

Verbal Communication During Cooperative Object Manipulation

Roy A Ruddle
Informatics Research Institute
School of Computing
University of Leeds, UK
+44 113 343 5430
royr@comp.leeds.ac.uk

Justin CD Savage
School of Psychology
Cardiff University, UK.
+44 292 087 4007
savagejc@cardiff.ac.uk

Dylan M Jones
School of Psychology
Cardiff University, UK
+44 292 087 4007
jonesdm@cardiff.ac.uk

ABSTRACT

Cooperation between multiple users in a virtual environment (VE) can take place at one of three levels, but it is only at the highest level that users can simultaneously interact with the same object. This paper describes a study in a straightforward real-world task (maneuvering a large object through a restricted space) was used to investigate object manipulation by pairs of participants in a VE, and focuses on the verbal communication that took place. This communication was analyzed using both categorizing and conversation analysis techniques. Of particular note was the sheer volume of communication that took place. One third of this was instructions from one participant to another of the locomotion and manipulation movements that they should make. Another quarter was general communication that was not directly related to performance of the experimental task, and often involved explicit statements of participants' actions or requests for clarification about what was happening. Further research is required to determine the extent to which haptic and auditory feedback reduce the need for inter-participant communication in collaborative tasks.

Categories and Subject Descriptors

I.3.6 [Computer Graphics]: Methodology and Techniques - *Interaction Techniques*. I.3.6 [Computer Graphics]: Three-Dimensional Graphics and Realism - *Virtual Reality*. H.5.2 [Information Interfaces and Presentation]: User Interfaces - *Input devices and strategies*.

General Terms

Algorithms, Measurement, Performance, Design, Experimentation, Human Factors.

Keywords

Virtual Environments, Object Manipulation, Verbal Communication, Piano Movers' Problem, Rules of Interaction.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CVE '02, September 30-October 2, 2002, Bonn, Germany.
Copyright 2002 ACM 1-58113-489-4/02/0009...\$5.00.

1. INTRODUCTION

Collaboration in a virtual environment (VE) can take place at one of three levels [8]. At the most basic level (Level 1), users co-exist in the VE, and can perceive and communicate with each other. At an intermediate level (Level 2), each user can individually modify the contents of a scene, but it is only at the most advanced level (Level 3) that users can simultaneously act on the same object. Cooperative object manipulation, where two or more users work together to manipulate virtual objects, is a classic example of Level 3 collaboration, and the types of task that have been studied include arranging furniture within a virtual room, solving a puzzle, and moving a ring along a virtual wire or a large object through virtual buildings [1, 5, 12, 14, 16, 17].

From a research perspective, object manipulation is ideal for studying the manner in which users can interact and perform practical tasks together in VEs, and this is the primary purpose of the present study. However, from an applied perspective, cooperative object manipulation also has great potential in domains such as simulation, training, design reviews and data exploration. Within simulation and training, VE systems can be used to mimic certain aspects of real-world operations. While current technology places limitations on the fidelity with which users can interact (e.g., the lack of locomotion devices and extended-range haptic feedback), astronauts can be trained in procedures for extravehicular activity even when not co-located, and manufacturing designers can gain insights into the ergonomic problems of a design by "being" virtual humans and simulating together operations of manual materials handling such as the installation of a dashboard into an automobile or the evacuation of a casualty on a stretcher. The role of collaboration in design reviews (e.g., in manufacturing [4]) and data exploration (e.g., in the oil and gas industry) is one of promoting interplay and the exchange of ideas between pairs or small groups of people. Here, by changing the process of interaction from being one-sided ("I do this, while you watch") to being truly cooperative, there is great potential for speeding up communication, ideas testing and information discovery.

Most experimental studies of cooperative manipulation have allowed users to verbally communicate with each other, but few of these studies have attempted to analyze that communication. Instead, these studies have tended to concentrate on objective measures of task performance (e.g., time taken) and subjective measures of "togetherness" and co-presence (e.g., [1, 17]). The present article reports findings of an experiment in which pairs of participants cooperated to move a bulky object through sections of a virtual building, a task known as the piano movers' problem

(e.g., see [7]). Objective data from this experiment, including measures of task performance and interaction behavior (e.g., degree of coordination of hand movements), and full details of the method used in the investigation are reported in [12]. The present article focuses on the verbal communication that took place between participants while they manipulated the virtual object, the data for which are previously unreported.

2. VERBAL COMMUNICATION

Participants have been allowed to communicate verbally in a variety of studies that have used collaborative virtual environments. The tasks performed in these studies ranged from object manipulation [5, 16], to solving puzzles [14, 15, 17], and searching small virtual buildings [3]. However, of these, it was only in the studies by Cotton [3] and Hindmarsh [5] that participants' verbal communication was recorded and analyzed.

The analysis of verbal communication can involve the measurement of the quantity of communication that takes place, as well as what and how people speak. Two general methodologies that can be used are based on the categorization of people's utterances (see [2]), and the detailed study of talk-in-interaction (conversation analysis; [13]). These two methodologies were used in the studies by Cotton et al. [3] and Hindmarsh et al. [5], respectively.

Categorization is best suited to situations where communication tends to follow set protocols, for example, in flying and military operations (see [2, 3]). Once categorized, the data can be analyzed in ways such as counting the number or proportion of statements that fall into each category.

Conversation analysis studies the order, organization and orderliness of verbal communication. The general aim is to identify patterns that exist in verbal communication, and build "collections" of instances that illustrate particular conversational phenomenon and the dynamics by which different people take turns to speak. Thus, close attention is paid to the content and form of the verbal communication. Conversation analysis lays out conventions for the transcription of verbal communication and has a basic unit of a turn. A turn is an utterance, and these occur in sequences (e.g., a greeting and return, or a summons and answer). An introduction to the methods used in conversation analysis can be found in [6].

The context of the following experiment is that of a collaborative task that would be trivial to perform in the real world, being performed in a VE. The task is defined as trivial because participants were shown how to move the object through the environment rather than having to determine whether it was possible for the object to be moved (participants only had to execute, not solve, the piano movers' problem). Despite this, participants frequently had difficulty performing the task whether they manipulated the object on their own in a VE [11] or in cooperation with another participant [12]. In the more difficult of the two versions of the piano movers' problem that were studied (the C-shaped VE; see below), twin-user manipulation took more than 50% longer than single-user manipulation.

One explanation for the difficulties that participants encountered is limitations in the sensory fidelity that existed in the VEs. For example, the graphical field of view was 50° (typical for a VE but much smaller than our view of the real world) and, due to the large range of movements that were required to manipulate the object, haptic feedback could not be provided (a possible work

around for this would be to use a haptic clutch; see [9]). This latter factor is likely to have been particularly important because it meant that participants found it difficult to determine the actions that each other were attempting to perform. To compensate, it seems likely that participants engaged in a greater amount of verbal communication. Thus, analyzing the communication that took place may allow us to identify general types of problem that occur in collaborative interaction and areas in which interfaces can be improved. It was this type of approach that led Hindmarsh et al. [5] to implement peripheral lenses within their graphical display and refine the methods by which users' actions were represented.

3. EXPERIMENT

Pairs of participants moved a large object through two types of VE. One contained two openings that were *offset* from each other, and the other was a *C-shaped* section of corridor. Participants' interactions, which were made via a pair of virtual humans that were situated in the VE, were integrated together using two different types of rule that were chosen to represent examples of symmetric and asymmetric action integration (for a discussion of this, see [12]). One of the rules allowed only the *synchronized* component of participants' manipulations to take place, and the other allowed the *mean*. A repeated measures design was used, with each pair of participants at different times performing the task using both rules of integration. Video recordings of participants' physical actions and verbal communication were made.

3.1 Method

3.1.1 Participants

Twenty participants (11 men and 9 women) took part in the experiment. Their mean age was 22.6 (SD = 4.5). All the participants volunteered for the experiment, were paid an honorarium for their participation, and had previously successfully completed another experiment in which they performed similar tasks, but in a single-user mode [11]. Participants performed the experiment in pairs that were divided into two groups to counterbalance the order in which the two rules of interaction (synchronized or mean) were used.

3.1.2 VE Application

The VE software was a C++ Performer application that was designed and programmed by the authors, and ran on an SGI Maximum IMPACT workstation. This drove the display for both participants in each pair, with the graphics frame buffer divided into two VGA outputs (the view for each participant) that were supplied to two 86 cm (34-inch) monitors. The application update rate was 15 Hz.

The participants in each pair stood back-to-back, facing a monitor, and separated by a wooden partition. Participants were allowed to talk to each other but could not see each other. All they could see was a view of the VE, which showed the walls and floor of the environment, and two virtual humans that were carrying a large object. Interior and plan views of the two environments that were used are shown in Figures 1, 2 and 3. For details of illustrative videos, see Appendix A.

The two virtual humans were the embodiments of the two participants within the VE. One of these virtual humans (Human 1, controlled by Participant 1) moved predominantly forwards, whereas Human 2 (controlled by Participant 2) moved predominantly backwards (see Figure 3). Each participant's

viewpoint was positioned 3 m behind the position of their virtual human, connected by an egocentric tether (an "over the shoulder" view). This type of view has been adopted in a number of other VE studies and allowed each participant to see their human's immediate surroundings in the VE, despite the impoverished field of view ($48^\circ \times 36^\circ$). The object was an abstract shape that comprised a 1.5 m long central limb, and two $0.5 \times 0.5 \times 0.5$ m stubs (e.g., see Figure 2).

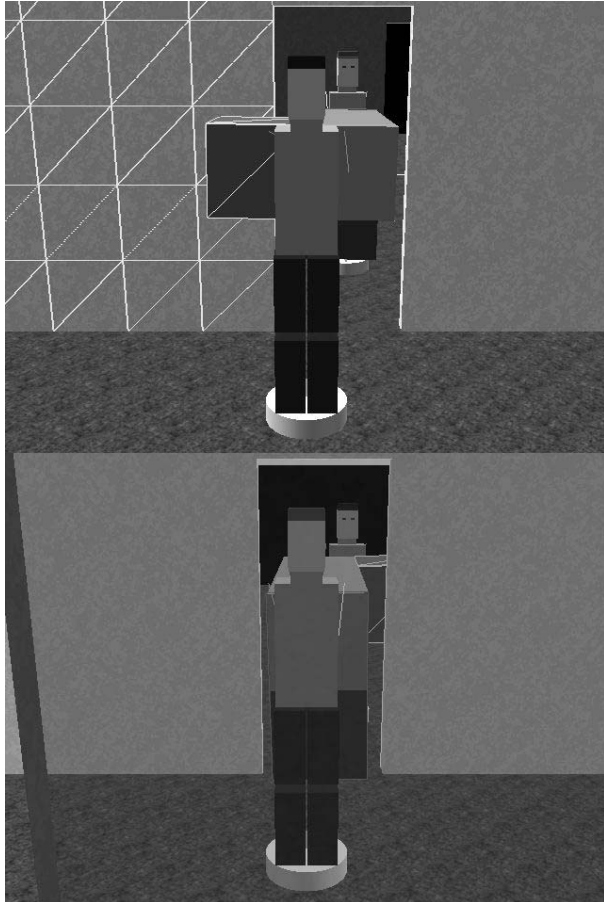


Figure 1. A view inside the offset VE showing the view seen by Participant 1 (top) and Participant 2 (bottom) of the same setting. The wireline highlighting indicates that a collision is taking place between the wall and the stub of the object held by Participant 1.

3.1.3 User Interface

Each participant held an interface prop, the position and orientation of which was tracked using a Polhemus Fastrak sensor. The props were small boxes ($100 \times 75 \times 40$ mm) that had the sensor mounted on the top, and four buttons. Two of the buttons allowed participants to move forward and backwards. The third button acted as a clutch that allowed participants to reposition and reorient the prop and, therefore, their hands without changing the position or orientation of the object. The fourth button was used to change the mode of the Fastrak sensor. When the button was held down, changes of the prop's orientation caused the participant's direction of view to be rotated. If the third and fourth buttons were held down simultaneously then the participant's virtual hand position remained fixed but their body

was repositioned according to their physical hand movements. This allowed participants to move their virtual humans directly in any direction and was particularly useful in the offset VE because it allowed the virtual humans to sidestep between the two openings. The flexibility provided by this type of interface produced a 33% reduction in interaction time compared with a basic interface in the single-user version of the task [11].

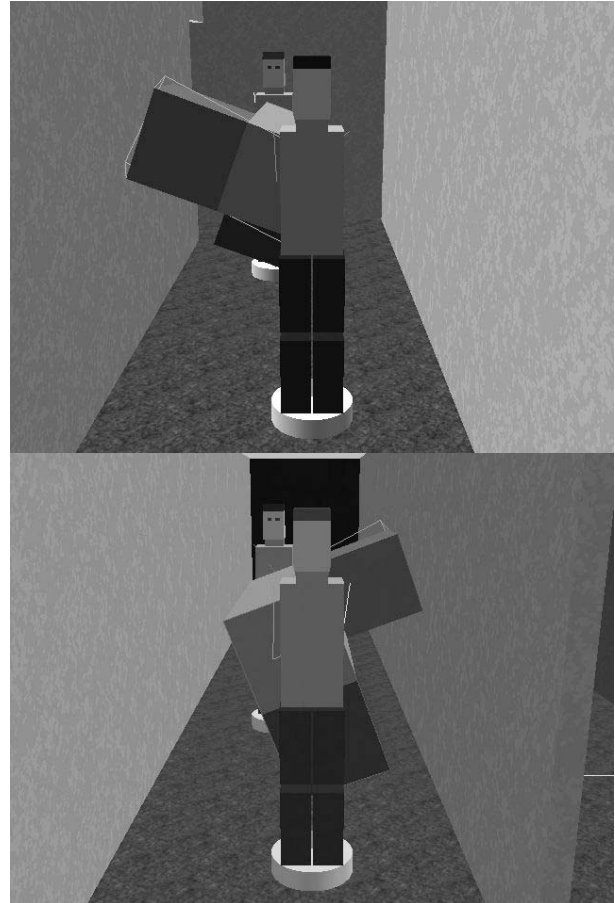


Figure 2. A view inside the C-shaped VE showing the view seen by Participant 1 (top) and Participant 2 (bottom) of the same setting. The wireline image of the end of the object shows where Participant 1 is trying to manipulate the object to (synchronized rule of interaction).

Throughout the duration of each trial, the two virtual humans grasped the two ends of the object. In general, there was 1:1 correspondence between the physical movements of a participant's hand and the movements that their virtual counterpart attempted to make. Exceptions to this occurred when the participant was using the clutch, sidestepping, or there was a collision. When the clutch was used the object and the virtual human both remained stationary while the participant physically changed the position and orientation of their hands. Sidestepping worked in the same way.

If the object collided with the environment then it was prevented from moving (it was not allowed to penetrate the walls, floor or ceiling) but a participant could still reposition their virtual human relative to the object. If one virtual human collided with the object then the other human could still move. If a virtual human

collided with the environment then a "slip" response algorithm (see [10]) allowed the virtual human to continue moving tangentially to the colliding surface, representing the fact that, in real life, it is trivial for people to avoid walking into walls.

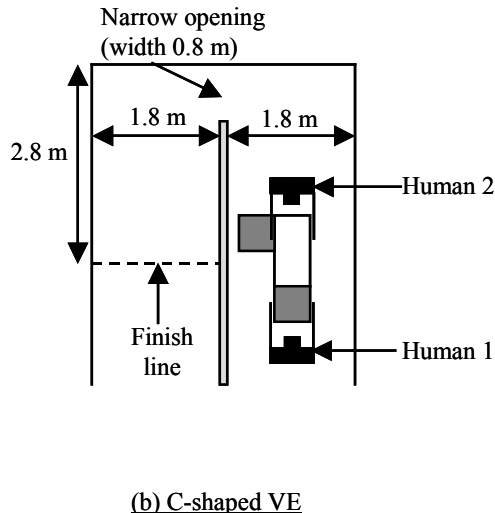
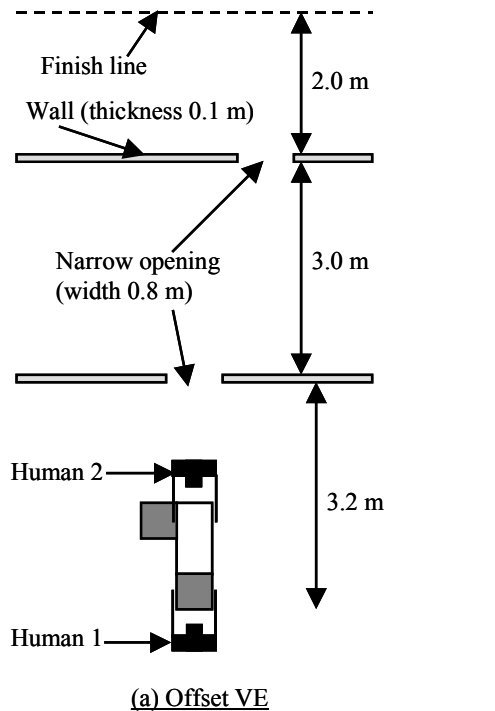


Figure 3. Plan views of the offset (a) and C-shaped VEs (b). In both cases, the ceiling was at a height of 2.4 m and the narrow openings were 2.0 m high. Human 1 moved forwards and was controlled by Participant 1. Human 2 moved backwards and was controlled by Participant 2.

The process of movement was broken down into two stages. First, the translatory movements of the two virtual humans were calculated by taking the mean or the dot product (mean and synchronized rules, respectively) of the raw movements of the

two humans, and this ensured that the two humans did not drift apart even though their speed or direction of movement usually differed. With mean movement, each participant could move both humans and the object through the VE, even if the other participant's human was attempting to remain stationary. However, with synchronized movement, progress would only be made if both participants moved their respective humans in non-opposing directions. The second stage of movement was object manipulation, and this was also calculated using the mean or dot product for both translation and rotation.

Clearly, there was usually a discrepancy between the manipulations that each participant attempted to make to the object, and those that actually took place. This was indicated using wireline, graphical feedback (see Figure 2).

3.1.4 Procedure

Participants were run in pairs and performed the experiment over two separate days. On the first day they performed trials in the offset VE and on the second they performed trials in the C-shaped VE. At the start of the first day, the experimenter demonstrated how to perform the piano mover's task, using a physical scale model of the object and the offset environment. Then the participants practiced moving the object through the offset VE in single-user mode. It is important to emphasize that all of the participants were already familiar with the experimental task because they had previously taken part in an experiment that studied single-user interaction for the piano mover's problem.

After the single-user practice, the participants performed trials in twin-user mode. First they performed three practice trials using one of the interfaces (e.g., synchronized), and then three practice trials using the other interface (e.g., mean). Then they performed six test trials using the first interface, and then six test trials with the second interface. Each set of test trials was split into two blocks of three. In each trial, a participant carried the object from the starting position until both virtual humans had crossed the finishing line, which was marked on the floor of each VE (see Figure 3).

The format of the second day was identical to the first, except the C-shaped VE was used. During all the practice trials, the experimenter gave advice on how to perform the task, but during the test trials the experimenter was silent. If participants had not completed a test trial after 600 s then the trial was terminated and they progressed to the next trial.

3.2 Results

The focus of this paper is on the verbal communication that took place between participants during their test trials with each combination of interaction rule (mean or synchronized) and VE (offset or C-shaped). To place these data in context, we first present an analysis of the time participants took to perform the trials. We then present analyses of the verbal data in terms of the quantity of communication that took place, the balance of that communication between the participants in each pair, categorization of turns in the communication, and the detailed content of the communication. As a first stage in performing these analyses, all of the verbal communication was transcribed onto spreadsheets. Each row in a spreadsheet contained a transcription of the words spoken in each utterance (a *turn*) and a number that identified the speaker (1 or 2). If both participants spoke together (overlapping speech) then the words spoken by both were transcribed and the identifier number was set to 3. Pauses in the communication were indicated by blank rows in the spreadsheets

and marked the end of one *period* of communication and the start of the next.

Two types of statistical analysis were used. The first of these was correlations of different dependent variables. The second was analyses of variance (ANOVAs) that treated the interaction rule, VE, and trial number as a repeated measures. Interactions from these ANOVAs are only reported where they were significant.

3.2.1 Trial Time

An ANOVA showed that participants took significantly less time to perform the trials in the offset VE than in the C-shaped VE, $F(1, 9) = 33.82, p < .01$, and significantly less time as the trials progressed, $F(5, 45) = 3.73, p = .01$. However, the difference between the two rules of interaction was not significant, $F(5, 45) = 0.64, p = .44$ (see Figure 4).

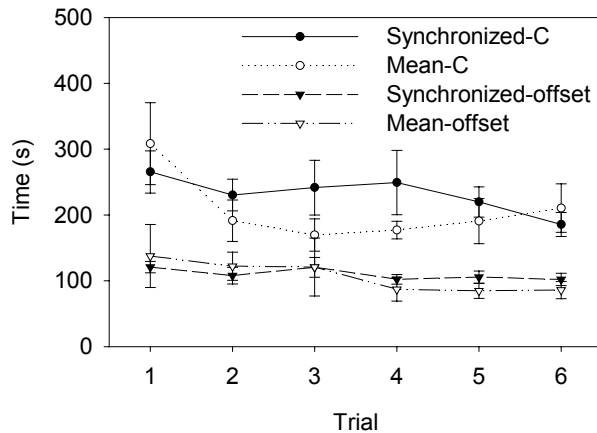


Figure 4. Time taken to perform the trials using the mean and synchronized rules in each VE.

3.2.2 Quantity and Rate of Communication

The three basic units of communication between participants were periods, turns, and words. The spreadsheet transcriptions identified each turn in the communication, the number of turns in each period, and the number of words in each turn. Correlation coefficients were calculated between the time taken in each trial and the three units of communication. The correlations were strongest between the time and number of turns, time and number of words, and the number of turns and words (see Table 1).

On average, there were 39 (offset VE) and 66 turns (C-shaped VE) in each trial, and in only one trial (offset VE) did no verbal communication take place at all. The average number of words spoken in each trial was 134 and 288 for the offset and C-shaped VEs, respectively. The time that participants took to perform the trials differed significantly between the two VEs and between trials. To take account of this, analyses were made of the rate of communication (number of turns and words per second) rather than the pure quantity of communication. The turns and words data produced similar results, so only the words data are reported here.

A repeated measures ANOVA showed that participants spoke significantly more words per second with the synchronized rule than with the mean rule, $F(1, 9) = 6.90, p < .03$, and significantly fewer words per second as the trials progressed, $F(5, 45) = 6.74, p < .01$. However, the difference between the two environments was not significant, $F(1, 9) = 2.68, p = .14$ (see Figure 5).

Table 1. Pearson's correlation coefficients between time, and the number of periods, turns and words spoken in each VE (N = 120; * $p < .05$; ** $p < .01$).

| Offset VE | No. periods | No. turns | No. words |
|-------------|-------------|-----------|-----------|
| Time | .37** | .86** | .94** |
| No. periods | - | .08 | .18* |
| No. turns | - | - | .94** |
| C-shaped VE | No. periods | No. turns | No. words |
| Time | .60** | .87** | .88** |
| No. periods | - | .37** | .33** |
| No. turns | - | - | .93** |

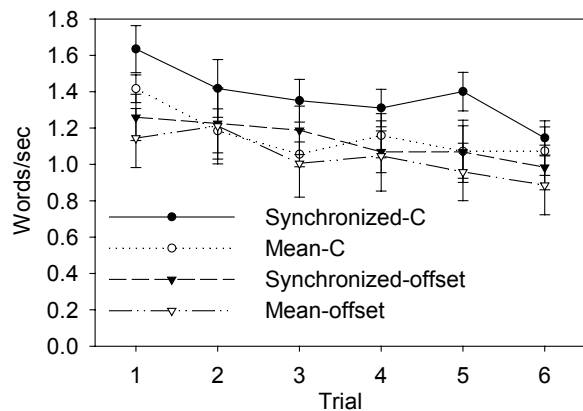


Figure 5. Rate of verbal communication (words/sec) while using the mean and synchronized rules in each VE.

3.2.3 Balance of Communication

The transcriptions contained details of the speaker(s) for each turn in the verbal communication. Of the 240 trials that took place, 91 contained no turns in which participants' speech overlapped, including the one in which no verbal communication at all took place. In the remainder of the trials, up to 29% of the turns were overlapping. An ANOVA showed that there were a significantly higher percentage of overlapping turns in the offset VE trials than in the C-shaped VE ($M = 5.0\%$ vs. 2.6%), $F(1, 8) = 25.31, p < .01$. However, there was no significant difference between the interaction rules, $F(1, 8) = 1.63, p = .24$, or across trials, $F(5, 40) = 0.78, p = .57$.

Analysis was also made of the proportion of turns taken by Participant 1 and 2 when they spoke individually (Participant 1 moved forwards through the VE and Participant 2 backwards). In this, the proportion of turns taken by Participant 1 was calculated as $100 * (\text{Participant 1 speaks individually}) / (\text{either participant speaks individually})$. An ANOVA showed that Participant 1 took significantly fewer turns as the trials progressed, $F(5, 40) = 2.46, p = .05$, with the proportion of turns taken by this participant falling from 47.7% (Trial 1) to 44.6% (Trial 6). The differences between the two environments and rules of interaction were not significant.

3.2.4 Categories of Communication

Participants' communication could have been categorized according to the content of each individual turn, or each sequence of turns. The former approach was chosen because many sequences involved an extended set of turns in which participants' communication covered many different topics.

The primary categories that were chosen were: (a) movement, (b) simple, and (c) other. In the *movement* category were turns that referred to the way in which participants manipulated the object (e.g., "blue up"; rotate the object so that the blue stub points upwards) or moved the virtual humans through the VE (e.g., "ok walk through" or "sideways"). These turns were subdivided according to whether the movement should be executed by *both participants* (e.g., "ok walk through"), only by the *speaker* (e.g., "I'll just move back a bit") or by the *speaker's partner* (e.g., "can you come forwards anymore?"). *Simple* turns were those that only contained words such as "ok", "good", "wicked", "done", "bugger", and "argh". These often represented the closure of a sequence, for example, an acknowledgement or an explicit statement that one stage of the manipulation task had been completed, and many were one-word utterances. A breakdown of the content of communication that was classed as "other" is shown in the following section. The categorization process was performed while viewing and listening to video recordings of participants' actions, and this helped to resolve ambiguities.

The percentage of turns falling into each category is shown in Figure 6. Separate ANOVAs were performed for each category and a summary of these is shown in Table 2. In the offset VE, a larger percentage of turns referred to both participants moving than in the C-shaped VE, but the opposite was true for turns that referred to the movements of only one participant (either the speaker or their partner). Similarly, with the synchronized rule a larger percentage of turns referred to both participants moving than with the mean rule, but the synchronized rule produced a significantly smaller percentage of turns that referred to movements of the speaker. Finally, there were a significantly larger percentage of other turns in the C-shaped VE than in the offset VE, and with the mean rule than with the synchronized rule.

3.2.5 Communication Content

The categorization data show that an average of 14.4 (offset VE) and 22.5 (C-shaped VE) turns in each trial referred to movement of the object or the virtual humans. This can be placed in the context of the experimental task by listing the nine manipulation operations that participants had to perform in the offset VE. To move the object through the first opening, participants had to (1) rotate the object, (2) move forward, (3) rotate, and (4) move forward again. They then had to (5) sidestep, and then (6-9) move-rotate-move-rotate to carry the object through the second opening and over the finish line. Examples of the exchanges that led to a large number of turns when participants were moving the object through just one opening of the offset VE are shown below. In both cases, only the movement turns are included. In Example 1, the participants take many attempts to maneuver the object through the opening, whereas Example 2 illustrates an occasion where participants gave a running commentary of their actions.

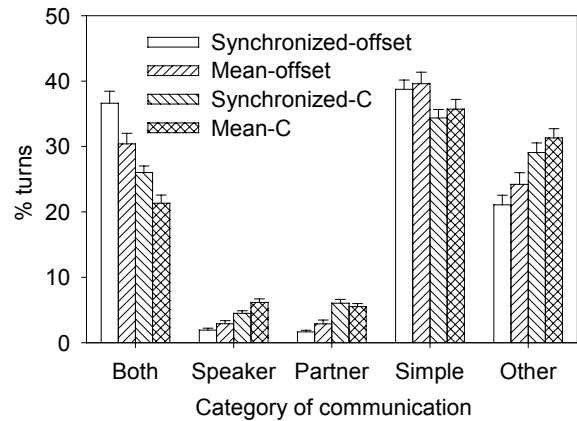


Figure 6. Percentage of turns falling into each category of communication.

Table 2. Fishers' F for ANOVAs of the turns falling into each category (* p < .05; ** p < .01). In none of the analyses was there an effect of trial. Significant interactions (p < .05) occurred between rule and trial for simple utterances, and VE and trial for other utterances.

| Type of utterance | VE, F(1, 8) | Rule, F(1, 8) |
|------------------------------|-------------|---------------|
| Movement (both participants) | 14.50** | 7.46* |
| Movement (speaker) | 17.01** | 6.32* |
| Movement (speakers' partner) | 11.86** | 0.65 |
| Simple | 0.35 | 0.25 |
| Other | 13.85** | 6.07* |

A quarter of participants' turns were categorized as "other." Investigation of the content of these turns indicated that they referred to four types of situation in particular. One of these was when the object was in collision with a wall, or one or other of the virtual humans (e.g., "it keeps hitting me"). Although collision feedback was provided in the form of graphical highlighting, it was sometimes obscured from the view of a participant by the VE's walls. Related to this were situations when a participant was unable to move because they were jammed up against a wall (e.g., "I can't go any further").

A second situation occurred when one participant was unable to determine the actions that the other participant was attempting to perform. This was characterized by explicit questions such as "what are you doing", "which way do you want it", "are you moving", and "are you coming towards me?"

The third situation referred to occasions when participants experienced difficulties of which the root cause was the limited field of view. Examples included "I can't see the door for a start", "I can't see you now so you've got to tell me what to do", and some of the problems associated with collision feedback (see above).

The fourth situation occurred when a participant tried to extend the reach of their virtual arms beyond the (realistic) limit set by the VE software. The participant was informed of the violation by a text message that was shown on their display, and the participant often also explicitly told their partner.

Example 1. Turns spoken while moving the object through one opening in the offset VE, during a trial that used the mean rule. Words spoken by the participants moving forwards and backwards are shown in bold and italics, respectively.

| Turn no. | Words spoken (movement turns only) |
|----------|---|
| 53 | right shall we... <i>are you turning it</i> |
| 57 | yeah go straight forwards |
| 59 | <i>go</i> |
| 62 | turquoise up |
| 66 | <i>right wait let's... let's</i> turquoise down turquoise down |
| 72 | <i>let me just go backwards</i> |
| 74 | <i>ok lets try it again</i> |
| 83 | still turning? |
| 85 | <i>ok its going going</i> |
| 88 | straight forwards |
| 92 | can you move... <i>move</i> |
| 93 | can you move forward a bit please |
| 101 | <i>well How about sidestepping</i> |
| 108 | ok I'm moving |

Example 2. Turns spoken while using the synchronized rule.

| Turn no. | Words spoken (movement turns only) |
|----------|--|
| 24 | <i>red up</i> then move red up |
| 26 | forwards <i>move</i> |
| 27 | stop blue up |
| 30 | <i>stop</i> move, stop, move move |
| 31 | <i>up up up up up up up</i> move move move move move move |
| 33 | <i>ok</i> move |
| 35 | <i>ok, let's bring the...</i> |
| 39 | stop forwards |

4. CONCLUSIONS

This study adopted a paradigm called the piano movers' problem to study cooperative object manipulation in CVEs. The basic problem was for pairs of participants to move a bulky virtual object through a restricted space. The present paper focuses on the verbal communication that participants performed to successfully complete the object manipulation task.

The first point that needs to be highlighted is the sheer quantity of communication that took place, averaging in excess of one word per second for the duration of the test trials. Although the rate of communication reduced as the trials progressed, something that is indicative of practice producing a reduction in the attentional

demands of the task, the quantity of communication remained high. This, it can be assumed, was to compensate for the reduction in the amount of sensory information present in VEs, compared with the real world, and in particular a lack of information about other participants' intended actions.

Participants' rate of communication was greater with the synchronized rule than with the mean rule. With the former, participants had to coordinate their actions if the task was to be completed but, despite the amount of communication that took place, objective measures such as the speed and direction of hand movements showed that coordination was poor (see [12]). Coordination was also poor with the mean rule, and it is likely that if participants had devoted more attention to this then the actions of one participant would have conflicted less often with those of the other (e.g., when each participant attempted to manipulate the object in a different manner, with the result that the object simply collided with the environment and, therefore, didn't move anywhere).

Previous research [5] has highlighted four particular problems that occur during collaborative object manipulation: (1) fragmentation of the workplace, (2) confusion repair, (3) making the implicit explicit, and (4) hidden perspectives (people can't always see what each other can see). The latter two were particularly prevalent in the present study.

Making the implicit explicit occurred in two main forms. First, a third of the turns in the verbal communication referred to movement of the object or the virtual humans. The quantity of these turns far exceeded the basic number of manipulation operations that had to be performed. Graphical feedback was provided to show the manipulations that each participant was attempting to make to the object (see Figure 2), but this was clearly less salient and required a much higher "cost" to process than if haptic feedback could have been provided. However, the implementation of haptic feedback in distributed environments has substantial technical hurdles that are centered on the requirements of the haptic update rate (typically 1 kHz) and the difficulty of maintaining consistency between multiple copies of a haptic scene graph [16]. Separate to this is *a priori* information about how another person is *going* to manipulate an object, which in the real world can be inferred from subtle changes in a person's physical body posture. Second, participants sometimes informed their partners that they were reaching too far, indicating that the other person should pause while the first person adjusted their virtual reach. This could be overcome by implementing 1:1 correspondence between participants' physical and virtual hand movements, according to a rule of interaction that we term *physical compatibility* (see [11]). However, this substantially complicates the implementation of the VE interface software.

Hidden perspectives are a familiar problem in VE systems, and in the present study were usually related to situations where either the object was in collision or a virtual human was unable to move any further because of a wall. The cause of hidden perspectives is typically identified as the impoverished field of view that exists with most forms of VE display, but in the present study the root cause lay elsewhere. The lack of haptic feedback meant that participants relied on visual information to be informed about collisions of the object or the other human. If this visual information was not available (e.g., due to an opaque barrier such as a wall) then participants compensated by explicit, verbal communication. However, the use of auditory collision feedback would reduce reliance on the visual channel and is likely to

remove the need for some of the verbal communication that took place.

Finally, the present study provides detailed data on the problems people experience when trying to perform trivial, real-world tasks in collaborative VEs. Research into tasks such as these provides useful guidance for the design of a wide range of collaborative systems, and into the complexities of human-human interaction in everyday life.

5. ACKNOWLEDGMENTS

This work was supported by grant GR/L95496 from the Engineering and Physical Sciences Research Council. We thank Amelia Woodward for her help with coding the verbal data.

6. REFERENCES

- [1] Basdogan, C., Ho, C., Srinivasan, M. A., & Slater, M. (2000). An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction*, 7, 443-460.
- [2] Bowers, C. A., Jentsch, F., Salas, E., & Braun, C. C. (1998). Analyzing communication sequences for team training needs assessment. *Human Factors*, 40, 672-679.
- [3] Cotton, J. E., & Lampton, D. R. (2000). Team communications in a virtual environment. *Proceedings of IEA 2000/HFES 2000 Congress* (pp 523-526). Santa Monica, CA: Human Factors Society.
- [4] Dailey, M., Howard, M., Jerald, J., Lee, C., Martin, K., McInnes, D., & Tinker, P. (2000). Distributed design review in virtual environments. *Proceedings of Collaborative Virtual Environments (CVE '00, 57-63)*. New York: ACM.
- [5] Hindmarsh, J., Fraser, M., Heath, C., Benford, S. & Greenhalgh, C. (2000). Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction*, 7, 477-509.
- [6] Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis: Principles, practices, and applications*. Cambridge, UK: Polity.
- [7] Lengyel, J., Reichert, M., Donald, B. R., & Greenberg, D. P. (1990). Real-time robot motion planning using rasterizing computer graphics hardware. *Computer Graphics*, 24, 327-335.
- [8] Margery D. M., Arnaldi B., and Plouzeau N. (1999). A general framework for cooperative manipulation in virtual environments. *Proceedings of Virtual Environments '99* (pp 169-178). New York: Springer.
- [9] McNeely, W. A., Puterbaugh, K. D., & Troy, J. J., (1999). Six degree-of-freedom haptic rendering using voxel sampling. *Proceedings of the 1999 ACM Conference on Graphics (SIGGRAPH '99, 401-408)*. New York: ACM.
- [10] Ruddle, R. A., & Jones, D. M. (2001). Movement in cluttered virtual environments. *Presence: Teleoperators and Virtual Environments*, 10, 511-524.
- [11] Ruddle, R. A., Savage, J. C., & Jones, D. M. (in press). Evaluating rules of interaction for object manipulation in cluttered virtual environments. *Presence: Teleoperators and Virtual Environments*.
- [12] Ruddle, R. A., Savage, J. C., & Jones, D. M. (in press). Symmetric and Asymmetric Action Integration During Cooperative Object Manipulation in Virtual Environments. *ACM Transactions on Computer-Human Interaction*.
- [13] Sacks, H. (1992). *Lectures on conversation* (Ed. G. Jefferson). Oxford: Blackwell.
- [14] Sallnäs, E., Rasmus-Gröhn, K., & Sjöström, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction*, 7, 461-476.
- [15] Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators and Virtual Environments*, 9, 37-51.
- [16] Slater, M., Steed, A., Crowcroft, J., Whitton, M. C., Brooks Jr. F. P., & Srinivasan, M., A. (2001). Final report of the Collaboration in Tele-immersive Environments project. Retrieved from the World-Wide Web March 19, 2002. <http://www.cs.ucl.ac.uk/research/vr/Projects/Internet2/Report/finalreport.pdf>
- [17] Wideström, J., Axelsson, A., Schroeder, R., Nilsson, A., Heldal, H., & Abelin, A. The collaborative cube puzzle: A comparison of virtual and real environments. *Proceedings of the Collaborative Virtual Environments Conference (CVE'00, pp. 165-171)*. New York: ACM.

7. APPENDIX A

Two MPEG videos support this submission, illustrating trials in the offset and C-shaped VEs (ruddle-offset.mpg and ruddle-c-shaped.mpg, respectively). These can be accessed from the Web page <http://www.comp.leeds.ac.uk/royr/video.html#cve02>. In each video, the view seen by Participant 1 (controlling the virtual human that is wearing the blue hat) is shown on top, and the view seen by Participant 2 (virtual human wearing the red hat) is shown below. The white line seen on the floor at the end of each video is the finish line. Due to an oversight in data recording, the wireline feedback that participants saw while manipulating the virtual object is not shown in the videos. Neither video contains sound.