

# Testbed Evaluation of Virtual Environment Interaction Techniques

Doug A. Bowman  
Department of Computer Science (0106)  
Virginia Polytechnic & State University  
Blacksburg, VA 24061 USA  
(540) 231-7537  
bowman@vt.edu

Donald B. Johnson, Larry F. Hodges  
Graphics, Visualization, and Usability Center  
Georgia Institute of Technology  
Atlanta, GA 30332-0280 USA  
(404) 894-8787  
{donny, hodges}@cc.gatech.edu

## 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 in order 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. This has been shown to produce measurable usability gains in a real-world VE application.

## 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 [6]. Users have difficulty controlling multiple degrees of freedom simultaneously, interacting in a volume rather than on a surface, and understanding 3D spatial relationships. These problems are magnified in an immersive VE, because standard input devices such as mice and keyboards cannot be used, and the display resolution is often low.

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

VRST 99 London UK

Copyright ACM 1999 1-58113-141-0/99/12...\$5.00

Therefore, the design of interaction techniques (ITs) and user interfaces for VEs must be done with extreme care in order to produce useful and usable systems. Since 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 for assessment of VE interaction. Testbeds are representative sets of tasks and environments. 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.

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 techniques for the tasks of travel and selection/manipulation of virtual objects. 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 need to be evaluated formally. Work has focused on a small number of "universal" VE tasks, such as travel [10, 15], and object selection and manipulation [12, 13].

Evaluation of VE interaction has for the most part been limited to usability studies [e.g. 3]. 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 [7]. 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 [e.g. 8]. 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 [11] 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 et al [14].

### 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 (see [2] for a more complete description of this framework). In this section we will briefly discuss pieces of this methodology relevant to the current work.

#### 3.1 Taxonomies

Our first step is to create a taxonomy of interaction techniques for the tasks in which we are interested. As an example, figure 1 shows a taxonomy for the tasks of selection & manipulation. We do this in two steps. First, we perform a task analysis using hierarchic decomposition, to partition the task into subtasks, of which there may be several levels. Second, for each of the lowest-level subtasks, we list technique components that accomplish that subtask.

For example, consider the task of modifying an object's color. We might partition this into three subtasks: select an object, select a color, and apply the color. For the color selection subtask, we could list components such as using RGB sliders, specifying a point in an RGB cube, or picking from a fixed palette.

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 determine 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 may also be important. Finally, task-related factors including spatial

orientation during navigation or expressiveness of manipulation often play a role.

Therefore, in our work, we have a broad definition of performance, and will attempt to measure multiple performance variables during testbed evaluation. For those factors which are not directly measurable, standard questionnaires (e.g. [9] for simulator sickness, [16] for presence) or subject self-reports may need to be used.

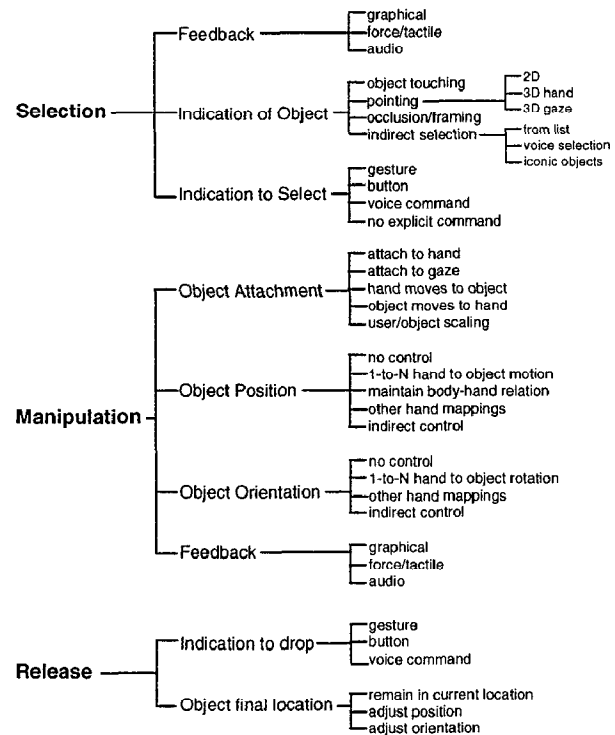


Figure 1. Taxonomy of selection/manipulation techniques

#### 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 (e.g. the required accuracy), environment (e.g. the number of objects), user (e.g. spatial ability), and system (e.g. stereo vs. 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.

#### 3.4 Application of Testbed Results

Testbed evaluation is not an end unto itself. Rather, it has the goal of producing applications with high levels of performance. In our methodology, applications specify their interaction performance requirements for each task in terms of the performance metrics that we have defined for that task (section 3.2). For travel, one application might need high levels of speed, while another is interested mainly in maintaining the user's spatial orientation.

In this way, we can use the results of testbed evaluation to match appropriate interaction techniques with each application. This

reflects the fact that each application has its own requirements, and that there is no set of techniques which will maximize performance for all applications and domains.

## 4. EXPERIMENTS

We present two experiments that bring together the components of the formal methodology. The first testbed was designed to evaluate selection and manipulation while 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 for each of the 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

The selection and manipulation testbed is composed of a selection phase, where the user selects the correct object from a group of objects, and a manipulation phase, where 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 blue box in the center of the array of cubes, and then place it within the two wooden targets in the manipulation phase. In certain trials, yellow spheres on both the selected object and the target specify the required orientation of the object.

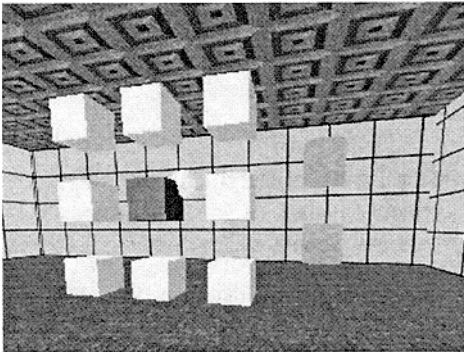


Figure 2. Trial setup in the selection/manipulation testbed

#### 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 (3 levels), the size of the object to be selected (2 levels), and the density of objects surrounding the object to be selected (2 levels). These seem 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 (2 levels – this corresponds to the accuracy required for placement). Second, the number of required degrees of freedom varied (2 levels), so that we could test the expressiveness of the techniques. The 2 DOF task only required

users to position the objects in the horizontal plane, while the 6 DOF task required complete object positioning and orientation. Finally, we varied the distance from the user at which the object must be placed (3 levels). Other outside factors, such as stereo vs. mono viewing, or the use of interactive shadows, could have been included but were not in order to maintain a manageable experiment size.

Response variables were the speed of selection, the number of errors made in selection, the speed of placement, and qualitative 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 10-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.

We required users to place the selected objects completely within the targets and within five degrees of the correct orientation on the 6 DOF trials. Graphical feedback told the user when the object was in the correct location.

Forty-eight subjects (31 males, 17 females) participated in the study. Each subject completed 48 trials, except for 3 subjects who did not complete the experiment due to dizziness or sickness. Subjects were allowed to practice the technique for up to five minutes before the experimental trials began. Subjects completed 4 blocks of 12 trials each, alternating between trials testing selection and manipulation.

Nine different selection/manipulation techniques, taken from our taxonomy [2], were compared in a between-subjects fashion. Thus, there were five subjects per technique. One technique was the Go-Go technique [13]. With Go-Go, the user can stretch her virtual arm much farther than her physical arm via a non-linear 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 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 HOMER technique is composed of ray-casting selection, moving the hand for attachment, and a linear mapping for positioning.

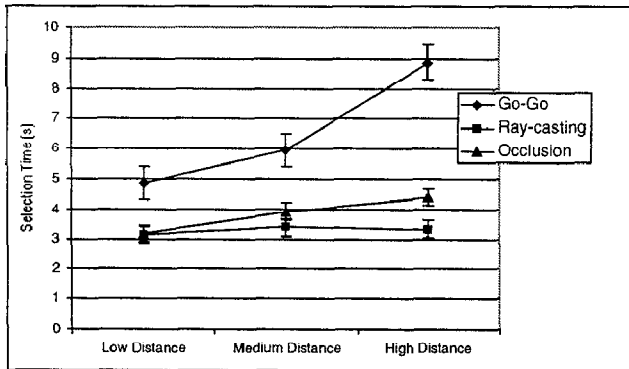
Subjects wore a Virtual Research VR4 HMD displaying biocular (non-stereo) graphics, and were tracked using Polhemus Fastrak trackers. Graphics were rendered on a Silicon Graphics Indigo2 MaxImpact. Input was given using a 3-button joystick.

#### 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, selection technique proved to be significant ( $f(2,42)=13.6, p < 0.001$ ). The Go-Go technique, which requires positioning the hand in 3D space (mean 6.57 seconds per trial), was significantly slower than either ray-casting (3.278 secs.) or occlusion selection (3.821 secs.), which are both basically 2D operations. There was no significant difference between ray-casting and occlusion.

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. Only Go-Go was significantly worse for selecting objects at a distance (figure 3). Also, the Go-Go technique benefits the most from larger object sizes as compared to ray-casting and occlusion selection.



**Figure 3. Interaction between selection technique and distance for the selection time measure**

It appears from this data that either ray-casting or occlusion is a good general-purpose choice for a selection technique. However, occlusion selection produced significantly higher levels of arm strain than ray-casting, because ray-casting allows the user to "shoot from the hip," while occlusion selection requires that the user's hand be held up in view. When selection takes a long time, or when selection is done repeatedly, 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 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. We found no significant effects of technique when attachment and manipulation techniques were considered separately.

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. Though this was not statistically significant, it indicates 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, 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 from the large f-value for degrees of freedom, this variable dominated the results, with the six degree of freedom task

taking an average of 47.2 seconds to complete and the two degree of freedom task taking 12.7 seconds on average.

Tech	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 1. Mean time (seconds) for manipulation task (\* one-to-one physical to virtual hand mapping)**

We also found a significant interaction between required accuracy and degrees of freedom, shown in table 2. The six degree of freedom 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 2 DOF task, indicating that the techniques we tested could produce quite precise behavior for this constrained task.

	2 DOFs	6 DOFs
Low Accuracy	11.463	40.441
High Accuracy	13.991	53.992

**Table 2. Interaction between required accuracy and degrees of freedom for manipulation time (seconds)**

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 degree of freedom task, makes intuitive sense. If the user is scaled to several times normal size, then a small 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 significant number of users.

However, scaled manipulation significantly increases the reported final level of dizziness relative to techniques where the user remains at the normal scale. Thus, an important guideline is that such techniques should not be employed when users will be immersed for extended periods of time.

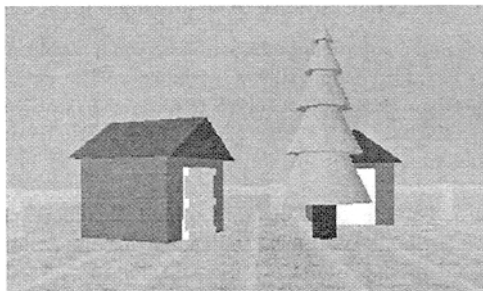
## 4.2 Travel Testbed

In the travel testbed, we implemented two search tasks that were especially relevant to our target application. Darken [5] characterizes 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 which has been visited before. If the user has developed a good cognitive map of the space and is spatially oriented, he should be able to return to the target.

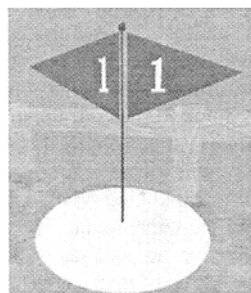
#### 4.2.1 Method

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 4).



**Figure 4. Example obstacles from the travel testbed experimental environment**

Targets for the search tasks were flags mounted on poles. Each target was numbered 1-4, 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 5). There were two sizes of this circle: a large one (10 meter radius) corresponding to low required accuracy, and a small one (5 meter radius) corresponding to high required accuracy.



**Figure 5. Target object from the travel testbed experimental environment including flag and required accuracy radius**

Each subject completed 24 trials – 8 trials in each of 3 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 4 naïve search trials and then 4 primed search trials. Before each trial, the flag number and color were presented to the user.

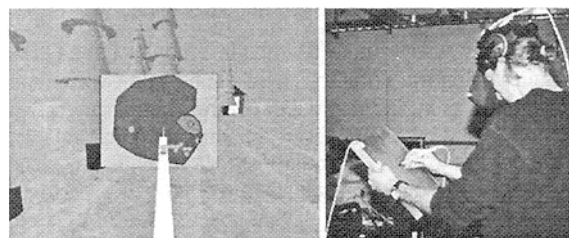
In the naïve search, the four targets were to be found in numerical order. Required accuracy was always at the low level, and targets were never visible from the user's starting location. During this phase, targets only appeared one at a time, at the appropriate trial. This was to ensure that subjects 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, they 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, since the subjects had already visited each target. Two factors were varied (within-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.

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 non-dominant hand. The map shows the layout of the environment and an icon indicating the user's position within the environment (figure 6, 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 6). 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.



**Figure 6. Virtual (left) and physical (right) views of the map dragging travel technique**

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 this first time would correspond to the time spent thinking about the task (cognitive effort to

remember where a target was last seen in the primed search task). We also obtained subjective user comfort ratings 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 completed the evaluation. Equipment used was the same as in the selection/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.

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

**Table 3. Mean times (seconds) for naïve search task**

For each of the three time measures (think time, travel time, and total time), the travel technique used had a statistically significant effect ( $p < 0.001$ ). The think time measure showed that the map dragging technique was significantly slower than all other techniques. This makes intuitive sense, since the map technique is based on the route-planning metaphor, where movement must be planned before it is carried out. The ray-casting technique (target specification) 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 can be seen directly in increased thinking time.

In the travel time measure, 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. It is also interesting that the Go-Go technique performed well here, but HOMER did not, since they are both manipulation-based travel techniques. The difference seems to be that HOMER requires an object to move about, while 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 multivariate analysis of variance (MANOVA), with technique as a between-subjects variable and visibility (2 levels) and required accuracy (2 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 since 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. Neither of the within-subjects factors was significant in predicting think time.

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

**Table 4. Mean times (seconds) for primed search task (\* normalized times: seconds per 100 meters)**

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. 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, since 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 moving directly to the area of the target, but then making small adjustments in order 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 not visible from the starting location. 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 seconds, as opposed to 23 seconds for trials in which the target was hidden.

We also performed an analysis that compared the two types of tasks. For this analysis, technique was again a between-subjects variable, while task was a within-subjects factor. We only considered the 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 seconds on average vs. 23 seconds for the primed search.

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 over others. Of the steering techniques, pointing is clearly the most versatile and flexible, since 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 discomfoting 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. 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. Performance on the primed-search would likely increase 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. 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.

## 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. Previously, we had implemented an immersive design system, 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. The initial implementation of our application [4] 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. A group of architecture students used the application and gave subjective usability ratings for various system tasks.

The results of the testbed experiments revealed a deficiency in our original choices of interaction techniques for this system. Based on these results, we replaced the Go-Go technique with the HOMER technique. As discussed above, we found that ray-casting exhibited better selection performance and that it was not significantly affected by object size or distance, which is important in the large gorilla habitat. We retained the pointing technique for travel since it proved to be one of the fastest and most favored techniques in our testbed. However, we also trained users extensively in the use of this technique with written and verbal instructions. A previous experiment [1] showed that users can more easily maintain spatial orientation (an important

requirement of this application) when they are aware of certain strategies, such as flying above the scene or moving through walls.

We performed a usability study with a second set of architecture students. Just as we did in the first version, we had the subjects answer questions and provide subjective ratings for their experiences. Both alterations to the application proved beneficial. Direct manipulation of objects with the Go-Go technique had been rated at 3.14 on a five-point scale and was the lowest rated of nine features in the initial implementation. After the change to HOMER, users ranked this feature the fourth most usable with a rating of 4.00. The pointing technique was rated 3.71 and eighth most usable in the initial system, but the addition of training raised its rating to 4.10, and its rank to second.

Though these results are subjective, they indicate that the use of our methodology, in particular testbed evaluation, produces measurable usability gains in a real-world VE application.

## 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, since 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. Since 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.

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.

## 8. ACKNOWLEDGMENTS

The authors would like to thank Don Allison, Jean Wineman, and Brian Wills for their work on the VR Gorilla Exhibit, and the VE group at Georgia Tech for their comments and support. Portions of this research were supported by a National Science Foundation Research Experiences for Undergraduates grant.

## 9. REFERENCES

- [1] Bowman, D., Davis, E., Badre, A., and Hodges, L. Maintaining Spatial Orientation during Travel in an Immersive Virtual Environment. *Presence: Teleoperators and Virtual Environments*, 8(6), 618-631, 1999.
- [2] Bowman, D. and Hodges, L. Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. *The Journal of Visual Languages and Computing*, 10(1), 37-53, 1999.
- [3] Bowman, D., Hodges, L., and Bolter, J. The Virtual Venue: User-Computer Interaction in Information-Rich Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 7(5), 478-493, 1998.
- [4] Bowman, D., Wineman, J., Hodges, L., and Allison, D. Designing Animal Habitats Within an Immersive VE. *IEEE Computer Graphics and Applications*, 18(5), 9-13, 1998.
- [5] Darken, R. and Sibert, J. Wayfinding Behaviors and Strategies in Large Virtual Worlds. *Proceedings of CHI*, 142-149, 1996.
- [6] Herndon, K., van Dam, A., and Gleicher, M. The Challenges of 3D Interaction. *SIGCHI Bulletin*, 26(4), October, 36-43, 1994.
- [7] Hix, D., Swan, J., Gabbard, J., McGee, M., Durbin, J., and King, T. User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment. *Proceedings of IEEE Virtual Reality*, 96-103, 1999.
- [8] Kaur, K. Designing Virtual Environments for Usability. Doctoral Dissertation, University College London., 1998
- [9] Kennedy, R., Lane, N., Berbaum, K., and Lienthal, M. A Simulator Sickness Questionnaire (SSQ): A New Method for Quantifying Simulator Sickness. *International Journal of Aviation Psychology*, 3(3), 203-220, 1993.
- [10] Koller, D., Mine, M., and Hudson, S. Head-Trackled Orbital Viewing: An Interaction Technique for Immersive Virtual Environments. *Proceedings of the ACM Symposium on User Interface Software and Technology*, 81-82, 1996.
- [11] Lampton, D., Knerr, B., Goldberg, S., Bliss, J., Moshell, J., and Blau, B. The Virtual Environment Performance Assessment Battery (VEPAB): Development and Evaluation. *Presence: Teleoperators and Virtual Environments*, 3(2), 145-157, 1994.
- [12] Pierce, J., Forsberg, A., Conway, M., Hong, S., Zeleznik, R., and Mine, M. Image Plane Interaction Techniques in 3D Immersive Environments. *Proceedings of the ACM Symposium on Interactive 3D Graphics*, 39-44, 1997.
- [13] Poupyrev, I., Billingham, M., Weghorst, S., and Ichikawa, T. 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, 1996.
- [14] Poupyrev, I., Weghorst, S., Billingham, M., and Ichikawa, T. A Framework and Testbed for Studying Manipulation Techniques for Immersive VR. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 21-28, 1997.
- [15] Ware, C. and Osborne, S. Exploration and Virtual Camera Control in Virtual Three Dimensional Environments. *Proceedings of the ACM Symposium on Interactive 3D Graphics*, in *Computer Graphics*, 24(2), 175-183, 1990.
- [16] Witmer, B. and Singer, M. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*. 7(3), 225-240, 1998.