A Methodology for the Evaluation of Travel Techniques for Immersive Virtual Environments

Doug A. Bowman, David Koller, & Larry F. Hodges

College of Computing

Graphics, Visualization, and Usability Center

801 Atlantic Drive

Atlanta, GA 30332-0280

(404) 894-8787

{bowman, koller, hodges}@cc.gatech.edu

ABSTRACT

We present a framework for the analysis and evaluation of *travel*, or viewpoint motion control, techniques for use in immersive virtual environments (VEs). The basic construct of this framework is a taxonomy of travel techniques, and we present a summary of three experiments mapping parts of the taxonomy to various performance measures. The complete framework, and the associated system called SVELTE, provide the experimenter with powerful built-in constructs and flexibility. Experiments within the SVELTE system can analyze not only the effects of different travel techniques, but also the effects of many outside factors simultaneously, and multiple performance measurements can be collected. This leads to a deeper understanding of the interaction and the technique(s) in question, as well as to broadly generalizable results. We also present an example experiment within this framework, which measures the user's ability to gather information while traveling through a virtual environment. Results indicate that, of the variables tested, the complexity of the environment is by far the most important factor.

INTRODUCTION

Human-computer interaction in three dimensions is not well understood (Herndon et al., 1994). In particular, little progress has been made in the comprehension and analysis of interaction within immersive virtual environments (VEs). In our research, we have been attempting to understand one of the most basic and universal interactions found in VE applications: *travel*. We define travel as the control of the user's viewpoint motion in the three-dimensional environment. This is distinguished from *wayfinding*, which is the cognitive process of determining a path based on visual cues, knowledge of the environment, and aids such as maps or compasses. Together, travel and wayfinding make up the overall interaction called *navigation*. In our work, then, we are studying the techniques which allow a user to move from place to place in a VE, and not the displays or other aids which help the user to find her way.

Travel is almost certainly the most common interaction in VE applications, apart from simple head motion. In most VE applications, the user must be able to move effectively about the environment in order to obtain different views of the scene and to establish a sense of presence within the 3D space. Therefore, it is essential that travel techniques be well-designed and well-understood if VE applications are to succeed. In most cases, travel is not an end unto itself; rather, it is simply used to move the user into a position where he can perform some other, more important task. Because of this, the travel technique should be easy to use, cognitively simple, and unobtrusive. It is not obvious at first glance whether a given technique meets these criteria, so formal evaluation and analysis are important.

Experiments to evaluate travel techniques formally, however, are difficult to design and perform. Custom software must be built, including the techniques themselves, the environments to be used, data collection routines, and experimental session management code. Moreover, the generalizability of the results can be questionable if techniques are compared for a single task and a single environment only. It is important to consider task, environment, and other outside characteristics when forming conclusions about the techniques in question. Thus, we are developing a system called SVELTE (Standard Virtual Environment Locomotion Technique Experiments) which allows formal experimentation with a relatively large number of independent and dependent variables. As we will describe, SVELTE takes care of many of the low-level tasks necessary to implement an experiment, allowing rapid prototyping and development for those interested in evaluating travel techniques.

Before discussing the system itself, we will mention some related work in this area. We will also present our formalized framework within which design and evaluation of travel techniques may be performed. After describing SVELTE, we will conclude by discussing an example experiment run within this framework.

RELATED WORK

A number of researchers have addressed issues related to navigation and travel both in immersive virtual environments and in general 3D computer interaction tasks. It has been asserted (Herndon, van Dam, & Gleicher, 1994) that studying and understanding human navigation and motion control (e.g. Schieser, 1986, Warren & Wertheim, 1990) is of great importance for understanding how to build effective virtual environment travel interfaces. Although we do not directly address the cognitive issues surrounding virtual environment navigation, this area has been the subject of some prior investigation and discussion (e.g. Wickens, 1995). Wayfinding issues have been the subject of studies by Darken and Sibert (1996a, 1996b). The use of maps, breadcrumbs, and landmarks were evaluated as tools for finding a path through a virtual environment. Also, a system has been proposed (Ingram & Benford, 1995) which attempts to replicate the classic urban wayfinding cues identified in "The Image of the City" (Lynch, 1960).

Various metaphors for viewpoint motion and control in 3D environments have also been proposed. Ware et al. (1988, 1990) identify the "flying," "eyeball-in-hand," and "scene-in-hand" metaphors for virtual camera control. As an extension of the scene-in-hand metaphor, Pausch et al. (1995) make use of a "World-in-Miniature" representation as a device for navigation and locomotion in immersive virtual environments.

Numerous implementations and studies of non-immersive 3D travel techniques have been described. Strommen compares three different mouse-based interfaces for children to control point-of-view navigation (Strommen, 1994). Mackinlay et al. describe a general method for rapid, controlled movement through a 3D environment (Mackinlay, Card, & Robertson, 1990).

Ware and Slipp assessed the usability of different velocity control interfaces for viewpoint control in 3D graphical environments (Ware & Slipp, 1991).

Mine (1995) offers an overview of motion specification interaction techniques. He and others (Robinett & Holloway, 1992) also discuss issues concerning their implementation in immersive virtual environments. Several user studies concerning immersive travel techniques have been reported in the literature, such as those comparing different travel modes and metaphors for specific virtual environment applications (e.g. Chung, 1992, Mercurio et al., 1990). Physical motion techniques have also been studied (e.g. Iwata & Fujii, 1996), including an evaluation of the effect of a physical walking technique on the sense of presence (Slater, Usoh, & Steed, 1995).

DESIGN AND EVALUATION FRAMEWORK

Given techniques for travel in immersive virtual environments, one could perform many experiments involving those techniques and come to some understanding of their effect on performance in certain applications. However, it is not entirely clear what determines the "performance" of a travel technique. Moreover, it would be difficult or impossible to determine which components of the techniques were significant in improving or lessening performance, and results from one application or task would not necessarily transfer to another. For this reason, we have devised a more formalized framework within which to evaluate virtual travel techniques. Stanney (1995) proposes that a taxonomy of interaction techniques is needed for "imposing order on the complex interactions between user, task, and system phenomena." The evaluation framework presented here includes such a taxonomy and an emphasis on outside factors which can influence user performance.

Taxonomy

In order to understand travel techniques and their effects more deeply, we need to categorize them and break them down into their lower-level components. Toward this end, we have developed a taxonomy of immersive travel techniques, which is presented in Figure 1. The taxonomy splits a technique into three components.

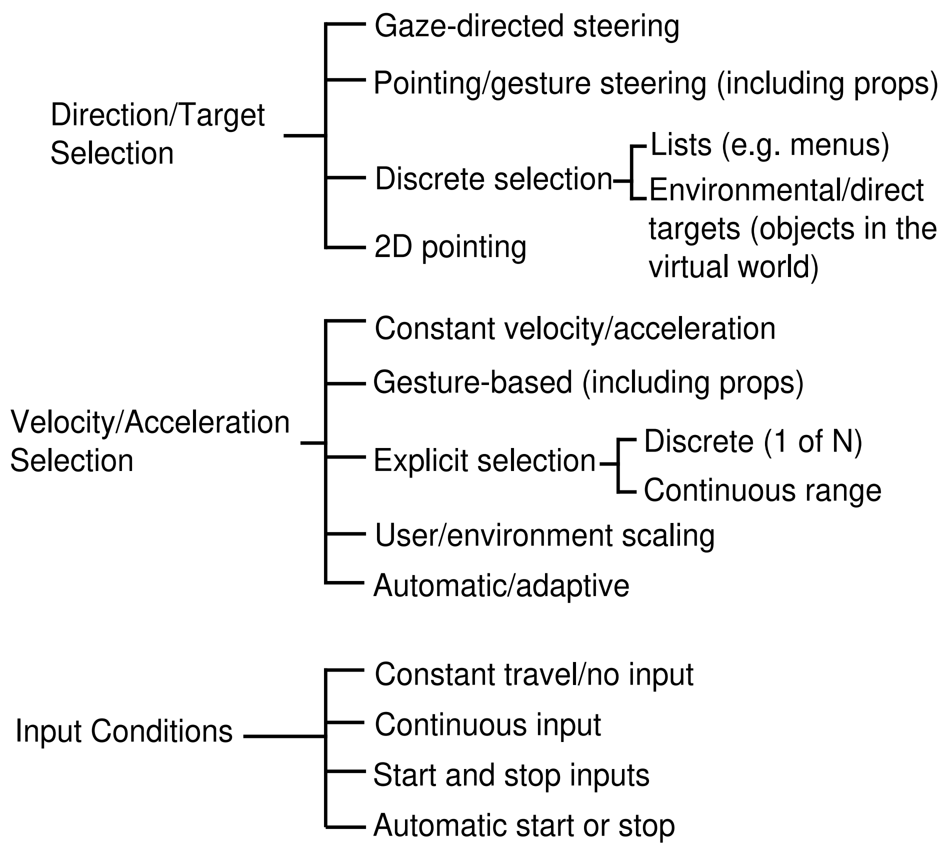

Figure 1. Taxonomy of travel techniques for immersive virtual environments.

Direction/Target Selection refers to the method by which the direction or object of travel is specified. Depending on whether control of direction is continuous or not, the user may either "steer" (choose a direction), or simply choose a target object. Gaze-directed steering, in which the user moves in the direction she is looking, and pointing, where the user points in the direction she wants to go, are two popular steering techniques. This section also lists techniques for discrete selection of a target object.

Velocity/Acceleration Selection techniques allow the user to vary the speed of travel. Many VE applications dispense with this entirely, and use a constant travel velocity. However, several techniques have been proposed, including continuous gestures to select velocity, the use of props such as foot pedals, or adaptive system-controlled speed.

The final component of a travel technique is the *Conditions of Input*. This refers to the input required by the system in order to begin, continue, and end travel. The user may be in constant motion, in which case no input may be required. Alternately, the system may require continuous input to determine the user's state, or simple inputs at the beginning and/or end of a movement. Again, this component may be under system control.

We do not claim that this taxonomy is complete, since many new techniques for controlling user motion are being designed. However, most current techniques fit into the taxonomy, at least at a high level. More importantly, by breaking a technique into three components, we can study them separately, and gain a greater understanding of differences in performance. A technique which is performing poorly may be improved by changing only one of the components, but this might not be recognized unless techniques are divided into their constituent elements.

This taxonomy also encourages the design of new techniques. By choosing a component (and an implementation of that component) from each section of the taxonomy, a travel technique may be created from its parts, and useful new combinations may come to light. Not all components will work with all others, because the branches of the taxonomy are not entirely orthogonal, but there are many opportunities for interesting designs.

For example, one might combine environmental target selection with gesture-based velocity selection, explicit start inputs, and explicit or automatic stop inputs. This would produce a technique that would allow a user to travel along a path from the current position to a specified object, using a high velocity on the less interesting parts and a slower speed at places of interest. The user could stop moving at any point along the path, or be stopped automatically when the target object was reached. Such a technique might be a natural fit for an immersive “tour” application, where there are certain known places that users wish to visit, but designers also wish that movement be under some degree of user control.

Quality Factors

There are few categories of virtual environment applications that are currently in use for productive, consistent work, but the requirements of these applications for travel techniques cover a wide range. Further, there are many new applications of VEs being researched, which also may require travel techniques with different characteristics. It is therefore impractical to evaluate travel techniques directly within each new application. Instead, we propose a more general methodology, involving an indirect mapping from travel techniques to a set of *quality factors*. Quality factors are measurable characteristics of the performance of a technique. With this indirect mapping, application designers can specify desired levels of various quality factors, and then choose a technique which best fits those requirements.

Our current list of quality factors for immersive travel techniques includes:

1. *Speed* (appropriate velocity)
2. *Accuracy* (proximity to the desired target)
3. *Spatial Awareness* (the user's knowledge of his position and orientation within the environment during and after travel)
4. *Ease of Learning* (the ability of a novice user to use the technique)
5. *Ease of Use* (the complexity or cognitive load of the technique from the user's point of view)
6. *Information Gathering* (the user's ability to actively obtain information from the environment during travel)
7. *Presence* (the user's sense of immersion or "being within" the environment due to travel)
8. *User Comfort* (lack of simulator sickness, dizziness, or nausea)

Again, this list may not be complete, but it is a good starting point for quantifying the effectiveness and performance of virtual travel techniques. In particular, we emphasize that speed and accuracy are not the only characteristics of a good travel technique, and in many applications may not be the most important. For example, the designer of an architectural walkthrough application might be most interested in high levels of spatial awareness, information gathering, and presence. By doing experiments within this framework that relate travel technique components to quality factors, we can identify techniques that meet those needs, and the results of the experiments will also be generalizable and reusable by designers of other applications.

Initial Experiments

Using this framework, we designed and ran three initial experiments on common VE travel techniques (Bowman, Koller, & Hodges, 1997). We wanted to show that generalizable results could be obtained without knowing the target application. These experiments produced useful data which is applicable in a variety of situations.

Spatial Awareness Experiment

Our first experiment focused on one of the more abstract quality factors: spatial awareness. We were interested in how immersive travel techniques would affect the user's knowledge of the three-dimensional environment around him. Specifically, we tested how various velocity and acceleration schemes altered the user's level of spatial awareness.

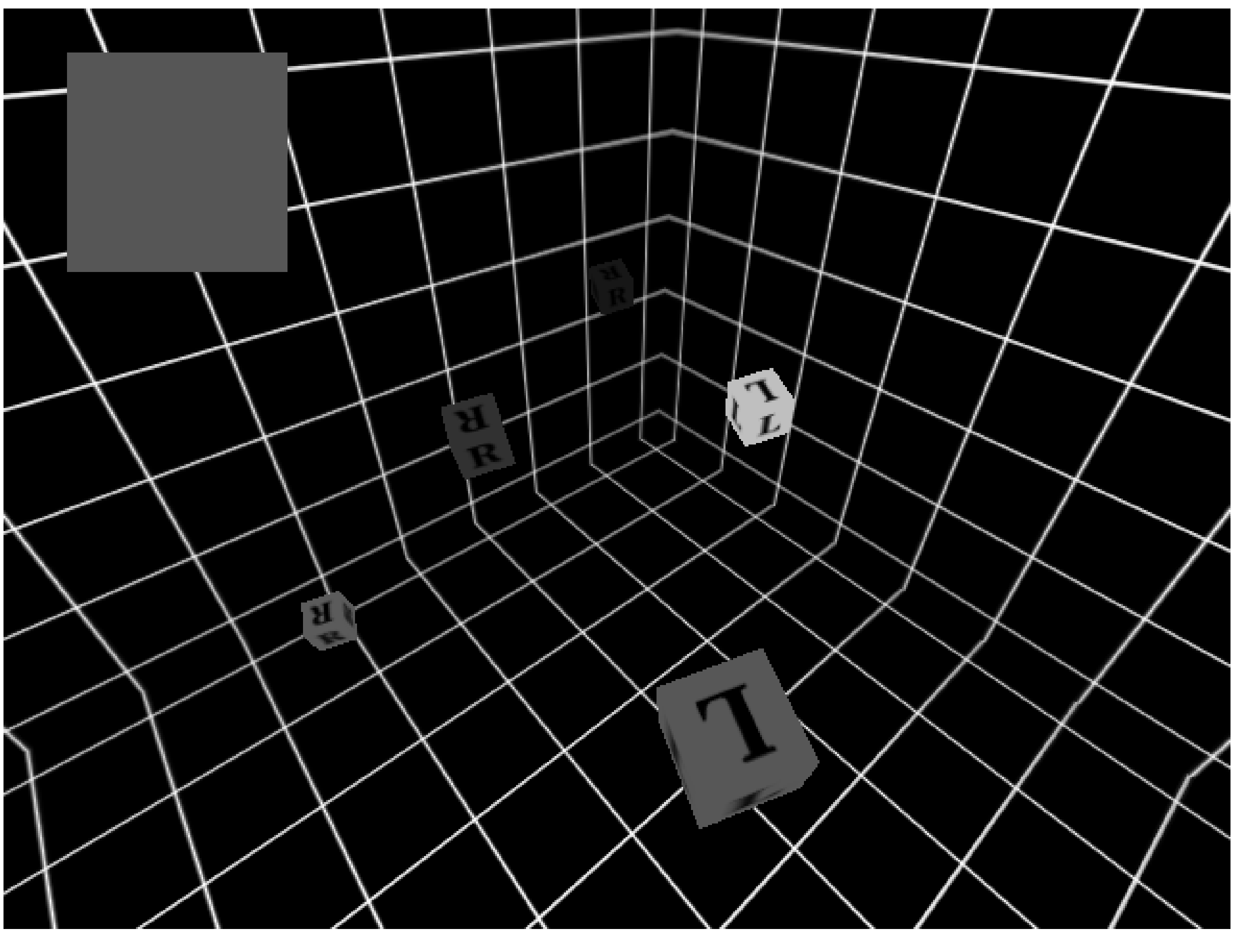

Figure 2. Environment for the Spatial Awareness Experiment. The stimulus is in the upper left corner.

The virtual environment for this experiment consisted of a set of cubes of contrasting colors, as seen in Figure 2. Users learned the locations of the cubes within the space, from both stationary and moving positions. In an experimental trial, the user was taken from the starting location to a new location, then shown a colored stimulus, matching the color of one of the cubes. We measured the user's spatial awareness by the time required to find the cube of that color. The subject proved she had found the correct cube by pressing either the left or right mouse button depending on the letter ("L" or "R") printed on the cube.

We contrasted four different velocity/acceleration techniques, each of which was system-controlled. The first two techniques used a constant velocity, one quite slow, the other relatively fast. We also implemented and tested a "slow-in, slow-out" technique, in which travel starts and ends slowly, with acceleration and deceleration in between. Finally, we tested an infinite velocity (also called "jumping" or "teleportation") technique, where users are taken immediately to the target location.

The results of the experiment showed that the level of spatial awareness was significantly decreased with the use of a jumping technique ($p < 0.01$). In fact, users were generally reduced to a simple search of the space after jumping from one location to another. This is a significant result, since many application designers might be tempted to use teleportation because of its speed and accuracy. The experiment shows that this is unwise unless some degree of user disorientation is acceptable in the target application. Surprisingly, none of the other techniques showed significant differences in performance: even up to relatively large velocities, users could maintain spatial awareness.

Absolute Motion Experiment

In the second experiment, we wanted to obtain some basic information about the speed and accuracy of two common steering techniques: gaze-directed steering, in which the direction of motion is determined by the user's gaze, and pointing, in which the user's hand orientation determines the direction of travel. Even though speed and accuracy are not always the most important considerations in a travel technique, they are still widely desirable. Once a target has been chosen, it is usually unacceptable to the user to move there slowly or imprecisely. We chose to compare gaze-directed steering with pointing because they are commonly used in many VE applications, and because they seem to be quite different in their focus: gaze-directed steering is simple but constraining, while pointing is expressive but also more complex.

The experimental task was quite simple. Users traveled using one of the techniques from a starting location to a target sphere. We varied the size of the sphere and the distance to the sphere. We hypothesized that gaze-directed steering might produce greater speed and accuracy than pointing, because of its simplicity and the relative stability of the head compared to the hand.

Although gaze-directed steering did produce slightly better times for this task, we found that there was no statistically significant difference between the two techniques. Users were able to travel very close to the optimal straight-line path between the starting and target locations whether gaze-directed steering or pointing was used. This was surprising, but also useful, information given the advantages of pointing shown by our next experiment.

Relative Motion Experiment

Rather than moving directly to an object in the environment, in this experiment users were required to move to a point relative to an object in the 3D space. This task is commonly used in applications such as architectural walkthrough. For example, suppose the user wishes to obtain a head-on view of a building so that it fills his field of view. There is no specific target object; rather, the user is moving relative to the building. In this experiment, the target was located on a line defined by a three-dimensional pointer, at a known distance from the tip of the pointer. Figure 3 shows the pointer and the target, although the target was not visible during experimental trials.

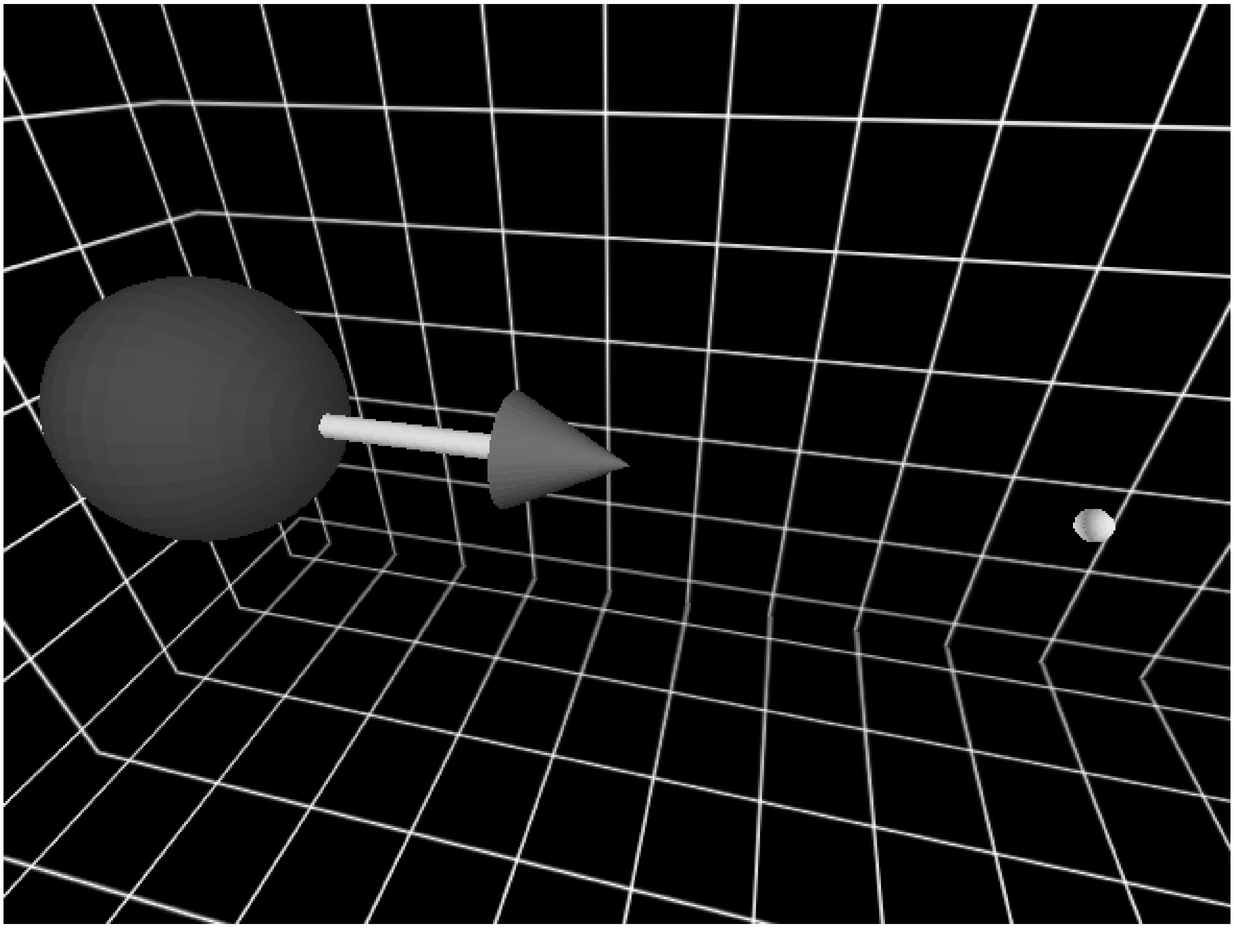

Figure 3. The Relative Motion Experiment Environment.

Again, we measured speed and accuracy for the gaze-directed steering and pointing techniques. With this task, however, we highlighted the main difference between the two techniques: that gaze-directed steering requires the user to be looking in the direction of motion, while pointing allows gaze and travel to be in different directions. Thus, users of the pointing technique could look at the pointer to judge their travel to the target location, while gaze-directed steering required users to look at the pointer, then look in the estimated target direction to travel, then look back to check their progress, and so on.

Indeed, the experiment showed that the pointing technique was significantly faster for the relative motion task ($p < 0.025$). When combined with results from the absolute motion experiment, we can conclude that pointing is a good general-purpose technique where speed and accuracy are important quality factors.

Expanded Framework

Although our initial set of experiments produced significant results in evaluations of some common VE travel techniques, we also noted that we were not able to capture a complete picture of the techniques from simple experimental designs. The problem was that our experiments studied the effects of a single factor only (travel technique), and did not consider other factors that might have an important effect on performance.

This is illustrated well by the absolute and relative motion experiments. Though they tested the same techniques (gaze-directed steering and pointing) and measured the same quality factors (speed and accuracy), they produced quite different results. In the case of absolute motion, the two techniques performed equally, but for a relative motion task, pointing showed more speed and accuracy. If we had studied these two tasks in the same experiment, we would have found that there was an interaction between technique and task. This illustrates the fact that a technique cannot in general be considered in isolation from the task for which it is to be used.

Similarly, characteristics of the environment may affect the performance of a travel technique. Consider the absolute motion experiment. In the environment that we used, there was only a single object (the target), visible at all times, with a straight-line path between it and the user. In this environment, gaze-directed steering and pointing produced the same results. However, if the environment had been full of distractor objects and obstacles that the user had to avoid to reach the target, the two techniques might have exhibited significantly different performance characteristics. Again, the technique cannot be considered in isolation from the environment in which it is to be used.

For these reasons, we have felt it necessary to expand our evaluation framework to include the multitude of other factors that can affect performance of virtual travel techniques. Rather than attempting to discern these dependencies in an ad hoc fashion for each experiment that is run, our expanded framework formalizes the notion that many variables contribute to the performance metrics. By explicitly including these variables in the framework, we can more easily choose what factors to control in an experimental setting, and choose values wisely for those variables which will be held constant. The expanded framework includes variables related to task, environment, user, and system characteristics.

Task Characteristics

For immersive travel, there are many factors related to the task that could conceivably affect performance of a travel technique. Some of these characteristics come directly from a consideration of the quality factors that we wish to measure, while others are simply differences between various situations. Some of the task characteristics that we consider are:

- Distance to be traveled
- Amount of curvature or number of turns in the path
- Visibility of target
- Number of degrees of freedom of motion required
- Accuracy required
- Complexity of the task; cognitive load induced on the user
- Information required of the user

For example, we could distinguish between the absolute and relative motion tasks described above using the visibility characteristic. The target is invisible in the relative motion task, meaning that other objects in the environment must be used to determine the location of the target.

Environment Characteristics

As we have noted, the environment in which the user travels can also have an effect on performance. The same task in different environments may produce strikingly different results on one or more of the quality factor measurements. We have identified characteristics such as:

- Visibility within the environment
- Number of obstacles or distractors
- Activity or motion within the environment
- Size of the environment
- Level of visual detail and fidelity

• Homogeneity (amount of variation) in the environment

• Structure

• Alignment with the standard axes

Varying one or more of these environment variables may have allowed us to see some significant differences between the gaze-directed steering and pointing techniques in the absolute motion experiment. For example, adding more distractor objects or greater activity in the environment may have caused the more cognitively simple gaze-directed steering technique to perform better.

User Characteristics

It is also important to consider the differences in users of VE applications when evaluating performance. This can be a significant factor in the performance of various techniques, because the designers of techniques often assume something implicitly about the people who will be using the techniques. Work in the field of user modeling (Kobsa & Wahlster, 1989) is quite relevant to this part of our framework. We are considering, among others, the following user characteristics:

• Age

• Gender

• Visual acuity

• Height

• Reach

• Ability to fuse stereo images

• Experience with VEs

• Experience with computers

• Technical / non-technical background

• Spatial ability

The importance of taking user characteristics into account became quite evident during a study we performed comparing various techniques for selecting and manipulating virtual objects (Bowman & Hodges, 1997). One technique (Poupyrev et al., 1996) causes the virtual arm to grow at an exponential rate as the physical arm is stretched out, allowing the user to reach objects at great distances. Our implementation of the technique set the exponent of the growth function to a constant such that the programmers could reach every object in the room. In the user study, most people liked this technique, but a few of our users had very short arms, and could not reach many of the objects at all. This caused them to become quite frustrated with this technique and to prefer other techniques that did not rely on reach.

System Characteristics

Finally, we have extended our framework to include aspects of the hardware or software used to realize the virtual environment application. It is quite possible that design decisions made by system implementors or hardware designers may affect the performance of techniques for virtual travel. However, just because these factors are not always under the control of the technique designer does not mean that they should not be considered in the design. For best performance, designers may need to create techniques which perform robustly under a wide variety of system conditions. The system characteristics we have identified include:

- Rendering technique
- Lighting model
- Frame rate
- Latency
- Display characteristics (stereo/mono, field of view, resolution, brightness, etc.)
- Collision detection
- Virtual body representation

These factors can cause differences in the usefulness of many interaction techniques. Studies on the effects of varying frame rate and latency for various tasks have been performed (Ware & Balakrishnan, 1994), but there is still much work to be done. For travel in particular, we surmised that the presence or absence of collision detection could have a large effect on a user's ability to gather information while traveling, as discussed in detail later in this paper.

Testbed Evaluation Philosophy

This greatly expanded experimental space allows us to more explicitly model the effects of various variables on the performance of the travel interaction, and to achieve more general results. However, this increased power also has a price: experiments are more difficult to plan and implement, and many more experimental trials may be needed to achieve statistical significance. Although one learns much more about the interaction and the techniques being studied with experiments such as these, it is unlikely that many researchers have the time or patience to design and run such tests for their travel techniques.

For this reason, we advocate a form of analysis that we term "testbed evaluation." A testbed is an environment in which many objects of the same type can be evaluated, even if they had not yet been proposed or created when the testbed is developed. A good analogy is a proving ground for automobiles. These are not just glorified race tracks; rather, they contain many different areas, each of which tests a different function of the car. There are cornering tests, braking tests, speed tests, and other evaluations. The performance of a car in these different areas can be looked at separately, or the car can be given a total score that represents its weighted performance in all areas. In a testbed for immersive travel techniques, then, we want to provide generality and a set of situations that test many aspects of a technique. The goal is to be able to predict the performance of a technique in almost any situation by inserting the parameters of the situation into a model of performance that was obtained by using the testbed.

To achieve this for travel techniques, we would need tasks and environments that tested all of the different external characteristics that we discussed in the previous section, which would be an extremely difficult task. It was attempted, at least in part, in the VEPAB project (Lampton et al., 1994), which provided some of the inspiration for the current work. Rather than try to determine the ultimate set of tasks and environments, then, we plan to offer an open support system which will allow experimenters to easily create tasks and environments which test the effects of many factors on any given travel techniques.

It should be noted that the testbed evaluation philosophy can be applied to the analysis of many different interaction techniques besides travel. In other work, we are carrying out a similar strategy for the evaluation of techniques for selection and manipulation of objects (Bowman & Hodges, 1997). We hope to build up performance models for each of these important interactions through the use of testbed experimentation.

THE SVELTE SYSTEM

Based on this expanded framework and philosophy of evaluation, we are proposing a software system which aids in experimental design, implementation, and analysis. This system is called SVELTE: **S**tandard **V**irtual **E**nvironment **L**ocomotion **T**echnique **E**xperiments. As its name suggests, and in keeping with the philosophy of testbed evaluation, the system will provide some common, predefined experimental situations within which any immersive travel technique can be tested and then compared with other techniques. These will be a common basis for the understanding of some of the basic characteristics of these various techniques.

However, we also recognize the fact that it is impossible to fully capture the essence of a technique's performance with only a few standardized tests, and that many other factors may determine the effectiveness of a given technique in a real-world application. Therefore, the system is flexible and open, allowing experimenters to design and implement their own experiments based on the framework presented above. In keeping with the framework, SVELTE has components that reflect our taxonomy, quality factors, and task, environment, user, and system characteristics.

The Taxonomy in SVELTE

The taxonomy of immersive travel techniques allows us to inspect and analyze the techniques in a more fine-grained manner. Similarly, in SVELTE, we wish to be able to consider each component of a technique separately. For this reason, designers will provide code and/or a description for a technique in three non-interacting parts, corresponding to the three branches of the taxonomy.

This approach allows a great deal of flexibility, and also the possibility of reuse. The system itself will have many common technique components, such as gaze-directed steering, built-in.

Designers of new technique components may also add them to the SVELTE system. In this way, it becomes a simple matter to change only a single component of a technique, and to mix and match parts that might not otherwise have been used together.

SVELTE will implement this general approach through the callback mechanism of the Simple Virtual Environment toolkit (Kessler et al., 1994). Given a description of the input conditions, the system will be able to define callbacks to begin and end the travel interaction. Within the starting and ending inputs, other callbacks can be executed each frame to select the direction, velocity, and acceleration of travel, based on the code and/or description provided by the technique's designer.

Quality Factors in SVELTE

The system also supports the notion of quality factors as measurements of the performance of a travel technique. As far as is possible in software, the system will have the ability to measure performance for quality factors within an experimental context. For example, speed is measured by the time from the beginning of travel until the assigned task is completed. For the more abstract quality factors, other measurement tools may be included to allow the experimenter to obtain a measurement which can be compared to results from other experiments. For instance, standard questionnaires for ease of use (e.g. Chin, Diehl, & Norman, 1988), presence (e.g. Slater, 1995), and simulator sickness (e.g. Kennedy et al., 1993) should be part of the system.

For some quality factors, such as spatial awareness or information gathering, the value taken as a measure for that factor may change depending on the situation. Although changing the measurement disallows comparison with previous results, it does provide more flexibility in the types of experiments that can be run. Therefore, SVELTE allows the experimenter to designate new values as the variables of interest, and to attach these values to quality factors. Furthermore, there may be quality factors that we have not considered which are important to certain classes of applications. These will also be accommodated by the system.

Outside Factors in SVELTE

Finally, SVELTE reflects our philosophy that it may be necessary to consider many factors simultaneously when evaluating the performance of an interaction technique. Therefore, it includes the idea of task, environment, user, and system characteristics that can be varied explicitly as part of an experiment. The experimenter can choose which characteristics should be varied and which levels of those characteristics should be used. For the factors which will be held constant, the experimenter can choose the constant value, or allow the system to assign a reasonable default.

Much of the difficulty in performing an experiment where many factors are varied simultaneously is found in designing the structure of experimental sessions, deciding on the number of trials and subjects that will be needed, and gathering and analyzing the large amount of data that is generated. Therefore, as much as possible, SVELTE will aid the experimenter in these tasks. The system can use basic knowledge from statistics and experimental methodology to suggest a design for the experiment and values for the number of subjects and the number of trials. For data that can be collected online, such as timing and positioning information, the system will be able to collect and organize that data, since it has knowledge of the type of experiment being run. This data can be placed into a format usable by common spreadsheet or statistical analysis packages.

System Architecture

Figure 4 shows a diagram representing our vision for the system architecture of SVELTE. It aims to provide the experimenter with complete access to design decisions, while also providing assistance in key areas, and using the framework to guide the experiment. It is our claim that such a system will provide:

• rapid implementation and prototyping of experiments,

• reuse of code and other components,

• flexibility that will accommodate new techniques, measurements, and outside factors,

• results which can be compared across experiments, and

• results which are widely generalizable.

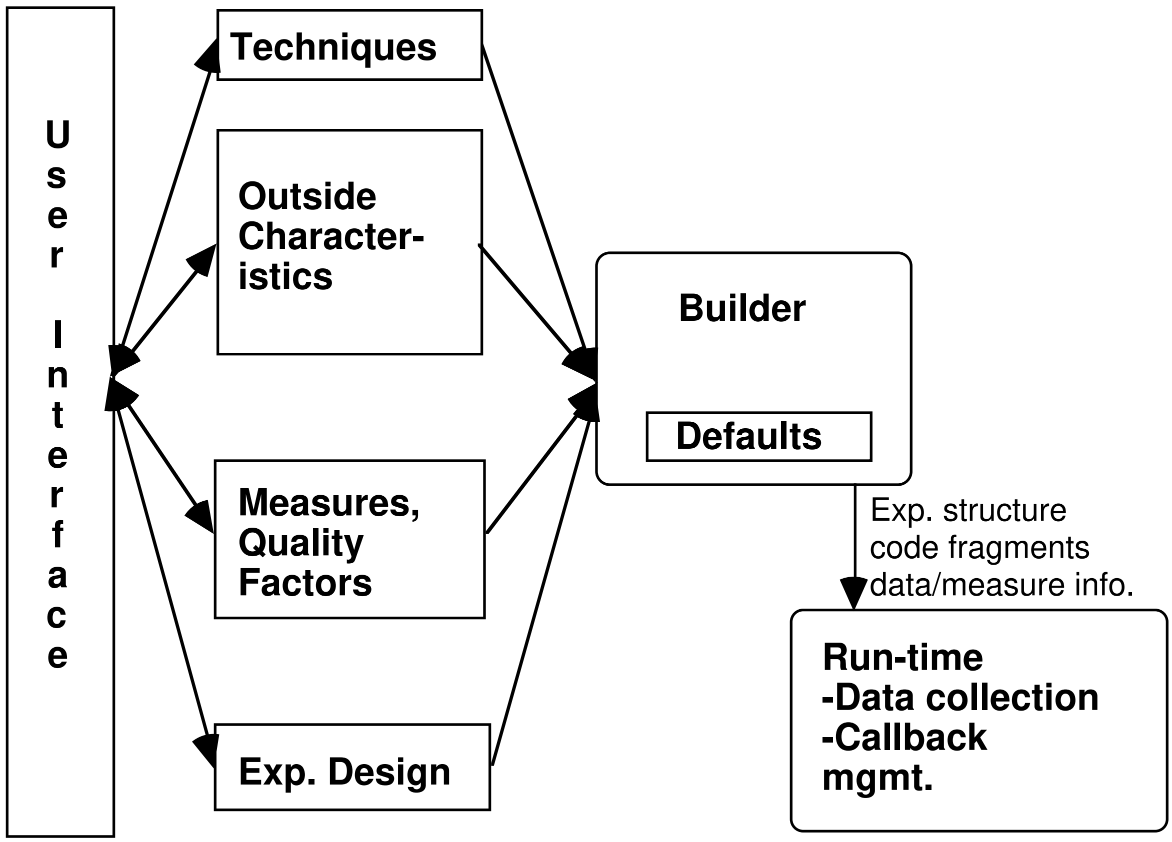

Figure 4. SVELTE System Architecture

INFORMATION GATHERING EXPERIMENT

In order to validate our evaluation methodology, and to extend the results of some of our previous experimental work in this area, we designed and ran a new experiment within our expanded framework. We hoped to isolate some important and general results, and to show the usefulness of considering a larger number of experimental variables simultaneously. Also, this experiment is one that could be designed and implemented using SVELTE.

Our focus was the effect of various steering techniques on the quality factor of information gathering. Information gathering is an important and interesting goal in many situations, and it is especially applicable to immersive virtual environments. Many of the major categories of VE applications, such as architectural walkthrough (e.g. Brooks, 1992), information visualization (e.g. Ingram & Benford, 1995, Bolter et al., 1995), simulation and training (e.g. Tate, Sibert, & King, 1997), and education (e.g. Allison et al., 1997), have a strong informational component. If the user is not able, for whatever reason, to focus on and remember important information, then the utility of the VE application is questionable.

There are many possible reasons why a user might not be able to gather as much information as is desirable, but the main factor is cognitive load. A famous result from cognitive psychology (Miller, 1956) shows the severe limitations on the capacity of working memory. When other influences force the person to use part of his working memory or other cognitive resources, information may be lost, or displaced (Baddeley, 1983). We wondered whether travel techniques induced cognitive load, and could therefore affect the amount of information that could be recalled by the user.

We chose to focus on the direction selection portion of the taxonomy, and to study gaze-directed steering and pointing techniques, as we had in two previous experiments. We also added a third technique, torso-directed steering, in which a tracker is attached to the user's torso, so that she travels in the direction her body is facing. We felt that these three represented a useful cross-section of commonly used techniques, and that there were some interesting tradeoffs among them.

For example, both pointing and torso-directed steering have the advantage that the user can look in one direction and move in another. This could be important when gathering widely scattered information. However, these techniques are also cognitively more difficult than gaze-directed steering, in which head orientation is the only thing the user must control. Torso-directed steering might be more natural (since it simulates the way we walk) and thus produce less cognitive load than pointing, but it also has the disadvantage that it can only be used to move in a horizontal plane, as the torso cannot comfortably be pointed up or down. We were quite interested to see how these tradeoffs affected a user's ability to gather information.

Looking at our expanded framework, however, we felt that there were several other factors that could influence information gathering ability while traveling through a virtual environment.

Therefore, we also chose one environment characteristic and one system characteristic to vary along with the travel technique. First, we felt that the complexity of the path through the environment might be quite important in the cognitive load induced upon a user. We captured this complexity characteristic in the *dimensionality* of the path. That is, some paths would be one-dimensional: straight and horizontal; others would be two-dimensional: still horizontal, but with turns; and still others would be three-dimensional: having turns and also vertical components.

Second, we hypothesized that the presence or absence of a collision detection feature might affect information gathering. If a user is focusing on information and not on the path he is traveling, he may move through a wall or other object. The effort required to move back through the object and back onto the desired path may use cognitive resources and displace information. With collision detection available, the user is kept near to the path, and is free to gather information without paying as much attention to the direction of motion. On the other hand, the use of collision detection may violate the mental model of the user, if the user has been told that he will keep moving as long as a button is pressed, for example. This also may induce cognitive loading. Therefore, we were interested to see how the use of collision detection would affect performance in this case.

Method

To measure the user's ability to gather information, we decided to use a memory task. Subjects traveled through corridors, using one of the three steering techniques. Corridors were used so that the user would have only a single, directed path through the environment, with no choices as to which path to take. The experiment used 1-, 2-, and 3-dimensional corridors, 3x3 meters in size, made up of straight segments, and employing only 90 degree turns. An outside view of a 3-dimensional corridor is seen in figure 5. Signs, each containing a single word, were located on the walls, ceilings, and floors of the corridors, as seen in figure 6. The words used were common, short, non-proper nouns and were randomly scattered through the corridor. Each corridor contained 12 signs. Subjects were instructed to minimize the amount of time spent in the corridor (the maximum time was 60 seconds, but a trial also ended if the subject reached the end of the corridor), and also maximize the number of words and locations of words that they could remember.

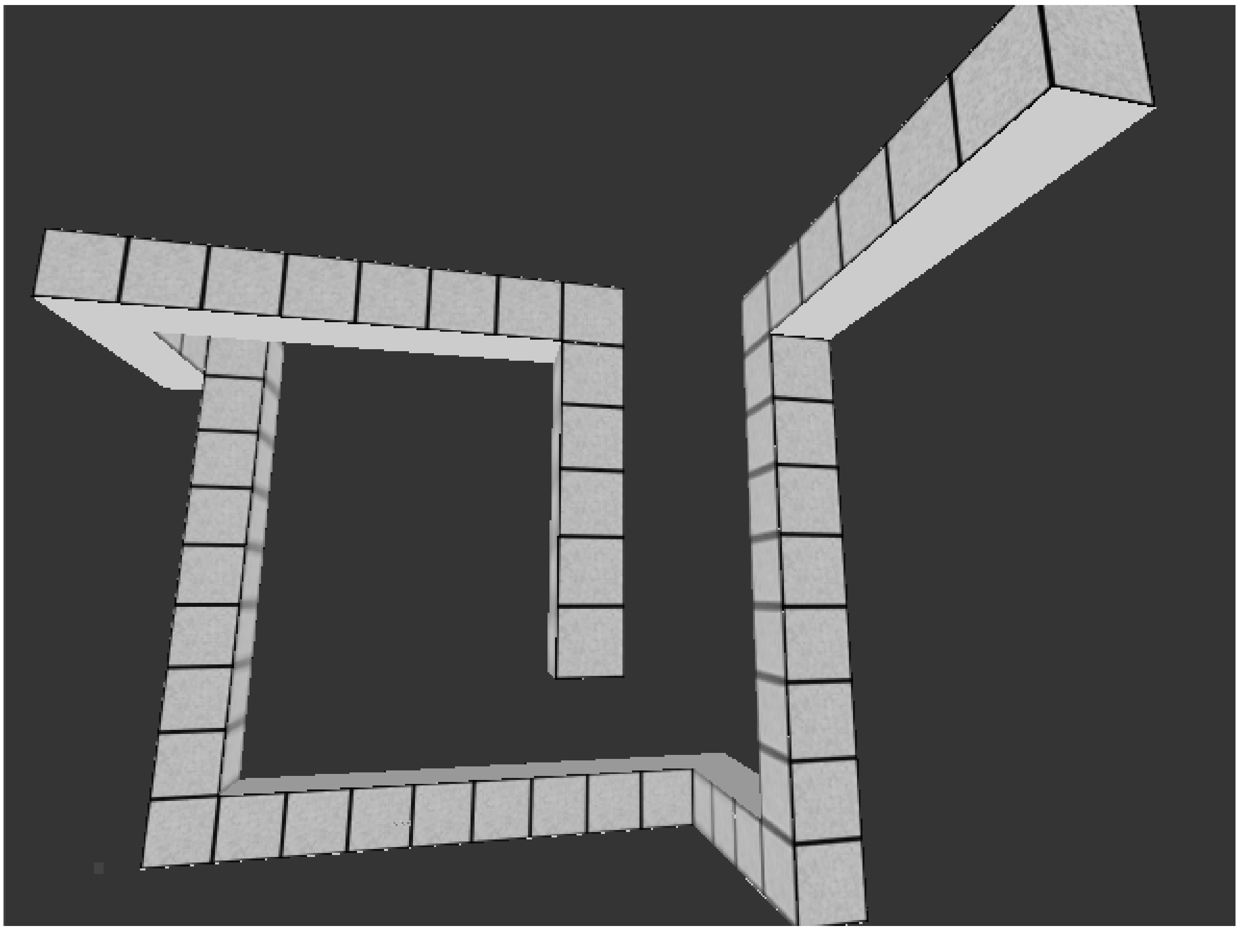

Figure 5. Outside view of a three-dimensional corridor.

Thus, we presented subjects with a very difficult, memory-overloading task. It has been shown that the limit of working memory is generally 7 plus or minus 2 chunks of information (Miller, 1956), and we were presenting 12 words and associated sign locations to the subject. Even if subjects could store both the word and location as a single chunk, and even if some words could be chunked together semantically or in some other way, the amount of information should still fill working memory. Therefore, if cognitive load is induced because of the travel technique, the dimensionality of the corridor, or the presence or absence of collision detection, we should observe that the amount of remembered information should decrease.

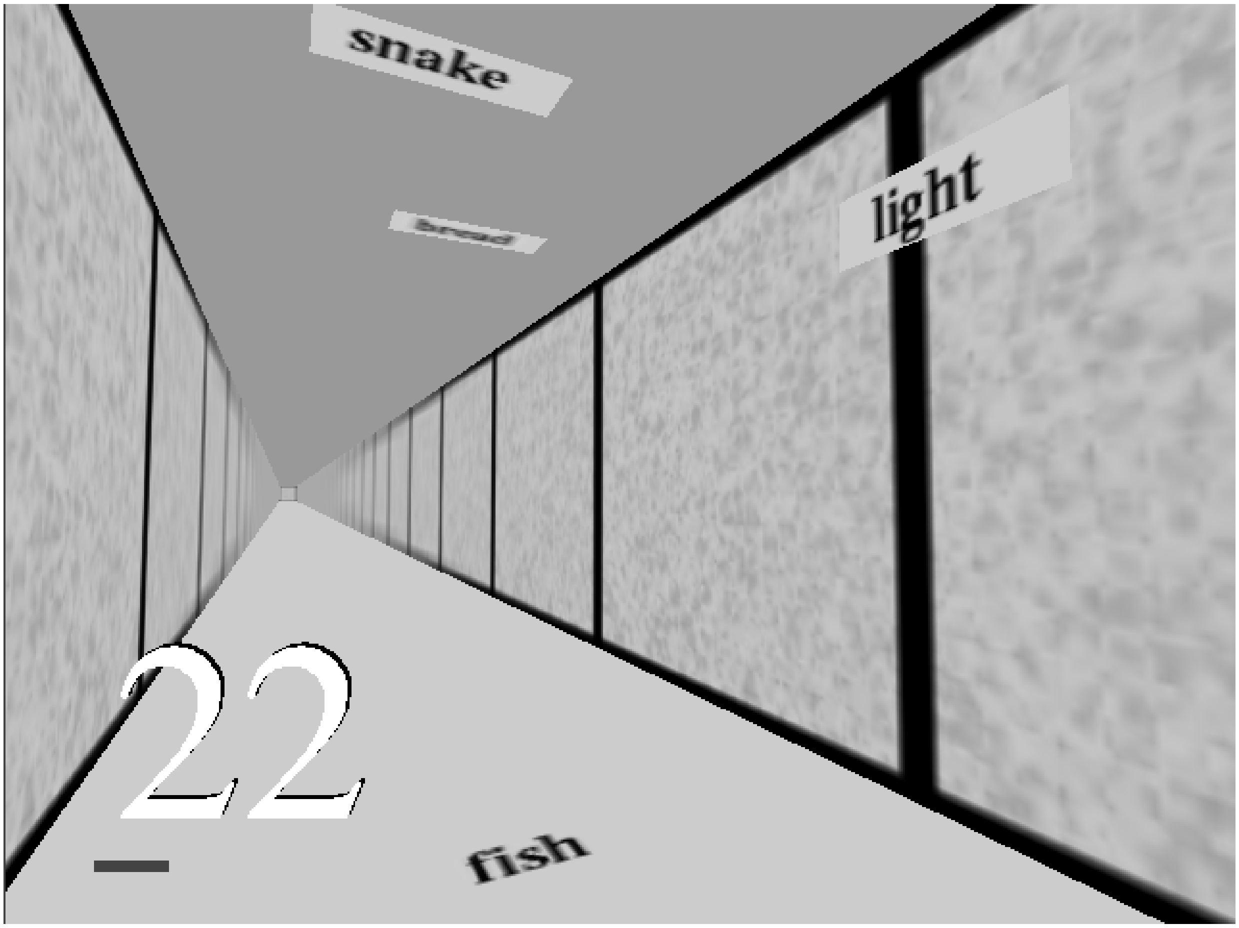

Figure 6. Interior of a corridor from the information gathering experiment.

In order to demonstrate their memory of the corridor, subjects indicated words and locations on a paper map of the corridor immediately after each trial. An example map is shown in figure 7. Subjects indicated the position of the sign along the corridor, the surface on which the sign was seen, and the word printed on the sign. If words were remembered without locations, or vice-versa, these could also be listed on the map.

For each of the steering techniques, the other two components of a complete travel technique were held constant. Velocity was 3.0 meters per second while traveling; subjects began travel by pressing and holding a button, and stopped by releasing the button. Collision detection was implemented by not allowing the user's eye point to pass through any of the walls of the corridors due to travel - if the user was moving and the eye would be at or beyond the wall in the next frame, travel was stopped.

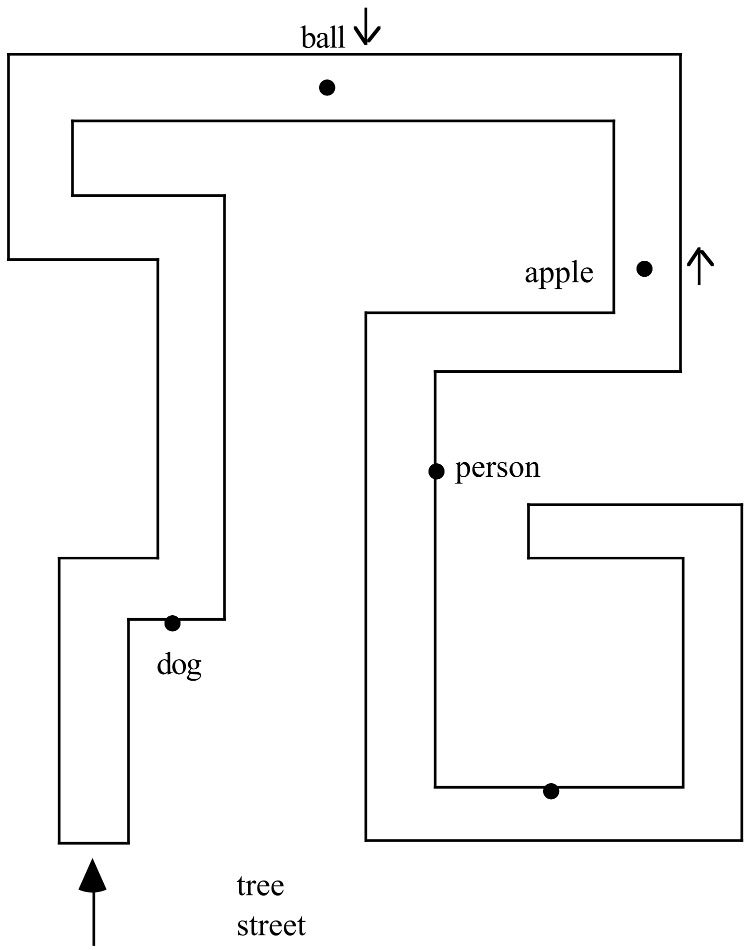

Figure 7. Example of completed corridor map with four word/location pairs, an unpaired location, and two unpaired words.

Table 1 presents a summary of the experiment. Each subject completed 16 trials: 6 each with the gaze-directed steering and pointing techniques, and 4 with the torso-directed steering technique. Within each technique, there were two trials of each dimensionality (the torso-directed technique can only be used in 1- and 2-dimensional environments), where one of the trials used collision detection and the other trial did not. Thus, each combination of the three variables (steering technique, dimensionality, and collision detection) was encountered once by each participant. Each subject traveled through each corridor exactly once, and the order of the corridors was different for each subject. To be less confusing for the subject, trials using a given technique were grouped together; however, we counterbalanced the order in which the techniques were seen. To eliminate effects of learning the techniques, subjects spent time in a "practice room" before each set of trials, where they practiced the use of the next steering technique.

	1-Dimensional		2-Dimensional		3-Dimensional	
	Collision Off	Collision On	Collision Off	Collision On	Collision Off	Collision On
Gaze-directed	X	X	X	X	X	X
Pointing	X	X	X	X	X	X
Torso-directed	X	X	X	X		

Table 1. Factor combinations considered in the information gathering experiment.

Twenty-six student volunteers (twenty-three males and three females) participated in the study. Two subject quit the experiment before completing it due to dizziness or nausea induced by the VR system. Each subject completed a pre-session questionnaire in which we gathered demographic data such as age, gender, eyesight, technical background, computer knowledge, and experience with immersive VR. Subjects wore a Virtual Research VR4 head-mounted display (HMD), and were tracked using Polhemus Fastrak or Isotrak II electromagnetic trackers. Input was given to the system with a three-button joystick. The system maintained a constant rate of 30 frames per second.

Results

The experiment measured various response variables related to the information gathering task. We measured the time spent in each corridor, the number of word/location pairs the subject got exactly right, and several variations of partially correct words and locations. Since we had instructed subjects to maximize several things simultaneously, we desired a single response variable that would encompass all of these values. The formula used for this overall score is: $1/3\ (60\text{-}t) + 3a + 2\ (b+c+d) + e + f + g$, where t=seconds spent in the corridor, a=number of word/location pairs exactly correct, b, c, and d represent responses that have two of three aspects (word, position, and surface) correct, and e, f, and g are responses where only one of the aspects are correct. This formula gives higher weight to the most correct responses, and rewards moving quickly through the corridor.

Using this metric as our response variable, we performed a 3-factor analysis of variance (ANOVA) on the data. Results were quite clear: the dimensionality of the corridor was extremely significant in affecting the score ($p < 0.01$), but travel technique and collision detection did not have a statistically significant effect. Further analysis using Duncan's test for comparison of means showed that the average score for each dimensionality was significantly different than the averages for the other two dimensionalities. Table 2 presents the average scores for each condition.

	1-Dimensional		2-Dimensional		3-Dimensional	
	Collision Off	Collision On	Collision Off	Collision On	Collision Off	Collision On
Gaze-directed	16.90	16.51	11.85	11.21	10.21	9.57
Pointing	15.57	16.68	10.36	10.85	9.33	9.38
Torso-directed	15.50	15.92	10.63	12.15		

Table 2. Average values of overall score for each tested treatment combination in the information gathering experiment. Higher scores are better.

We also performed some further analysis of the data in order to find other interesting relationships between our three independent variables and performance of the information gathering task. First, we wondered whether any learning was occurring during the trials themselves. There are two possible areas where learning could take place: the subject could learn how to use the travel technique better, and thus improve his score in later trials, or the subject could refine his strategy for performing the task in order to better maximize the overall score. We plotted learning curves for each of the orderings of techniques (necessary since the number of trials depended on the technique), and found no significant improvement in score over time for any of the orderings, implying that neither the use of the technique nor the task strategy changed much as the trials progressed.

Second, we also performed a 3-factor ANOVA with total time per trial as the response variable, in order to see which variables had an effect on the speed with which users moved through the corridors. The results here were synonymous with the previous ANOVA: dimensionality was the only significant factor ($p < 0.01$). Thus, as the dimensionality of the path increased, time spent in the corridor increased. Most subjects finished the 1-dimensional corridors quickly, while 2- and 3-dimensional corridors often took the entire 60 seconds.

Finally, we examined the demographic data collected in the questionnaire for any trends relating this information to performance of the information gathering task. Unfortunately, our user population was not as diverse as we had hoped. 23 of 26 participants were male, 24 were undergraduate students, and 24 characterized themselves as "fairly familiar" or "very familiar" with computers. However, there was a fairly even split between those who had never experienced immersive VR (16 subjects) and those who had used a VE system previously (10 subjects). Among those who completed the experiment, the more experienced participants had a slightly higher average score per trial (13.2 vs. 11.5). This is not a statistically significant result, but may show that users with even a single experience using a VE application were more focused on the task and not distracted by the technology itself or the feel of the system.

Discussion

The results of the information gathering experiment were somewhat surprising, as we had expected that different steering techniques would produce different levels of cognitive load, and thus significantly different overall scores. We found, though, that path dimensionality was the only significant variable, and that it dominated the results. However, this does not mean that we learned nothing about the nature of the travel techniques in question.

On the contrary, we noted many important characteristics of the various techniques that help us to explain the lack of significant differences from the experiment. First of all, as we noted previously in our absolute motion experiment (Bowman, Koller, & Hodges, 1997), novice users tend to emulate gaze-directed steering with pointing (by keeping their hands pointed in the direction of their gaze) unless there are large rewards for doing otherwise (as in the relative motion experiment). We saw this again in the current experiment, and also noted the same characteristic with the torso-directed steering technique. This fact quite possibly led to the lack of significant differences between the techniques. We hypothesize that users more familiar with the techniques would be able to use them more advantageously (e.g. look to the side as you move forward with the pointing or torso-directed steering techniques). Given enough expert users of the techniques, it would be interesting to include the experience level of users as another independent variable.

Also, as we stated at the beginning of this section, each technique contains certain tradeoffs. Gaze-directed steering is simple, but less expressive. Pointing and torso-directed steering allow more general travel by decoupling the gaze from the direction of motion, but also have more cognitive complexity. If both sides of tradeoffs such as these are equally important, then no significant difference between techniques will be seen. In order to further examine these tradeoffs, we would need experimental tasks that test the limits of both sides. For example, a task and environment which required looking in a direction different than the direction of motion and which induced cognitive load could produce interesting results for these three techniques.

This experiment also showed the usefulness of our evaluation framework and methodology. Before the experiment began, it was not clear what factors would lead to significant performance differences. However, because of our expanded framework, we were able to identify three different factors which we felt could be important in an information gathering task. Had we considered only travel techniques in isolation, this experiment might not have revealed any significant results. Because we varied several factors, however, we were able to identify a characteristic with an extremely significant effect on performance.

We found no statistically significant information about the effects of the use of the collision detection feature. However, several subjects did comment to the experimenter that they felt that it was easier to move through the space and perform the task when this feature was enabled. This in itself should encourage designers to include this characteristic in their systems. On some trials, though, the use of collision detection turned out to be detrimental because of an implementation issue.

Collisions in VEs without haptic devices cannot occur as they do in the real world, where we get force feedback and are physically constrained from moving through the object. In a virtual environment, we can stop the user's virtual travel at the wall, but we cannot prevent him from then moving his head through the virtual wall without violating the rule that physical head motion should result in the appropriate scene change in the VE. When this happens, the collision detection system is of course activated, and the user is prevented from moving again until his head moves back to the correct side of the wall. This is a simple implementation concept, but is non-obvious until it is experienced.

Finally, we observed that our subjects had several different strategies for performing the experimental task. Some focused on time, and raced through the environment as quickly as possible, memorizing a few words and locations along the way. Others were much more deliberate, stopping at each sign or cluster of signs to try to commit them to memory. Still others developed hybrid schemes. Subjects also differed in what they attempted to remember. Some consistently recalled the first 3 or 4 words and locations (the primacy effect), while others focused on the last things they saw in the corridor (the recency effect). A third group simply wrote down as many words as they could, then tried to match them to locations on the map.

All of these dissimilar strategies may have affected our ability to get significant results. We could have imposed a strategy on the user by instructing them explicitly to perform the task a particular way, and perhaps seen less variability. On the other hand, users will also have differing methods in real applications, and we should be searching for interaction techniques which perform robustly under a variety of strategies. In this sense, it is correct to allow the user flexibility in determining the most appropriate tactic for the task at hand.

CONCLUSIONS AND FUTURE WORK

In this work, we have designed and proposed a formalized methodology and framework for the evaluation of travel techniques for immersive virtual environments. These structures fulfill a need for an environment in which orderly analysis of VE interactions and associated characteristics can be performed. The SVELTE system should provide a standard software base which will allow experimentation and design within this framework, and produce generalizable and reusable results.

Our future work in this area includes of course the implementation of SVELTE, and its use in a variety of experimental settings. Also, we would like to extend the ideas of this system to encompass other important user interactions for immersive virtual environments, such as selection, manipulation, and issuing commands. Our goal is a suite of tools that allow for rapid development of experiments involving these interactions, simplified execution of the experiments themselves, and aids for analyzing the results.

We have also shown the utility of the framework and the testbed evaluation philosophy through an experiment measuring the ability to gather information while moving through a VE. We have shown that path complexity is an extremely important factor for such tasks, and have gained a deeper understanding of the characteristics of three important steering techniques.

Experiments always raise more interesting questions than they answer, and we feel that there is much more to be learned about the important task of gathering information within a virtual environment. We would like to explore other types of environments, more diverse travel techniques, including techniques for modifying velocity and acceleration, and system characteristics such as the use of a virtual body representation. The SVELTE system should allow us to examine these and other factors more quickly and completely.

ACKNOWLEDGMENTS

The authors would like to acknowledge our experimental participants for their time and effort. They also thank Neff Walker, Ben Watson, and Don Allison for their valuable advice and feedback.

REFERENCES

Allison, D., Wills, B., Hodges, L., & Wineman, J. (1997). Gorillas in the Bits. *Proceedings of the Virtual Reality Annual International Symposium*, 69-76.

Baddeley, A. (1983). Working Memory. *Philosophical Transactions of the Royal Society*

London B, *302*, 311-324.

Barfield, W., Zeltzer, D., Sheridan, T., & Slater, M. (1995). Presence and Performance Within Virtual Environments. In W. Barfield & T. Furness (Eds.), *Virtual Environments and Advanced Interface Design*. Oxford University Press.

Bolter, J., Hodges, L., Meyer. T., & Nichols, A. (1995). Integrating Perceptual and Symbolic Information in VR. *IEEE Computer Graphics and Applications*, *15*(4), 8-11.

Bowman, D. & Hodges, L. (1997). An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. *Proceedings of the Symposium on Interactive 3D Graphics*, 35-38.

Bowman, D., Koller, D., and 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.

Brooks, F. et al (1992). Final Technical Report: Walkthrough Project. Report to National Science Foundation.

Chin, J., Diehl, V., & Norman, K. (1988). Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. *Proceedings of CHI*, 213-218.

Chung, J. (1992). A Comparison of Head-tracked and Non-head-tracked Steering Modes in the Targeting of Radiotherapy Treatment Beams. *Proceedings of the Symposium on Interactive 3D Graphics*, 193-196.

Darken, R. & Sibert, J. (1996a). Navigating in Large Virtual Worlds. *The International Journal of Human-Computer Interaction*, *8*(1), 49-72.

Darken, R. & Sibert, J. (1996b). Wayfinding Strategies and Behaviors in Large Virtual Worlds. *Proceedings of CHI*, 142-149.

Darken, R. & Sibert, J. (1993). A Toolset for Navigation in Virtual Environments. *Proceedings of the Symposium on User Interface Software and Technology*, 157-165.

Herndon, K., van Dam, A., & Gleicher, M. (1994). The Challenges of 3D Interaction. *SIGCHI Bulletin*, *26*(4), 36-43.

Ingram, R. & Benford, S. (1995). Legibility Enhancement for Information Visualization. *Proceedings of Visualization '95*, 209-216.

Iwata, H. & Fujii, T. (1996). Virtual Perambulator: A Novel Interface Device for Locomotion in Virtual Environment. *Proceedings of the Virtual Reality Annual International Symposium*, 60-65.

Kobsa, A., Wahlster, W. (Eds.) (1989). *User Models in Dialog Systems.* Berlin, Springer.

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.

Kessler, D., Kooper, R., Verlinden, J, & Hodges, L.. (1994). The Simple Virtual Environment Library Version 1.4 User's Guide. Graphics, Visualization, and Usability Center Technical Report GIT-GVU-94-34.

Lampton, D., Knerr, B., Goldberg, S., Bliss, J., Moshell, J., & Blau, B. (1994). The Virtual Environment Performance Assessment Battery (VEPAB): Development and Evaluation. *Presence*, *3*(2), 145-157.

Lynch, K. (1960). *The Image of the City*. Cambridge, MA, MIT Press.

Mackinlay, J., Card, S., & Robertson, G. (1990). Rapid Controlled Movement Through a Virtual 3D Workspace. *Proceedings of SIGGRAPH*, in *Computer Graphics*, *24*(4), 171-176.

Mercurio, P., Erickson, T., Diaper, D., Gilmore, D., Cockton, G., & Shackel, B. (1990). Interactive Scientific Visualization: An Assessment of a Virtual Reality System. *Proceedings of INTERACT*, 741-745.

Miller, G. (1956). The Magical Number Seven Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, *63*(2), 81-97.

Mine, M. (1995). Virtual Environment Interaction Techniques. UNC Chapel Hill Computer Science Technical Report TR95-018.

Pausch, R., Burnette, T., Brockway, D., & Weiblen, M. (1995). Navigation and Locomotion in Virtual Worlds via Flight into Hand-held Miniatures. *Proceedings of SIGGRAPH*, In

Computer Graphics, *29*(4), 399-400.

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 Symposium on User Interface Software and Technology*, 79-80.

Robertson, G., Card, S., & Mackinlay, J. (1993). Information Visualization Using 3D Interactive Animation. *Communications of the ACM*, *36*(4), 57-71.

Robinett, W. & Holloway, R. (1992). Implementation of Flying, Scaling, and Grabbing in Virtual Worlds. *Proceedings of the Symposium on Interactive 3D Graphics*, 189-192.

Schieser, E. (1986). Principles of Navigation. In C. Belove (Ed.). *Handbook of Modern Electronics and Electrical Engineering*. Wiley.

Slater, M., Usoh, M., & Steed, A. (1995). Taking Steps: The Influence of a Walking Metaphor on Presence in Virtual Reality, *ACM Transactions on Computer-Human Interaction*, *2*(3), 201-219.

Stanney, K. (1995). Realizing the Full Potential of Virtual Reality: Human Factors Issues that Could Stand in the Way. *Proceedings of the Virtual Reality Annual International Symposium*, 28-34.

Stoakley, R., Conway, M., & Pausch, R. (1995). Virtual Reality on a WIM: Interactive Worlds in Miniature. *Proceedings of CHI*, 265-272.

Strommen, E. (1994). Children's Use of Mouse-Based Interfaces to Control Virtual Travel. *Proceedings of CHI*, 405-410.

Tate, D., Sibert, L., & King, T. (1997). Virtual Environments for Shipboard Firefighting Training. *Proceedings of the Virtual Reality Annual International Symposium*, 61-68.

Ware, C. & Balakrishnan, R. (1994). Reaching for Objects in VR Displays: Lag and Frame Rate. *ACM Transactions on Computer-Human Interaction*, *1*(4), 331-356.

Ware, C. & Jessome, D. (1988). Using the Bat: A Six-Dimensional Mouse for Object Placement. *IEEE Computer Graphics and Applications*, *8*(6), 65-70.

Ware, C. & Osborne, S. (1990). Exploration and Virtual Camera Control in Virtual Three-

Dimensional Environments. *Proceedings of the Symposium on Interactive 3D Graphics*, 175-183.

Ware, C. and Slipp, L. (1991). Using Velocity Control to Navigate 3D Graphical Environments: A Comparison of Three Interfaces. *Proceedings of the Human Factors Society 35th Annual Meeting*, 300-304.

Warren, R. & Wertheim, A. (Eds.) (1990). *Perception and Control of Self-motion*. Hillsdale, NJ, Erlbaum.

Wickens, C. & Baker, P. (1995). Cognitive Issues in Virtual Reality. In W. Barfield & T. Furness (Eds.), *Virtual Environments and Advanced Interface Design*. Oxford University Press.