Doug A. Bowman
bowman@vt.edu
Department of Computer Science
Virginia Polytechnic Institute & State
University

Donald B. Johnson
Larry F. Hodges
{donny, hodges}@cc.gatech.edu
Graphics, Visualization, and Usability
Center
Georgia Institute of Technology

Testbed Evaluation of Virtual Environment Interaction Techniques

Abstract

As immersive virtual environment (VE) applications become more complex, it is clear that we need a firm understanding of the principles of VE interaction. In particular, designers need guidance in choosing three-dimensional interaction techniques. In this paper, we present a systematic approach, testbed evaluation, for the assessment of interaction techniques for VEs. Testbed evaluation uses formal frameworks and formal experiments with multiple independent and dependent variables to obtain a wide range of performance data for VE interaction techniques. We present two testbed experiments, covering techniques for the common VE tasks of travel and object selection/manipulation. The results of these experiments allow us to form general guidelines for VE interaction and to provide an empirical basis for choosing interaction techniques in VE applications. Evaluation of a real-world VE system based on the testbed results indicates that this approach can produce substantial improvements in usability.

1   Introduction

Applications of immersive virtual environments (VEs) are becoming both more diverse and more complex. This complexity is not only evident in the number of polygons being rendered in real time, the resolution of texture maps, or the number of users immersed in the same virtual world, but also in the interaction between the user(s) and the environment. Users need to navigate freely through a three-dimensional space, manipulate virtual objects with six degrees of freedom, or control attributes of a simulation, among many other things.

However, interaction in three dimensions is not well understood (Herndon, van Dam, & Gleicher, 1994). Users have difficulty controlling multiple degrees of freedom simultaneously, interacting in a volume rather than on a surface, and understanding 3-D spatial relationships. These problems are magnified in an immersive VE, because standard input devices such as mice and keyboards may not be usable (if the user is standing, for example), the display resolution is often low (limiting the ability to display text, for example), and 3-D depth cues may be in conflict with one another (accommodation and convergence, for example).

Therefore, the design of interaction techniques (ITs) and user interfaces for VEs must be done with extreme care to produce useful and usable systems. Because there is a lack of empirical data regarding VE interaction techniques,

we emphasize the need for formal evaluation of ITs, leading to easily applied guidelines and principles.

In particular, we have found testbed evaluation to be a powerful and useful tool to assess VE interaction. Testbeds are representative sets of tasks and environments, and the performance of ITs can be quantified by running them through the various parts of a testbed. Testbed evaluations are distinguished from other types of formal experiments because they combine multiple tasks, multiple independent variables, and multiple response measures to obtain a more complete picture of the performance characteristics of an IT, and because they produce application-independent results.

In this paper, we present our experience with this type of evaluation. We will begin by discussing related work and the design and evaluation methodology of which testbed evaluation is a part. Two testbed experiments are presented, evaluating interaction techniques for the tasks of travel and selection/manipulation of virtual objects. The results of these experiments were applied to the design of a complex VE application. We conclude with a discussion of the merits of this type of evaluation.

2   Related Work

Most ITs for immersive VEs have been developed in an ad hoc fashion or to meet the requirements of a particular application. Such techniques may be very useful, but they need to be evaluated formally. Work has focused on a small number of "universal" VE tasks, such as travel (Koller, Mine, & Hudson, 1996; Ware & Osborne, 1990), and object selection and manipulation (Pierce et al., 1997; Poupyrev, Billinghurst, Weghorst, & Ichikawa, 1996).

Evaluation of VE interaction has for the most part been limited to usability studies (for example, Bowman, Hodges, & Bolter, 1998). Such evaluations test complete applications with a series of predefined user tasks. Usability studies can be a useful tool for the iterative design of applications, but we feel that lower-level assessments are necessary due to the newness of this research area.

Another methodology that has been applied to VE interaction is usability engineering (Hix et al., 1999). This technique uses expert evaluation, guidelines, and multiple design iterations to achieve a usable interface. Again, it is focused on a particular application and not ITs in general.

A number of guidelines for 3D/VE interaction have been published (such as Kaur (1998)). Guidelines can be very useful to the application developer as an easy way to check for potential problems. Unfortunately, most current guidelines for VEs are either too general and therefore difficult to apply, or taken only from experience and intuition and not from empirical results.

Testbeds for virtual environments are not new. The VEPAB project (Lampton et al., 1994) produced a battery of tests to evaluate performance in VEs, including tests of user navigation. Unlike our work, however, the tasks involved were not based on a formal framework of technique components and other factors affecting performance. The most closely related work to the current research is the manipulation assessment testbed (VRMAT) developed by Poupyrev, Weghorst, Billinghurst, and Ichikawa (1997).

3   Methodology

How does one design and validate testbeds for VE interaction? It is important that these testbeds represent generalized tasks and environments that can be found in real VE applications. Also, we need to understand ITs at a low level and standardize the measurement of performance. For these reasons, we base our testbeds on a systematic, formal framework for VE interaction techniques (Bowman & Hodges, 1999). In this section, we will briefly discuss pieces of this methodology that are relevant to the current work.

3.1 Taxonomies

The first step in creating a formal framework for design and evaluation is to establish a taxonomy of interaction techniques for each of the universal interaction
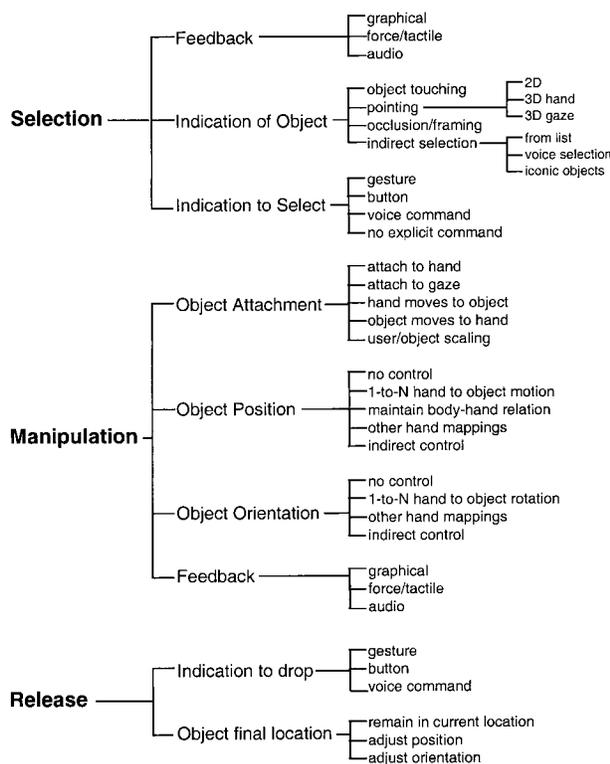
Figure 1. *Taxonomy of selection/manipulation techniques.*

tasks. (Note the word *taxonomy* because we will employ both of its accepted meanings: "the science of classification," and "a specific classification.") Taxonomies partition the tasks into separable subtasks, each of which represents a decision that must be made by the designer of a technique. In this sense, a taxonomy is the product of a careful task analysis. For each of the lowest-level subtasks, technique components (parts of an interaction technique that complete that subtask) may be listed. Figure 1 presents a taxonomy for the tasks of selection and manipulation, including two levels of subtasks, and multiple technique components for each of the lowest-level subtasks. We have also created two taxonomies for the task of travel.

The taxonomies must come from a deep and through understanding of the interaction task and the techniques that have been proposed for it. Therefore, some initial qualitative evaluation of techniques and/or design of new techniques for the task is almost always required before a useful taxonomy can be constructed.

Let us consider a simple example. Suppose the interaction task is to change the color of a virtual object. (Of course, this task could also be considered a combination of other interaction tasks: select an object, select a color, and give the "change color" command.) A taxonomy for this task would include several subtasks. Selecting an object whose color is to change, choosing the color, and applying the color are subtasks that are directly task-related. On the other hand, we might also include aspects such as the color model used or the feedback given to the user, which would not be applicable for this task in the physical world, but which are important considerations for an IT.

We do not claim that any given taxonomy represents the "correct" partitioning of the task. Different users have different conceptions of the subtasks that are carried out to complete a task. Rather, we see our taxonomies as practical tools that we use as a framework for design and evaluation. Therefore, we are concerned only with the utility of a taxonomy for these tasks, and not its "correctness." In fact, we have developed two possible taxonomies for the task of travel, both of which have been useful in determining different aspects of performance. Rules and guidelines have been set forth for creating proper taxonomies (Fleishman & Quaintance, 1984), but we felt that the categorical structure of these taxonomies did not lend itself as well to design and evaluation as the simple task analysis, because they do not allow guided design or evaluation at the subtask level.

Taxonomies have many desirable properties. First, they can be verified by fitting known techniques into them in the process of categorization. Second, they can be used to design new techniques quickly, by combining one component for each of the lowest-level subtasks. More relevant to testbed evaluation, they provide a framework for assessing techniques at a more fine-grained level. Rather than evaluating two techniques for the object-coloring task, then, we can evaluate six components. This may lead to models of performance that allow us to predict that a new combination of these

components would perform better than either of the techniques that were tested.

3.2 Performance Metrics

Quantifying the performance of VE interaction techniques is a difficult task, because performance is not well defined. It is relatively simple to measure and quantify time for task completion and accuracy, but these are not the only requirements of real VE applications.

VE developers are also concerned with notions such as the naturalism of the interaction (how closely it mimics the real world) and the degree of presence the user feels. Usability-related issues such as ease of use, ease of learning, and user comfort will also be important to an interface's success. Finally, task-related performance, such as spatial orientation during navigation or expressiveness of manipulation, is often required.

We should remember that the reason we wish to find good ITs is so that our applications will be more usable, and that VE applications have many different requirements. In many applications, speed and accuracy are not the main concerns, and therefore these should not always be the only response variables in our evaluations.

Also, more than any other computing paradigm, virtual environments involve the user—his or her senses and body—in the task. Thus, it is essential that we focus on user-centric performance measures. If an IT does not make good use of the skills of the human being, or if it causes fatigue or discomfort, it will not provide overall usability despite its performance in other areas. In this work, then, we will base our evaluations on multiple performance measures that cover a wide range of application and user requirements.

Therefore, in our work, we have a broad definition of performance, and we will attempt to measure multiple performance variables during testbed evaluation. For those factors that are not objectively measurable, standard questionnaires (for example, Kennedy, Lane, Berbaum, and Lilienthal (1993) for simulator sickness, and Witmer and Singer (1998) for presence) or subject self-reports may need to be used.

3.3 Outside Factors Influencing Performance

The interaction technique is not the sole determinant of performance in a VE application. Rather, there are multiple interacting factors. In particular, we have identified four categories of outside factors that may influence performance: characteristics of the task (such as the required accuracy), environment (such as the number of objects), user (such as spatial ability), and system (such as stereo versus biocular viewing).

In our testbed experiments, we consider these factors explicitly, varying those we feel to be most important and holding the others constant. This leads to a much richer understanding of performance. Often there are too many possible outside factors to evaluate in a single experiment. In this case, pilot studies can help to eliminate some factors.

3.4 Testbed Evaluation

Our experimental evaluations of VE interaction techniques have taken many forms, from simple observational user studies (Bowman & Hodges, 1997), to usability evaluation (Bowman, Hodges et al., 1998), to formal experiments (Bowman, Koller, & Hodges, 1997). However, none of these methods is able to examine the wide range of task conditions as well as produce quantitative, general results. Therefore, we propose the use of testbed evaluation as the final stage in the analysis of interaction techniques for universal VE interaction tasks. This method addresses the issues discussed above through the creation of testbeds—environments and tasks that involve all of the important aspects of a task, that test each component of a technique, that consider outside influences (factors other than the interaction technique) on performance, and that have multiple performance measures.

As an example, consider a proving ground for automobiles. In this special environment, cars are tested in cornering, braking, acceleration, and other tasks, over multiple types of terrain, and in various weather conditions. Task completion time is not the only performance

variable considered. Rather, many quantitative and qualitative results are collected, such as accuracy, distance, passenger comfort, and the user's perception of the "feel" of the steering.

3.5 Application of Testbed Results

Testbed evaluation produces a set of results that characterize the performance of an interaction technique for a specified task. Performance is given in terms of multiple performance metrics, with respect to various levels of outside factors. These results become part of a performance database for the interaction task, with more information being added to the database each time a new technique is run through the testbed.

Testbed evaluation is not an end unto itself. Rather, it has the goal of producing applications with high levels of performance. Thus, the last step in our methodology is to apply the performance results to VE applications, with the goal of making them more useful and usable. To choose interaction techniques for applications appropriately, we must understand the interaction requirements of the application. We cannot simply declare one best technique, because the technique that is best for one application will not be optimal for another application with different requirements. For example, a VE training system will require a travel technique that maximizes the user's spatial awareness, but will not require a travel technique that maximizes point-to-point speed. On the other hand, in a battle-planning system, speed of travel may be the most important requirement.

Therefore, applications need to specify their interaction requirements before the correct ITs can be chosen. This specification will be done in terms of the performance metrics that we have already defined as part of our formal framework. Once the requirements are in place, we can use the performance results from testbed evaluation to recommend ITs that meet those requirements. These ITs, having been formally verified, should increase the user's performance levels and the application's usability.

4 Experiments

We present two experiments that bring together the components of the formal methodology. The first testbed is designed to evaluate selection and manipulation techniques, and the second is for travel techniques. Each testbed is a set of tasks and environments that measure the performance of various combinations of technique components and outside factors for multiple performance metrics.

Both testbeds were designed to test any technique that could be created from its respective taxonomy. However, exhaustive testbeds would be too immense to carry out. Therefore, our testbeds have been simplified to assess conditions based on a target application (see section 5). Nevertheless, the tasks and environments are not biased towards any particular set of techniques, and others can be tested at any time with no loss of generality. For both testbeds, the tasks used are simple and general.

4.1 Selection and Manipulation Testbed

We designed and implemented a limited testbed that can evaluate selection and manipulation techniques in a number of what we consider to be the most important conditions. The analysis of importance is based on our experiences with real applications, our more informal study of selection and manipulation, and the requirements of our target application.

The testbed was designed to support the testing of any technique that can be created from the taxonomy. The tasks and environments are not biased towards any particular set of techniques. We have evaluated nine techniques, but others can be tested at any time with no loss of generality.

In the selection phase, the user selects the correct object from a group of objects. In the manipulation phase, the user places the selected object within a target at a given position and orientation. Figure 2 shows an example trial. The user is to select the darker box in the center of the 3×3 array of boxes, and then place it between the two wooden targets in the manipulation
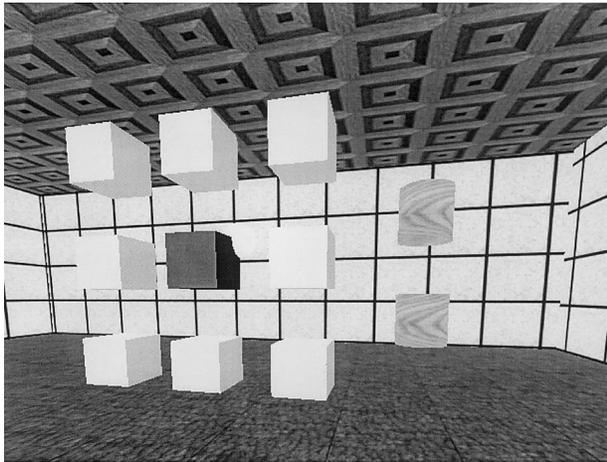
Figure 2. *Trial setup in the selection/manipulation testbed.*

phase. In certain trials, yellow spheres on both the selected object and the target determine the required orientation of the object.

4.1.1 Method. Three within-subjects variables were used for the selection tasks. We varied the distance from the user to the object to be selected (three levels: 3, 6, or 10 m from the user), the size of the object to be selected (two levels: 0.4 or 1.0 m cubes), and the density of objects surrounding the object to be selected (two levels: 0.4 or 1.0 m between cubes). Based on pilot studies, we believed these to be some of the most important factors in determining speed, accuracy, ease of use, and comfort for selection techniques.

The manipulation phase of the task also involved three within-subjects variables. First, we varied the ratio of the object size to the size of the target (two levels: the target was either 1.5 or 3.75 times the size of the object), which corresponds to the accuracy required for placement. Second, the number of required degrees of freedom varied (two levels: two or six DOFs), so that we could test the expressiveness of the techniques. The two-DOF task only required users to position the objects in the horizontal plane, whereas the six-DOF task required complete object positioning and orientation. Finally, we varied the distance from the user at which the object must be placed (three levels: 3, 6, or 10 m).

Other outside factors, such as stereo versus mono viewing, or the use of interactive shadows, could have been included, but were not in order to maintain a manageable experiment size. It is possible that these or other outside factors could interact with some of our experimental variables, but this experiment does test the most-accessible task parameters using an average VE system.

We required users to place the selected objects completely within the targets and within five degrees of the correct orientation on the six-DOF trials. Graphical feedback told the user when the object was in the correct location.

Response variables were the speed of selection, the number of errors made in selection, the speed of placement, and subjective data related to user comfort. Comfort was measured in the areas of arm strain, hand strain, dizziness, and nausea. After a practice session and each block of trials, the subjects gave a rating for each of these factors on a ten-point scale. Each subject also took a standardized test of spatial ability. Finally, we gathered demographic information about our subjects, including age, gender, handedness, technical ability, and VE experience via a questionnaire.

Forty-eight subjects (31 men, 17 women) participated in the study. Each subject completed 48 trials, except for three subjects who did not complete the experiment due to dizziness or sickness. Subjects were allowed to practice the technique for up to five minutes in a room filled with furniture objects before the experimental trials began. Subjects completed four blocks of twelve trials each, alternating between trials testing selection and manipulation.

Nine different selection/manipulation techniques, taken from our taxonomy (figure 1), were compared in a between-subjects fashion. Thus, there were five subjects per technique. One technique was the Go-Go technique (Poupyrev et al., 1996). With Go-Go, the user can stretch her virtual arm much farther than her physical arm via a nonlinear physical-to-virtual hand distance mapping. The other eight techniques were created by combining two selection techniques (ray casting and occlusion), two attachment techniques (moving the hand to the object and scaling the user so the hand
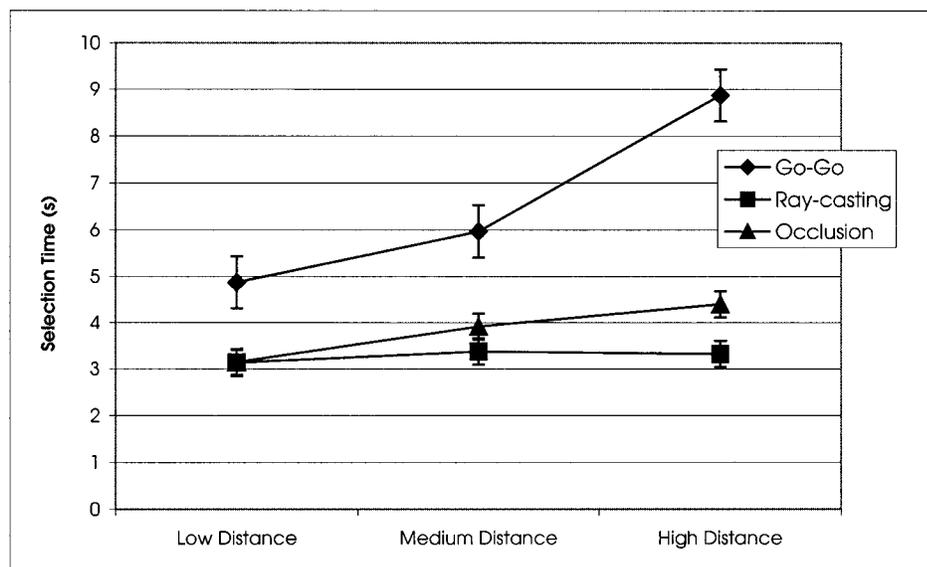
Figure 3. *Interaction between selection technique and distance for the selection time measure.*

touches the object), and two positioning techniques (linear mapping of hand motion to object motion and the use of buttons to move the object closer or farther away). Some of these combinations correspond to published interaction techniques. For example, the Direct HOMER technique (Bowman & Hodges, 1997) is composed of ray-casting selection, moving the hand for attachment, and a linear mapping for positioning.

Subjects wore a Virtual Research VR4 HMD displaying binocular (nonstereo) graphics at VGA resolution and were tracked using Polhemus FASTRAK trackers. Graphics were rendered on a Silicon Graphics Indigo2 MaxImpact. Input was given using a three-button joystick.

4.1.2 Results and Analysis. This complex experiment necessarily has a complex set of results. However, there are several major findings that emerge from the data. We performed a repeated measures analysis of variance (MANOVA) for both the selection and manipulation tasks.

First, results for selection of objects matched most of the experience that we had in our earlier informal study. Selection technique proved to be significant ($f(2,42) =$

13.6, $p < 0.001$), with the Go-Go technique (mean 6.57 sec. per trial) proving to be significantly slower than either ray casting (3.278 sec.) or occlusion selection (3.821 sec.) in post-hoc comparisons (LSD and Bonferroni). There was no significant difference between ray casting and occlusion. This is because selection using ray casting or occlusion is essentially a 2-D operation, whereas the Go-Go technique requires users to place the virtual hand within the object in three-dimensional space.

We also found significant main effects for distance ($p < 0.001$) and size ($p < 0.001$), with nearer and larger objects taking less time to select. There were also several interesting significant interactions. As shown in figures 3 and 4, the effects of distance and size varied depending on the selection technique being used ($p < 0.001$ in both cases). Figure 3 shows that selection time for the Go-Go technique increases with distance, while the other two selection technique times remain approximately constant, regardless of object distance. Figure 4 indicates that the Go-Go technique benefits much more from larger object sizes as compared to ray casting and occlusion selection.

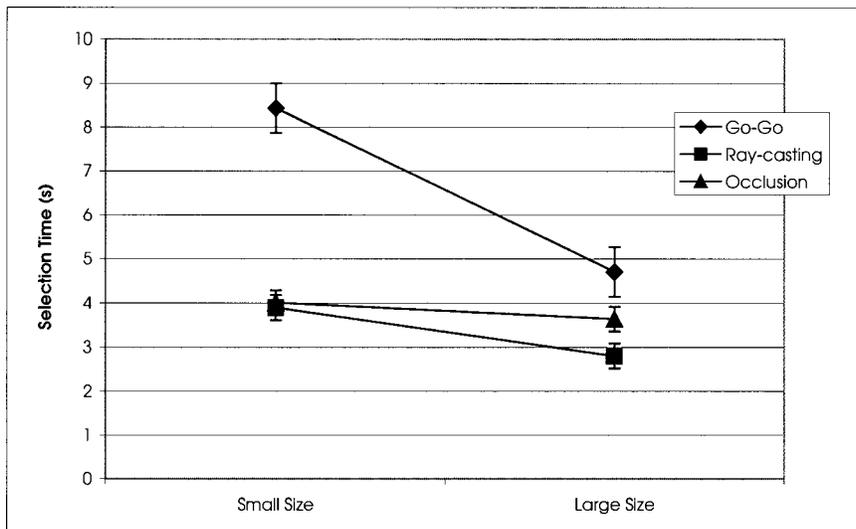We found that the number of errors made during

Figure 4. *Interaction between selection technique and object size for selection time measure.*

selection (errors included both selecting the wrong object and selecting no object) was significantly affected by both distance ($p < 0.001$) and size ($p < 0.001$). Interestingly, however, selection technique had no significant effect on errors.

It appears from this data that either ray casting or occlusion is a good general-purpose choice for a selection technique. However, this is tempered by our findings with regard to user comfort. We found that selection technique had a high correlation to the reported final level of user arm strain (after all trials had been completed, approximately thirty minutes of use). Occlusion selection produced significantly higher levels of arm strain than did ray casting, because ray casting allows the user to "shoot from the hip," whereas occlusion selection requires that the user's hand be held up in view. When selection takes a long time, as in the case of small or faraway objects, this can lead to arm strain of unacceptable levels.

The results for manipulation time were more difficult to interpret. Once the object had been selected, many of the techniques produced similar times for manipulation. (Table 1 shows the results for the nine techniques.) We found no significant effects of technique when attachment and manipulation techniques were considered separately. We did find a significant main effect for technique ($f(8,36) = 4.3$, $p < 0.001$), where technique is the combination of selection, attachment, and manipulation components. The only combinations that were significantly worse than others were the two combinations that combined ray casting with the attachment technique that scales the user, and this was likely due to poor implementation, from our observations of users.

One interesting fact to note from table 1 is that, for each pair of techniques using the same selection and attachment components, the technique using indirect depth control (button presses to reel the object in and out) had a faster mean time. Although this was not statistically significant, it suggests that an indirect, unnatural positioning technique can actually produce better performance. These techniques are not as elegant and seem to be less popular with users (based on informal questioning), but, if speed of manipulation is important, they can be a good choice.

All three of our within-subjects variables proved significant. Distance ($f(2,72) = 18.6$, $p < 0.001$), required accuracy ($f(1,36) = 19.6$, $p < 0.001$), and degrees of freedom ($f(1,36) = 286.3$, $p < 0.001$) all had significant main effects on manipulation time. As can be seen

Table 1. *Mean Time (Seconds) for Manipulation Task (*1:1 Physical-to-Virtual Hand Mapping)*

Technique	Selection	Attachment	Manipulation	Time (s)
1	Go-Go	Go-Go	Go-Go	26.551
2	Ray-casting	Move hand	Linear mapping	32.047
3	Ray-casting	Move hand	Buttons	30.970
4	Ray-casting	Scale user	Linear mapping*	40.683
5	Ray-casting	Scale user	Buttons	39.851
6	Occlusion	Move hand	Linear mapping	31.800
7	Occlusion	Move hand	Buttons	22.537
8	Occlusion	Scale user	Linear mapping*	24.780
9	Occlusion	Scale user	Buttons	20.528

Table 2. *Interaction Between Required Accuracy and Degrees of Freedom for Manipulation Time (Times in Seconds)*

	Two DOFs	Six DOFs
Low Accuracy	11.463	40.441
High Accuracy	13.991	53.992

from the large f-value for degrees of freedom, this variable dominated the results, with the six-DOF task taking an average of 47.2 sec. to complete and the two-DOF task taking 12.7 sec. on average.

We also found a significant interaction between required accuracy and degrees of freedom, shown in table 2. The six-DOF tasks with a high accuracy requirement (small target size relative to the size of the object being manipulated) were nearly impossible to complete in some cases, indicating that we did indeed test the extremes of the capabilities of these interaction techniques. On the other hand, required accuracy made little difference in the two-DOF task, indicating that the techniques we tested could produce quite precise behavior for this constrained task.

Unfortunately, these data cannot answer the question of whether there is a qualitative difference between the two-DOF and six-DOF tasks. Does the time to complete the two-DOF task have a constant slope regardless of the required accuracy, or is its upward slope simply of lower magnitude than that of the six-DOF task? In other words, does adding more degrees of freedom to a manipulation task create a different type of task, or does it simply add more of the same type of difficulty? The best way to answer these questions would be to include a middle condition with three degrees of freedom, and we propose this as future work. We can get some idea of the importance of this interaction by looking at these data on a log scale (figure 5). This graph does not appear to show an interaction, and thus we suggest that degrees of freedom may be additive, and not qualitatively different. This may be a fruitful topic for further research.

All of the significant results reported above have observed statistical power (computed using alpha = 0.05) of 0.92 or greater.

Finally, we found a demographic effect for performance. Males performed better on both the selection time ($p < 0.025$) and manipulation time ($p < 0.05$) response measures. Spatial ability and VE experience did not predict performance.

The lowest mean times were achieved by techniques using occlusion selection and/or the scaling attachment technique (techniques 7, 8, and 9). The fact that the scaling technique produces better performance, especially on the six-DOF task, makes intuitive sense. If the user is scaled to several times normal size, then a small
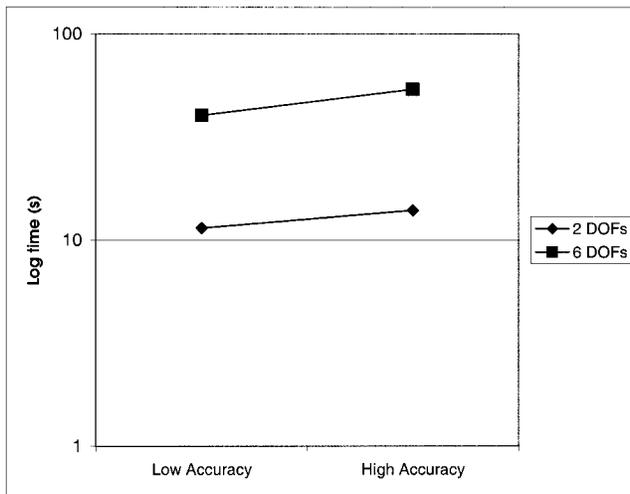
Figure 5. *Logarithmic scale graph of interaction between degrees of freedom and accuracy.*

physical step can lead to a large virtual movement. That is, users can translate their viewpoint large distances while manipulating an object using this technique. Therefore, on the difficult manipulation tasks, users can move their viewpoint to a more advantageous position (closer to the target, with the target directly in front of them) to complete the task more quickly. We observed this in a large number of subjects. However, scaled manipulation substantially increases the reported final level of dizziness (average 2.5 on the ten-point scale) relative to techniques where the user remains at the normal scale (average 1.3).

4.1.3 Guidelines for the Design of Selection and Manipulation Techniques. This testbed evaluation, because it included a large number of techniques, outside factors, and performance metrics, resulted in several general guidelines and principles that can be applied to the design of selection and manipulation techniques. We list some of the most important guidelines here:

- Use ray-casting (two-DOF) techniques if speed of remote selection is a requirement.
- Ensure that the chosen selection technique inte-

grates well with the manipulation technique to be used.
- If possible, design the environment to maximize the perceived size of objects.
- If the application allows, use manipulation tasks requiring the user to control fewer degrees of freedom.
- Provide general or application-specific constraints or manipulation aids.
- Avoid repeated, frequent scaling of the user or environment.
- Use indirect depth manipulation for increased efficiency and accuracy.

4.2 Travel Testbed

In the travel testbed, we implemented two search tasks that were especially relevant to our target application. Darken and Sibert (1996) characterize the two as *naïve search* and *primed search*. Naïve search involves travel to a target whose location within the environment is not known ahead of time. Primed search involves travel to a target that has been visited before. If the user has developed a good cognitive map of the space and is spatially oriented, she should be able to return to the target. We would also like to test exploration, in which the user is simply moving about with no specific target, but it would be very difficult to quantify performance on such an open-ended task.

We created a medium-sized environment (one in which there are hidden areas from any viewpoint, and in which travel from one side to the other takes a significant amount of time). The size of the environment could be varied if this was deemed an important outside factor on performance, but we left it constant in our implementation. We also built several types of obstacles that could be placed randomly in the environment. These included fences, sheds, and trees (figure 6).

Targets for the search tasks were flags mounted on poles. Each target was numbered 1 through 4 and had a corresponding color. Each target also had a circle painted on the ground around it, indicating the distance within which the user would have to approach to complete the search task (figure 7). The circles were of two

Figure 6. *Example obstacles from the travel testbed experimental environment.*

Figure 7. *Target object from the travel testbed experimental environment including flag and required accuracy radius.*

sizes: a large one (10 m radius) corresponding to low required accuracy, and a small one (5 m radius) corresponding to high required accuracy.

4.2.1   Method. Each subject completed 24 trials: eight trials in each of three instances of the environment. Each environment instance had the same spatial layout, but different numbers and positions of obstacles and different positions of targets. In each environment instance, the user first completed four naïve search trials and then four primed search trials. Before each trial, the flag number and color were presented to the user.

In the naïve search, the four targets were to be found in numerical order. Required accuracy was always at the low level, and targets were never visible from the user's starting location. During this phase, targets appeared only one at a time, at the appropriate trial. This was to ensure that subjects would not see a target before its trial, thus changing a naïve search to a primed search. The first trial began at a predefined location, and subsequent trials began at the location of the previous target.

In the primed search trials, the subjects returned to each of the four targets once, not in numerical order. During these trials, all targets were present in the environment at all times, because the subjects had already visited each target. Two factors were varied (wit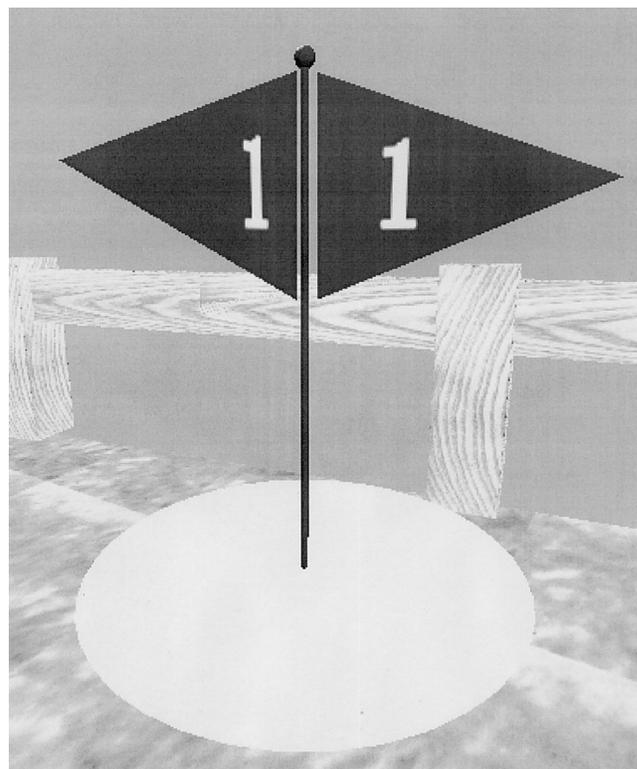hin subjects) during these trials: first, we varied whether the target could be seen from the starting position of the trial (visible/invisible); second, we varied the required accuracy using the radii around each target. Each of these variables had two levels, and therefore there were four possible combinations, and one trial of each of these combinations during each environment instance.

For each subject, we measured the total time taken to complete each trial (broken into two parts: the time between the onset of the stimulus and the beginning of movement, and the actual time spent moving). We assumed that the first time would correspond to the time spent in mental processing (perception of the stimulus and environment, and cognitive effort to remember where a target was last seen in the primed search task). This is not entirely accurate, as wayfinding activities undoubtedly continue after a subject's travel has begun. Therefore, the absolute measurements here are not
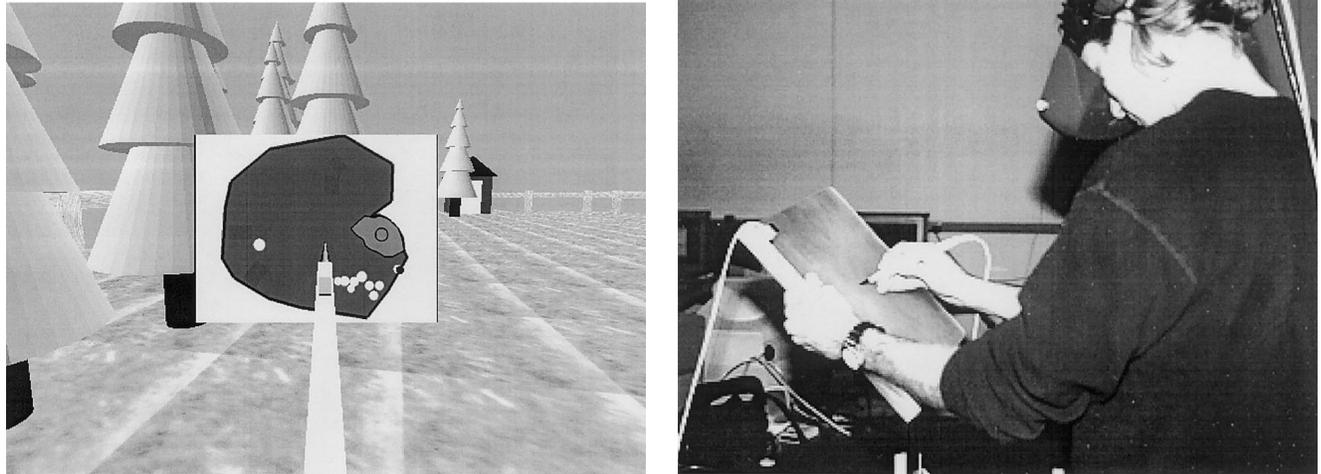
Figure 8. *Virtual (left) and physical (right) views of the map-dragging travel technique.*

meaningful, but the relative differences between techniques may give some indication of the amount of perceptual/cognitive processing necessary to move to a certain location or in a certain direction using a technique. We have labeled this measure *think time* in the analyses to follow. We also obtained subjective user comfort ratings, spatial ability scores, and demographic information, just as we did in the selection and manipulation testbed.

Forty-four subjects participated in the experiment. Four subjects did not complete the experiment due to sickness or discomfort, and two subjects did not complete the experiment due to computer problems. Thus, 38 subjects (32 men, 6 women, mean age 19.7) completed the evaluation, meaning that each technique was used by at least five subjects.

Seven travel techniques were implemented and used. Travel technique was a between-subjects variable. Three were steering techniques: pointing, gaze-directed, and torso-directed. These techniques use tracked body parts (hand, head, and torso, respectively) to specify the direction of motion. Two were manipulation-based travel techniques, one based on the HOMER technique and another on the Go-Go

technique. These techniques use object-manipulation metaphors to move the viewpoint by grabbing the world or an object, and then using hand movements to move the viewpoint around that position. Finally, we implemented two target-specification techniques. In the ray-casting technique, the user pointed a virtual light ray at an object to select it and then was moved by the system from the current location to that object. The map dragging technique involved dragging an icon on a two-dimensional map held in the nondominant hand. The map shows the layout of the environment and an icon indicating the user's position within the environment (figure 8, left). Using a stylus, the user can drag this icon to a new location. When the icon is released, the user is flown smoothly from the current location to the corresponding new location in the environment. Both the stylus and the map have both physical and virtual representations (figure 8). This technique was one of the travel metaphors used in our target application at the time. With both the ray-casting and map techniques, the user could press a button during movement to stop at the current location.

Equipment used was the same as in the selection/

Table 3. *Mean Times (Seconds) for Naïve Search Task*

Technique	Think time	Travel time	Total time
Gaze-directed	2.16	18.28	20.44
Pointing	2.20	22.33	24.53
Torso-directed	2.77	27.00	29.77
HOMER	4.20	37.66	41.86
Map dragging	29.54	52.39	81.93
Ray casting	1.86	34.95	36.81
Go-Go	3.29	21.48	24.77

manipulation testbed, except that a stylus was used instead of the joystick.

4.2.2 Results and Analysis. We performed a one-way analysis of variance (ANOVA) on the results for the naïve search task, with travel technique as a between-subjects variable. Table 3 gives the results for the naïve search task for each technique.

For each of the three time measures (think time, travel time, and total time), the travel technique used had a statistically significant effect ($p < 0.001$). We also performed post-hoc comparisons of techniques (LSD and Bonferroni), and found that for the think time measure the map-dragging technique was significantly slower than all other techniques. This makes intuitive sense, because the map technique is based on the target-specification metaphor, in which movement must be planned before it is carried out. The ray-casting technique also has this property, but selection of a single object is much faster than planning an entire route. With the other techniques, movement could begin immediately. However, because the difference is so large, we feel that there may be another factor at work here. The map technique requires users to mentally rotate the map so that it can be related to the larger environment. This mental rotation induces cognitive load on the user, which may cause them to be unsure of the proper direction of movement. The increased cognitive load may be reflected in the increased thinking time.

In the travel time measure, using the same post-hoc tests, we found that the pointing and gaze-directed

steering techniques and the Go-Go technique were significantly faster than HOMER, ray casting, and map dragging. The torso-directed steering technique was significantly faster than HOMER and map dragging. In general, then, steering techniques performed well at this task because of their directness and simplicity. The torso-directed technique performs slightly worse. We believe this is purely a function of mechanics. The user of the torso-directed technique must physically move his entire body to change direction. It is also interesting that the Go-Go technique performed well here, but HOMER did not, because they are both manipulation-based travel techniques. The difference seems to be that HOMER requires an object to move about, whereas the Go-Go technique allows the user to simply grab empty space and pull himself forward. Again, the map-dragging technique performed poorly. It is simply not suited for exploration and naïve search, because it assumes the user has a distinct target in mind.

For the primed search task, we performed a MANOVA, with technique as a between-subjects variable and visibility (two levels) and required accuracy (two levels) as within-subjects variables. Travel times were normalized relative to the distance between the starting point and the target. (This was not necessary for the naïve search task because subjects in that task had no knowledge of the location of the target and thus did not move in straight lines.) Table 4 presents a summary of results for this task. We do not list results for the two levels of required accuracy independently, because this factor was not significant in any of our analyses.

Results for think time mirrored the naïve search task. Again, technique was significant ($p < 0.001$), with the map dragging technique significantly slower in post-hoc comparisons (LSD and Bonferroni) than all other techniques, for the same reasons given above. Neither of the within-subjects factors was significant in predicting think time.

Technique was significant for the travel time measure ($p < 0.001$). Here, we found that pointing and gaze-directed steering, because they are direct and simple, were significantly faster than HOMER, ray-casting, and the map technique. Again, these techniques allow the user to form a direct mapping between the desired di-

Table 4. *Mean Times (Seconds) For Primed Search Task (*Normalized Times: Seconds per 100 Meters)*

Technique	Invisible think time *Visible think time*	Invisible travel time* *Visible travel time**
Gaze-directed	1.69 *1.49*	10.52 *4.70*
Pointing	2.30 *2.03*	10.20 *5.61*
Torso-directed	2.95 *1.40*	22.87 *5.81*
HOMER	3.85 *2.67*	26.34 *13.81*
Map dragging	20.58 *14.01*	25.07 *18.97*
Ray casting	2.09 *1.92*	29.69 *13.72*
Go-Go	2.66 *1.72*	17.55 *7.36*

rection of motion and the action that needs to be taken (look or point in that direction).

We also found that visibility of the target from the starting location was significant here ($p < 0.001$). Trials in which the target was visible averaged 12 sec., as opposed to 23 sec. for trials in which the target was hidden.

The map technique performed badly, but it was only significantly worse than gaze-directed steering, pointing, and Go-Go. We had expected that the map would be useful for the primed search, because it allows users to specify the location of the target and not the direction from the current location to the target. However, this assumes that the user understands the layout of the space, and that the technique is precise enough to let the user move exactly to the target. In the experiment, the size of the target was not large enough, even in the low-required-accuracy condition, to allow precise behavior with the map technique. We observed users mov-

ing directly to the area of the target, but then making small adjustments to move within the required range of the target. However, the best results with the map occurred in trials with low required accuracy and a target that was not visible from the starting location.

Another technique that we expected to perform well in the primed search task was ray casting, because it allows the user to move directly to a target. This should especially hold in cases in which the target is initially visible. We believe these results were not found due to our implementation of targets as flags. The flagpoles were very thin, and thus impossible to select at any distance. The flags themselves were larger, but due to the size of the environment might appear very small from the starting location. Thus, users of the ray-casting technique often had to select an intermediate target in order to get close enough to select the flag.

We also performed an analysis that compared the two types of tasks. For this analysis, technique was again a between-subjects variable and task was a within-subjects factor. We considered only those trials in which the target was initially visible and the required accuracy was low, to match the naïve search trials. For the travel time measure, we found that task was significant ($p < 0.001$), with the naïve search taking 30 sec. on average versus 23 sec. for the primed search.

For the think time measure, task was not significant, but we did find a significant interaction between task and technique ($p < 0.025$). This interaction is due to the fact that the amount of think time for the map technique drops significantly for primed search trials (figure 9—error bars have been omitted in the figure for readability), whereas think time for the other techniques remains approximately the same. This indicates that subjects had learned the layout of the space and were more confident in the map-dragging task because they knew the area in which the target was.

For each of the significant results reported above, the observed statistical power was 0.987 or greater, with alpha = 0.05.

Our evaluation showed that, if the most important performance measure is speed of task completion, steering techniques are the best choice. Users also seem to prefer these techniques to others. Of the steering tech-
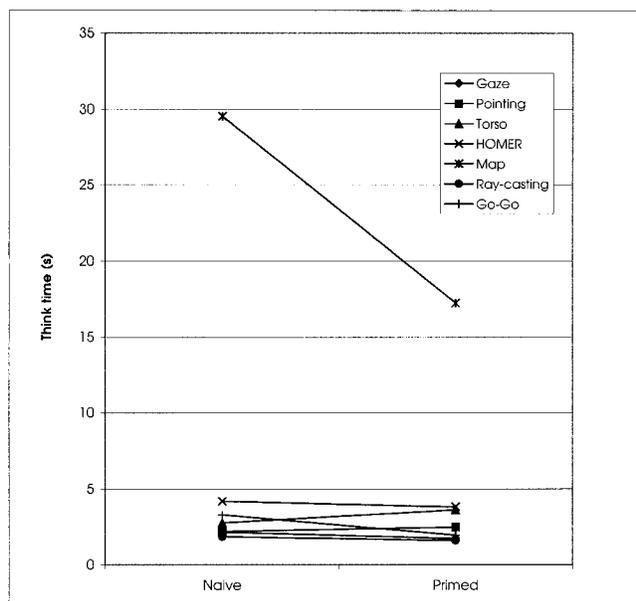
Figure 9. *Interaction between task type and technique for think time on search tasks.*

niques, pointing is clearly the most versatile and flexible, because it allows comfortable and efficient changes in direction. The Go-Go technique also performed well in this study with respect to speed. However, upon analysis of our comfort-rating measures, we found that Go-Go produced arm strain, dizziness, and nausea in some users when used as a travel technique. This suggests that viewpoint movement using hand-based manipulation may be discomforting to users because it is so different from the normal methods of movement. Gaze-directed steering also produced some significant discomfort (mainly dizziness), likely because it requires rapid and repeated head movements. The visual scene lags behind head movements due to tracker latency, so these could be the cause of discomfort. Of the seven techniques, only pointing and ray casting produced no significantly high discomfort levels.

As discussed above, the map technique was the most disappointing technique in this study. It seems to be well suited for low-precision, goal-directed travel. We believe that this technique would have performed better if the required accuracy had been lower on certain trials. It would probably also benefit from the use of a "view-

up" map as opposed to a standard "north-up" map (Darken & Cevik, 1999). Performance on the primed-search would likely improve because of its egocentric nature. However, we have other reasons for using a north-up map, including the fact that it is a fixed frame of reference within a dynamic environment, and thus may facilitate learning of the spatial layout more quickly (Wickens & Baker, 1995). The map technique is also useful for other tasks, such as object manipulation, and so we do not believe that this technique should be removed from consideration as a result of its performance in this evaluation.

Finally, we also noted that user strategies (Bowman, Davis, Badre, & Hodges, 1999) were important in this experiment. No collision detection was implemented in the experimental environments, so users could move through objects if desired. In certain cases, this was highly advantageous, (for example, when the flag was just on the other side of a large fence). We noted that subjects using this strategy performed better on the primed search task, because they could take a straight-line path to the target. We also observed that certain techniques afford this strategy more than others. Steering techniques in general do not afford this, as they more closely mimic natural movement. Subjects using steering techniques generally went around obstacles. More unnatural techniques such as map dragging, Go-Go, HOMER, and ray casting seem to suggest to the user that the VE does not work in the same manner as the physical world, and that therefore moving through objects is allowed. This represents another benefit of so-called "magic" techniques.

4.2.3 Guidelines for the Design of Travel Techniques. The travel testbed also produced some important guidelines that should be taken into consideration when designing travel techniques for VE applications. These include:

- Use steering techniques for generality and efficiency in search tasks.
- If target-based techniques are used, design the environment or application so that targets are large and less accuracy of movement is required.

- Avoid manipulation-based travel techniques in applications where travel is frequent and users experience long exposure times.
- "Magic" techniques not based on a natural movement metaphor may afford more efficient, straight-line movement, but these should not be used in applications in which realism of movement is a requirement.

5 Application of Results

The most important test of the validity of testbed evaluation is its usefulness in informing the interaction design of real-world VE applications. Prior to our two experiments, we had implemented an immersive design system called the Virtual Habitat (Bowman, Wineman, Hodges, & Allison, 1998), which used an accurate model of the gorilla habitat at Zoo Atlanta. The application allowed the user to move about and modify the habitat for the purpose of environmental design education. User tasks included exploring the environment; playing audio annotations; choosing an appropriate terrain model; moving trees; creating and placing new trees, rocks, and grass as design elements; and positioning and orienting visitors' viewpoints into the habitat. Clearly, this is a complex VE application encompassing the tasks of travel, selection, manipulation, and system control. If a redesign of this system based on the results of the testbed evaluations caused increases in usability, the benefits of testbed evaluation would be validated.

First, we need to understand the tasks and interaction requirements of this application. There are essentially two different travel tasks that the user of the Virtual Habitat might wish to perform. First, general exploration of the environment needs to be supported. In this type of travel, the user is simply looking around, getting a feel for the layout, size, and features of the VE. For this purpose, a travel technique must be intuitive to the user, so that the focus can be on the environment and not on the technique. It must also allow continuous changes to the trajectory of motion, so that the user can instantaneously make course corrections. In terms of the performance metrics we have described for travel, a

technique for exploration requires high levels of spatial awareness and information gathering. Ease of learning, ease of use, presence, and user comfort will also be important. Speed and accuracy are not requirements for such a technique.

Second, users may wish to travel to specific locations in the environment to obtain information. This type of travel has an explicit goal and direction, and is therefore unlike the exploration described above. It also has different requirements; in particular, speed and accuracy will be quite important, because we do not wish to require the user to wait to get the desired information, and we want the user to be able to move accurately to the location of the information. Because the user's focus is on the destination and not the path, spatial awareness and information-gathering ability during travel may not be as important. Such a technique will still require moderately high levels of ease of use and user comfort.

The application needs one or more techniques for selection, including a standalone technique to select audio annotations for playback and a technique to select objects for manipulation in the immersive design component. These techniques may be the same, or they may be individually considered, as was the case with the travel techniques. It is more likely here that we can find a single selection technique to do the job, because the requirements for both tasks are similar. In general, we need a technique that can be used at a reasonable distance, and which is quite intuitive and easy so that users can focus on the task at hand. In terms of performance metrics, the application requires high levels of accuracy of selection, ease of use, and user comfort, with speed also being a main consideration.

Finally, we need one or more manipulation techniques with which to accomplish the immersive design tasks (moving visual elements, for example). We need expressive techniques that can be used to place objects at any location, but that are also well constrained and easy to use. An additional consideration is that the manipulation technique integrates well with the selection and travel techniques that are chosen. Expressiveness (the range of positions and orientations in which an object can be placed), accuracy of placement, and ease of use will be the most important requirements for design-

ers, and speed and user comfort will be secondary considerations.

The initial implementation of this application (Bowman, Wineman et al., 1998) used both the pointing and the map techniques for traveling. Users could select and manipulate objects directly with the Go-Go technique and indirectly on the virtual map. We performed a usability evaluation on this implementation of the system. Students from a class on environmental design used the Virtual Habitat to redesign the gorilla exhibit for a class presentation. We collected subjective usability ratings from each of the students. These ratings were on a five-point scale and covered all of the features of the system. The results of this evaluation are shown in the right column of table 5.

The results of the travel testbed showed that our initial design iteration actually met the application's performance requirements well. We found that speed and other metrics on both the exploratory and the directed travel tasks was best with continuous steering techniques, such as pointing. Although this was intended in the previous design iteration to be used for exploration, it appears to be well suited to the performance requirements of the goal-directed travel task. User comfort was not a major factor in the testbed experiment, but the pointing technique performed well in this category.

In our earlier usability study, the map-dragging technique was rated subjectively higher than the pointing technique. However, we noted some problems with it, and these problems were verified in the testbed evaluation. Most notably, users often did not know which direction to drag the user icon to move to a given location. In the usability study, we found that certain users were better with the map technique than others, and we hypothesized that these people were able to do the mental rotations of the map necessary to determine direction. Therefore, we left the map-dragging technique in place in the final design, but encouraged users to utilize it only after they were quite familiar with the spatial layout of the habitat.

A related usability problem that we found in the initial design iteration concerned the loss of spatial orientation on the part of users. Users often became lost or disoriented, especially after using the pointing technique

to fly in a direction other than that of their gaze. Some users also had difficulty relating the static map information to the dynamic environment. These are exactly the concerns addressed by our spatial orientation experiment (Bowman et al., 1999). In that evaluation, we found that subjects who used advanced strategies for maintaining orientation had the best performance. Therefore, in the final design iteration for this application, we modified our written and verbal instructions in order to train users in these strategies. Strategies relevant to the Virtual Habitat include 3-D overview (fly up above the environment to get a survey view), backing in (moving backwards to a destination so that it is placed in the context of previously visited areas), proprioceptive pointing (reminding oneself of the location of known objects by pointing), stop and look (pausing to look around at the current location), and path retracing (moving again along previously traveled paths, often from a different direction). Users are not likely to use all of these strategies, but using one or more of them could increase spatial orientation.

The selection and manipulation testbed confirmed our informal observations of the Go-Go technique. It is not well suited for selection of objects that are small and/or far away. Moreover, it was the lowest rated of the techniques in our usability study, due to the frustration people had with selecting distant objects. The testbed results showed that the HOMER technique was the best fit for the performance requirements specified above for selection and manipulation. It can select objects well at long distances, and ray casting is quite easy to use and speedy. The manipulation component of HOMER is very expressive and also easy to use and moderately fast, according to the empirical results. HOMER was not near the top of the rankings for manipulation time in our study, but, as stated above, speed of manipulation is not a key performance requirement of the Virtual Habitat.

When the interaction design was finalized, a new usability study was performed under similar circumstances and using the same evaluation metrics (interviews and usability ratings). In this way, we compared the usability of a system designed using intuition and observation to that of a system implemented based on formal evalua-

Table 5. *Mean Usability Ratings (Standard Deviations in Parentheses) for the Virtual Habitat Interaction Design (*Features Changed Since the Previous Iteration)*

Usability categories	Final implementation	Initial implementation
Selecting annotations*	4.70 (0.45)	N/A
Changing terrain	4.20 (0.76)	4.21 (1.15)
User movement with stylus*	4.10 (0.89)	3.71 (1.11)
Tablet: dragging user icon to move*	4.10 (0.74)	4.21 (0.81)
Direct object manipulation*	4.00 (0.35)	3.14 (1.18)
Tablet: object creation	4.00 (0.71)	4.43 (0.53)
Moving viewpoints	3.55 (0.94)	4.20 (0.84)
Tablet: object manipulation	3.50 (1.00)	3.86 (0.94)
Moving viewpoint barriers*	3.40 (1.39)	4.10 (1.02)
Tablet: general interaction	2.90 (0.89)	3.86 (0.90)

tion and design methods. This study would validate the use of our formal design and evaluation methodology if increased usability were found.

Five user sessions were held, lasting from sixty to ninety minutes each. During the session, the users were instructed on the use of the techniques, allowed to explore the virtual habitat, and shown how to use the various information and design tools. Each user or group of users spent twenty to forty minutes using the design tools to modify the design of the gorilla habitat. Subjects were members of an undergraduate design class with experience in both traditional and computer-aided design. At the end of the session, each user or group was asked for their comments and observations on the system, as well as a set of usability ratings on the various features of the application. These ratings again were on a five-point scale, with a value of 5 representing high usability. A summary of the results is presented in table 5, including average usability ratings and standard deviations for each of the system's features.

The most important result from the table is that our application of the results of formal design and evaluation had positive results on reported usability. This is most easily seen for the direct object manipulation feature, which was changed from the Go-Go technique to the HOMER technique, and which received a much higher usability rating in the final iteration. This is de-

spite the fact that this group of users seemed to have a lower baseline rating overall. (For all unchanged components, the average usability rating was lower than the corresponding rating from the initial iteration.) Also, ray casting proved to be very easy to use as a selection mechanism for the audio annotations, receiving the highest rating of any feature. Although we did not measure the usability of the Go-Go technique for annotation selection in the previous study, it was the source of many verbal usability complaints by users.

Second, we note that the reported usability of the pointing technique was improved in the final iteration. Although the implementation of this technique did not change, the training given to users in the proper use of this technique was modified. Both written and verbal instructions were given to users telling them how to use this technique to maintain spatial orientation (for example, flying upwards to get a survey view of the environment). This result validates our earlier finding that the training of specific strategies can have an effect on overall performance.

The map-dragging technique for travel was rated highly, but slightly lower than the rating from the previous iteration. Again, this is consistent with other features that remained unchanged. Therefore, the additional training in strategies for spatial orientation did not increase the usability of this technique, again vali-

dating our earlier findings. Strategy sophistication can increase performance with steering techniques, but performance using target-specification techniques is relatively constant no matter what strategies are used. Also, fewer of the strategies are possible when using the map-dragging technique.

The comment of one subject is particularly enlightening with regard to the travel techniques used in this system. Although the map technique performed poorly in the testbed evaluation and is not useful on its own, it can be a good complement to a steering technique. The subject stated that he would not rate the map technique highly, except that it worked well in conjunction with the pointing technique. This leads to the general principle that multiple, redundant interaction techniques should sometimes be used to improve usability.

On the whole, this usability study provided an unequivocal endorsement of our methodology. The use of the formal design and evaluation framework, testbed evaluation, and application of results based on performance specification caused a measurable increase in usability.

6   Discussion

Testbed evaluation does have disadvantages relative to more-traditional assessment methods. It is generally more time-consuming, more costly to implement, and requires more experimental subjects. Testbed experiments produce complex sets of data that may be difficult to analyze. However, the benefits outweigh the disadvantages.

Reusability is one important advantage of testbed evaluation. If new techniques for a given interaction task are developed, they may be run through the testbed for that task and compared against previously tested techniques.

Second, because a testbed uses multiple variables, the data that is generated is more complex. This often leads to interesting interactions between variables that would not have emerged otherwise.

Third, the testbeds give us the ability to produce predictive models of performance within the design space

defined by a taxonomy. Because we partition techniques into components, we obtain performance results at the component level rather than at the level of the complete technique. Thus, we may be able to predict the performance of a combination of components that were not evaluated directly. In doing this, we do not sacrifice generality, because components are always assessed as part of a complete technique.

For both interaction tasks, we showed that none of the techniques performed best in all situations. Rather, performance depends on a complex combination of factors including the interaction technique and characteristics of the task, environment, user, and system. Therefore, applications with different attributes and interaction performance requirements may need different interaction techniques.

7   Conclusions and Future Work

In this paper, we have shown that testbed evaluation can be an effective and useful method for the assessment of interaction techniques for virtual environments. Our experiments, using multiple independent and dependent variables and a broad definition of performance, demonstrate the rich and complex characteristics of VE interaction. Simple experiments would not reveal this complexity. We have validated the testbed approach by applying its results to a real-world VE application and measuring usability gains as a direct result.

In the future, we would like to extend this approach to make it more rigorous and systematic. Although our testbeds were based on a formal design and evaluation framework, we currently do not have any way to verify their coverage of the task space, that is, the extent to which they test all of the important aspects of a task. The ability to state this definitively would increase the descriptive power of the testbed experiments.

One important outside factor not addressed specifically in these experiments was the user's level of expertise. The testing of expert users could produce quite different results. There are very few potential users of most VE applications who could be considered experts, so the current results are useful, but an understanding

of how performance changes over time would have an added value.

We also plan to make the testbeds and experimental results more readily available to VE developers and researchers. The environments and tasks themselves are designed to be reusable for any interaction technique, so their dissemination could be useful as new techniques are developed. The results of the testbeds are complex, and not easily applied to VE systems. A set of guidelines based on the results is part of the answer to this problem, but we feel that it would also be useful to create an automated design guidance system that suggests interaction techniques by matching the requirements of a VE application to the testbed results.

Finally, we would like to compare this methodology to others, such as usability engineering. These approaches are quite different, but both have the goal of increasing the performance (including usability) of VE applications. It would be interesting to compare the costs and benefits of applying these two methods.

Acknowledgments

References

Bowman, D., Davis, E., Badre, A., & Hodges, L. (1999). Maintaining spatial orientation during travel in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments, 8*(6), 618–631.

Bowman, D., & Hodges, L. (1997). An evaluation of techniques for gabbing and manipulating remote objects in immersive virtual environments. *Proceedings of the ACM Symposium on Interactive 3D Graphics,* 35–38.

———. (1999). Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *The Journal of Visual Languages and Computing, 10*(1), 37–53.

Bowman, D., Hodges, L., & Bolter, J. (1998). The Virtual Venue: User-Computer Interaction in Information-Rich Virtual Environments. *Presence: Teleoperators and Virtual Environments, 7*(5), 478–493.

Bowman, D., Koller, D., & Hodges, L. (1997). Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. *Proceedings of the Virtual Reality Annual International Symposium,* 45–52.

Bowman, D., Wineman, J., Hodges, L., & Allison, D. (1998). Designing animal habitats within an immersive VE. *IEEE Computer Graphics and Applications, 18*(5), 9–13.

Darken, R., & Cevik, H. (1999). Map usage in virtual environments: Orientation issues. *Proceedings of IEEE Virtual Reality,* 133–140.

Darken, R., & Sibert, J. (1996). Wayfinding behaviors and strategies in large virtual worlds. *Proceedings of CHI,* 142–149.

Fleishman, E., & Quaintance, M. (1984). *Taxonomies of Human Performance: The Description of Human Tasks.* Orlando: Academic Press.

Herndon, K., van Dan, A., & Gleicher, M. (1994, October). The challenges of 3D interaction. *SIGCHI Bulletin, 26*(4), 36–43.

Hix, D., Swan, J., Gabbard, J., McGee, M., Durbin, J., & King, T. (1999). User-centered design and evaluation of a real-time battlefield visualization virtual environment. *Proceedings of IEEE Virtual Reality,* 96–103.

Kaur, K. (1998). *Designing virtual environments for usability.* Unpublished doctoral dissertation, University College London.

Kennedy, R., Lane, N., Berbaum, K., & Lilienthal, M. (1993). A simulator sickness questionnaire (SSQ): A new method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203–220.

Koller, D., Mine, M., & Hudson, S. (1996). Head-tracked orbital viewing: An interaction technique for immersive virtual environments. *Proceedings of the ACM Symposium on User Interface Software and Technology,* 81–82.

Lampton, D., Knerr, B., Goldberg, S., Bliss, J., Moshell, J., & Blau, B. (1994). The virtual environment performance assessment battery (VEPAB): Development and evaluation. *Presence: Teleoperators and Virtual Environments, 3*(2), 145–157.

Pierce, J., Forsberg, A., Conway, M., Hong, S., Zeleznik, R., & Mine, M. (1997). Image plane interaction techniques in

3-D immersive environments. *Proceedings of the ACM Symposium on Interactive 3D Graphics,* 39–44.

Poupyrev, I., Billinghurst, M., Weghorst, S., & Ichikawa, T. (1996). The Go-Go interaction technique: Non-linear mapping for direct manipulation in VR. *Proceedings of the ACM Symposium on User Interface Software and Technology,* 79–80.

Poupyrev, I., Weghorst, S., Billinghurst, M., & Ichikawa, T. (1997). A framework and testbed for studying manipulation techniques for immersive VR. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology,* 21–28.

Ware, C., & Osborne, S. (1990). Exploration and virtual camera control in virtual three dimensional environments. *Computer Graphics, 24*(2), 175–183.

Wickens, C., & Baker, P. (1995). Cognitive issues in virtual reality. In W. Barfield & T. Furness, (Eds.), *Virtual Environments and Advanced Interface Design.* New York: Oxford University Press.

Witmer, B., & Singer, M. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.