem with the cost/benefit framework is that it assumes a fixed capacity user: namely that a particular user puts the same cognitive resources into a task under different communication conditions. Without making that assumption, the calculation of relative costs and gains becomes unwieldy at best. Yet intuitively, it seems likely that people talking over a poor telephone line can achieve successful communication by increasing the effort they put into the task. If we compared communication effectiveness under good and poor telephone conditions, we might find no difference simply because the user is compensating for the poor conditions, and relaxing somewhat under the good conditions.

One way of dealing with this, is to impose tasks that are so demanding cognitively, that the user does not have the resources to step up his or her performance under the poorer conditions. Another is to look at relative performance over the course of the task: under the less demanding communication conditions, we might expect to find better performance early on in the task, even if performance under the more demanding conditions catches up later on. In the studies which follow, we tried both these techniques: we used tasks which were very demanding (study 2), and we analysed performance at earlier and later time points in the tasks (study 3). This way, we hoped to be able to identify when and how the Workspace improves communication effectiveness.

Undemanding text-based collaboration: Slide task.

This task required participants to generate and agree on short textual summaries about aspects of their work. We expected few benefits for adding the Workspace.

System

An audio link was supplied using headsets with microphones to allow "hands-free" operation. Audio output was provided by a small speaker at each location. The Workspace was the "Wscrawl" program running

under X Windows, on HP 300 series UNIX workstations connected via an ethernet LAN. "Wscrawl" offered a shared writing and drawing surface like a whiteboard: it enabled people at different workstations, to write or draw concurrently with near-instantaneous transmission of information. Figure 1 shows "Wscrawl" outputs. The input device was a stylus and tablet and the results of writing and drawing activity were displayed on a monitor. The size of the monitor was 16 inches and the resolution of the writable area 1002 by 727 pixels. Users were free to write in any area of the bitmapped screen. There was no input control regime, so that different users could concurrently enter and erase inputs. Material could be deleted by selection of an eraser. There were no restrictions on what material could be erased, so people could delete material that had been created by someone else. To allow both the participants and the analyst to identify the writer, users were each allocated a different colour and all their inputs appeared in that colour. Participants' cursors were not visible to one another, so they had to make a small mark in order to gesture at drawn material. Although a pointer is available with Wscrawl, to simplify operation we told subjects only to use draw and erase modes. These modes were selected from a menu. The task only required the generation of a single screen of output, so that users were restricted from creating new "pages". If the "page" was filled, and more space was required then existing material had to be deleted.

FIGURE 1 ABOUT HERE

Method

Sixteen participants (aged 24-34) worked in pairs on two tasks, which involved the joint production of a summary sheet or slide. The topics to be summarised were "The output of a Research Lab. should be:" and "A good human software interface should be:". Subjects were asked to list a minimum of 7 prioritised points on each within a time limit of 7 minutes. All participants worked in the Hewlett Packard research laboratory at Bristol, UK, so these were suitable work-related topics, and participants were familiar with producing such summaries for overhead slides. The participants all came from research groups that were independent of the team developing and evaluating the workspace software. Each pair of participants together created one slide under audio only (A) conditions, and the other under audio plus Workspace (A+W) conditions. The order of the two communication modes and tasks was counterbalanced. In the audio mode, one member of the pair was allocated the role of "writer", and was asked to write the agreed summary on the electronic

Workspace. We did this to ensure that there was the same output device in both conditions, although in the audio only condition, only the writer could see what was on the Workspace. In the audio plus Workspace (A+W) condition, participants had the same audio link but also could use a Shared Workspace which both could see and write on.

There were 3 outcome measures to assess communication effectiveness: (a) The time partners took to reach agreement on their summary slide; (b) Number of post hoc changes: This was based on the number of changes to their agreed product, which the pair chose to make on the working day after they had completed the task. We assumed that the more effective communication had been during the task, the fewer changes the pair would make subsequently. When making post-hoc changes, the pairs met face-to-face and edited a hard copy of their summaries. There was no time limit for this follow-up activity; (c) Independent ratings of product quality. The initial summaries were rated by 5 independent judges (managers who were used to producing and evaluating such summaries) according to whether one would be able to tell, for example, whether a laboratory satisfied each of the criteria listed. Vague or evaluative rather than descriptive criteria, should have been given low ratings, and the judges evaluated hardcopy versions of each pair's handwritten output.

We also analysed aspects of the oral interaction between pairs in each condition. Details are given below.

Finally, we obtained subjective measures from the participants using rating scales and a short questionnaire. The questions we asked were: (i) how satisfied were you with what you as a team produced? (ii) how happy were you with the cooperation between you and your partner? (iii) how well-suited was the system to the task? (iv) if you had to do this task again which system would you prefer? Each question except (iv) was answered by placing a mark on a rating scale. We also noted spontaneous user errors and comments about the system during and after each task. These are discussed in the Conclusions.

Results

Typical outputs are shown in Figure 1. The prioritised list is shown on the left-hand side of the screen. Initially the pair who generated this slide proceeded by dividing the Workspace in two and noting tentative suggestions on the right hand side. Having verbally prioritised some of the items they concurrently entered their final choices in the left hand column.

Outcome

We obtained no effects for the outcome measures in a series of three way ANOVAs with the variables medium (A or A+W), order (A+W first or second), and role (writer or partner in the audio only condition): (a) Time to agreement: Nearly all pairs used the full time available, so the data were not analysed. (b) Subjects' post hoc alterations: There was no significant difference between A and A+W media in number of alterations made. (c) Judges' ratings of quality of summary: There was no significant difference between media, in quality ratings, nor in number of summary points generated in the two conditions.

Communication Content

We then analysed oral behaviour. We analysed differences in the amount that people talked by looking at the total number of turns, and found only one significant difference between A and A+W conditions: a medium x order interaction ($F(1,6) = 7.38$, $p < 0.05$). Users performed better on their second task irrespective of medium.

We then went on to look at the content of the turns. Here is an example for the audio only condition:

```
Example: Slide Task - Audio only.


1. A: Outputs of a Research Lab, uh, documents
2. N: documents - oh ya - lots of documents
3. A: patents
4. N: oh patents
5. A: prototypes
6. N: prototypes - ah how about technology?
7. A: ok
```

This shows that typical behaviour was quickfire exchanges in which participants presented "ideas" for criteria "ah how about technology?", "what about videos?", and these were often accepted without discussion or refinement. In A and A+W combined, suggestions for ideas were the most common category of utterance, comprising 43% of all turns. Two raters (EG and SJW) independently judged whether turns contained

"ideas" and inter-rater agreement was high: Kappa = 0.89, $p < 0.001$ (Cohen, 1960).

To see whether number of turns containing ideas differed between A and A+W, we carried out a 3 way ANOVA on the number of such turns, with the variables medium (A or A+W), order (A+W first or second), and role (writer or partner in A). Figure 2 shows there was a significant effect of medium ($F(1,12) = 17.28$, $p < 0.01$), with more such turns mentioning ideas in the audio condition, and a significant medium x order interaction ($F(1,12) = 15.84$, $p < 0.01$). Post-hoc tests showed most turns with ideas were produced in A when that came second, with little difference between the other three means. This result suggests that arriving at agreement about ideas took fewer turns and hence was more efficient in A+W but we have already noted that there was no objective difference in the quality of final product in the two media.

FIGURE 2 ABOUT HERE

We also looked for signs of costs of using the Workspace: People had to organise where they wrote so as to avoid overwriting, as evidenced by remarks such as "Did we agree on parting the page?", "Do you take the upper half?" There were also problems in achieving joint attention, such as the following example of an A+W exchange:

```
Example: Slide Task: Audio and Workspace


1.R: I could put up here, you see I could put
2.A: Up where?
3.R: Here
4.A: I can't see where here is
5.R: Up there
6.A: B - oh yeah - so we'll rub out
```

There were 11 examples of organisational activity and failed mutual attention in the A+W transcripts, and as expected none in the A condition.

Questionnaire data

We conducted 3 way ANOVAs with the variables: Order (A first or second), Medium (A or A+W) and Role (whether the subject was writer or partner in the audio only condition.) There were no differences between A and A+W either for perceived co-operativity or the suitability of the system in the two communication conditions. For the overall satisfaction with the group product however there was an effect: there was a significant interaction ($F(1,12) = 12.46$, $p < 0.01$) between order and medium: higher satisfaction scores were given for the second task, whether this was A or A+W. For the final question about system preference, 12 subjects preferred A+W, 2 were undecided and 1 audio (Chi square, $p < 0.05$). We were unable to get a response from the final subject.

Discussion

The objective measures from this study fail to show that people communicated more effectively in the A+W condition than in A alone. This is as we expected. The task was relatively undemanding and even with audio alone people could presumably remember the ideas suggested while they constructed their lists and remember the sequence of items while prioritising. The Workspace may therefore have offered few benefits but some organisational or attentional costs. Our null results with the measure of post hoc alterations were at least in part due to the fact that some collaborators continued to discuss the task after their joint session: some people also rejected their initial collaborative effort and started the post-hoc activity anew. Hence this measure of communication effectiveness was not very useful. Nevertheless, our other outcome measures also failed to show any difference between conditions.

Consistent with the null results for the outcome measures, subjects' perceptions of communication under the two conditions seemed to be very similar. Although they would choose to use the A+W system instead of audio alone (possibly simply a sign of a bias as mentioned in the introduction), they did not strongly believe that it was better suited to the task at hand.

The Workspace could in principle help people to reach agreement in the slide task, by allowing a shared visual record of the points raised by each partner, their priorities, and the consensus reached. It appears from our data, however, that people performed the task just as well without these aids. The slide task may have

been easy enough to complete successfully using an audio channel alone; indeed perhaps even face-to-face communication would show no advantages over the audio condition for this type of task. Alternatively, it may have been that the costs of managing the use of space and achieving joint attention counteracted the benefits offered by the Workspace for recording and organising ideas. There was a suggestion of Workspace benefits, with fewer turns being taken discussing ideas in A+W implying easier reference to, and recording of ideas. There was no evidence however, that this benefit led to higher quality outputs with A+W, maybe because of these costs in using the Workspace.

Demanding text-based collaboration: text-editing task

This time we used a much more demanding text-based task, in which people had to scroll through a document several pages in length, identifying passages which needed to be edited, agreeing on the editing, and annotating the text with their changes. Our expectation was that for this task the benefits of having shared visual reference, and a shared record of changes would outweigh the costs of using the Workspace collaboratively.

System

Audio communication was the same as in the previous study, but in this study we used a different type of whiteboard and a pen and tablet which allowed easier writing. The software was installed on 2 PCs (HP-vectra RS/25C), with a monitor screen size of 20 inches and resolution of 480 by 640 pixels. Communication between computers was via an ethernet LAN. The software ANNOTATOR ran under Microsoft 3.0. Participants annotated the document by hand, using a WACOM Tablet with a cordless pen. The results of their efforts were displayed on the PC monitor. The use metaphor was a transparent overlay over underlying text, so that participants entered their annotations on the overlay. They were able to erase both their own and other people's annotations, but not the original underlying typewritten text. Sample outputs are shown in Figure 3. As in Study 1, both participants could write, draw and erase concurrently. There was no control regime to sequence or restrict inputs and transmission and display of inputs was almost instantaneous. The software allowed the user(s) to scroll through the documents using standard Windows scroll bars. In contrast

with the first study, participants were continuously able to see each other's cursors in blue (writer) and red (partner). Pointing was achieved by holding the pen near, or lightly touching the surface of the Tablet. Also in contrast to Study 1, write and erase modes were selected from constantly displayed buttons rather than accessed indirectly by a menu selection. The shape of the cursor when writing was an arrow and when erasing a square, to give feedback about the current mode the user was in.

FIGURE 3 ABOUT HERE

Method

Sixteen participants (aged 18-54 years), all administrative staff at Hewlett Packard Laboratories Bristol, worked in pairs under both A and A+W conditions for 2 different tasks, in counterbalanced order. The tasks involved editing documents. We used articles about 500 words long (10-11 paragraphs) on matters of general interest that were written for 9-year-olds. We modified three of the later paragraphs so that they had a much higher reading age and stood out from the rest of the text as being more complex to read and understand. Reading age was calculated using the Automated Readability Index program which is part of the UNIX Writer's Workbench. Examples of the texts are given in the appendix, with the discrepant paragraphs marked with asterixes. Participants were asked to identify these paragraphs and jointly edit them so that the reading age was uniform across the document.

Initially each participant of the pair was given 3 minutes private reading time to read through the article. The pair then had 12 minutes to work jointly at the editing under either A or A+W conditions. As before, in the A condition, one participant was the writer, and under A+W conditions both participants could write on the screen.

As in study 1, participants gave subjective ratings, we analysed aspects of the oral interaction and had several outcome measures. Our outcome measures were (a) Reading age of the modified paragraphs after the participants had identified and edited them, (b) Number of paragraphs completed, and (c) Time taken to complete the editing of each paragraph.

Results

Figure 3 shows an edited paragraph when both participants have access to the Workspace. The central paragraph is the one that has been modified. Both participants have made alterations to the paragraph, as indicated by the different handwriting, and they also used brackets and arrows to indicate the placement of selected parts of the text.

Outcome Measures

All pairs identified the paragraphs which needed to be edited. In contrast to the first study, the outcome measures showed significant effects[1].

(a) Reading age: We calculated the reading age of each edited paragraph. We then divided this figure by the reading age of the corresponding original, so that a low score indicated effective editing. We conducted a 3 way ANOVA with the variables medium (A or A+W), paragraph number (first, second or third), and order (A+W first or second). The only significant effect was an interaction between order and medium ($F(1,6) = 20.99$, $p < 0.01$), which showed that participants performed better on their second task irrespective of medium. Of course there was also a main effect for paragraph number because more people completed earlier than later paragraphs.

(b) Number of paragraphs edited: We conducted a 2 way ANOVA on number of paragraphs completed, with the variables order and medium. The only significant effect was an order x medium interaction ($F(1,6) = 18.00$, $p < 0.01$): fewer paragraphs were completed with A+W when that came first, with little difference between the other means (Tukey post-hoc tests $p < 0.05$, for A+W first versus A+W second, with the no other comparisons significant). This suggests difficulty in adjusting to Workspace use when the task is unfamiliar.

FIGURE 4 ABOUT HERE

(c) Time to complete first paragraph: All pairs completed their editing of the first paragraph, and we conducted a 2 way ANOVA on the time to complete this in whole minutes, with the variables order and medium. The only significant effect was again an order x medium interaction ($F(1,6) = 16.20$, $p < 0.01$).

[1] Given the large age range of our participants we conducted analyses to compare "old" and "young" pairs, defined with reference to median age. There were no age differences on any of the measures.

As Figure 4 shows, participants were slowest on A+W when that came first, but quickest when A+W came second, with the A conditions intermediate. Again post-hoc tests showed differences only between A+W first and second ($p < 0.05$).

(d) Time to complete second paragraph: For those who completed their editing of the second paragraph (7 pairs in A, and 7 in A+W, with those who did not complete included in the analysis as missing data), there was a significant effect of order ($F(1,6) = 13.50$, $p < 0.05$), of medium ($F(1,4) = 13.50$, $p < 0.05$) and an order x medium interaction ($F(1,4) = 37.50$, $p < 0.01$). Figure 4 shows that completion time was quicker for A+W when that came second rather than first, post-hoc, $p < 0.01$). There was also a difference between A and A+W for people receiving A+W second ($p < 0.01$).

Both results on paragraph completion time indicate problems in using A+W when the task is unfamiliar but advantages for A+W with familiar tasks. Too few pairs completed the third paragraph to make analysis appropriate.

Communication Content

To compare the oral interaction in the A and A+W conditions, we counted the total number of turns in each and carried out a 2 way ANOVA with the variables medium (A or A+W) and order. We found no significant differences in total number of turns in each condition. We also counted the number of turns in which characteristics of the paragraphs were mentioned. Agreement between two coders (EG and SJW) was high: Kappa = 0.93, $p < 0.001$). Our expectation was that the Workspace would make reference to paragraphs and discussion of modifications more efficient. They would therefore take fewer turns discussing paragraphs and fewer turns overall. However, there were no significant differences between A and A+W. We present an example of an A+W conversation, in which quotes ("") indicate that people are referring to specific words from the original text.

`Example: Text editing - Audio and Workspace`

`1.A: well, how-if we can include some of the stuff that's already written here.`

`2.B: Yeh. ''noticed that'' -`

`3.A: ''that'' -`

4.B: you could put -

5.A: instead of-instead of ''migratory animal''-so if you -

6.B: you could put ''it''. -

7.A: if you crossed this out.

8.B: Yeh.

9.A: And put-and put ''a bird'' there, instead of that.

10.A: Ha hee. Alright -

However, there did seem to be costs in using the Workspace: inspection of the transcripts suggested that participants in A+W spent a good deal of time discussing scrolling through the document. We therefore counted the number of turns in which navigation through the document was the topic, and there were indeed significantly more in A+W, with a mean of 14.0, compared with 6.5 in A ($F(1,11) = 7.60$, $p < 0.05$). Here is an example of the negotiation that is involved in joint scrolling.

Example: Text editing - Audio and Workspace

1.A: That's fine. Do you want to scroll down? It's just the last one
now, isn't it?

2.B: What's it begin with?

3.A: ''Because the Tudors''.

4.B: Just takes so long. There it is.

5.A: Ah. A bit more.

6.B: A little bit more.

7.A: A bit more. Ha ha. A bit more.

.... 7 more turns about navigation .....

15.A: that's the whole of the sentence-paragraph.

16.B: Right.

This negotiation constituted 7% of the total number of turns in the whole dialogue. These types of problems have been noted elsewhere with systems that allow several users concurrent access to scrolling (Stefik, Foster, Bobrow, Kahn, Lanning and Suchman, 1988).

Questionnaire data

Finally, the subjective data: As for study 1, we conducted a 3 way ANOVA with the variables order (A first or second), medium (A or A+W) and role (writer or partner in A condition). There were significant effects of role for question 1, product satisfaction, ($F(1,12 = 6.81$, $p < 0.05$) and for question 3, system satisfaction ($F(1,12) = 5.73$, $p < 0.05$) with more satisfaction expressed by the writer. There were no significant effects for question 2 (cooperation satisfaction). In answer to the final question about system preference, all participants preferred A+W.

Discussion

The Shared Workspace clearly provided more benefits for this task than it did for the slide task, but there were costs too. Initially, collaborators found it hard to scroll jointly through the document, but once they had mastered this and become more familiar with the text-editing task, they worked much more efficiently under A+W than under A conditions. This is shown by the faster paragraph completion times when participants had the Workspace second. Despite this, the style of oral interaction did not seem different in the two conditions, with the content of turns appearing to be very similar in both. The exception to this was turns about navigation which appeared more frequently in A+W and the scrolling mechanism seemed difficult for users to master together. The satisfaction ratings showed few overall differences apart from role: participants preferred the role of writer in the audio only condition. This is not surprising; it was hard for the partner to keep track of what had been recorded, as evidenced by monitoring remarks to the writer such as "What have we got so far?" Furthermore, under these conditions, the partner was perhaps able to contribute relatively little.

Design collaboration: Furniture task

Finally, we looked at the impact of adding a Workspace for a design task with a high visual/spatial component. We expected benefits for the Workspace here.

System

We used the same system as in Study 1.

FIGURE 5 ABOUT HERE

Method

Pairs of participants were asked to agree on the arrangement of differently-shaped pieces in a 2D space that was too small to accommodate them all. The tasks were presented as arranging "pieces of furniture on a floor plan". Each member of the pair was given a different set of pieces and the same floorplan. The two sets of pieces for one of the tasks are shown in Figure 5. Each piece had a score, and the aim was to arrange pieces so as to maximise joint total score. Hence people had to share information about the shape and size of their own pieces, and the score allocated to each. The pieces were also designed to encourage collaboration: High scores could be obtained by combining complementary pieces from each other's sets. The letters S, T, and C are abbreviations for Sofa, Table and Chair. The numbers on the external parts of the shapes represent dimensions and the internal numbers represent the number of points awarded if the piece is successfully inserted. Figure 6 shows a completed floor plan. People were given a maximum of 12 minutes to complete each task. As before, there were two tasks, and each of 8 pairs (ages 25-36) performed one task under audio only conditions, and another under audio plus Workspace conditions.

FIGURE 6 ABOUT HERE

As in the previous studies we analysed the oral interaction between participants, and obtained satisfaction measures. Our outcome measures were (a) Time to complete the task; (b) Post hoc alterations as in Study 1, (c) Scores obtained for furniture placed after 7, 10 and 12 minutes.

Results

Figure 6 shows a completed floor plan for the A+W condition. Here the participants have depicted their individual sets of pieces around the edge of the plan, before beginning to insert them. It was typical for users with the Workspace to gesture at the pieces and at spatial locations.

Outcome measures

First we analysed the outcome measures: (a) Time to agreement: Again nearly all pairs chose to use the full time available, so we did not analyse this measure further. (b) Post-hoc alterations. As in Study 1, and for apparently similar reasons, this outcome measure was not very successful, and there were no differences between conditions. (c) Scores obtained: We coded each pair's score after 7 minutes, 10 minutes, and at the end of the 12 minutes allowed. Figure 7 shows that there was no difference between A and A+W conditions after 12 minutes, but the A+W group showed a clear advantage early in the task. We performed a 3 way ANOVA on scores, with the variables order (A first or second), medium (A or A+W) and time (after 7, 10 or 12 minutes). There was an almost significant effect of medium ($F(1,6) = 5.30$, $p = 0.61$), with a tendency for higher scores in the A+W condition. The trial x medium interaction was significant ($F(2,12) = 3.97$, $p < 0.05$), with A+W showing much higher scores in the early trials. For example, after 7 minutes the mean score in A+W was 102.3, whereas for A it was only 49.9 (Tukey post-hoc, $p < 0.05$). However by 12 minutes, scores were almost equivalent: A+W scored 132.4, and A scored 132.0 (Post hoc $p = 1.00$, not significant). With A, subjects increased their scores by 164% in the last five minutes compared with only a 29% increase in the same period for the A+W group. This suggests that A+W offered strong advantages early in the task, but that audio users caught up later. We discuss possible reasons for this below.

FIGURE 7 ABOUT HERE

Communication content

How did oral behaviour compare in the two conditions? The following audio transcript shows two people trying to identify one of the pieces in Figure 6.

Example: Furniture Task - Audio

1.A: right - can you describe it to me a little bit

2.F: the 12 x 10 one - the 12 is the height the 10 is the base ok

3.A: how will it fit into the 12 x 10 square I put on my screen? if I start from the top left corner

4.F: top left corner ok

```
5.A: can I go along 10?

6.F: yes go along 10

7.A: ok - alright and now I am on the right hand side?

8.F: no actually - on the top right hand corner - right?

9.A: ya - top right hand corner

....... 26 more turns about the same piece .........

36.F: so that's what you got, that's a big piece - it's worth 40 points - -

     - ok
```

The description of a single piece has taken 36 turns and required various clarifications by the person who cannot see it (Eg Turns 3, 5, 7). The pattern of interaction was very different when the Workspace was added:

```
   Example: Furniture Task - Audio and Workspace

1. A: right now, I got some objects that can fit in a 10 x 12 which is

      worth 40 points - do you want me to scribble it in?

2. F: yes --- 10 x 12 --- I've got a 8 x 12

3. A: go 10 x 12 that fit in there that is worth 40 points

4. F: ok yours is more than mine - keep going

5. A: I am not quite sure how to do it but here we go ----- ya, that's not

      very good - and that's worth 40
```

The description of a similar piece by the same pair took just 5 turns, and required no clarifications, because people were able to draw it (as indicated by Turns 1, 4, 5). We coded the number of turns in which pieces of furniture were described (agreement between 2 coders EG and SJW: Kappa = 0.78, $p < 0.001$). Consistent with the above example, a 3 way ANOVA performed on the variables order, medium and role showed there were many more turns involving description in A than in A+W ($F(1,12) = 28.01$, $p < 0.001$). The mean number of turns involving description was 54.3 in A, but only 35.1 in A+W, suggesting it took much longer to identify and refer to pieces in the A condition.

Given that the pairs with Workspace took fewer turns to identify pieces, and achieved higher scores early on in the task, yet continued to work for the full 12 minutes allowed, and why was their advantage over A

not maintained throughout the task? The following example offers some clues:

Example: Furniture Task - Audio and Workspace

1. B: ok - nothing else we are filled
2. A: yes and now we can try and optimise it if we want now -what is happening in this area here?
3. B: in this area - this is - nothing exists - actually I am going to -
4. A: if we can use the time to fill it properly

It seems here that A+W participants were using the extra time and turns to check and optimise their solutions. In the above session, this checking behaviour began on turn 120 and took 86 further turns, and only added 16 out of a total of 156 points, so that the last 40% of turns only contributed 10% of points. This type of activity may have led to the effects described earlier, with A+W participants obtaining many fewer points late in the task. Another possibility is that there were ceiling effects: 63% of A+W users scored greater than 90% possible points compared with only 25% of A users achieving this. On a more complex task, there may have been greater differences between the two media conditions.

One feature of the Workspace was that people could monitor the actions of their partner directly. In contrast, with audio, people had to check orally what the other person was doing: "so you are now filling in the bottom left hand corner of the top right hand room -is that right?' In addition people with audio had to inform their partner about progress through the task "we have completely filled the first rectangular room with 79". Again this was unnecessary when there was shared visual information. The inability to monitor the other person's actions with audio only, also produced instances of co-ordination failure in which one person told another to carry out an action they had already done:

Example: Design Task - Audio

1. F: that's right just put three of them in
2. A: I already did

Figure 6 also shows the use of the Workspace as a way of maintaining context throughout the task. In the Figure, participants have each sketched their set of pieces around the edges of the plan prior to inserting them. They indicate when a piece has been entered into the plan and hence is no longer available by putting an "X" through the depiction of the piece. The material entered into the plan also makes it possible to visually determine how much space remains, as well as the design decisions that have already been made, without having to directly converse with the other participant.

Again we looked at the costs of using the Workspace. As with the slide task we found examples of people having to organise the Workspace explicitly: "I'll sketch out my furniture above and you can sketch out yours below" and also problems of achieving mutual attention: "you can't see my cursor at the moment can you?". The 12 instances of these problems in organising attention and Workspace only occurred in the A+W condition.

Questionnaire data

As before, scores were entered in a 3 way ANOVA with variables: Order (A first or second), Medium (A or A+W) and Role (whether the subject was writer or partner in the audio only condition.) There were no significant main effects either for product satisfaction or co-operation. However there were large condition differences for the perceived suitability of the system for the task. There were significant main effects for medium ($F(1,12) = 18.59$, $p < 0.01$): with more satisfaction with A+W, and for order ($F(1,12) = 7.63$, $p < 0.05$): with more satisfaction when A+W came second rather than first. There were no significant interactions. Again more people preferred the A+W system to A alone (13 vs 3) (Chi squared, $p < 0.01$).

Discussion

As we predicted, people performed better with Workspace: they believed that A+W was more suitable for the task than A alone, and their verbal interaction was markedly different in the two conditions. In the audio only condition, there were more turns talking about pieces. This is consistent with the possibility that people found it harder to achieve reference and hence establish common ground in the audio only condition. As expected it was easier in A+W (a) to achieve shared understanding without extensive verbal descriptions and clarifications, because shapes could be drawn rather than just described orally, and (b) to know when

shared understanding had been achieved. In addition reference to the pieces and their locations could be achieved by gesture rather than complex descriptions. We also saw use of the workspace to record progress through the task and hence maintain context. For the furniture task, the gains from using the Workspace appear to be substantially greater than the costs.

Questionnaire Ratings; The Three Studies Compared

In order to avoid any bias associated with using new technology, we made comparisons between questionnaire ratings in the three studies, each of which involved that new technology. Any bias should be the same in all three studies, showing itself as an overall preference for A+W. This did appear: In answer to the question 'Which system did you prefer?', there was a clear overall preference for A+W (chi sq=61.57, df=2, $p < 0.0001$).

Next we attempted to factor out overall preferences. Should we detect greater preferences for A+W in some tasks rather than others, we could be confident that these were associated with a genuine perceived preference for that technology when performing that particular task. We therefore performed a 4 way ANOVA on each of the 3 questionnaire scores, with the variables experiment (1, 2 or 3), order (A first or second), role (writer or partner), and medium (A or A+W). The 3 questionnaire scores were for product satisfaction, cooperation satisfaction, and system satisfaction.

Figure 8 shows satisfaction for the two measures for which we obtained differences, namely product and system satisfaction, for each of the tasks.

FIGURE 8 ABOUT HERE

For question 1 (product satisfaction), there was a significant main effect for experiment ($F(2,36) = 8.20$, $p < 0.01$), with the highest scores in study 3 (furniture task), and the lowest in study 2 (text editing). There was also a significant order x medium interaction ($F(1,36) = 6.69$, $p < 0.05$), with the highest satisfaction for A+W when that came second, and the lowest for A when that came first (Tukey post-hoc, $p < 0.02$).

For question 2 (cooperation satisfaction), there were no significant effects, perhaps because partners had already established satisfactory working patterns through their daily working contact.

As Figure 8 shows, for question 3 (system satisfaction), overall satisfaction was higher in the slide than

text editing tasks, but satisfaction for the furniture task was critically dependent on the medium used. Results of the ANOVA showed that there were significant main effects for experiment ($F(2,36) = 4.05$, $p < 0.05$), with the highest scores in study 1 (slide task) and the lowest in study 3 (furniture task), for order ($F(1,36) = 6.55$, $p < 0.05$), with higher scores when A+W was second, for role ($F(1,36) = 5.62$, $p < 0.05$), with higher scores given by the writers, and for medium ($F(1,36) = 21.62$, $p < 0.001$), with higher scores for A+W. There was also a significant experiment x medium interaction ($F(2,36) = 3.41$, $p < 0.05$), with higher scores for A+W compared with A in the furniture (Post hoc, $p < 0.0005$).

These satisfaction scores reflect our expectations that working under audio only conditions would be most difficult in study 3 (furniture), but relatively easy in study 1 (slide). However, when the Workspace was added in the furniture task we saw large perceived benefits.

Final discussion and conclusions

Early work evaluating the effects of different media on communication failed to show benefits for adding visual information to audio for important classes of work-related tasks, such as information transmission and collaborative problem-solving (Chapanis, 1975; Reid, 1977; Williams, 1977). In contrast, this study shows that visual information in the form of a Shared Workpace can be useful for this class of task, although the effects depend on the presence of certain critical task parameters: in the slide task none of our outcome measures showed differences between A and A+W conditions, and in the text-editing task A+W only showed an advantage over A when A+W came second. It was only in the furniture task that A+W showed a clear overall advantage on the outcome measures. Neither was the addition of a Workspace necessarily always perceived as helpful by participants. In all three studies participants would choose to use A+W in preference to A, but it was only for the furniture task that people clearly thought A+W was better suited to the task than audio alone. Overall, these results indicate that Workspaces will be most usefully applied to tasks possessing a strong graphical component. There may also be advantages in tasks featuring complex textual material such as documents, where there are multiple changes to be made to the underlying drawing or document.

These results indicate applications where Workspaces might be successfully employed. Our data also offer

experimental evidence for some of the early claims about Workspace benefits (Bly, 1988; Minneman and Bly, 1988; Tang, 1991; Whittaker et al., 1991). The task-specific effects also reinforce the conclusion that the application of novel technology may not always bring expected communication benefits (Noll, 1992; Reid, 1977). The results also validate our method of comparing novel technologies with existing ones: If novel technology does not always bring expected benefits we should be able to specify those tasks where there are indeed improvements over existing technology and our method enables us to identify these situations. A theoretical implication is that any account of mediated communication should address these effects of task-specificity: Why did we find some advantages for visual information where early experimental research failed? A key reason may lie in the type of visual information that we provided. In our studies we supplied visual information about the task materials rather than information about other participants visual facial expressions and eye-gaze (Nardi et al., 1992).

Even for visual information about task materials, however, we found task-specific effects. What is the basis of these task-specific effects and how can they be explained in terms of communication processes? In our analysis of conversational content, we focussed on three critical aspects of communication: (i) Tracking focus of attention; (ii) Retaining the context; and (iii) Achieving joint reference. For each of these three, the A and A+W media of communication offered different costs and benefits according to the task.

Our results suggest that for the slide and text-editing tasks, joint reference was achieved using much the same style of interaction under both A and A+W conditions. Although there were differences according to media in the numbers of turns mentioning ideas in study 1, with agreement achieved faster in A+W, there was no sign that participants actually went about the task in a different way. Nor were there outcome differences when the audio channel was supplemented by a Shared Workspace. Moreover, for the text-editing task, when the Workspace was still unfamiliar, there seemed to be heavy costs involved in scrolling jointly through the document. Only when both the editing task and the use of Workspace had become more familiar, were the benefits of the Workspace in achieving joint reference realised. For the furniture task, in contrast, a different and more efficient style of interaction emerged when participants tried to achieve joint reference in A+W as compared with A. Collaborators could draw their pieces for each other, pieces could be identified in fewer turns, and then gestured at. The overall result was that high performance scores were obtained more quickly.

The other two aspects of work-related conversation, tracking the other's attention and retaining the context, can sometimes be achieved in A+W without intentional communication between participants. Evidence for Workspace benefits was less clear here, but the importance of attention monitoring was revealed in both slide and furniture tasks. Here the failure to provide a constantly visible cursor led to attentional breakdowns in A+W. These breakdowns were often caused by one participant gesturing at, and verbally referring to an object or spatial location, but forgetting that this gesture was not visible to the other participant. There were few such problems in the text editing task, because we provided a constantly visible cursor.

As far as retaining context is concerned, the Shared Workspace offers the obvious advantage of providing a permanent record of what has been achieved so far, and to some extent, of what has still to be done. The benefits were clearest in the furniture task: participants could see the inserted pieces and those pieces still to be placed were often drawn on the screen around the edges of the plan. Yet whether this potential gain actually helps, depends on the cognitive demands of the task. It appears from our results of Study 1, that collaborators on the slide task managed to retain the relevant context just as well with, and without, the Shared Workspace.

We also found that there are particular costs associated with novel media. People have to learn new strategies for co-ordinating activity and attention, whereas these regimes are conventionalised in speech on the telephone. In all three studies we observed user behaviours that indicated the costs of adapting to the novel A+W technology, such as organising space. If, however, for a given task and novel medium, the benefits are sufficiently great, then communication techniques will be devised by participants to reduce the costs of the novel technology. The uptake of novel technology is also crucially dependent on this cost/benefit analysis and here our studies may underestimate potential benefits. While our tasks were representative of workplace activity, effects of task novelty were found in the text-editing study; benefits only occurred when subjects had specific experience of the task. The fact that benefits were found only for familiar tasks augurs well for the uptake of Workspaces in real work settings however, because in such settings the work and the communication task is likely to already be highly familiar.

Other ways in which reductions in cost can occur is from improved design. We make suggestions for improved design, based on: users' problems with early prototypes, their comments during and after the tests, and our informal observations of their behaviour during testing. First, menu-driven selection of options such

as writing or erasing seemed to be less efficient than push-button selection from a set of options which were continually present on the screen. Another option might be to place both write and erase options on the pen itself: so that pen-inversion or a switch on the pen could be used to change modes, with feedback to indicate current mode. A second crucial point is that both participants need to see the other's cursor at all times: Pointing should always be available as the default mode, i.e. it should not require the user to make a selection. The cursor not only indicates to a person's collaborator where he or she is working, without this having to be communicated deliberately, but it can also prevent collaborators accidentally writing at the same location, because it offers strong cues about the other participant's focus of attention. In addition, all Workspace actions should be propagated with as short a transmission time as possible, and synchronisation preserved between voice and gesture/action. These requirements are necessary to support both mutual attention and gesture, and prevent misunderstandings arising because participants have different views of the Workspace. These misunderstandings can arise because the relevant updates have not been presented at all locations and thus are not shared (Tatar et al., 1991). The quality of the stylus is also important, and poor quality may have reduced use of the Workspace in our studies. Again many of our users' comments concerned this problem and those comments suggested that this presented more difficulties for writing than for drawing. Both legibility and writing size were cited as problems. Furthermore, writing on a tablet, and having output appear on a separate screen, introduces a level of indirection in hand eye co-ordination. This indirection may have severe implications for the quality of handwriting or drawing, if there are significant networking transmission delays. Finally, in the text-editing task, collaborators found it quite difficult to navigate jointly through the document using the scrolling mechanism, and a page-based organisation would probably have been easier for them to handle. Many of the same design suggestions are made by Minneman and Bly (1991) and Tang (1991) for drawing and design tasks. Here we have replicated the results with multiple users for different classes of task including those in which text plays a role.

The study also shows the complexity of evaluation methodology. In our studies we used three types of measure: objective outcome measures, analyses of participants' oral interactions, and subjective ratings. Two of our outcome measures, time to completion and post hoc alterations, though justifiable in principle, proved to be unsatisfactory in practice. Collaborators tended to fill the time available, and they sometimes found it hard to make post hoc changes to their earlier version rather than to start afresh. Nevertheless,

in each of our studies we did have outcome measures which appeared to be appropriate. In addition to the outcome measures, the interaction analyses and the subjective ratings provided useful evidence about the value of adding the Workspace to audio communication in order to explain our effects. No one type of measurement on its own would have allowed us to draw valid conclusions, but together they help us to identify just when and how the addition of a Shared Workspace to an audio link is helpful, and hence define situations where novel technology would be most beneficial.

Future work should also address how Workspaces might function in extended interactions and how they might be integrated with asynchronous communications technology. The benefit of permanent record may well be much greater in collaborations that extend over multiple interactions over periods of weeks or months. Here the results of synchronous Workspace activity could serve as a "trail" of actions and decisions taken in earlier interactions. Further work should investigate the use of Workspaces in such long term collaborations. At present mediated communications technology is either synchronous (phone, video-conference) or asynchronous (email, FAX, letters), but little work has been done to integrate the two. Our users' comments after the text-editing task reinforced the utility of making real-time changes to a shared permanent record, that lasts over multiple conversations. They also pointed to ways in which the Workspace might reduce the number of interactions in collaborations extending over time. The Workspace supports a shared visual record of changes. Current FAX and phone technology does not allow synchronous updates to visual materials. The lack of such updates may compel participants to exchange their own privately annotated versions of documents or notes after the discussion of changes, to check for consistency. Again further studies of extended collaboration should test whether using the Workspace in this way does indeed reduce the total number of separate synchronous and asynchronous interactions. This could pave the way to better integration of synchronous and asynchronous communications technologies.

Acknowledgements

We offer thanks to Patrick Hyland of Hewlett Packard Laboratories for writing the software ANNOTATOR, to David Frohlich, Patrick Hyland, Robin Jeffries, Phil Stenton, and Lyn Walker for comments on early drafts and to Crystal Drew for transcription.

References

[1] BLY, S. (1988). A use of drawing surfaces in different collaborative settings. *Proceedings of the Conference on Computer Supported Co-operative Work*, 250-256.

[2] CHAPANIS A., OCHSMAN, R. B., PARRISH, R. N., & WEEKS, G. D. (1972) Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors*, 14, 487-509.

[3] CHAPANIS, A. (1975) Interactive Human Communication. *Scientific American*, 232, 34-42.

[4] CLARK, H. H., & BRENNAN, S. (1991). Grounding in communication. In RESNICK, L. B., LEVINE, J., & TEASLEY, S. D., *Perspectives on socially shared cognition.* Washington: APA Press

[5] COHEN, J. A. (1960). A co-efficient of agreement for nominal scales *Educational and Psychological Measurement*, 20, 37-46.

[6] GALE, S. R. (1989). The vision project. Technical Memo: Hewlett-Packard Laboratories, Bristol, U.K., HPL-BRC-TR-89

[7] HARKNESS, R. C. (1977). *Technology assessment of telecommunications-transportation interactions.* Menlo Park, CA: Stanford Research Institute.

[8] HEATH, C. & LUFF, P. (1991). Disembodied conduct: Communication through video in a multi-media environment. *Proceedings of Conference on Computer Human Interaction*, 99-103.

[9] JOHANSEN, R. (1984). *Teleconferencing and Beyond.* McGraw-Hill.

[10] KELLEHER, K., & CROSS, T. (1985). *Teleconferencing: linking people together electronically.* London: Prentice Hall.

[11] KRAUT, R. (1989). Telecommuting: The trade-offs of home work. *Journal of Communication*, 39, 19-47.

[12] KRAUT, R., EGIDO, C., & GALEGHER, J. (1988). Patterns of contact and communication in scientific research collaboration. *Proceedings of the Conference on Computer Supported Co-operative Work*, 1-12.

[13] MCCARTHY, J., MILES, V., MONK, A., HARRISON, M., DIX, A., & WRIGHT, P. (1991). Four generic communication tasks which must be supported in electronic conferencing. *SIGCHI*, 1, 41-44.

[14] MINNEMAN S., & BLY, S. (1991). Managing a trois: a study of a multi-user drawing tool in distributed design work. *Proceedings of Conference on Computer Human Interaction*, 217-224.

[15] NOLL., A. M. (1992). Anatomy of a failure: Picturephone revisited. *Telecommunications Policy*, 307-316.

[16] NARDI, B., SCHWARZ, H., KUCHINSKY, A., LEICHNER, R., WHITTAKER, & SCLABASSI, R. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. *Proceedings of Conference on Computer Human Interaction*, 327-334.

[17] O'CONAILL, B., WHITTAKER, S., & WILBUR, S. (in press). Conversations over video-conferences: An evaluation of video-mediated interaction. To appear in *Human Computer Interaction*.

[18] OVIATT, S., & COHEN, P. R. (1989). The effects of interaction on spoken discourse. *Proceedings of the Association of Computational Linguistics*, 126-134.

[19] REID A. A. L. (1977). Comparing telephone with face-to-face contact. In I. Pool (Ed.), *The social impact of the telephone*. Cambridge, Mass.: MIT Press.

[20] ROOT, R. (1988). Design of a multi-media vehicle for social browsing. *Proceedings of the Conference on Computer Supported Co-operative Work*, 25-38.

[21] RUTTER, D. R. & ROBINSON, B.R. (1981). An experimental analysis of teaching by telephone. *Progress in Applied Social Psychology*, 15, 143-178.

[22] SHORT, J., WILLIAMS, E. & CHRISTIE, B. (1976). *The Social Psychology of Telecommunications*. London: Wiley.

[23] SPROULL, L., & KIESLER, S. (1991). *Connections*. Cambridge: MIT Press.

[24] STEFIK, M., FOSTER, M., BOBROW, D., KAHN, K., LANNING, K., & SUCHMAN, L. (1988). Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. In I. GRIEF (Ed.), *Computer Supported Co-operative Work*. San Mateo, CA.: Morgan Kaufmann.

[25] SUCHMAN, L. (1985). *Plans and situated actions*. Cambridge: Cambridge University Press.

[26] TATAR, D., FOSTER, G., & BOBROW, D. (1991). Design for conversation: Lessons from COGNOTER. *International Journal of Man-Machine Studies*, 34, 185-211.

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

[28] WHITTAKER, S. J., BRENNAN, S., & CLARK, H. H. (1991). Co-ordinating activity: An analysis of computer supported co-operative work. *Proceedings of Conference on Computer Human Interaction*, 361-367.

[29] WHITTAKER, S. J., & O'CONAILL B. (1993). An evaluation of video-conferencing. *Adjunct proceedings of Conference on Computer Human Interaction*, 73-74.

[30] WILLIAMS, E. (1977) Experimental comparisons of face-to-face and mediated communication. *Psychological Bulletin*, 84, 963-976.