Where Are You Pointing? The Accuracy of Deictic Pointing in CVEs

Nelson Wong and Carl Gutwin
Computer Science Department, University of Saskatchewan
110 Science Place, Saskatoon, SK, S7N 5C9, Canada
nelson.wong, carl.gutwin @usask.ca

ABSTRACT
Deictic reference – pointing at things during conversation – is ubiquitous in human communication, and should also be an important tool in distributed collaborative virtual environments (CVEs). Pointing gestures can be complex and subtle, however, and pointing is much more difficult in the virtual world. In order to improve the richness of interaction in CVEs, it is important to provide better support for pointing and deictic reference, and a first step in this support is to determine how well people can interpret the direction that another person is pointing. To investigate this question, we carried out two studies. The first identified several ways that people point towards distant targets, and established that not all pointing requires high accuracy. This suggested that natural CVE pointing could potentially be successful; but no knowledge is available about whether even moderate accuracy is possible in CVEs. Therefore, our second study looked more closely at how accurately people can produce and interpret the direction of pointing gestures in CVEs. We found that although people are more accurate in the real world, the differences are smaller than expected; our results show that deixis can be successful in CVEs for many pointing situations, and provide a foundation for more comprehensive support of deictic pointing.

Author Keywords
Pointing, gestures, avatars, CVEs.

ACM Classification Keywords
H.5.3 [**Group and Organization Interfaces**]: CSCW

General Terms
Design, Human Factors, Experimentation

INTRODUCTION
Pointing at things to indicate them to others – that is, deictic reference – is ubiquitous when humans communicate about objects and other people in a shared environment [21]. Deictic pointing allows verbal communication to be much more efficient, by freeing people from the need to construct verbal descriptions that uniquely identify an object [25].

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

Deictic gesture in real-world communication is often a subtle and nuanced action. There are several important stages to a successful pointing gesture [24]: determining that the viewer will be able to see both the gesture and the referent; staging the gesture so that the viewer is prepared to see and interpret it; producing the gesture itself, and holding the gesture while coming to mutual agreement with the observer about what is being pointed at. Through all these stages, the pointing gesture is linked with verbal utterances; and people can use all of the aspects of the situation – orientation, expectation, talk, and gesture – to determine the referent. Despite this complexity, people are experts at producing and interpreting deictic gestures, and the practice is such a natural part of communication that most people only notice its absence.

The ubiquity of this form of communication in the real world led designers of some of the earliest groupware systems to incorporate support for deixis. In these systems, simple telepointers allow people to create complex pointing gestures. The success of telepointers in groupware suggests that pointing should also play an important role in collaborative virtual environments (CVEs) [24]. However, pointing in CVEs can be much more difficult than it is in either the real world or 2D groupware [17, 24]. A key difference is that CVEs present different views of the scene to each person, whereas groupware systems generally provide the same representation – that is, when a person positions their pointer on an object in a 2D scene, the pointer is over that object for the observer as well. This greatly simplifies pointing in groupware – but the divergent views of CVEs mean that observers must interpret the pointing gesture, resulting in a more difficult problem.

Each of the stages of a pointing gesture becomes more difficult in a CVE. It is difficult to imagine another person's view in a CVE [16], and so it is difficult for the actor to determine whether the observer will be able to see the gesture; in addition, the narrow field of view of most CVEs means that less of the world can be seen by the observer [16, 17]. Second, control over avatar pointing in CVEs is dramatically less expressive than pointing in the real world. The many degrees of freedom available to a person in the real world are usually reduced to simple commands or rudimentary controls, and pointing can usually only be directed towards specific objects in the environment. The lack of expressive power means that the subtle staging and

preparation actions that orient an observer to a real-world gesture are impossible to recreate in a CVE. Third, pointing gestures in CVEs often happen through symbolic commands, which removes the progressive and gradual production of the gesture that can provide rich information [15]. Fourth, visual and perceptual differences between CVEs and the real world make pointing actions more difficult to interpret: lack of depth cues, low resolution, artificiality of avatars, and altered perspective and field of view all make it more difficult to determine the precise direction and orientation of pointing.

Some of these problems have been discussed by previous CVE researchers, who have proposed a variety of techniques to improve pointing (e.g., wide-angle views to aid in seeing and orienting to others' gestures). One type of solution goes beyond natural pointing to explicitly augment the gesture – e.g., avatar arms that stretch out toward the target [12], visual highlights on the target [16], or visual 'laser beams' attached to avatar arms (e.g., Second Life).

Although augmented pointing is an appropriate solution in some situations (and has been shown to help in resolving certain referential problems [12]), it may not be the best solution in all cases. While it can be beneficial in object-focused interaction where indicating specific items is the main goal, it is not designed for other situations such as showing areas, paths, and general directions. Also, augmented pointing may be distracting to others (e.g., arms or lasers extending through the view), may reveal one's location unnecessarily (e.g., in a combat game), may require additional controls (e.g., as Fraser et al. discuss [12]), and may be too specific for some purposes (e.g., with a general 'over that way' gesture).

Therefore, even though augmented pointing is an important possibility in CVEs, it is also important to improve more natural deictic pointing gestures as well [24] – that is, to improve people's ability to carry out deictic pointing in CVEs in the same ways that they do in face-to-face settings.

Improving natural pointing involves supporting all of the different stages of the pointing gesture as introduced above. The first step, however, must be to determine whether the most basic element of pointing will work in virtual environments: that is, whether the direction of a pointing gesture can be accurately determined by both the person pointing and the person observing the gesture. This is fundamental to deictic pointing, since if direction cannot be established, then the other stages of pointing cannot play their roles successfully. This paper therefore focuses on interpreting the direction of pointing. We do not discount the importance of mutual orientation, preparatory staging, or gradual production of a gesture; however, the most fundamental part of a deictic gesture is the directional reference itself, and we must first determine whether or not this most fundamental part can be successful.

The issue of interpreting direction in pointing has not been studied in detail before. To investigate this problem, we carried out two studies. The first was an informal observational study to look at how people construct and interpret 'distant pointing' gestures in a real-world setting. This study showed that people produce a variety of gestures when pointing at distant objects, and that different tasks require different levels of pointing accuracy. However, there is no information available about exactly how accurately people can point in a CVE, and so it was still unclear whether the kinds of gestures seen in our study could be translated to that environment.

Therefore, our second study (a controlled experiment) focused specifically on assessing people's ability to interpret pointing direction in both CVEs and the real world. The most important result of this study was that although people were less accurate in interpreting direction in a CVE than in the real world, this difference was less than expected, and was small enough that several types of real-world pointing gestures seen in the first study would be possible in a CVE.

Our results are the first to provide an empirical answer to the question of "can you tell where I'm pointing?", and suggest that natural deictic pointing can be effective in CVEs. By identifying that different kinds of deixis require different levels of accuracy, and by empirically determining what levels of accuracy are possible with naturalistic pointing, we help designers determine whether and when they need to provide enhanced pointing techniques, and when natural pointing will be sufficient. This work is a step toward improving richness and expressiveness in CVEs.

BACKGROUND: POINTING IN COLLABORATION

Referring Expressions

There is a large body of research on referring expressions and the use of pointing for indicating objects (e.g., [18, 21]). This is called deictic reference – using pointing gestures with words like "this" and "that." How people communicate with deictic pointing depends on various factors: the environment, objects within the environment, bodily conduct [18], and mutual knowledge [7, 22]. People generally use other types of verbal communication along with pointing gestures. However, there are situations where it is difficult to rely on verbal descriptions to convey ideas [22] (e.g., indicating objects with hard-to-describe shapes). In this case, pointing gestures can become more important. Nevertheless, the meaning of the gesture can remain unclear without accompanying talk [6].

Pointing in Co-located Physical Settings

When people use pointing gestures in the real world, they are usually in the same physical area. In such co-located settings, gestures have been shown to be important in several different collaborative situations [2, 18, 27].

Gestures play a significant role in communication and collaboration. For example, pointing gestures were found to be useful in directing a group's attention to a common location when a group of people work together [27]. While pointing gestures are normally used to indicate physical

items in face-to-face settings, they can also be used to refer to imaginary objects, abstract concepts, and even other gestures [2, 23]. For example, people often point to the same place in the air as an earlier gesture was made, in order to refer to the previous gesture. In many situations, pointing gestures are used to support speech to convey information in daily activities. However, there is also evidence showing that the opposite can be true [18]: pointing gestures sometimes have the primary role in communication, and speech is used to enhance the meaning of the gestures, rather than the other way around.

Prior research has uncovered considerable complexity in a simple deictic gesture. As summarized by Moore and colleagues [24], there are several issues to consider.

- *Accountability*. A key concept in human visual communication is that actions are *accountable* – that is, they are constructed so as to be understandable to others.
- *Observability*. The production of social actions is observable by others, who can use the information 'given off' by the production to understand the action.
- *Orientation*. People understand that actions are observable, and therefore they often orient the production of their actions so that the information they provide can be seen and used by others.
- *Projectability*. People are able to predict the boundaries and timing of communicative acts, and use these predictions in coordinating talk. Gestures are also projectable: as Moore et al state, "by observing the early stages of a pointing gesture, recipients can begin to predict in advance in what general direction it will go and what the likely referent will be" ([24], p. 272).

From these concepts, a deictic pointing gesture can be divided into four stages: orientation, preparation, production, and holding. Across all of these stages is the goal of achieving mutual understanding, between the speaker and the listener, about what is being referred to [8].

1. *Mutual orientation*. The producer of the pointing gesture must determine whether the observer will be able to interpret the gesture – e.g., whether they can see both the pointing action and the target [16, 24].
2. *Preparation and staging*. The producer may make preparatory actions that indicate to the observer that a pointing gesture is going to be made, allowing them to orient themselves and prepare to interpret it.
3. *Production of the gesture*. Pointing gestures are not immediate, and the gradual production of the action allows people to predict (as described above) the general direction of the gesture before it is completed [15, 24].
4. *Holding*. Once the gesture is produced, it has not succeeded until mutual understanding of the referent has been achieved. Therefore, the gesture's producer must hold the gesture [18] until they are sure that the observer has seen it, and until they are sure that mutual understanding has been reached [24].

Although the first three stages are important (and may well suffice in some situations to establish understanding), it is the last stage (holding) that provides the best information about what is being pointed to, and the fallback state if understanding has not been established earlier.

Pointing in Distributed Virtual Settings

The ubiquity of gestures in the real world has led to considerable research into distributed gestures as well. Gestures in distributed settings have been shown to be useful in improving collaborative task performance [10] and in establishing common ground [20]. When people use gestures to collaborate in distributed environments, the gestures are often represented via embodiments (visual representations of people in groupware systems) that can be as simple as a 2D telepointer [14, 15], or as complex as a video image [26, 28, 29] or a 3D avatar [3, 4].

Various kinds of pointing have been widely studied [1, 9, 15, 19]. However, as discussed above, the majority of this work focuses on pointing where both parties see the same representation of objects and pointer (or telepointer). In contrast, 3D CVEs present more difficult problem because divergent views mean that people see different representations of the pointing gesture in relation to the objects in the environment.

CVEs are a kind of distributed environment that is characterized by their use of 3D models for the world and people's avatars. CVEs are now commonly used for games (e.g., World of Warcraft) and social communities (e.g., Second Life). These CVEs are detailed worlds where people can interact, communicate, and carry out shared activities. As CVEs become more popular in distributed communication and collaboration, their usability also becomes more important. Research has been carried out to explore a wide variety of issues around CVE usability and use, including relationships between CVEs and the physical world [5], interactions that occur inside CVEs [9, 16, 17], and effects of CVE visual structures on interactions [11].

However, there are a number of limitations to current virtual worlds in terms of human-to-human interaction – even current game worlds are "much less advanced in terms of their *interactional* sophistication" ([24], p. 267) in areas such as turn-taking, referential pointing, and awareness [24]. In particular, researchers have noted several problems that complicate the production of deictic gestures.

Fraser, Hindmarsh, and colleagues [12, 16] identify several problems with pointing in CVEs. For example, they discuss how mutual orientation is difficult to achieve because of narrow fields of view in CVEs, and because of the difficulty of determining what others can see in the world. Their experiments show that pointing at objects that are not close by can lead to frequent misunderstandings that require extensive verbal interaction to repair.

Moore and colleagues [24] point out other problems. First, pointing actions in CVEs are often initiated with a command action in the interface (e.g., "\point"), which often means that they are produced immediately, and do not

provide the gradual information available from gestures in the real world. Second, pointing in virtual worlds is often restricted to defined objects in the world, greatly reducing the expressiveness of people's pointing actions. Fraser and Benford [11] provide a deep analysis of this problem, showing how object-based pointing leads to communication failures in collaborative work.

Various techniques and design guidelines for improving interaction with objects in CVEs have been suggested [12, 17]. For example, researchers have suggested several techniques for improving the problem of mutual orientation: wide-angle peripheral lenses to increase a user's field of view; a visible view frustum to better indicate what others can see; or elongated arms for improved pointing visibility. Other techniques are meant to improve the determination of the referent: for example, object highlighting to enhance awareness of what others are looking at; or the ability to look through another person's view to determine what they can see [17].

This research provides a strong basis for understanding pointing in the real world and the problems of supporting natural pointing in CVEs. However, there is still little work that looks at the types of pointing gestures produced in collaborative situations, and no research on how precisely people can determine what others are pointing at. The investigation of these questions is the subject of the two studies described in the next sections.

STUDY 1: OBSERVATIONS OF DISTANT POINTING

In order to determine issues that are important for pointing in CVEs, we carried out an informal observational study to look at the way that people point (and interpret pointing gestures) in the real world. We focus on the final stage of the pointing process (*holding*, as described above) – that is, once a pointing gesture has been produced, the process of coming to agreement about what is being indicated.

There were three issues under investigation in this study: what kinds of pointing gestures people produce for different types of tasks; whether people can in fact determine what others are pointing at; and how communicative richness affects the way that people use gestures.

Methods

Eight participants were paired and asked to perform a series of activities involving pointing at distant objects. The study was carried out in a fifth-floor room that had large windows overlooking a city. There were three types of tasks, all involving pointing at targets outside the windows (e.g., the buildings shown in Figure 1). Participants who performed pointing gestures were called *actors*, and the participants who interpreted the pointing gestures were called *observers*.

In the first task, the actor was given photographs of targets outside the building, and asked to point to the targets. Once the actor was holding their pointing gesture, the observer was asked to determine what object in the world the actor was pointing at. The pair did twelve of these tasks, and then switched roles for twelve additional trials. There were three kinds of targets: specific objects, general areas, and paths. The targets could be either directly visible, partially occluded, or completely out of view (targets that were too far to see or blocked by other objects). We also asked actors to indicate the target in three different ways: pointing gesture only, pointing plus speech, and pointing plus written notes.

Figure 1. Part of the view from the study room.

Figure 2. Actor (right) and observer (left) during a task.

In the second task, the actor was asked to point at ten different objects of her choice. No verbal describing was allowed; the observer was asked to determine each targets as quickly as possible.

In the third task, participants collaboratively decided upon five different locations outside the building to hide imaginary objects. This task allowed us to see how directional interpretation fit within pointing gestures that are allowed to occur naturally.

The experimenter observed the ways that pointing gestures were used and were interpreted by the participants; we also video-recorded the experiment for further analysis, and reviewed the video after each study session.

Observations

The participants used pointing gestures frequently. We categorized the ways that they pointed, types of gestures, and reactions of the observers, based on the types and visibility of the targets. These observations provided initial insights into the three questions underlying the study.

Types of pointing gestures for different tasks

People produced a wide range of pointing gestures, but different gestures were used for different kinds of targets. For example, when indicating a plainly-visible object, actors would use an extended-arm pointing gesture (Figure 2). When pointing at more general areas, people would use a variety of gestures including motions with an open hand, or movement of a pointing finger to circle the area. In

addition, the less rich the communication channel was, the more complex the pointing gestures were.

Ability to determine what actors were pointing at

Success in identifying the referent varied depending on the complexity of the target. Observers were able to identify obvious targets immediately, but performed less well when targets were in a group, hard to describe, or partially occluded. When the targets were landmarks or obvious objects that were visible in the indicated direction, observers had no problem with identification. However, the difficulty of identifying targets dramatically increased when the targets were unobvious. The varying difficulty of determining a referent suggest that there are multiple requirements for specificity in producing pointing gestures. We identified three canonical situations from the study that have different accuracy requirements.

Pointing at an object in a group. Precise pointing appeared to be more important when pointing at an object that is near or within a group of similar objects. People often had trouble identifying targets that were not obvious: for example, when someone pointed at a particular car in a full parking lot, it was difficult for observers to identify the target (see [12] for a similar example in a CVE where accuracy is required to differentiate between look-alike objects). The problem arises both because of the density of the objects, and the difficulty of disambiguating the objects using speech. People would often have to guess at the target within the cluster using a linear search. The more accurate a pointing gesture can be, the less work will be required in these situations.

Pointing at distinct objects. Pointing precision was much less important when targets were distinct or easy to describe. People were able to easily identify targets such as a car, when the car was the only vehicle in a parking lot; in these cases a general directional gesture and the phrase "the car" was usually enough for the observer to correctly identify the right object. In these situations, pointing accuracy was not a major issue; people needed only the general directions of the targets in order to successfully identify them through the verbal cues.

Pointing at out-of-view targets. When targets were out of view, pointing accuracy also appeared to be less important. People tended to rely on the general directions for targets that they could not see from their view. For example, when people wanted to indicate a parking lot that behind other buildings, they might point at the direction of the parking lot and say, "somewhere over there."

How communication richness affects pointing

There was a clear relationship in our study between the richness of the communication channel and the type and complexity of pointing gestures. When communication channels were more restricted (written notes or no verbal communication), gestures were more detailed. For example, when only gestures were allowed, participants would form shapes with their hands or draw the outline of the targets more often. When communicating with gestures and written notes, participants needed to constantly switch between pointing and writing, which interrupted the flow of communication. Not surprisingly, participants preferred the condition where they used both gestures and speech.

Other observations

Observer attention to pointing gestures. Observers did not necessarily look directly at the actors and the pointing gestures. When an actor performed a simple pointing gesture, (e.g., raised arm and extended index finger as shown in Figure 2), the observers often appeared to only see the gesture in their peripheral field of view. In these cases, the observers focused on the targets rather than the simple gestures, although they were always able to determine the general direction of the pointing gesture.

Pointing as an initiator of communication. Pointing gestures were not only used by actors to indicate targets, but also served as a signal to initiate communication. Even though observers did not look at all pointing gestures, they had immediate responses to all of them. They looked at the indicated directions or the gestures as soon as the actors started pointing, and even before the associated verbal communication. These results echo earlier work showing that gesture can aid in the management of turn-taking and the coordination of talk [18].

Overall, these results show that people are clearly able to use 'held' pointing gestures to identify referents in the environment, and that the degree of precision needed for the gesture depends on how difficult it is to describe the target using the available verbal channel. This suggests that some kinds of pointing do not need the high degree of accuracy that we know from prior work will be difficult to achieve. However, there is no information available about whether even moderate pointing accuracy is possible in a CVE. This information is critical to understanding how the types of pointing seen in Study 1 can be supported, and therefore we carried out a second study that looked at people's accuracy in determining the direction of pointing, and how accurately people could produce those gestures themselves.

STUDY 2: INTERPRETING POINTING DIRECTION

The second experiment (a controlled study) assessed people's ability to interpret the direction of a pointing gesture. This study looked at how accurately people could determine what others were pointing at (using 'held' gestures alone), how accurately people could point at objects themselves, and how interpretation of pointing direction differed between the real world (RW) and a CVE. Given the difficulties that have been reported for CVE-based pointing, our goal was to determine how well people could carry out a fundamental part of pointing – interpreting the direction of a gesture.

As discussed earlier, we focus on one stage of pointing, but the stage that is most critical for the overall communication. Without the ability to communicate direction in the holding

stage, the other stages of pointing (mutual reference, staging, and production) cannot be successful.

We designed the study to investigate two main questions:
Q1. How well can people determine the direction of a pointing gesture, both in the RW and in a CVE?
Q2. How accurately can people point at things themselves?

In addition, we were interested in the effects of three additional factors:
Q3. Does distance to the target affect accuracy (either for the producer of the gesture or the observer)?
Q4. Does the observer's location affect interpretation?
Q5. Does field-of-view width affect pointing?

Experimental Design

The experiment had five factors: two *task types* (production of the pointing gesture or interpretation of the gesture), two *environments* (RW and CVE), two *distances* (600cm or 300cm to the targets), two *field-of-view widths* (85° or 120°, only used in the CVE), and two *observing locations* (behind or beside, only used for interpretation tasks). These conditions and the specific values were chosen based on informal tests in our environment, and a pilot study.

Figure 3 shows the differences between 'behind' and 'beside' in the real world and the CVE. The two locations were chosen to ensure that both the targets and the actor's arm could be seen by the observers in a single view – that is, we provided situations where mutual orientation [16] was already established. There were 15 trials per condition, with the first five trials marked as training and not analysed. Table 1 summarizes the conditions used in the study. The experiment used a within-participants design; condition order was counterbalanced using a Latin square design.

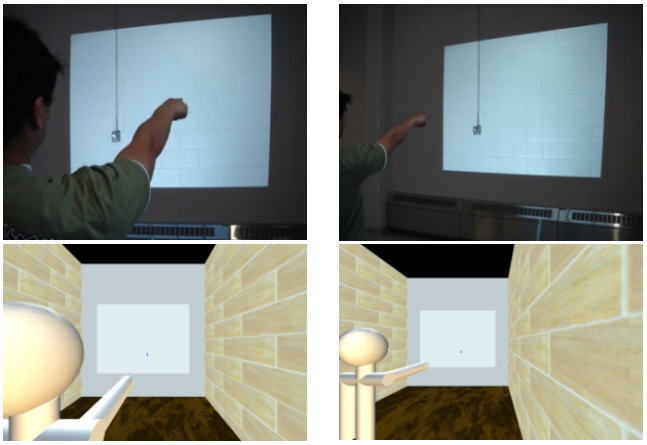

Figure 3. Observer's views: behind (left) and beside (right).

		CVE		RW	
		Far	Near	Far	Near
Gesture Production	Small FoV	10	10	10	10
	Large FoV	10	10		
Gesture Interpretation	Behind	10	10	10	10
	Beside	10	10	10	10

Table 1. Number of test trials in each experimental condition.

We asked participants to focus on accurately determining the target of the pointing gesture, rather than speed. We collected accuracy data from all tasks, as described below. In addition, a post-study questionnaire was used to collect subjective data at the end of the experiment.

Settings

The experiment was conducted in two environments. In the RW setting (a 750cm x 400cm room), a 1024 x 768 projector displayed targets on a 400cm-width wall. The projected area was 300cm x 225cm. The image was horizontally centred on the wall and 100cm above the ground. The locations where the participants and experimenter stood are shown in Figure 4. For production tasks, participants stood at location A (300cm from the wall) and B (600cm). For pointing interpretation tasks, the experimenter stood at A and B, while participants stood at C, D, E, and F. Although the study room was not as large as the setting in Study 1, people used the same 'distant pointing' gestures in Study 2 as they did in Study 1.

In the CVE, we created a room that replicated the real world setting: the virtual room was the same size as the real room, and participants placed their avatars in the same locations as were used in the real room. For gesture production tasks, participants used a mouse to control the avatar's arm movement. The avatar was at location A and B. For the pointing interpretation task, the participant used the mouse to control the camera (the observer's view) at locations C, D, E, and F. The avatar in the role of the actor was located at A and B.

Participants and Apparatus

Ten university students (6 male and 4 female) participated in the study. The mean age of the participants was 24, and all were regular computer users. The CVE used in the study was custom software built using C# and the XNA framework. The system ran on a standard Windows PC, and used a 22-inch LCD monitor.

Tasks

The study used two tasks: gesture production, in which participants were asked to point as accurately as possible at a given target; and gesture interpretation, in which they were asked to determine the direction of another person's held pointing gesture.

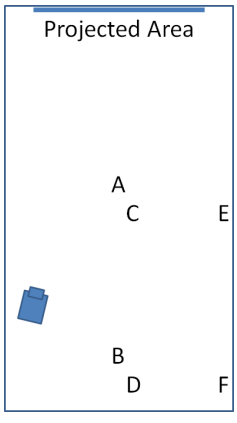

Figure 4. Top view of experimental setting.

Task 1: production of gestures

Participants were asked to point at the centre of targets that appeared on the wall in front of them. Targets appeared at random locations, one at a time. In RW, participants pointed with a laser pointer; in the CVE, the participants controlled their avatar's arm with the mouse (Fig. 5 and 6).

In RW, participants were first given a laser pointer and were asked to practice with it until they could point consistently. They then stood at the required position in the room, and pointed at targets with their arm held straight. Participants were told not to turn the laser pointer on until they were confident that it was aiming at the target. When the laser was switched on, the experimenter recorded the location of the laser dot.

Figure 5. First-person view in the real world and the CVE.

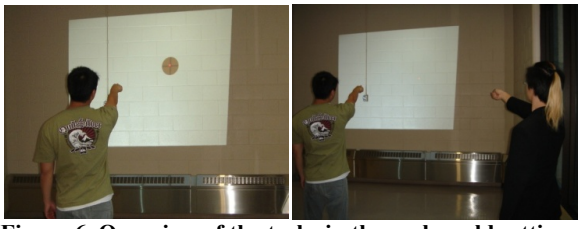

Figure 6. Overview of the tasks in the real world setting: production at left, interpretation at right.

In the CVE, a similar procedure was followed, except that participants used the mouse to control the avatar's arm direction, and clicked the mouse button to complete each trial. On each mouse click, a red dot appeared on the virtual wall to provide the same feedback about where the user had pointed as was given in the real world.

Task 2: interpretation of the direction of pointing gestures

For this task, participants were asked to observe an actor who pointed at locations on the front wall, and then determine what location was being pointed at (see Figure 6). With the participant turned away, the actor produced and held a straight-arm pointing gesture using a visible target; when the pointing gesture was ready, the target was hidden and the observer turned around. The experimenter was the actor in both settings.

In the real-world version of the task, participants used a laser pointer (on at all times) to indicate where on the wall they thought the actor was pointing; this location was recorded by the experimenter. The CVE version of the task was equivalent, but adapted to the desktop setting similar to the description of the first task.

Results

We recorded all target locations and the locations where the participants pointed. Using these data, we calculated the angular error of each task (i.e., the difference in angle between imaginary lines drawn from the actor's shoulder to the target, and from the shoulder to the participants' recorded location). Angular error must be used as the measure of performance, rather than absolute error, because it is not affected by distance from the targets, and thus results can be comparable across different distances. Results are organized below based on the five research questions specified earlier.

Q1: Can people determine where others are pointing?

Using all conditions in the gesture interpretation task, the mean angular error was 5.5°. Analysis of variance (ANOVA) showed a significant main effect of environment ($F_{1,9}$=7.04, $p<0.05$), with errors in RW (mean 5.1°) less than in the CVE (5.9°).

There was also a significant interaction between environment and distance ($F_{1,9}$=7.38, $p<0.05$); as shown in Figure 7, the difference between the CVE and RW was much more pronounced at 300cm than at 600cm. There was no interaction between environment and observation location ($F_{1,9}$=3.04, $p=0.12$).

Q2: Can people produce accurate pointing gestures?

Using all data in the gesture production task, the mean angular error was 3.1°. ANOVA showed a main effect of environment ($F_{1,9}$=31.56, $p<0.001$), with RW pointing (mean 1.8°) substantially more accurate than pointing in the CVE (mean 4.4°) (see Figure 7). In addition, production of pointing gestures overall was more accurate than interpretation; ANOVA showed a main effect of task ($F_{1,9}$=169.47, $p<0.001$).

Q3: Does distance to the target affect accuracy?

An ANOVA using data from both tasks showed a main effect of distance ($F_{1,9}$=12.31, $p<0.01$). However, the ordering of the two conditions was surprising: when standing 600cm from the target, error was always less than when standing at 300cm (see Figure 7). As reported above, there was also a significant interaction between distance and environment (with distance having more of an impact on interpreting pointing in the CVE).

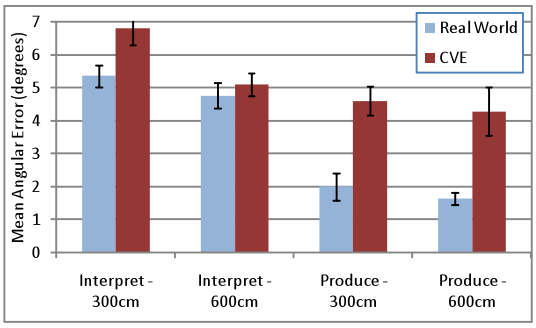

Figure 7. Mean error by environment, task, and distance.

Q4: Does the observer's location affect interpretation?

An ANOVA showed a significant main effect of location on interpretation accuracy ($F_{1,9}$=14.32, $p<0.01$). When observers stood behind the actor, error was less (4.9°) than when standing beside (6.1°). No interaction was found between location and environment ($F_{1,9}$=3.04, $p=0.12$).

Q5: Does field-of-view width affect pointing?

ANOVA on the gesture-production task did not show a significant main effect for field-of-view width ($F_{1,9}$=1.53,

p=0.25), and no interaction between FoV and distance ($F_{1,9}$=1.84, p=0.21). The mean error of the 85° view was 4.18°, and of the 120° view was 4.71°.

Questionnaire responses

All participants reported having more confidence in doing both tasks in RW as compared to the CVE; participants were also unanimous in stating that the tasks were more difficult in the CVE. Most participants (7 of 10) reported having more confidence when observing from behind the actor as opposed to beside.

DISCUSSION

The main findings from the study are:

- Participants could produce and interpret pointing gestures more accurately in the real world than in the CVE, with a larger difference for producing gestures;
- The difference between the environments for interpreting pointing direction was much smaller than expected – only 1.4° at 300cm, and only 0.33° at 600cm.
- Errors were larger (by approximately one degree) when people were nearer to the target.
- Observers were more accurate when interpreting a pointing gesture from behind than from beside (a difference of 1.13°).
- The different fields of view available in the CVE made little difference in producing pointing gestures.

Although the differences between the real world and the CVE were significant (with people performing better in the real world) the most striking feature of the study results is that the actual differences between the two environments are relatively small. We expected the real world to be dramatically better for pointing than a CVE, but the differences were less than we expected. To put the differences into real-world terms, Figure 8 compares the error from the two environments, for the interpretation task. At 300cm from the target, people would be able to identify referent objects that are 50cm apart in the real world, but in a CVE, referent objects would have to be 53cm apart.

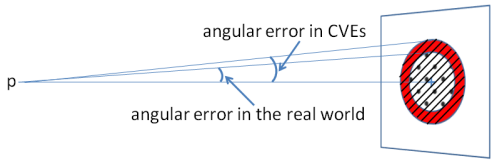

Figure 8. Comparison of error zones in RW and CVE.

Explanations for the results

What caused the differences between real world and CVE?

First, and not surprisingly, people found it much easier to point in the real world than to control the avatar's arm in the CVE. One participant commented "[in the] real world, I just found I have more control over what I was doing and felt more confident in doing it." Although participants did not have any major difficulties with the input techniques, controlling the avatar in the CVE was definitely more difficult than moving one's own arm.

Second, there are perceptual limitations of the CVE that may have reduced performance. As has been discussed in previous work, a CVE displayed on a desktop monitor presents a much poorer visual environment than that of the real world in terms of depth cues, stereo vision, realism, resolution, field of view, and proprioception. Participants commented that it was more difficult to assess angles and depth in the CVE: for example, one person stated "the distances are not as real compared to the real world; in the virtual world, I couldn't feel the distance difference."

Why was pointing more accurate from far away?

We were surprised that people were more accurate when the target was further away. There are two possible explanations for this result. First, we asked participants to aim at the centre point of each target, and distance does not affect the size of a point. Therefore, nearer targets do not have the advantage of appearing bigger. Second, parallax and the distance between the actor's eye and their shoulder may have an effect on accuracy. Near objects have larger parallax than faraway objects for both pointing and observing. The difference in angle between the eye-to-target line and the shoulder-to-target line becomes smaller as the actor moves further from the target.

Parallax can also serve to explain the fact that people were more accurate in producing gestures than in interpreting them, and the finding that people were more accurate when observing from behind than from the beside. The observer has a greater separation from the origin of the pointing gesture than does the actor, and this separation is greatest when the observer is beside the actor.

Design Implications

Here we list lessons and implications that are raised by our two studies; some of these are novel, and others reinforce conclusions that have been reported by other studies.

1. Deictic pointing has varying accuracy requirements (S1). The precision required for different pointing gestures varied with the difficulty of the referential task. This means that designers can support different kinds of pointing with different mechanisms; for example, augmented pointing techniques (e.g., laser beams) may not be required for several types of deictic pointing.

2. CVEs should support multiple types of pointing (S1). Our results agree with previous work, e.g., Goodwin's observations [13], that people created a wide variety of different pointing gestures that depended on the task difficulty and the features of the environment. To allow this richness in CVEs, designers should provide much more expressivity than what is currently available.

3. Reduced communication richness in CVEs may increase requirements for accurate pointing (S1). The relationship between pointing accuracy and communication richness in Study 1 suggests that in lower-richness CVEs (e.g., chat-based communication in environments like Second Life),

the difficulty of constructing referential statements puts the onus on pointing gestures to carry the reference.

4. The importance of peripheral vision (S1). Several situations in our first study involved people looking at the target instead of the pointing gestures, but clearly maintaining an awareness of the gesture in peripheral vision. As reported before, narrow fields of view in current CVEs are a major problem when people need to manage their orientation to different parts of the environment [12].

5. Natural pointing in CVEs can be successful (S2). Our results show that people can interpret others' pointing gestures in a CVE almost as well as they can in the real world. Given that many types of pointing gesture do not require high precision, our results strongly suggest that naturalistic deictic reference can be used to a much greater degree than has been seen in current CVEs. In particular, both general directional pointing, and more specific pointing where the target is relatively easy to disambiguate through speech, should be possible in CVEs using natural pointing. However, these results are limited to situations where mutual orientation has already been established.

6. Pointing in CVEs is still less accurate than RW (S2). Although people were more accurate overall for production than for interpretation, producing gestures in the CVE was considerably less accurate than in the real world. There are several ways in which people could be assisted in this task; one that we will study in future work is the importance of other objects in the scene to help people estimate distance when producing a pointing gesture.

7. Compressed field-of-view does not aid accuracy (S2). The process of mutual orientation, the ability to see both gesture and target, peripheral awareness of gestures, and the visibility of pointing actions are all made difficult or impossible by restricted fields of view. Our results underscore previous work suggesting that CVEs should be equipped with very wide fields of view, but show that accuracy cannot be supported simply by compressing the view [12]. Multi-monitor displays are a more likely solution to this problem, as they provide a true FoV increase.

Supporting Other Stages of Deictic Pointing

Our results provide a clear indication that the most fundamental stage of deixis – directional interpretation of held gestures – can work in CVEs. This means that we must now consider how to support the rest of the pointing process. Our experiences in the two studies suggest that four issues will be critical to comprehensive support for these other stages in deictic communication:

Gradual and fine-grained production of gestures. Prior work suggests that much of the value of gestures comes through the accountable and projectable process by which they are produced. In order to support these qualities, people must be able to produce rich and detailed gestural movements in CVEs. This means that designers must consider ways to improve expressivity far beyond the command-based pointing that is common in current virtual worlds. One direction we are currently exploring is tracking real-world arm motions as input for CVE gestures.

Input mappings for natural pointing. While object-based pointing can be accomplished by clicking the targets (e.g., in Second Life), natural pointing requires additional input mapping for the arm. In our studies, we allowed people to change pointing direction with the mouse; however, the mouse is often used for other purposes in a CVE (e.g., view control), and so further work will be needed to find ways of adding degrees of freedom to the interface without making CVEs overly complex or removing basic functions like view control. We are currently exploring different input mappings with various input devices such as joysticks, trackballs, and gamepads.

Improving perceptual richness in CVEs. More work is needed to determine the effects of perceptual factors on pointing and other gestures in CVEs, and to find ways of overcoming the limitations of virtual environments. Several strategies show promise: we are currently working on improving depth cues (e.g., adding objects of known size to the environment), on improving subtlety and detail in the rendering of avatar arm movement (to better show the expressive movements that can be produced with the control methods described above), and on the effects of different camera and view angles on the problem of head-to-shoulder parallax.

Integration of augmented and natural pointing. In situations where natural pointing remains difficult in CVEs, augmented pointing techniques will continue to be valuable. One of our goals for future work is to find ways of integrating natural and augmented pointing, allowing people to carry out both kinds of pointing without losing the subtlety and richness evident in natural pointing.

CONCLUSION

Pointing is a natural and expressive part of human gestural communication, but despite the wealth of research into pointing and deixis, current CVEs do not yet provide good support for pointing. One of the requirements that must be addressed, before designers can think about the rich and subtle process of deictic pointing, is whether people will be able to interpret the direction of a pointing gesture in a CVE. To investigate this fundamental aspect of pointing in collaboration, we carried out two studies. The first study identified different accuracy requirements for different types of deictic pointing, and showed that people produce a wide variety of gestures for different tasks. The second study showed that interpretation of gestures – unexpectedly – is almost as accurate in CVEs as it is in the real world. These results suggest that deictic pointing can work well in CVEs, and that there is great potential for supporting the full process of gesturing. Our findings help designers determine whether and when they need to provide enhanced pointing techniques, and when natural pointing will be sufficient.

In future, we will study richer control over expressive pointing, add depth cues for distance interpretation, and carry out further work on the issue of parallax. In addition, we will observe more realistic dyadic collaboration in CVEs to confirm and extend the results discussed here.

REFERENCES

1. Bangerter, A. 2004. Using Pointing and Describing to Achieve Joint Focus of Attention in Dialogue, *Psychological Science*, 15, 6, 415-418.
2. Bekker, M., Olson, J.S., and Olson, G.M. 1995. Analysis of gestures face-to-face design teams provides guidance for how to use groupware design. *Proc. DIS'95,* 157-166.
3. Benford, S., Bowers, J., Lennart, E.F., Greenhalgh, C., and Snowdon, D. 1995. User embodiment collaborative virtual environments. *Proc. CHI'95*, 242-249.
4. Bowers, J., Pycock, J., and O'Brien, J. 1996. Talk and embodiment collaborative virtual environments. *Proc. CHI '96*, 58-65.
5. Bowers, J., O'Brien, J., and Pycock, J. 1996. Practically accomplishing immersion: Cooperation and for virtual environments. *Proc. CSCW '96*, 380-389.
6. Clark, H.H. 2003. Pointing and Placing. S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet*. Hillsdale NJ: Erlbaum. 243-268.
7. Clark, H.H., and Brennan, S.E. Grounding Communication. 1991. L.B. Resnick, R.M. Levine, & S.D. Teasley (Eds.). *Perspectives on socially shared cognition*. Washington, DC: APA. 127-149.
8. Clark, H., and Wilkes-Gibbes, D., Referring as a Collaborative Process, *Cognition*, 22, 1986, 1-39.
9. Duchowski, A., Cournia, N., Cumming, B., McCallum, D., Gramopadhye, A., Greenstein, J., Sadasivan, S., and Tyrrell, R. 2004. Visual deictic reference a collaborative virtual environment. *Proc. ETRA*, 35-40.
10. Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E.M., and Kramer, A. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction,* 19, 273-309.
11. Fraser, M., and Benford, S. 2002. Interaction effects of virtual structures. *Proc. CVE 2002,* 128-134.
12. Fraser, M., Benford, S., Hindmarsh, J., and Heath, C. 1999. Supporting awareness and interaction through collaborative virtual interfaces. *Proc. UIST'99,* 27-36.
13. Goodwin, C. 2003. Pointing as Situated Practice. S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet*. Hillsdale NJ: Erlbaum. 217-241.
14. Greenberg, S., Gutwin, C., and Roseman, M., 1996. Semantic Telepointers for Groupware. *Proc. OzCHI '96*, 54-61.
15. Gutwin, C., and Penner, R. 2002. Improving Interpretation of Remote Gestures with Telepointer Traces, *Proc. CSCW'02,* 49-57.
16. Hindmarsh, J., Fraser, M., Heath, C., Benford, S., and Greenhalgh, C. 1998. Fragmented Interaction: establishing mutual orientation virtual environments. *Proc. CSCW'98*, 217-226.
17. Hindmarsh, J., Fraser, M., Heath, C., Benford, S., and Greenhalgh, C. 2000. Object-Focused Interaction in Collaborative Virtual Environments, *ToCHI,* 7, 4, 477-509.
18. Hindmarsh, J., and Heath, C. 2000. Embodied reference: A study of deixis workplace interaction. *J. Prag.* 32, 12, 1855–1878.
19. Kelly, J., Bell, A., and Loomis, J. 2004. Perception of Shared Visual Space: Establishing Common Ground in Real and Virtual Environments, *Presence*, 13, 4, 442-450.
20. Kirk, D., Rodden, T., and Fraser, D.S. 2007. Turn it this way: grounding collaborative action with remote gestures. *Proc CHI'07*, 1039-1048.
21. Kita, S. (Ed.). 2003. *Pointing: Where language, culture and cognition meet*. Mahwah, NJ: Erlbaum.
22. Krauss, R.M., and Fussell, S.R. 1990. Mutual knowledge and communicative effectiveness. In, J. Galegher, R.E. Kraut & C. Egido (Eds.), *Intellectual Teamwork: Social and Technical Bases of Collaborative Work*. Hillsdale, NJ: Erlbaum. 111-145.
23. McNeill, D., Cassell, J., and Levy, E.T. 1993. Abstract deixis. *Semiotica,* 95(1/2), 5-19.
24. Moore, R., Ducheneaut, N., and Nickell, E. 2007. Doing Virtually Nothing: Awareness and Accountability in Massively Multiplayer Online Worlds, *JCSCW*, 16, 3, 265-305.
25. Pechmann, T., and Deutsch W. 1982. The Development of Verbal and Nonverbal Devices for Reference. *Journal of Experimental Child Psychology* 34, 330-341.
26. Tang, A., Boyle, M., and Greenberg, S. 2005. Display and Presence Disparity Mixed Presence Groupware. *JRPIT*, 37, 2, 71-88.
27. Tang, J. 1991. Findings from Observational Studies of Collaborative Work, *IJMMS*, 34, 2, 143-160.
28. Tang, J.C., and Minneman, S.L. 1990. VideoDraw: a video interface for collaborative drawing, *Proc. CHI'90*, 313-320.
29. Tang, J., and Minneman, S. 1991. VideoWhiteboard: video shadows to support remote collaboration, *Proc. CHI'91,* 315-322.