“Where Are You Pointing At?” A Study of Remote Collaboration in a Wearable Videoconference System

Martin Bauer
University of Stuttgart
Computer Science Dept.
Breitwiesenstrasse 20 –22
70565 Stuttgart, Germany
MP.Bauer@studbox.uni-stuttgart.de

Gerd Kortuem
University of Oregon
Dept. of Computer Science
Deschutes Hall
Eugene, OR 97403 USA
+1 541 346 1381
kortuem@cs.uoregon.edu

Zary Segall
University of Oregon
Dept. of Computer Science
Deschutes Hall
Eugene, OR 97403 USA
+1 541 346 0488
zs@cs.uoregon.edu

Abstract

This paper reports on an empirical study aimed at evaluating the utility of a reality-augmenting telepointer *in a wearable videoconference system. Results show that using this telepointer a remote expert can effectively guide and direct a field worker's manual activities. By analyzing verbal communication behavior and pointing gestures, we were able to determine that experts overwhelmingly preferred pointing for guiding workers through physical tasks.*

Keywords
Wearable computing, CSCW, video conferencing, telepointer, augmented reality, empirical study.

1. Introduction

Remote collaboration is one of the most promising applications of wearable computing technology. Past studies of mobile collaborative systems have shown that field workers make repairs more quickly and accurately when they have a remote expert helping them [5,6,7].

From experience with desktop-based groupware systems like MMConf [1], CoLab [8] or GroupWeb [2] we know that *telepointers* are an effective means for users to establish a common understanding during computer-mediated conversation. A telepointer is a mouse pointer controlled by a remote user that allows pointing at objects in a shared window. Without telepointers, users have to rely on verbal description, since they cannot simply say ‘that one’. Instead they have to give a precise description of the object they refer to, for example ‘the yellow round hole in the upper right corner'.

In this paper, we report the results of an experimental study conducted at the University of Oregon that aimed at evaluating the utility of a *reality-augmenting telepointer* in a wearable videoconference system. We use the term *reality-augmenting telepointer* to describe a telepointer that is displayed on a wearable computer's head-mounted display and is used for pointing at real-world objects in the wearable user[1]'s physical space.

We conducted this experiment in order to compare remote collaboration with and without reality-augmenting telepointer. In particular, we were interested if and how users employ pointing gestures in combination with verbal explanations when guiding another user's manual activities.

Results show that by using a reality-augmenting telepointer a remote user can effectively guide and direct a wearable user's activities. By analyzing verbal communication behavior and pointing gestures, we were able to determine that experts overwhelmingly preferred pointing for guiding workers through physical tasks. In addition, we were able to determine that the majority of verbal instructions contained deictic references like ‘here’, ‘over there’, ‘this’, and ‘that’. Because deictic references are mostly used in connection with and in support of gestures, this finding is a strong indication that participants heavily relied on a combination of pointing. gestures and verbal communication.

While there are many interesting theoretical issues related to communication behavior, our work was also motivated by practical issues. This study was carried out within the larger context of the Netman project [3,4], which is targeted at developing collaborative wearable systems for the support of network technicians.

[1] We use the term *wearable user* to describe a user wearing a wearable computer.

 DOI: 10.1109/ISWC.1999.806696

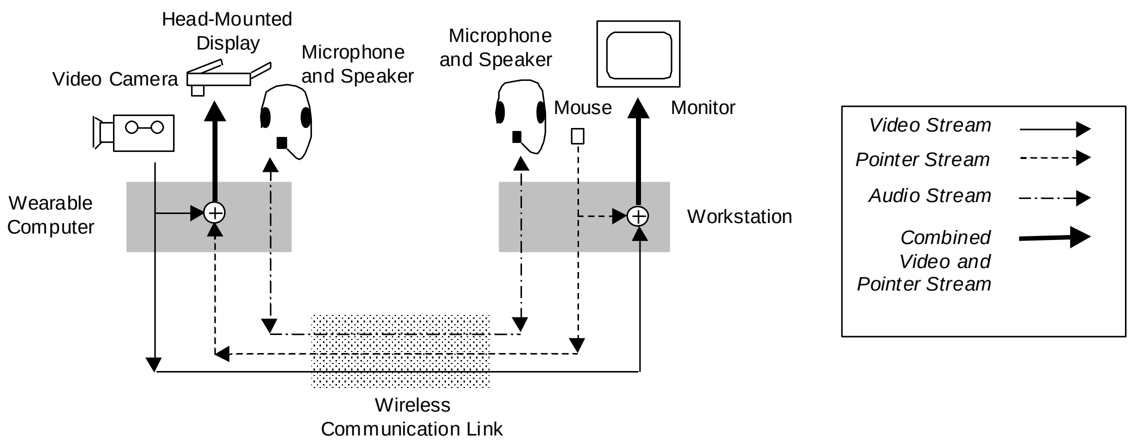


Figure 1. Implementation of Reality-Augmenting Telepointer

2. The System

We have built a wearable system for real-time video collaboration that integrates a telepointer. Because of its ability to augment a wearable user's view of the surrounding physical world, we use the term *reality-augmenting telepointer*. This system is a hybrid system enabling office-based *experts* to collaborate with highly mobile *field workers* using wearable computers. The wearable computer has a head-mounted display (HMD) to which a video camera is attached. The camera is pointing away from the user in the approximate direction of the user's gaze and captures images of the task area. The video image is displayed on the HMD as well as sent over a wireless network link to the remote expert's desktop computer. The expert, in turn, controls a telepointer the (x,y) coordinates of which are sent back to the wearable computer where they are used to display the telepointer on the wearable display (see Figure 1). In order to compensate for the wearable user's head movements, we provide the remote expert with an image-freezing feature: pressing the mouse button freezes the most current video frame on both the wearable HMD and the desktop computer. The expert can then move the telepointer without having to follow the wearable user's head movements.[2]

This arrangement is well suited for "over the shoulder" consultations. The expert can quickly provide assistance by pointing at objects in the wearable user's environment while giving verbal explanations like "connect this cable here". Similarly, the expert can guide the worker through a solution, and watch the novice attempt it himself.

[2] The system was implemented using Microsoft NetMeeting.

3. The Study

3.1. Overview

In order to evaluate the effectiveness of reality-augmenting telepointing, we conducted an experimental study where pairs of test users used our wearable videoconference system to perform a set of artificial tasks. Of each pair, one user played the role of an *expert* or *instructor* guiding the other user, the *worker* or *apprentice*, through a set of predefined tasks. The worker was using the wearable computer with HMD as described above, while the expert gave instructions watching and directing the worker's actions from his desktop computer.

For this study, we used a combination of controlled experiment, protocol analysis, and user questionnaires. Our data includes videotapes, notes taken by experimenters during sessions, post-experimental coding of videotapes, and questionnaires.

3.2. Experimental Hypotheses

Our primary interest in this study was to find out what effect reality-augmenting telepointing has on collaboration. In particular, our goal was to compare outcomes and communication patterns with and without telepointing. Additionally, we studied how image freezing was used by participants.

In order to clarify our expectations, we formulated several hypotheses.

H1. Pointing will be the most decisive communication element. In other words, the expert will use pointing gestures first and foremost and support them by verbal explanations, not the other way around.

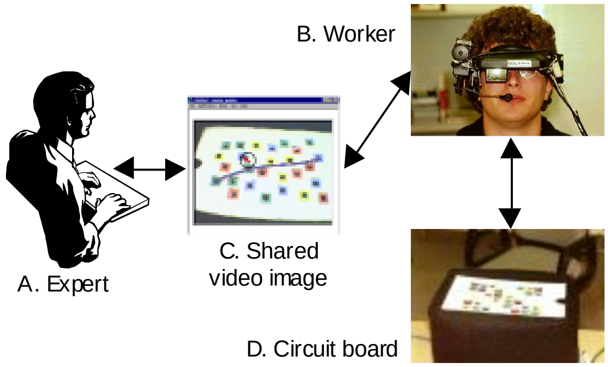


Figure 2. Experimental Setup

H2. The combination of telepointing and verbal communication is faster than verbal communication alone.

Hypotheses H1 and H2 are based on our expectation that telepointing, if implemented adequately, allows the expert to more easily and more precisely direct the worker's actions. As corollary, we put forth the next hypothesis:

H3. Language will be used in the same way as in unmediated co-located situations: In 'difficult' situations, it is used to augment pointing gestures. In less difficult situations it will be used less often.

Our believe that freezing is an effective means to cope with rapid head and camera movements is formulated in hypotheses H4:

H4. The expert will use freezing quite often to compensate for the worker's head movement.

Finally, we expect that pointing leads to a decrease in *communication errors*. We define an *error* as a situation where the worker performs an action different from the expert's instructions or where the result of an action does not match the given instruction.

H5. There will be less *communication errors* and misunderstandings when pointing is used in combination with speech compared to speech-only conditions.

3.3. Method

3.3.1. Participants. Nine pairs, or 18 people altogether, took part in our study. Seven participants were female, eleven were male. All participants were university students majoring in Computer Science or had similar computer science background. All had normal color vision and were free of other uncorrected vision problems that could limit their ability to use a HMD.

3.3.2. Experimental Design. In this study, we used a set of artificial tasks that were derived from the network maintenance scenario of the Netman project [4]. The tasks simulate a situation where an expert or instructor remotely guides a technician through the process of wiring network equipment, such as a network closet full of routers and bridges.

In the study, participants were working in pairs in building a *mockup* circuit (see Figure 2). The "expert" using the desktop computer was given a wiring plan that had to be completed by the "worker" The mockup circuit had to be plugged together on top of a cardboard box (see Figure 2 D.). We could replace the top part of the box with different cardboard templates. The templates had square holes cut into them that were arranged in different ways. Instead of real cables, we used colored cords.

We designed six different templates representing six "difficulty levels" (see Figure 3). The templates differed in the coloring and the arrangement of the holes. One pair of templates arranged the holes as a grid; another pair used a structured arrangement, while holes of the final pair were arranged chaotically. Within each pair, one template was color-coded with yellow, red, blue and green holes, while the second template used no colors at all. The cables were color-coded as well in white, yellow, blue/red, and blue/green.

Our expectation was that the structured templates were more suited for verbal descriptions, while the unstructured ones would force the participants to use pointing gestures more often. Similarly, color-coded holes and cables made it easier for participants to refer to them using verbal descriptions.

For the controlled experiment, we used a within-groups experimental design in which we compared task performance under three conditions:

A. *No-pointing:* participants collaborated through-shared video and audio.

B. *Mainly Pointing:* participants collaborated through shared-video, audio, and pointing gestures, but were asked to use pointing as often as possible, using speech only as a last resort.

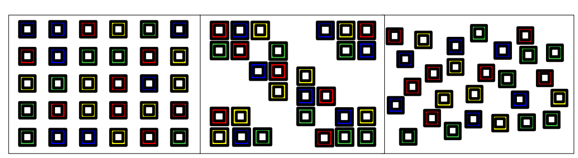


Figure 3. Three circuit templates representing different difficulty levels.

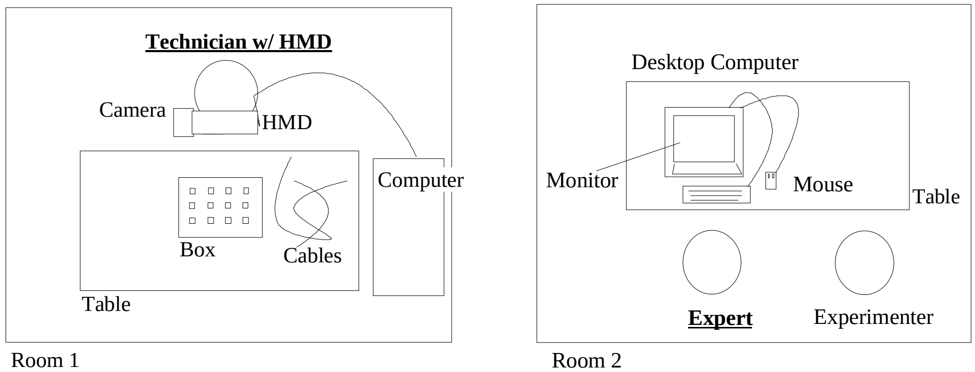


Figure 4. Experimental Setup

C. *Speech & Pointing:* participants collaborated through shared-video, audio, and pointing gestures, in any way they felt most comfortable with.

We did not run tests without image freezing, because our initial observations indicated that movements of the head-mounted camera made it very difficult to use the telepointer without image freezing.

3.3.3 Setup. Each member of a pair was working in a separate room. There was no direct communication; all communication was mediated through our prototype system. Figure 4 shows a plan of the setup in the two rooms seen from above.

The worker was standing or sitting in front of a table, whatever he/she felt was more comfortable. The box, on which the circuit was to be built, was placed on the table. The cables could be found next to it. During the test, the worker had to plug the cables into the holes guided by the expert sitting next door in front of a desktop computer. The expert was given a wiring plan for each subtask. During the test, he/she had to instruct the worker how to set up the circuit according to the wiring plan.

For practical reasons, we had the HMD and the video camera connected to a desktop computer instead of a wearable computer.

3.3.4. Procedure. Each test session, which lasted for approximately 90 minutes, was divided up into six parts:

Introduction: In each test session, both participants were first given a handout containing some information about the goals of the usability study and a short description of the tasks that they were going to perform. Then both participants were shown around both work areas and they had the opportunity to ask questions about the test. We tried to make sure that they had understood how the test was going to be conducted. We left it to each pair to decide who wanted to work with the HMD first. We switched roles about three-fourth through the test, so that all participants had a chance to see both sides.

We had an experimenter/investigator in each room. It was their job to help the participants with the equipment, e.g. to put on the HMD, to prepare the different tasks, and to make sure that the expert had the right wiring plan. During the experiment, they filled out protocol forms that provided us with some information for the evaluation of the study. In addition to the protocols and questionnaires, we videotaped the experiment for later evaluation. We had two cameras: one was directed at a monitor showing the image of the head-mounted display; the other, recorded as picture in picture, was pointed at the wearable user and the task area.

Warm-up Task: To start out, we gave the participants a warm-up task that enabled them to familiarize themselves with the equipment. The worker looked around in the room and the expert picked an object (e.g. the telephone). He/she then froze the picture and pointed at the object while the wearable user had to identify the object. The users were asked to repeat that a couple of times until they felt comfortable enough to continue.

Section 1: In the first real section of the test, the participants were asked to collaborate under no-pointing conditions, i.e. by audio and video only, and without the telepointer[3] (Condition A). Thus, the expert had to describe verbally how the circuit was to be set up while watching the worker's actions on his screen. Two wiring templates were used: a template with a white grid, and one with a structured arrangement of color-coded holes.

[3] To ensure that pointing was not used, we disconnected the mouse from the expert's computer.

Section 2: For the second section of the test, the participants were asked to collaborate under Condition B (pointing-only), i.e. using telepointing as often as possible while keeping verbal communication to a minimum. The same templates were used as in the first section, but with different wiring plans.

Section 3: In the third section, the same participants were asked to accomplish a set of tasks under Condition C (speech & pointing), i.e., they could use whatever combination of verbal communication and pointing gestures they felt most comfortable with. All the templates were used that had not been used in section 1 and 2.

Section 4: After the third section, the participants changed roles. The warm-up task was repeated to familiarize them with the equipment they had to use for their new roles. In the fourth section, they repeated the third section (Condition C, speech & pointing) with switched roles. The same templates, but different wiring plans were used.

After Section 2 and 4, participants were asked to fill out a questionnaire containing questions regarding the usefulness of telepointing in general and our implementation in particular. Most importantly, participants were asked to rate their preferences of using verbal directions only, pointing only, or a combination of verbal directions and pointing gestures.

4. Findings

4.1. Terminology

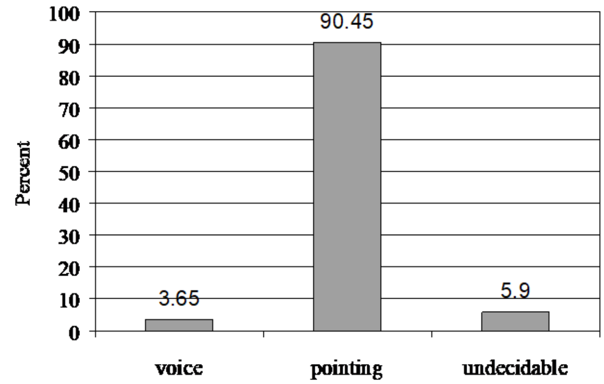
In order to classify the expert's actions we introduce the following terminology:

Pointing Act: a pointing act is a situation in which the expert uses the telepointer to point at a particular hole.

Referential Speech Act: A referential speech act is an expression or utterance used by the expert to identify in which hole the worker has to plug a cable. We distinguished between three different types of referential speech acts:

- An *absolute reference* identifies a hole by its absolute location, as exemplified by the phrase 'the second hole to the left of the lower right corner'.
- A *relative reference* identifies a hole by referring to another object, i.e. hole or cable. It is exemplified by the phrase 'above the yellow cable'
- A *deictic reference* identifies a hole by means of context dependent phrases like 'here', 'there', or 'this' and 'that', as used in "plug it in here" or "put it over there". Deictic references cannot stand alone, but are only used in connection with and in support of pointing gestures.

Freezing Act: A freezing act is a situation where the expert freezes the video image by pressing the mouse button.



Graph 1. Importance of Communication Elements for Condition C

4.2. Pointing Acts

Pointing gestures were used very often throughout the experiment. In Condition C (speech & pointing), there were a total of 576 cases in which the worker had to connect a cable to a hole. Out of this theoretical maximum, pointing acts were used in 572 (99.31%) cases.

Yet, more interesting is the question which communication element was most decisive for the worker's decision to connect a cable to a particular hole. In some cases, the expert would direct the worker by simultaneously giving a verbal command and making a pointing gesture. In other cases, however, the expert would use either speech or gestures, but not both.

We define the *decisive element* as the communicative element that first establishes the identity of the object being referred to by the expert, so that the worker can act upon this information. In order to determine the decisive element, we analyzed the information that was available at the exact moment when the worker started moving towards the target.

We found that in Condition C, that is when users were given a choice whether to use speech or pointing, in 521 (90.45%) cases, a pointing act could be determined as the

Table 1. Frequency of Pointing and Referential Speech Acts for Condition C[4]

	Absolute References	Relative References	Deictic References	Pointing
Grid Template	35.43%	5.55%	38.88%	98.61%
Struct. Template	32.64%	7.65%	36.78%	99.30%
Chaos Template	17.7%	2.07%	55.89%	99.30%
Overall	28.59%	5.09%	43.87%	99.07%

decisive element; in 21 (3.65%) cases it was a speech act and in 34 (5.9%) cases it couldn't be decided (Graph 1). This is a strong indication that users overwhelmingly relied upon pointing and that they recognized the telepointer as a tool that helped them accomplish a given task more efficiently.

It is interesting to note that the different template types (grid, structured or chaotic arrangement) that were intended as difficulty levels did not have a large effect on the use of pointing (Table 1). All template types show a very high number of pointing acts around 99%.

4.3. Referential Speech Acts

Table 1 lists the frequencies with which experts used different types of speech acts. Overall, users clearly preferred deictic references, which were used in 43.87% of all cases. Since deictic references can only be used simultaneously with and in support of pointing gestures, this result again indicates the users' strong preference for pointing.

Contrary to our expectations, the template style (grid, structured or chaos) did not have a great impact on communication behavior. Although the deictic references are slightly higher for the chaos template (55.89%) than for grid (38.88%) and structured template (36.78%), the difference is not as big as expected.

4.4. Image Freezing Acts

Freezing was used in 56.77% of all cases (see Table 2). The usage of freezing differed considerably between the groups. As shown in Graph 2, there were two pairs that

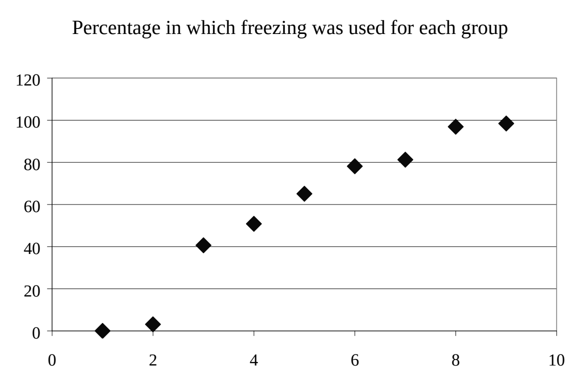
[4] The numbers for grid and structured templates have been adjusted to account for the fact that twice as many chaos templates were used in Condition C than grid and structured templates.

Table 2. Frequency of Pointing and Freezing Acts for Condition C

	Absolute Number	Absolute Percentage	Relative. Percentage
Pointing Acts (total)	572	99.31%	100%
Freezing Acts (total)	330	56.77%	100%
Freezing Acts (short)	21	3.67%	6.36%
Freezing Acts (long)	228	39.58%	69.09%
Freezing Acts (permanent)	79	13.71%	23.94%

practically didn't use it at all (0%-3%), two groups used it all the time (96%-98%) and the other five groups fell somewhere in between (41 %-81%).

One possible explanation for this large variance in the use of image freezing may be a bug in our implementation, of which we became aware only about halfway through the experiment. Occasionally, it caused the frozen image of wearable and desktop system to be off by one frame. In such a case, when the expert pointed at a particular hole, the telepointer on the worker's display would actually point at another hole or be slightly off. Amazingly some participants were able to identify the problem and recover from it. Others, however, were visibly confused. Some users who experienced these problems may have decided to stop using freezing entirely. This is supported by the fact that one group stopped using freezing abruptly between two tasks. In total, we counted 20 out of about 576 pointing acts, in which the pointer was noticeably off.



Graph 2. Image Freezing Acts per Group in Percent

Not everyone used freezing in the same way. Some participants froze the image only for a very short time: 6.36% of all freezing acts were less than two seconds. The majority of freezing acts lasted longer than 2 seconds (69.09%) There were even a few pairs that developed the ‘permanent freeze’ method. They froze the image once at the beginning of each subtask and kept it frozen until the last cable was plugged in. We had an amazingly high percentage (23.94%) of permanent freezing acts. Some of the users who did not use the image freezing feature at all, “implemented” their own way of freezing: the wearable user kept the head still, so the camera image would remain stable at all times.[5]

4.5. Task Completion Time

Although we got the impression during the experiment that telepointing speeds up task completion considerably, - with only nine pairs participating in the study - we couldn’t hope for reliable numbers regarding a possible “speedup”.

There are a lot of factors involved in human communication, which may account for the variance in completion time that we got for the pairs of participants. In addition, since we measured times by hand and task completion times were rather short (about one to two minutes for one wiring plan), systematic errors may have been introduced making our results less reliable.

Still, pairs relying almost exclusively on telepointing appeared to be the fastest. We plan additional tests to measure the influence of pointing on task performance.

4.6. Communication Errors

Owing to the simplicity of the tasks, task completion rate was 100%. However, workers sometimes made mistakes by not following or misunderstanding the experts directions. For example, occasionally the expert would instruct the worker to connect a cable to one hole, but the worker would actually plug it into a different hole. Throughout the entire experiment, experts were very effective in detecting workers’ errors by watching the shared-video image. In general, errors were detected and corrected as soon as they occurred.

Apart from the problem with freezing that we discussed in 4.4, we identified the following reasons for communication errors:

- Vague language: Especially in Condition A (no-pointing) we noticed that verbal directions given by the expert were oftentimes difficult to understand. In Condition C (pointing), the need to use complicated verbal descriptions was reduced, since users could simply point at holes instead of having to describe them. Informal observations suggest that the number of errors in Condition C was much lower than in Condition A. However, we currently do not have enough data to substantiate this claim.
- Poor audio quality: There were a number of misunderstandings that were caused by poor audio quality. We noticed especially low volume, long delays, and lost packets as contributing factors. Part of this problem was due to the audio equipment, especially the headsets; others were caused by how Microsoft NetMeeting handles sound. In the questionnaire, two participants noted audio as a severe problem.
- Poor image quality: The blue/red and blue/green cables were hard to distinguish on the video, so they were sometimes confused. We counted 12 cases in which the worker picked up and used a wrong cable. Except for one group, the participants noticed the problem and corrected the error. Similarly, the white cable is very hard to see on the white templates, especially if the holes are not color-coded. Therefore, the expert was sometimes not able to identify what the worker was doing and to verify if the worker plugged the cable into the correct hole. However, experts were in all cases aware of this problem and were able to resolve it by asking the worker for clarification.
- Low video frame rate: The video frame rate in our experiment never exceeded three frames per second. Thus, rapid head and camera movements resulted in jerky video that made it difficult for the users to collaborate. This resulted in an increase in meta-communication in which the expert instructed the worker to look at a particular object or in a particular direction.

4.7. User Feedback

The results of the questionnaire show that all participants thought the system was useful and enjoyed using it. They gave our telepointer implementation an average of 4.5 on a scale from 1=useless to 5=useful with a standard derivation of 0.6. The difficulty of both pointing (in the role of the expert) and figuring out what was being pointed at (in the role of the worker) was rated as 1.8 on a scale from 1=easy to 5=difficult. The standard derivation was 1.0 and 0.7 respectively.

[5] This behavior was only possible because our artificial tasks did not require the worker to walk or look around much.

5. Discussion

For the most part, the results of this study matched our expectations. The results show that by using a reality-augmenting telepointer a remote user can effectively guide and direct a wearable user's activities. The analysis of verbal communication behavior and pointing gestures clearly indicates that experts overwhelmingly used pointing for guiding workers through physical tasks. While the use of pointing reached 99%, verbal instructions were used considerably less. In more than 20% of all the cases experts did not use verbal instructions at all, but relied on pointing alone instead. Indeed, we would sometimes observe a pair rush through the tasks with hardly any verbal communication at all.

The majority of verbal instructions contained deictic references like 'here', 'over there', 'this', and 'that'. Because deictic references are mostly used in connection with and in support of gestures, this finding is a strong indication that participants naturally combined pointing gestures with verbal communication, much the same they do in face-to-face conversations.

We were not able to prove or disprove our second hypothesis, which states that the combination of pointing and verbal communication is faster than verbal communication alone. Although we saw indications during the experiment that pointing decreases task completion time, we do not have enough data at this point to back up our claim. The same is true for an expected decrease in communication errors.

At the onset of this experiment, we thought that the head movements of the wearable user might represent a severe impediment for effective telepointing. However, the image-freezing feature we implemented proved to be quite effective. The image-freezing feature was used on average in close to 60% of all cases.

We were surprised to find that two pairs used 'permanent freezing', a method in which the picture was frozen at the beginning of the task and would not be released until the entire task was completed. The reason that some participants were able to use permanent freezing was that our experiment did not call for extensive head movements. The cable box fits entirely into the camera view, so that head movements were almost unnecessary. In more realistic field situation, permanent image freezing is no longer possible.

6. Conclusion

We have shown that a reality-augmenting telepointer is an effective means for enhancing remote collaboration of wearable users. Participants of our study overwhelmingly used the telepointer to guide a wearable user through a series of physical activities. We see it as an indication of the utility of the reality-augmenting telepointer that participants in our study rarely, if ever used phrases like: "Which hole do you mean?" or "Where are you pointing at?" Such questions would have indicated serious problems in our design.

The system, the evaluation of which we presented here, is a hybrid system for collaboration of users with desktop and wearable computers. As our next goal, we are looking at modifying our system for direct collaboration of two or more wearable users.

Acknowledgments

The results presented in this paper would not have been possible without the generous support of the Human-Interface Group at the University of Oregon. In particular we would like to thank Jay Schneider, Sarah Douglas, Eck Doerry and Ted Kirkpatrick.

References

[1] Crowley, Terrence; Milazzo, Paul; Baker, Ellie; Forsdick, Harry, and Tomlinson, Raymond. MMConf: An Infrastructure for Building Shared Multimedia Applications. Proceedings of Computer Supported Cooperative Work (CSCW) 1990.

[2] Greenberg, Saul and Roseman, Mark. GroubWeb: A WWW Browser as Real Time Groupware. In Companion Proceedings of the ACM SIGCHI'96 Conference on Human Factors in Computing Systems; 1996.

[3] Kortuem, Gerd. *Some Issues in the Design of User-Interfaces for Collaborative Wearable Systems* Position paper for VRAIS98 Workshop on Interfaces Wearable Computers, March 15, 1998, Altlanta, GA.

[4] Kortuem, Gerd; Segall, Zary, and Bauer, Martin. NETMAN: The Design of a Collaborative Wearable Computer System. MONET: Mobile Networks and Applications; 1999

[5] Kraut, Robert E.; Miller, Mark D., and Siegel, Jane. Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication. Computer Supported Cooperative Work '96 Cambridge, MA, USA: ACM; 1996.

[6] Siegel, J., Kraut, R.E.., John, B.E., Carley, K.M. An Empirical Study of Collaborative Wearable Systems, CHI Companion 1995, Denver, Colorado, 312-313.

[7] D. Siewiorek, A. Smailagic, L. Bass, J. Siegel, R. Martin and B. Bennington Adtranz: A Mobile Computing System for Maintenance and Collaboration. Proc. of the 2nd International Symposium on Wearable Computers; 1998 Oct; Pittsburgh, PA

[8] Stefik, M.; Foster, G.; Bobrow, D. G.; Kahn, K.; Laning, S., and Suchman, L. Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings. Computer Supported Cooperative Work: A Book of Readings. Edited by Irene Greif. San Mateo, CA: Morgan Kaufmann. 1988.