In ACM Transactions on Information Systems, Vol. 11, No. 3, July 1993, Pages 239-265.

Evaluating 3D Task Performance for Fish Tank Virtual Worlds[1]

Kevin W. Arthur Kellogg S. Booth
University of British Columbia

Colin Ware
University of New Brunswick

Abstract

"Fish tank virtual reality" refers to the use of a standard graphics workstation to achieve real-time display of three-dimensional scenes using stereopsis and dynamic head-coupled perspective. Fish tank VR has a number of advantages over head-mounted immersion VR which make it more practical for many applications. After discussing the characteristics of fish tank VR, we describe a set of three experiments conducted to study the benefits of fish tank VR over a traditional workstation graphics display. These experiments tested user performance under two conditions: (a) whether or not stereoscopic display was used and (b) whether or not the perspective display was coupled dynamically to the positions of a user's eyes. Subjects using a comparison protocol consistently preferred head-coupling without stereo over stereo without head-coupling. Error rates in a tree tracing task similar to one used by Sollenberger and Milgram showed an order of magnitude improvement for head-coupled stereo over a static (non-head-coupled) display and the benefits gained by head-coupling were more significant than those gained from stereo alone. The final experiment examined two factors that are often associated with human performance in virtual worlds: the lag (or latency) in receiving and processing tracker data and the rate at which frames are updated. For the tree tracing task, lag had a larger impact on performance than did frame update rate, with lag having a multiplicative effect on response time. We discuss the relevance of these results for the display of complex three-dimensional data and highlight areas requiring further study.

CR Categories and Subject Descriptors: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism, I.3.6 [Computer Graphics]: Interaction Techniques.
General Terms: Human Factors
Additional Keywords and Phrases: Virtual Reality, Virtual Worlds, Head-Coupled Display, Stereopsis.

---

[1] This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under various operating grants and a postgraduate fellowship, by the British Columbia Advanced Systems Institute under a research fellowship and a visiting fellowship, and by the University and Province of British Columbia.

Authors' addresses: K. W. Arthur, Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599-3175, USA; email: arthur@cs.unc.edu; K. S. Booth, Department of Computer Science, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada; email: ksbooth@cs.ubc.ca; C. Ware, Faculty of Computer Science, University of New Brunswick, Fredericton, New Brunswick E3B 5A3, Canada; email: cware@unb.ca

1 Introduction

Recently, considerable interest has been shown in the area of virtual reality (VR). The idea was first introduced by Ivan Sutherland in the late 1960's when he developed a research prototype of a head-mounted display [33]. It was popularized almost two decades later when off-the-shelf systems started to become practical for virtual reality [4][14]. The underlying motivation in virtual reality is to realistically present three-dimensional virtual worlds to a user so that he or she perceives and interacts with them naturally, thus borrowing from built-in human abilities that evolved from our normal dealings with the three-dimensional world that surrounds us every day.

Current virtual worlds are still quite limited in terms of the realism they achieve. This is due in part to the resolution limits of current head-mounted displays and to certain temporal artifacts that seriously affect human perception. While much can be expected from future hardware advances, such problems are not likely to disappear for some time and hence there is a need for careful study of these effects if task performance using current and soon-to-be-available virtual worlds is going to be acceptable.

Almost all of the current VR research assumes the use of head-mounted displays. But many of the benefits of virtual worlds can be achieved using conventional display technologies if a few augmentations are made to the hardware and software [8][13][24][34]. A high-resolution workstation monitor supplemented with hardware for stereoscopic viewing and a head-tracking device for adjusting the perspective transformation to a user's eye position can simulate the appearance of stable 3D objects positioned behind or just in front of the screen. This is not the full virtual world produced by a head-mounted display, but it affords a number of practical advantages that make it an attractive alternative to the immersion VR produced by a head-mounted display. We have adopted the term "fish tank VR" to describe such systems because the experience is similar to looking at a fish tank inside which there is a virtual world.

In the next two sections we define our terminology and describe some of the advantages of fish tank VR. We then summarize important issues regarding the effective use of this type of display technology and survey previous related work. With this as background, three experiments are described in the main body of the paper. The first two experiments measured the usefulness of fish tank VR for 3D tasks often performed on graphics workstations. This was done using subjective user preference tests in the first experiment and objective task performance studies in the second experiment to compare fish tank VR with a variety of other viewing conditions. The third experiment assessed the effect of lag and frame rate on user performance for fish tank VR.

In the first two experiments, fish tank VR was found to be of significant value compared to standard display techniques, with head-coupling being more important than binocular vision. Both are superior to non-head-coupled monocular viewing. The third experiment showed that lag is more important than frame rate in determining user performance and that frame rate itself is probably not important except for the lag it produces; these results probably apply to immersion VR as well. The paper concludes with a discussion of the results and suggestions for future research.

2 Virtual Reality Terminology

Many of the terms that are used in discussing virtual reality are over-loaded. This section provides a definition of the terms as they are used in this paper, with appropriate references to the literature and to related terms.

Virtual reality is the presentation of a 3D scene of sufficient quality to evoke a perceptual response similar to a real scene. The 3D scene is often called a *virtual world* and the term "virtual reality" is commonly abbreviated “VR”, a practice we will adopt throughout this paper. The first use of the term is usually credited to Ivan Sutherland.

The number of times per second that an image or scene is presented to the eye(s) is the *refresh rate* of the display. High-speed repetition is required to produce persistence of vision. The image may or may not change each time it is presented. The *frame rate* or *update rate* is the number of times per second that an image or scene is changed. Baecker emphasized the difference between refresh rate and frame rate in his survey paper on dynamic raster graphics [2]. He observed that a double-buffered display need not have the same refresh and frame rates and that refresh rates on the order of 60 frames per second are needed to achieve persistence of vision whereas update rates of only 10 frames per second are often adequate to provide the illusion of a smoothly changing scene. One can expect similar rates to be acceptable for VR systems, although the effects of rapid head movements and lag may call for faster updates.

The delay between the time that new viewing parameters are determined by a user’s movement and the time that the virtual world is seen to change in response to the new viewing parameters is an instance of *lag* or *latency*. A low frame rate implies a certain degree of lag, but a high frame rate alone is no guarantee that lag is not present. There are many places within a VR system where lag can occur. Only some of these are in the graphics pipeline. Many others occur in the measuring or processing of the viewing parameters prior to entering the graphics pipeline.

A *monocular* display presents an image or scene to only a single eye. A *binocular* display presents a scene or image to both eyes. If the presentation has different images for each eye, each appropriate for the position of the respective eye, the display is *stereoscopic*, a term derived from the Greek word for solid or three-dimensional.

A *head-mounted* display presents images to one or both eyes through the use of small displays located on or near the head with appropriate lenses so that the images are seen as if viewing the world through glasses. A *head-coupled display* is one in which the calculation of a viewing transformation, usually including perspective, is based on the position of a user’s head. Other commonly used terms for this type of display include *head-tracked display*, used by Deering [8] and others, and *viewpoint dependent imaging*, used by Fisher [13]. The head-coupling technique has been employed in various guises since the early 1980’s.

Head-coupling assumes a fixed relationship between the position and orientation of a user’s head and the position(s) and orientation(s) of the user’s eye(s). A more accurate approach would involve tracking eye movements because the perspective transformation for an eye is actually determined by the position of the optical center of the eye, not just the position of the head. The difference between the optical center of the eye and the center of rotation of the eye makes this harder than head tracking unless a simplifying assumption is made that the eyes are looking in the general direction of the center of the screen (an assumption we make). Few practical systems track eye movements and only experimental systems track eye focus, which is required for correct depth-of-field calculations.

Immersion VR uses a head-mounted display to totally surround the user in a virtual world. When it works, it is a very effective technique. The term *presence* is sometimes used for the effect created by immersion VR.

Fish tank VR is the use of a regular workstation display to provide a less ambitious virtual world where the effect is limited to the volume of space roughly equivalent to the inside of the display monitor. Fish tank VR is also referred to by the names of the combination of techniques used to produce it, *head-coupled stereo display*. Another term that is sometimes used is *virtual holographic display*. Deering cites the earliest reference to a holographic-like display as being the Krell teaching machine in the 1956 film *Forbidden Planet*, where the character Morbius created a 3D “live” image of his daughter in a volume display much like today’s holographic displays. An even earlier example might be the fish tank virtual world that appeared in the film *The Wizard of Oz*, in which the wicked witch spies on Dorothy and her companions using a crystal ball 3D holographic display.

3 Factors Affecting Performance in Fish Tank VR

A monitor-based display cannot provide the immersion experience possible with head-mounted displays. But head mounted displays suffer from a number of difficulties that are not nearly as bothersome in a standard workstation display. Fortunately, immersion is not necessary for all applications. Only a few applications, such as entertainment or building walk-through systems, really need the full power of immersion VR [4]. There are a number of good reasons to settle for something less.

Cost is one reason. Immersion VR is still quite expensive. Many current systems use two high-performance workstations to compute images, one for each eye [3][30]. Cost will not always be a problem, however. So this alone does not rule out immersion VR.

A more serious problem with immersion VR is the effect of separating a user from the real world. This results in the loss of conventional interaction devices such as the keyboard or mouse. More importantly, it provides the sensation of being in only the computer-generated world, so a "handler" is required to be present to guard against physical injury to an immersion VR user. "See-through" head-mounted displays solve this problem to a degree, but the illusion of immersion is then diminished. With fish tank VR, a user sees both the virtual world within the display and the surrounding environment in which he or she is working. There is no need for a handler. Because of the small amount of extra equipment required, a fish tank VR workstation can be part of the office, just as the conventional workstation is part of the office.

Virtual reality may be the display technique of choice in the future, but there are some technical problems that must be solved before that happens. We discuss three types of factors that are of particular relevance to fish tank VR: spatial display artifacts, temporal display artifacts and the use of auxiliary depth cues to enhance the sense of three-dimensionality in a virtual world.

3.1 Spatial display artifacts

The *resolution* of the display is an important consideration. In immersion VR, computer generated images are viewed through wide angle optics to provide a wide field of view to the eyes. The small LCD screens that are typically used in head-mounted displays suffer from low resolution that is made worse by the fact that the optical elements must stretch the image non-linearly [19][28]. In typical head-mounted display systems the resolution is such that a pixel will subtend approximately 12 minutes of arc. Viewing a high resolution monitor with a 30 degree field of view results in 2 minutes of arc per pixel, which is close to the resolution limits of the human eye. While head-mounted display technology can be expected to improve in the coming years, it may be some time before the angular resolution of immersion VR will match that of monitor-based displays.

Depth-of-field is another factor that virtual reality systems have difficulty coping with. When our eyes fixate on an object in the real world, they converge inward and the focal length of the lenses adjusts so that the fixated object is in focus, while objects at different depths are not in focus. When viewing stereoscopic displays the eyes will converge according to a fixation point in the scene, but they must always focus on the image plane, which is at a fixed depth. Hence all objects in the scene will appear in focus, regardless of depth. This effect can be distracting and degrades the level of realism exhibited by the virtual scene.

If the system had knowledge of where the eyes were fixating it could render objects differently according to their depth, blurring objects to appear out-of-focus. It is not possible to do this correctly without directly measuring the focal length of a user's lens or the convergence of a user's eyes. One can provide approximate depth-of-field effects, however, if some assumptions are made about where a user is fixating. In fish tank VR it is natural to limit the working scene to the region near the screen surface because objects which are far away from the screen in depth will be clipped when a user's head moves. Parts of the scene that are further in the background can be drawn to simulate an out-of-focus image using a variety of blurring techniques. This technique is less useful for immersion VR because it is harder to make such an assumption about where a user is fixating.

3.2 Temporal display artifacts

Inaccuracies resulting from timing delays due to communication and processing time either in the graphics pipeline or prior to that in the computation of the viewing parameters for the virtual world must be dealt with in any VR system. The two main problems are related to the overhead required to change the image of the virtual world presented to a user.

Lag is inherent in all instances of human-machine interaction, including virtual reality and telerobotics. In telerobotics systems, the human operator is physically separated from the mechanical robot performing the task. Slow transmission lines, low bandwidth and slow mechanical linkages lead to greater lag than is typical in VR. Lag in six degree-of-freedom trackers arises from delays in transmitting data records, hardware or software delays while processing the data to perform smoothing or prediction, and additional processing time spent in the main display loop prior to sending scene description data through the graphics pipeline.

Although lag is recognized as an important factor in all VR interfaces and work has been done on techniques to compensate for it, there has been little experimental study of the perceptual and performance effects of lag in practical virtual world applications.

A low *frame rate* will not only contribute to lag, but if it is very low the image distortions resulting from head movement will no longer be smooth and the scene will appear jittery. This is a standard problem in computer graphics, usually caused by a scene complexity that exceeds the processing capacity of the graphics pipeline. What is different in VR is that the changes required in the image are a direct result of a user's natural movements while viewing and these may make the user much more sensitive to jitter than for a regular workstation.

3.3 Auxiliary 3D cues

Also contributing to a convincing illusion of three-dimensionality in the display is the use of traditional techniques from computer graphics for providing depth cues. Particular techniques include shading, both Lambertian and specular, and the use of shadows to suggest the shape and relative positions of objects. The perspective projection alone provides depth information as the extent of objects in the x and y directions is scaled with z (depth). These and other well known cues are described in most standard graphics texts [15][16] [25][26].

4 Previous Work

This section surveys related research that is directly related to the work reported here. A more general discussion of human factors issues for virtual reality can be found in the survey article by Ellis [10] and the collection of papers in [11].

4.1 Fish Tank Virtual Reality

Some of the first work on this type of display took place in the early 1980's. Fisher used a fixed monitor with head-tracking to provide different views to the observer by way of precomputed images stored on a video disc [13]. Diamond, et al., generated head-coupled perspective wire-frame images with head positions obtained by using a video camera to track a light bulb placed on a user's head [9]. Venolia and Williams describe a similar system using a Polhemus tracker and stereo shutter glasses to provide stereoscopic images with head-coupled perspective [34]. They propose a system using precomputed perspective images, and tracking only the horizontal movements of a user to minimize the number of images which must be computed.

Deering [8] discussed the technical components required to create a high quality 3D display on a monitor using head-coupled perspective and stereoscopic display. He provided a thorough discussion of the mathematics required to derive the correct viewing transformations for each eye and the need to have accurate estimates for each of the parameters that correlate the centers of the eyes with head position and orientation as measured by most head trackers. Deering takes into account a number of effects in his derivation, including the distortions due to the curvature and thickness of display screen.

McKenna [24] reported on experiments conducted with three types of monocular displays, all employing head tracking. The first was a stable high-resolution monitor with head-coupled perspective. The next was the same monitor but with a tracker attached to it in addition to the head tracker; the display changed accordingly as the monitor was tilted or swiveled as well as when the user's head moved. The third was a small hand-held LCD screen that could be freely moved and rotated; again the display depended on the screen position and the head position. McKenna reports on an experiment to evaluate performance with the stationary head-coupled high-resolution monitor system. Subjects were given a positioning task under three different perspective viewing conditions which were (a) a static view, (b) a view that changes with mouse position, and (c) a head-coupled view. Subjects performed best with the head-coupled perspective. The mouse-controlled condition generally decreased performance.

A similar type of experimental study, this time using immersion VR, was performed by Chung [6]. Subjects performed a task under various head-coupled and non-head-coupled conditions using a 6D mouse, joystick, Spaceball, or no device at all for interaction. The task was taken from an application in radiation therapy treatment. The goal was to optimally position and orient a beam in 3D so that it covers a tumor while avoiding the surrounding healthy tissue. The subjective rankings obtained showed no significant differences between the head-tracked and non-head-tracked modes, although there were differences within these two groups. The best overall rating was for a constrained head-tracked viewing mode with no additional input device, referred to as "orbital" mode.

Ware and Osborne report on the experimental study of different metaphors for navigation through virtual worlds [38]. Their experimental system used a fixed non-head-coupled monitor display and a six degree-of-freedom hand-held tracker. Although the results showed no clear winner, they suggested which metaphors may be better for different navigation tasks.

4.2 Spatial and temporal accuracy

There are two methods for assessing spatial and temporal inaccuracies in VR systems. One is by making physical measurements on the devices that are being used and thereby obtaining calibration data that can be incorporated into the VR system. The other is to measure the effect that these inaccuracies have on the performance of VR users. Both are important.

Liang, Shaw and Green [21] measured the lag inherent in the hardware processing performed by the Polhemus IsoTrak device and found a delay of approximately 110 msec (newer trackers from Polhemus have improved lag, reportedly less than 20 msec). They suggest the use of a predictive Kalman filter to compensate for lag, as do Friedmann, et al. [17]. One must be careful, however, that the prediction method does not introduce undesirable artifacts into the data. In particular, overshoot or amplification of the sensor noise can occur. These artifacts become apparent when the prediction interval is chosen to be too large. Based on their experience with Kalman filtering, Liang, Shaw and Green recommend that prediction not be performed using an interval more than three times the sampling interval of the device. For example, with a sampling rate of 20 Hz, a prediction interval of no more than 150 msec would be tolerable.

MacKenzie and Ware [22] have studied the effect of lag on a 2D Fitt's law target selection task. Analysis of response times and error rates showed that a model in which lag has a multiplicative effect on Fitts's index of difficulty accounts for 95% of the variance in the data. This is better than alternative models which propose only an additive effect for lag.

4.3 Auxiliary 3D cues

Atherton and Caporael assessed the quality of 3D images by subjective rankings to determine the degree of polygonalization required to adequately approximate a sphere and the relative effectiveness of three commonly used shading models [1]. Meyer, et al., used subjective pair-wise comparisons to assess the quality of radiosity-based global illumination models for rendering 3D scenes [23]. Further work at Cornell has examined the use of shadows to provide depth cues for interactive 3D tasks [36][37].

5 A Testbed for Studying Fish Tank VR

Hardware and software for testing fish tank VR in a realistic setting must have a number of properties that follow from a consideration of the factors that affect VR performance. This section discusses these requirements and describes a hardware and software testbed that has been implemented for fish tank VR studies. The testbed was used to conduct all of the experiments described in this paper.

5.1 Stereo display

We use the StereoGraphics Crystal Eyes stereo system connected to a Silicon Graphics IRIS workstation with the ability to display 120 frames per second. The shutter glasses are synchronized to the monitor refresh and provide an effective 60 Hz frame rate to each eye. The monitor we were using, like most color workstation monitors, suffers from a longer than ideal phosphor decay time. This results in faint ghosting of the left eye image to the right eye and vice versa. We minimized this effect by choosing colors with only a small green component because the green phosphor has the longest decay time on this type of monitor.

For our experiments, we used a 4D/240VGX workstation. Our tests with a Crimson/VGX and an Indigo Elan also gave adequate frame rates, but less powerful versions of the IRIS did not prove very satisfactory. The software also runs on various IBM RS/6000 workstations; none of the ones we use has a monitor capable of 60 Hz stereo, but this is apparently possible given an appropriate monitor and synchronization hardware [7].

5.2 Head-coupled perspective

A necessary component for rendering a 2D image of a 3D scene is the determination of the correct projection to use for transforming points and geometric primitives in 3-space into points in 2-space. Most commonly, orthographic or perspective projections are employed. Perspective projection scales the horizontal and vertical coordinates linearly with depth, and hence provides depth cueing. Conventional computer graphics display applications, in the absence of head tracking capability, employ a perspective projection according to a single viewpoint, usually translated from the center of the screen perpendicular to its surface. Non-perpendicular viewing projections have been discussed in the literature [5], but are seldom used except in flight simulators where one or more of the display surfaces is mounted at an angle to the principal direction of view [29].

In fish tank VR systems, a user's head is tracked so that by moving his or her head, a user can view the scene from different angles. This requires computing the correct off-axis perspective projection to accommodate the corresponding eye positions. The effect of head-coupled perspective is illustrated by the four screen photographs in Figure 1. In all cases the program is displaying the same 3D model of an automobile positioned at the center of the screen, level with respect to the monitor. Two different perspective projections and two corresponding camera angles are employed (resulting in four photographs). Only in the two photographs where the camera position matches the perspective projection does the object appear three-dimensional and undistorted. In the other two photographs, where the camera position does not match the perspective projection, the object appears distorted.

5.3 Spatial and temporal accuracy in head tracking

Inaccuracies in the measurement and processing of head tracking data can result in delays or noise which in turn effect the calculation of the off-axis perspective transformation. Accuracy is especially important in conjunction with stereo. If the artifacts are serious enough, head-coupling will not be effective and stereo-induced motion may occur. The scene may appear to deform and twist as the viewer moves.

We use a Shooting Star Technology ADL-1 tracker to provide head position and orientation records. No smoothing or prediction techniques are applied in hardware or in software. The ADL-1 uses mechanical linkages with potentiometers at the joints to provide fast readings. This device has a lag of less than 3 msec, which is shorter than the lag induced by other factors such as the time taken to read the tracker input buffer and to update the image. The rated absolute positional accuracy of the ADL-1 is 0.51 cm and its rated repeatability is better than 0.25 cm. Our head-coupled perspective software was calibrated once for all users and hence some residual inaccuracies in the display due to variation in eye position relative to head position will be present.

The rate of screen updates is inversely related to the complexity of the 3D scene being displayed. The maximum rate attainable when using alternate-eye stereo shutter glasses is 60 Hz. A scene of moderate complexity will easily reduce this to around 10 Hz. For our experiments comparing head-coupling and stereo, we desired as high a frame rate as possible. With careful adjustment of scene complexity we were able to achieve the 60 Hz update rate with our test scenes on a Silicon Graphics 4D/240VGX.

In one experiment, frame rate was a manipulated condition. The testbed includes a provision for this. The program's main display loop can be maintained at a frame rate of 30 frames per second by synchronizing to the monitor's refresh rate. Synchronization accuracy was verified with timing calls to the system, both during calibration tests and during the actual experiments.

When frame rate was a manipulated condition in an experiment, frame rates of 15 Hz and 10 Hz were generated by repeating the display loop within the program two or three times, respectively. Thus to produce an effective 15 frames per second using the double-buffered display system, each frame would be refreshed twice before swapping frame buffers to update the image. Frame rates of 30, 15 and 10 Hz generate frame interval times of 33.3 msec, 66.7 msec and 100 msec, respectively.

When tracker lag was a manipulated condition in an experiment, it was simulated by buffering the in-coming tracker records for a number of frame intervals. The tracker data is always requested by the program just after a new frame interval begins and the device responds within less than a frame time. Thus the data does not affect the current frame, but does affect the next frame. We treated the time half-way into the frame interval as the time at which a user perceives the new frame. The user does not effectively see the tracker update until approximately 1.5 frame intervals after the tracker record is requested. Hence a frame rate of 30 Hz produces a frame interval of 33.3 msec and an inherent lag of 1.5×33.3 msec $= 50$ msec. Buffering the tracker data by one extra frame time will produce a total lag of 83.3 msec. Similarly, a frame rate of 15 Hz introduces an inherent lag of 100 msec and a lag of 166.6 msec when tracker data is buffered for one frame time. During the experiments, the actual average frame times and total lags were measured. During none of the trials were the measurements found to be more than 3 msec away from their intended values.

5.4 Auxiliary 3D cues

Because stereo and head-coupling are by no means the only factors affecting 3D fidelity, we endeavored to provide as many additional depth cues as possible within the design constraints. The following is a discussion of the qualities that were used in our experimental testbed.

The term “vection” is usually used to refer to the feeling of self movement when a large field display is moved with respect to an observer. Recent evidence indicates that the effect can be achieved with even a small field of view [18]. Howard and Heckman suggest that one of the important factors in eliciting vection is the perceived distance of a moving visual image, with images that are perceived as furthest away contributing the most. In our experiments, we desire the observer to perceive the monitor as a window into an extensive space. We created a background consisting of a random field of objects computed as though they were an infinite distance from the observer. To give the illusion of depth-of-field we chose discs with blurred edges as our background objects. The discs are not intended to be focussed on; they are intended to give a feeling of spaciousness when objects in the foreground are fixated. We call this a *vection background.*

Shading from one or more light sources is an important depth cue in 3D scenes. This was used in the first experiment where we constructed two 3D scenes to investigate subjective evaluation of viewing conditions (see Figure 3). The first scene consisted of a pink sphere with smooth shading and specular highlights, positioned above a striped ground plane. Smooth *shadows* were computed and displayed on the ground plane. The second scene consisted of a bent tube object, again with smooth shading and specular highlights. Both scenes were lit by a directional light source and an ambient source.

5.5 Physical configuration and general experimental procedure

The experiments were conducted using the workstation and testbed software described earlier in this section. The physical configuration of the equipment consisted of the monitor mounted so that the screen center was level with the subject’s eyes. The mouse and its pad were positioned comfortably for the subject and the ADL-1 head tracker was mounted above the monitor by a wooden frame strapped to the sides of the monitor (see Figure 4). Throughout the experiments the subjects always wore the stereo glasses and the head tracking equipment. This was to avoid effects related solely to the physical factors of wearing the equipment.

Five viewing conditions were used in the experiments. These are shown schematically by the diagrams in Figure 2. In the non-stereo conditions the same image was presented to both eyes. In the binocular non-stereo condition the perspective was computed for a viewpoint midway between the two eyes. In the monocular viewing condition the right eye position was used and the subject was asked to close or cover the left eye. In the fixed viewpoint condition the perspective was established as in the head-coupled binocular non-stereo condition, according to the head position at the start of the trial. Subjects were asked to move their heads around for all conditions in order to assess the value of head-coupling.

6 Experiment 1: Subjective Impression of 3D

To allow subjects to subjectively compare the relative effectiveness of head-coupling and stereopsis we implemented an experimental protocol which involves comparison of randomly selected pairs of conditions. The first experiment used this protocol to obtain subjective rankings for the five viewing conditions when looking at a stationary three-dimensional scene.

6.1 Procedure

Subjects were presented with one of the two scenes described earlier (the sphere or the tube). For each trial, two of the five viewing conditions were selected and subjects were asked to compare the two and decide which type of viewing gave them a better perception of 3D space. The subjects were told to alternate between the two conditions by pressing the space bar, until they decided which condition they preferred and then to click on a mouse button corresponding to one of the two conditions. We questioned subjects after the experiment on their opinions regarding the quality of the different viewing conditions.

6.2 Design

There were 10 pairwise comparisons of the 5 viewing conditions. These 10 comparisons were performed twice each, once for the sphere scene and once for the bent tube scene. A trial block consisted of these 20 trials in random order. Each subject was presented with two blocks (a different ordering was used for each block).

Following the comparison trials, subjects were asked for their comments in an effort to gain further insight into the results. The following set of questions was presented.

> **All of the following questions relate to the quality of the 3D spatial impression.**
> *Is head-coupling as important, more important or less important than stereo?*
> *Is the combination of head-coupling and stereo better than either alone?*
> *Is head-coupling alone worthwhile? (If you had the option would you use it?)*
> *Is stereo alone worthwhile? (If you had the option would you use it?)*
> *Is head-coupling with stereo worthwhile? (If you had the option would you use it?)*
> *Do you have other comments on these methods of displaying 3D data?*

Seven subjects were used in this experiment, four of whom were familiar with high performance graphics systems.

6.3 Results

The data resulting from comparisons with the sphere scene showed no systematic differences from the data resulting from the bent tube scene and so these data were merged. Table 1 summarizes the combined results from all subjects. Each entry in the numbered columns corresponds to a pair of viewing conditions, and the values are the percentages of trials in which the row condition was preferred over the column condition. Hence corresponding percentages across the diagonal sum to 100%. For example, the value 89% in row 4, column 2 means that condition 4 was preferred over condition 2 in 25 out of 28 possible responses (4 responses from each of 7 subjects) and the value of 11% in row 2, column 4 accounts for the other 3 responses in which condition 2 was preferred over condition 4. The most significant result apparent from the data is that head-coupling without stereo was preferred over stereo alone by a wide margin of 91% to 9% (averaging the monocular and binocular results).

The last column in Table 1 shows for each viewing condition the percentage of times it was preferred over all the trials in which that condition was present. The values in the column sum to $n/2 \times 100\% = 250\%$, where the number of viewing conditions $n = 5$ in our experiment. Head-coupled display without stereo (both monocular and binocular) was preferred somewhat more often than head-coupled display with stereo.

The responses to the questions also showed strong preference for head-coupling. All users said that they would use it for object visualization if it were available. When asked to compare the importance of head-coupling with stereo, two of the seven subjects stated that they thought stereo was more important than head-coupling. However, these same subjects preferred the head-coupling in the direct comparison task. One subject complained about the awkwardness of the apparatus and pointed out that this would be a factor in how often it would be used.

7 Experiment 2: A Graph Tracing Task

In the second experiment, we took a more rigorous approach, and measured subjects' performance on a 3D tree tracing task. Subjects were given a task very similar to one designed by Sollenberger and Milgram [32] who tested the ability of observers to perceive arterial branching in brain scan data under different viewing conditions.

Two trees consisting of straight line segments were constructed in 3-space and placed side-by-side so that a large number of the branches overlapped (see Figure 5). One leaf of one of the trees was distinguished and the subject was asked to respond as to whether the leaf was part of the left tree or part of the right tree. In each case, the leaf chosen was the one whose x coordinate was nearest the center of the screen.

In each experimental trial, the subject was presented with the scene and asked to click on the left or right mouse button depending on whether the distinguished leaf appeared to belong to the left or the right tree. The bases of the two trees were labeled on the screen with a triangle (the left tree) and a square (the right tree). The corresponding left and right mouse buttons were similarly labeled with a triangle and a square as an additional aid to help subjects remember the labeling.

The trees were recursively defined ternary trees. A trunk of 8.0 cm was drawn at the base of the tree, connected to the root node. Nodes above the root were defined recursively, with the horizontal and vertical positions of the children placed randomly relative to the parent. There were three levels of branches above the root, resulting in 27 leaves for each tree. The following recurrence relation gives a precise specification for one tree. This assumes a right-handed coordinate system with y pointing upwards.

$$X_{base} = Y_{base} = Z_{base} = 0.0$$

$$VerticalSpacing_{root} = 8.0 \text{ cm}$$
$$HorizontalSpacing_{root} = 8.0 \text{ cm}$$

$$X_{root} = X_{base}$$
$$Y_{root} = Y_{base} + VerticalSpacing_{root}$$
$$Z_{root} = Z_{base}$$

$$VerticalSpacing_{child} = 0.7 \times VerticalSpacing_{parent}$$
$$HorizontalSpacing_{child} = 0.7 \times HorizontalSpacing_{parent}$$

$$X_{child} = X_{parent} + HorizontalSpacing_{child} \times Rand()$$
$$Y_{child} = Y_{parent} + VerticalSpacing_{child} \times (1.0 + 0.25 \times Rand())$$
$$Z_{child} = Z_{parent} + HorizontalSpacing_{child} \times Rand()$$

The function $Rand()$ returns a uniform random number in the range $[-1, +1]$. The two trees constructed for each trial were displayed side-by-side separated by a distance of 1.0 cm.

The visual complexity of the trees was tested beforehand, with the goal of making the task difficult enough that depth perception was a factor, but not so difficult that a significant number of errors would be made by a typical subject. This resulted in the specific parameters that were selected.

The experiment employed the same five viewing conditions as in Experiment 1 and subjects wore the stereo glasses and head tracking equipment throughout the experiment. Ten undergraduate and graduate students, most of whom had experience with computer graphics workstations, served as subjects for the experiment. They were instructed that their error rates and response times were being recorded and that they should be most concerned with making as few errors as possible.

7.1 Design

A new pair of random trees was generated for each trial. Trials were given in groups of 22 with the viewing condition being constant within the group. The first two trials of each group were designated as additional practice trials to familiarize the subject with the condition. A trial block consisted of all 5 groups given in a random order, and the entire experiment consisted of 3 such blocks, resulting in a total of 60 trials in each of the 5 experimental conditions. A practice group of 10 trials (two in each condition) was given at the start of the experiment. A stereo test scene was presented to each subject prior to the experiment to verify the subject's ability to use stereopsis to perceive depth.

7.2 Results

The results from Experiment 2 are summarized in Table 2. The timing data shows that the head-coupled stereo condition was the fastest, but that head-coupling alone was slow. There are significant differences at the 0.05 level between conditions 3 and 5 and between conditions 4 and 5, by the Wilcoxon Matched Pairs Signed Ranks Test. The only other difference that is significant is between conditions 4 and 1.

The error data in Table 2 provide more interesting results, with errors ranging from 21.8% in the static non-stereo condition without head-coupling to 1.3% for the head-coupled stereo condition. All of the differences are significant in pairwise comparisons except for the difference between conditions 3 and 4, the two head-coupled conditions without stereo.

8 Experiment 3: Effects of Lag and Frame Rate

In a head-coupled display system the lag in the display update arises from two primary sources. The first is the delay in receiving and processing physical measurements from the tracker. The second lag comes from the delay between receiving the tracker values and updating the display. That is, the time required to compute and render the scene using a perspective projection that takes into account the latest tracker measurements. This second lag is a function of frame rate. There is usually a third component which is also present due to variations in system load. In our experiment we eliminated this as a factor by restricting network access to the workstation during the experiment.

8.1 Procedure

The Sollenberger-Milgram tree tracing task was used again for this experiment. All of the experimental trials employed head-coupling and stereo viewing (condition 5). Subjects were informed that the accuracy of their responses and their response times would be recorded. They were instructed to perform the task as quickly as they could without seriously sacrificing the accuracy of their responses. Note that this is different from the instructions given to subjects in Experiment 2, where error rate was considered most important. The reason for this change of focus is due primarily to the low level of difficulty of our task and the fact that the trials were always performed under head-coupled stereo viewing. We reasoned that measuring response times would be most relevant when dealing with the addition of temporal artifacts and that error rates would not vary significantly as presumably a large degree of depth perception can still be obtained through stereopsis and motion (even in high lag conditions where the motion is not coupled accurately with head movements).

The subjects were ten graduate students, all of whom had some prior experience using graphics workstations.

8.2 Design

Frame rates of 30 Hz, 15 Hz and 10 Hz and lags of 0, 1, 2, 3 and 4 frame intervals were used, making 15 conditions in total. Table 3 shows the total lag times resulting from these values.

Subjects were presented with 15 blocks of 22 trials, with lag and frame rate kept constant within blocks. The first two trials in each block were designated as practice trials to enable a subject to become familiar with the block's lag and frame rate. A block of 22 practice trials with moderate lag and frame rate (15 frames per second and 233.3 msec lag) was given at the start of the experiment. A stereo test scene was presented to each subject prior to the experiment to verify the subject's ability to use stereopsis to perceive depth.

8.3 Results

Figure 6 shows a plot of average response times for each of the 15 experiment conditions. The horizontal axis measures the total lag time, and the points are marked according to the different frame rates. Response time was found to be exponential with respect to total lag, and ranged from 3.14 to 4.16 seconds. On average, subjects responded incorrectly in 3.4% of the trials. The distribution of errors across conditions showed no distinguishable pattern.

An analysis was performed to compare the effect of lag over frame rate. Three models were tested using linear regression on the 15 averaged points. The models were

Model 1: $\log time = C1 + C2 \times Lag$

Viewing Condition	1	2	3	4	5	All
1 Picture		43%	4%	0%	7%	13%
2 Stereo only	57%		7%	11%	0%	19%
3 HC monocular	96%	93%		29%	61%	70%
4 HC binocular	100%	89%	71%		68%	82%
5 HC & stereo	93%	100%	39%	32%		66%

Table 1: Pairwise comparison results from Experiment 1. The five percentages in the numbered columns of each row correspond to the frequency a particular row condition was preferred over each of the other conditions. The last column shows, for each viewing condition, the frequency that it was preferred over the other conditions in all of the trials in which it was present.

Viewing Condition	Time (sec)	% Errors
1 Picture	7.50	21.8
2 Stereo only	8.09	14.7
3 HC monocular	8.66	3.7
4 HC binocular	9.12	2.7
5 HC & stereo	6.83	1.3

Table 2: Experiment 2 timing and error results

Frame Rate			Tracker Lag		
FR (Hz)	Frame time	Base Lag	# frames	Lag	Total Lag
30	33.3	50.0	0	0.0	50.0
30	33.3	50.0	1	33.3	83.3
30	33.3	50.0	2	66.6	116.6
30	33.3	50.0	3	100.0	150.0
30	33.3	50.0	4	133.3	183.3
15	66.6	100.0	0	0.0	100.0
15	66.6	100.0	1	66.6	166.6
15	66.6	100.0	2	133.3	233.3
15	66.6	100.0	3	200.0	300.0
15	66.6	100.0	4	233.3	333.3
10	100.0	150.0	0	0.0	150.0
10	100.0	150.0	1	100.0	250.0
10	100.0	150.0	2	200.0	350.0
10	100.0	150.0	3	300.0	450.0
10	100.0	150.0	4	400.0	550.0

Table 3: Experiment 3 conditions (all times are in msec)

Model 2: $\log time = C1 + C2 \times FrameInterval$

Model 3: $\log time = C1 + C2 \times Lag + C3 \times FrameInterval$

The effectiveness of the regression fit to the data can be measured from the coefficient of determination r^2, which measures the fraction of the variance which is accounted for by the regression model. The r^2 values for the three models are as follows

Model 1: $r^2 = 0.50$

Model 2: $r^2 = 0.45$

Model 3: $r^2 = 0.57$

The best fit regression line for Model 1 is plotted in Figure 6.

9 Discussion

The results of the experiment suggest a number of conclusions including evidence of subjective user preference, differences in objective task performance and definite effects due to spatial and temporal artifacts.

9.1 User preference

The strong preferences expressed in Experiment 1 by most subjects for head-coupled viewing over stereo viewing and the enthusiastic response of viewers to the concept of head-coupled display suggest that applications which involve viewing 3D scenes on graphics workstations may benefit significantly from the addition of head-tracking hardware and software. The apparent inconsistency of subjects preferring head-coupled non-stereo over head-coupled stereo is probably due to the slight ghosting which was visible in the stereo display because of slow phosphor decay.

9.2 3D task performance

Experiment 2 provides objective evidence that head-coupled stereo can aid users in the perception of a complex 3D object. Here the evidence shows that both head-coupling and stereo contribute to performance, although head-coupling helps to a much larger extent.

Given that our error rates for head-coupling were near zero it is difficult to tell the extent of the performance increase with head-coupling and stereo combined, because the error rate cannot fall much further. To obtain a better measure of the value of head-coupled stereo over head-coupling alone would require making the task more difficult, or taking a different measure of performance which is not bounded as error rate was for our task.

Overall, the error rates obtained are lower than those obtained by Sollenberger and Milgram [32], but the pattern is strikingly similar despite the differences in the stimulus trees, the viewing conditions and the experimental protocols. Both our study and the study by Sollenberger and Milgram found motion to be more important than stereo, even though their motion was simple rotation of the object whereas ours resulted from head-coupling. Both studies found combined motion and stereo to be more effective than either in isolation.

It can be argued that the improvements seen with head-coupling in the tree tracing task are not due to the head-coupling as such, but rather to the motion-induced depth [35]. Our current evidence does not counter this objection. However, it is likely that the image motion produced by dynamic head-coupled perspective is less distracting than techniques such as rocking the scene back and forth about a vertical axis, which is commonly done in commercial molecular modeling and volume visualization packages.

In the results from Experiment 3, the best response times are approximately half of the best response times for Experiment 2. This discrepancy is likely due to two factors. The subjects were given slightly different instructions for Experiment 3 because response time was more important than error rate. In Experiment 2, subjects may have been more careful to minimize errors at the expense of response time. This is supported by the fact that in Experiment 2 with head-coupled stereo the average error rate was 1.3%; in Experiment 3 this grew to 3.4% on average, with no distinguishable pattern between the low and high lag conditions.

9.3 Spatial and temporal accuracy

The results from Experiment 3 suggest some interesting conclusions regarding the effects of lag and frame rate on the performance of 3D tasks.

The three regression models provide an indication of the relative effects of lag and frame rate. The third regression model which incorporates both lag and frame rate accounts for a large portion of the variance ($r^2 = 0.57$), as expected. The fact that the model only accounts for 57% of the variance is likely due to the random nature of the selection of trees, resulting in a wide range of difficulty among trials.

Model 1, which involves lag only, accounts for more of the variance than does model 2, which involves frame rate only (for model 1, $r^2 = 0.50$ and for model 2, $r^2 = 0.45$). The values suggest that a large part of the variance is attributable to lag alone and that lag is likely a more important factor than frame rate.

Given the data describing performance fall-off as lag increases, it is useful to obtain some measure of what level of lag becomes prohibitive (for simplicity we will consider only lag here and not frame rate). Specifically we would like to know the lag value which makes performance under head-coupling worse than the performance would be without head-coupling at all. We can compare the results from Experiments 2 and 3 to obtain an approximate lag cut-off value, by finding a point on the regression line in Figure 6 where response time is the same as for the static viewing condition.

The analysis becomes more complicated because our range of response times for Experiment 3 was lower than that for Experiment 2 due to the differing instructions given to subjects, as was discussed in the previous section. In Experiment 2 we found a best case response time of 6.83 seconds, whereas in Experiment 3 under the same conditions the response time was 3.25 seconds, which is a factor of 2.10 less. The Experiment 2 average response time for static viewing was 7.50 sec. If we scale this by the same factor of 2.10, we find that it corresponds to a response time of 3.58 seconds under the conditions of Experiment 3. From the plot of the first regression model (Figure 6), this corresponds to a total lag of 210 msec. This suggests that for tasks similar to the tree tracing task, lag above 210 msec will result in worse performance, in terms of response time, than static viewing.

The error rates for Experiment 3 remained low over all conditions, averaging to 3.4%, in contrast to Experiment 2 where the number of errors rose significantly in the non-head-coupled conditions. This suggests that even in the presence of large lags and low frame rates, head-coupling provides some performance improvement. This is not surprising, however, because the effects are likely due to depth from motion; while we do not have the data to verify this, it is likely that the performance is similar to what it would be if the scene were moving independent of the user, even without head-coupled viewing.

Systems that use predictive methods such as Kalman filtering must make a compromise between the size of the prediction interval and noise artifacts that become worse as this interval increases. Introducing prediction into a system will effectively flatten out the low-lag portion of the curve in Figure 6, and hence there will be a cut-off point beyond which the lag becomes unacceptable because of the filtering artifacts.

9.4 Applications

Our results regarding the tree tracing task should be applicable to other tasks that require tracing of lines in 3-space. Sollenberger and Milgram used this task to investigate how doctors looking at 3D brain scan images could trace the paths of blood vessels. Similar tasks arise in molecular modeling, and in the domain of software engineering where we may wish to trace object dependencies between software modules represented as networks in 3D space [12][27].

10 Future work

The studies reported in this paper are part of an on-going investigation of factors affecting human performance on 3D graphics workstations. A number of interesting questions remain.

10.1 Combining Head-Coupling with Motion Control

In our evaluation of the effectiveness of head-coupled stereo displays we have used scenes that are stable in space. In the past, applications have used yoked or automatic rotation of the scene to provide depth cues. Head-coupled perspective provides a related effect which is arguably more natural because we are accustomed to moving our head to obtain better views of scenes. However, it is not clear whether this argument applies to all scenes. In particular, when looking at a small object we are more likely to rotate the object than move our head about the object if we want to see the object's sides. One interesting possibility is to combine the techniques of motion with head-coupled perspective. One could have the scene rotate about its center, with the axis and direction of rotation defined in a counteractive way by the observer's head motion, and so obtain a more complete view of the object than possible with head-coupled perspective alone. We have implemented this technique, but have not conducted any formal studies of its effectiveness.

10.2 Spatial and temporal accuracy

Another area requiring further study from a human factors standpoint is to determine what level of spatial registration is sufficient for effective performance, and how registration errors affect the accuracy of depth judgements.

Our study of the effects of lag and frame rates provides useful guidelines for employing fish tank VR in applications. The effects of temporal artifacts in immersion displays likely have similar characteristics, but one would expect the effects to be more extreme due to the wide field-of-view provided to the eyes.

We have dealt only with the artifacts present when tracking head position and orientation. Presumably inaccuracies in tracking hand position will also degrade performance, but it is difficult to predict what the effect of two or more lags in a VR system will be.

In Experiment 3, two of the ten subjects reported slight dizziness afterwards. This is surprising because there is little sense of immersion in fish tank VR; one would expect much more significant problems in immersion VR. Motion sickness is a serious problem for VR interfaces and requires careful study with experimental subjects who are experienced in recognizing the symptoms.

10.3 Auxiliary 3D cues

In our experiments, we have been concerned primarily with the depth cues provided by head-coupled perspective and stereoscopy. Comparisons of these techniques with other computer graphics methods for depth cueing are important because efficiency trade-offs must be made when implementing real-time 3D graphics applications. Possible studies include conducting experiments to compare human perception of wire frame objects and shaded objects, and objects with and without specular highlights, both with and without head-coupled perspective.

11 Acknowledgements

The authors are indebted to Michael Deering of Sun Microsystems for comments on early versions of the experiments and the history of fish tank virtual worlds. The reference to the Krell holographic display is due to him. William Chernoff of Shooting Star Technology (Burnaby, British Columbia) provided advice on using the ADL-1 head tracker. We would like to thank the referees for their valuable comments which aided in clarifying the presentation and in particular for comments regarding the optics in head-mounted displays.

A preliminary account of Experiments 1 and 2 was presented at the INTERCHI '93 conference [39].

References

[1] Peter R. Atherton and Linda R. Caporael, A subjective judgment study of polygon based curved surface imagery, Proceedings of CHI '85 (April, 1985), 27-34.

[2] Baecker, Ronald M. Digital video display systems and dynamic graphics. Proceedings of SIGGRAPH '79. In *Computer Graphics, 13*, 2 (August 1979), 48-56.

[3] Blanchard, Chuck, Scott Burgess, Young Harvill, Jaron Lanier, Ann Lasko, Mark Oberman, and Michael Teitel. Reality built for two: A virtual reality tool. Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics. Special issue of *Computer Graphics* (1990), 35-36.

[4] Brooks, Frederick P., Jr. Walkthrough – A dynamic graphics system for simulating virtual buildings. Proceedings of the 1986 ACM Workshop on Interactive 3D Graphics. Special issue of *Computer Graphics* (1986), 9-21.

[5] Carlbom, Ingrid and J. Paciorek. Planar geometric projections and viewing transformations. *ACM Computing Surveys, 10* (December 1978), 465-502.

[6] Chung, James C. A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams. Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics. Special issue of *Computer Graphics* (1992), 193-196.

[7] Codella, Christopher, Reza Jalili, Lawrence Koved, J. Bryan Lewis, Daniel T. Ling, James S. Lipscomb, David A. Rabenhorst, Chu P. Wang, Alan Norton, Paula Sweeney, and Greg Turk. Interactive simulation in a multi-person virtual world. Proceedings of CHI '92 (1992), 329-334.

[8] Deering, Michael. High resolution virtual reality. Proceedings of SIGGRAPH '92. In *Computer Graphics, 26*, 2 (July 1992), 195-202.

[9] Diamond, R., A. Wynn, K. Thomsen and J. Turner. Three-dimensional perception for one-eyed guys. *Computational Crystallography.* Oxford, Clarendon Press (1982).

[10] Ellis, Stephen R. Nature and origins of virtual environments: A bibliographic essay. *Computing Systems in Engineering, 2*, 4, (1991), 321-347.

[11] Ellis, Stephen R. (editor) with M.K. Kaiser and A.J. Grunwald. *Pictorial communication in virtual and real environments,* Taylor and Francis (1991).

[12] Fairchild, K.M., E.E. Poltrock and G.W. Furnas. SemNet: Three-dimensional graphic representations of large knowledge bases. In *Cognitive Science and Its Applications for Human-Computer Interaction.* Raymond Guindon (editor), Lawrence Erelbaum (1988), 201-233.

[13] Fisher, Scott S. Viewpoint dependent imaging: An interactive stereoscopic display. *Processing and Display of Three-Dimensional Data*, S. Benton (editor), Proc. SPIE, 367 (1982), 41-45.

[14] Fisher, S. S., M. McGreevy, J. Humphries, and W. Robinett. Virtual environment display system. Proceedings of the 1986 ACM Workshop on Interactive 3D Graphics. Special issue of *Computer Graphics* (1986), 9-21.

[15] Foley, J.D. and A. Van Dam. *Fundamentals of Interactive Computer Graphics*, Addison-Wesley Publishing Company (1982).

[16] Foley, J. D., A. Van Dam, S. K. Feiner, and J. F. Hughes *Computer Graphics: Principles and Practice*, The Systems Programming Series, Addison-Wesley Publishing Company. (1990).

[17] Friedmann, Martin, Thad Starner, and Alex Pentland. Device synchronization using an optimal linear filter. Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics. Special issue of *Computer Graphics* (1992), 57-62.

[18] Howard, I.P. and T. Heckman. Circular vection as a function of the relative sizes, distances and positions of two competing visual displays. *Perception, 18*, 5 (1989), 657-665.

[19] Eric M. Howlett. Wide angle orthostereo. Proc. SPIE, Stereoscopic displays and applications (1990).

[20] Jacob, Robert J.K. What you look at is what you get: Eye movement-based interaction techniques. Proceedings of CHI '90 (1990), 11-18.

[21] Liang, Jiandong, Chris Shaw, and Mark Green. On temporal-spatial realism in the virtual reality environment. In *Proceedings of ACM UIST '91* (1991), 19-25.

[22] MacKenzie, I. Scott and Colin Ware. Lag as a determinant of human performance in interactive systems. Proceedings of INTERCHI '93 Conference on Human Factors in Computing Systems (April 1993).

[23] Meyer, Gary W., Holly E. Rushmeier, Michael F. Cohen, Donald P. Greenberg and Kenneth E. Torrance, An experimental evaluation of computer graphics imagery, *ACM Transactions on Computer Graphics, 5*, 1 (January, 1986), 30-50.

[24] McKenna, Michael. Interactive viewpoint control and three-dimensional operations. Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics. Special issue of *Computer Graphics* (1992), 53-56.

[25] Newman, W. M. and R. F. Sproull. *Fundamentals of Interactive Graphics.* McGraw-Hill (1972).

[26] Newman, W. M. and R. F. Sproull. *Principles of Interactive Computer Graphics*, Second Edition. McGraw-Hill (1979).

[27] Robertson, G.G., J.D. Mackinlay and S.K. Card. Cone trees: animated 3D visualizations of hierarchical information. Proceedings of CHI '91 (1991), 189-194.

[28] Robinett, Warren and Jannick P. Rolland. A computational model for the stereoscopic optics of a head-mounted display. *Presence, 1*, 1 (1992), 45-62.

[29] Schachter, Bruce J. (editor). *Computer Image Generation.* John Wiley and Sons, New York (1983).

[30] Shaw, Chris, Jiandong Liang, Mark Green, and Yunqi Sun. The decoupled simulation model for virtual reality systems. Proceedings of CHI '92 (1992), 321-328.

[31] Smith, K.U. *Delayed Sensory Feedback and Behaviour.* W.B. Saunders, Philadelphia (1962).

[32] Sollenberger, Randy L., and Paul Milgram. A comparative study of rotational and stereoscopic computer graphic depth cues. Proceedings of the Human Factors Society 35th Annual Meeting, San Francisco (1991), 1452-1456.

[33] Sutherland, Ivan. A head-mounted three dimensional display. Proceedings of the Fall Joint Computer Conference, AFIPS Conference Proceedings, 33 (1968), 757-764.

[34] Venolia, D. and L. Williams. Virtual integral holography. Extracting Meaning from Complex Data: Processing, Display, Interaction, Proc. SPIE, 1259 (Santa Clara, CA, February 1990), 99-105.

[35] Wallach, H. and D.H. O'Connell. The kinetic depth effect. *Journal of Experimental Psychology, 45* (1953), 205-217.

[36] Wanger, Leonard. The effect of shadow quality on the perception of spatial relationships in computer generated images. Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics. Special issue of *Computer Graphics* (1992), 39-42.

[37] Wanger, Leonard R., James A. Ferwerda, and Donald P. Greenberg. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications, 12*, 3 (May 1992), 44-59.

[38] Ware, Colin and Steven Osborne. Exploration and virtual camera control in virtual three dimensional environments. Proceedings of the 1990 ACM Symposium on Interactive 3D Graphics. Special issue of *Computer Graphics, 24*, 2, (1990) 175-183.

[39] Ware, Colin, Kevin W. Arthur, and Kellogg S. Booth. Fish Tank Virtual Reality. Proceedings of INTERCHI '93 Conference on Human Factors in Computing Systems, (April 1993), 37-42.

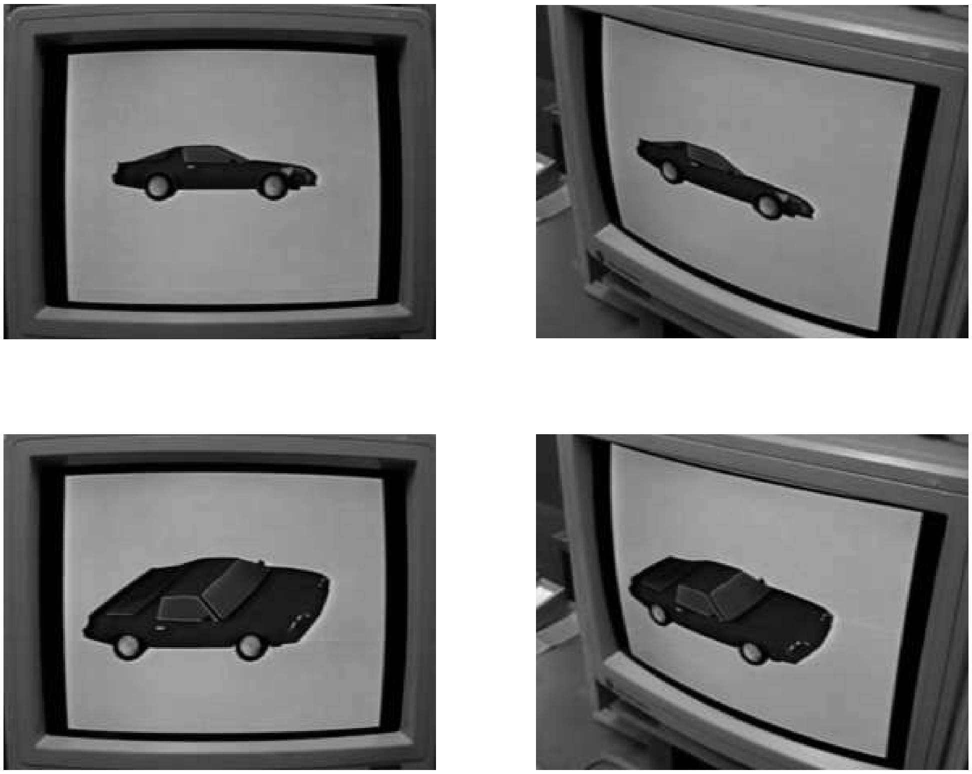

Figure 1: An illustration of the effects of head-coupled perspective. The program is displaying images of a car positioned in 3-space at the center of the screen. In the top row the image is computed with an on-axis perspective projection and in the bottom row with an off-axis projection. The left column shows the screen when viewed from a position on-axis and the right column shows the screen when viewed from a position off-axis. Only in the top-left and bottom-right photographs does the perspective projection match the viewing position, resulting in a realistic image which does not appear distorted.

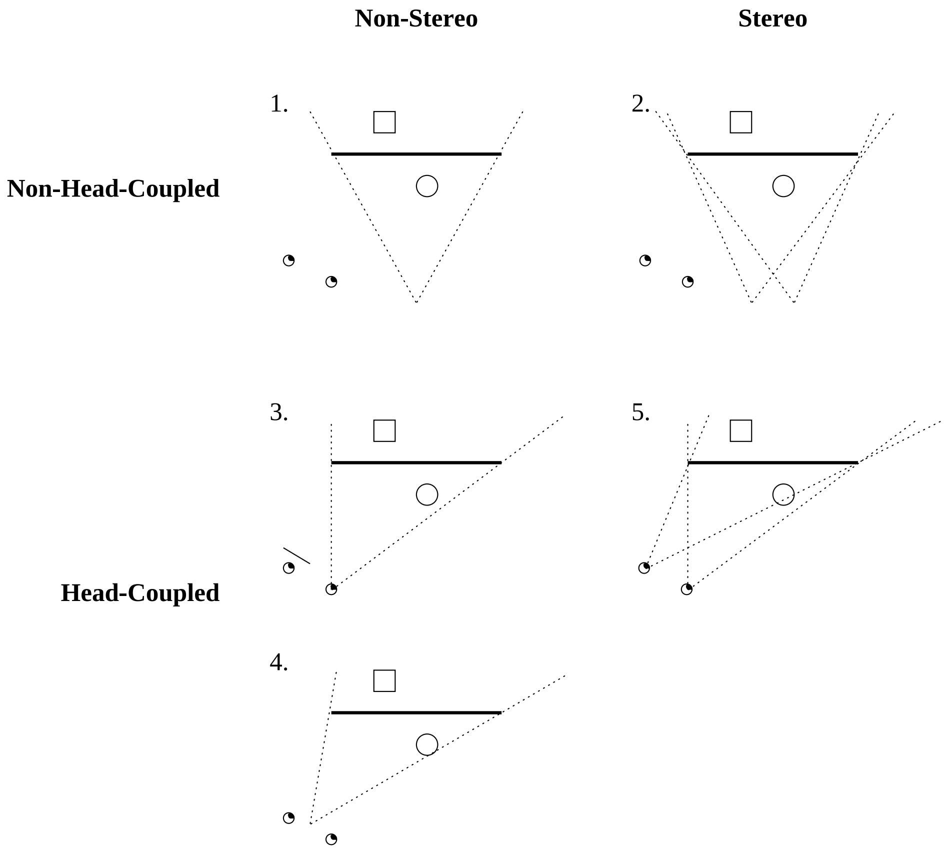

Figure 2: The five viewing conditions used in Experiments 1 and 2. In each of the diagrams the image plane is represented by the bold horizontal line, and virtual objects are shown in front of and behind the screen. The dotted lines indicate the perspective projections employed, each defined by an eyepoint and the corners of the screen.

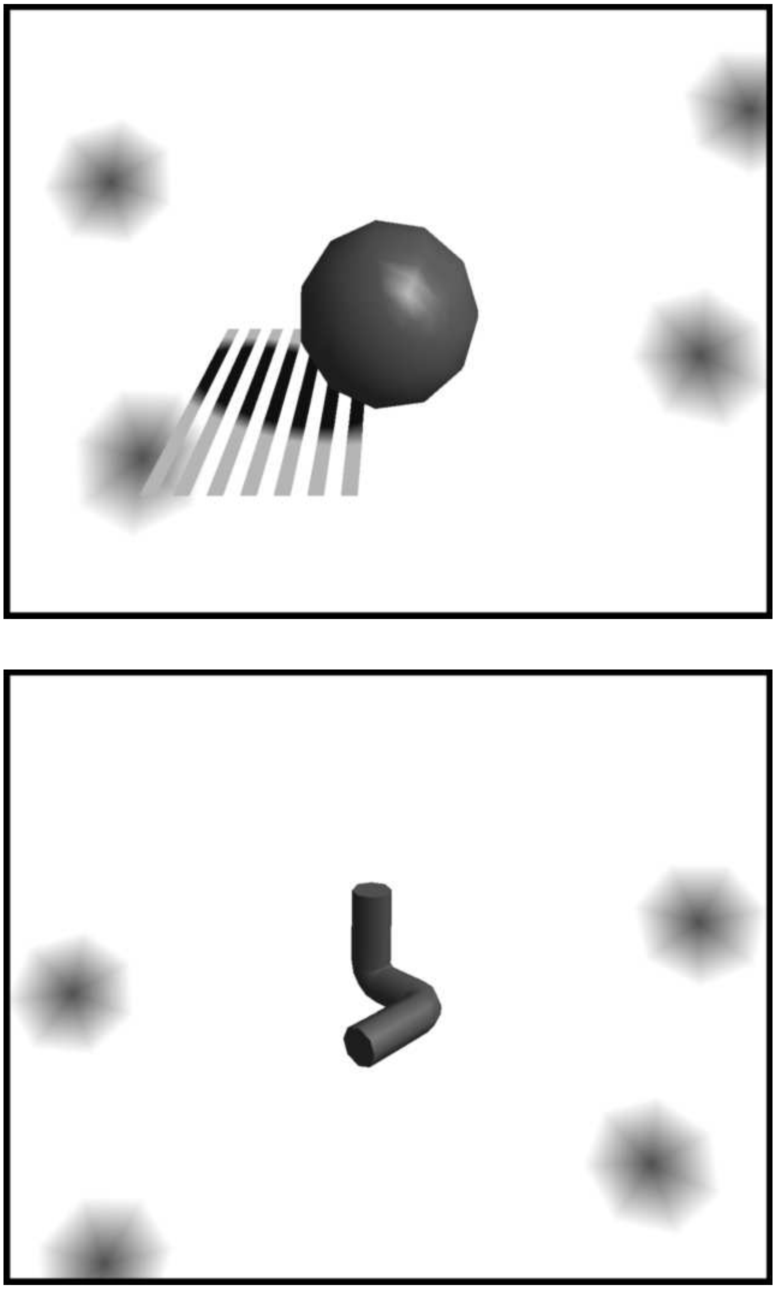

Figure 3: The sphere and bent tube displays used in Experiment 1. Hardware lighting was used to achieve the specular reflection. The blurry cast shadow was pre-computed and subsequently texture-mapped onto the floor. The colours and background have been modified for black and white reproduction.

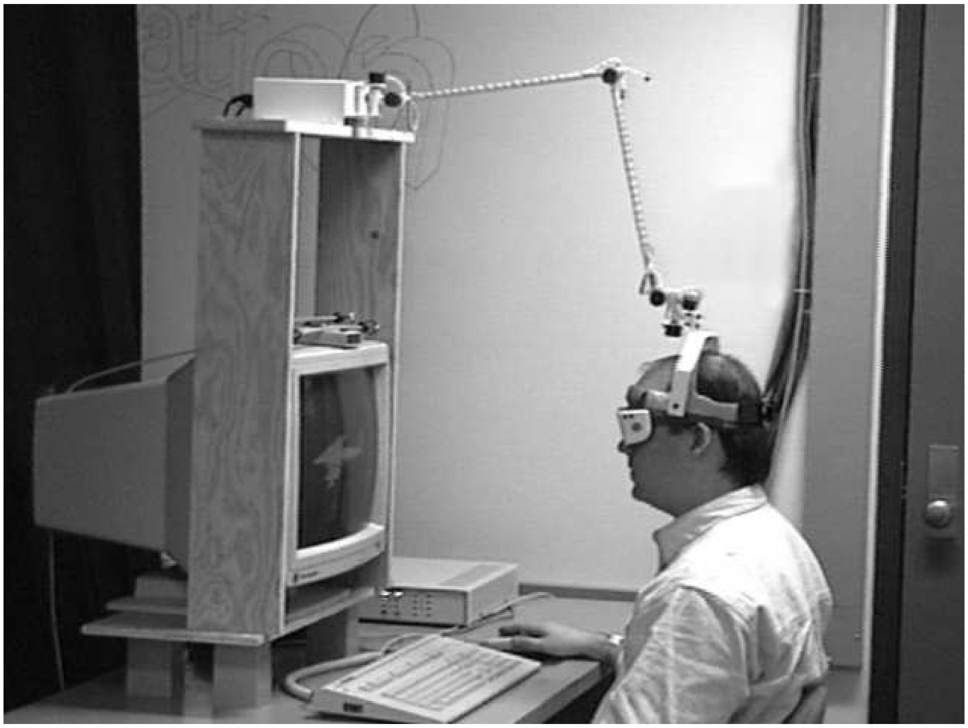

Figure 4: The fish tank VR system. The subject's head position is measured by the ADL-1 mechanical tracker. StereoGraphics glasses are worn to provide different images to the left and right eyes, and the display monitor is synchronized with the glasses to provide an effective 60 Hz to each eye.

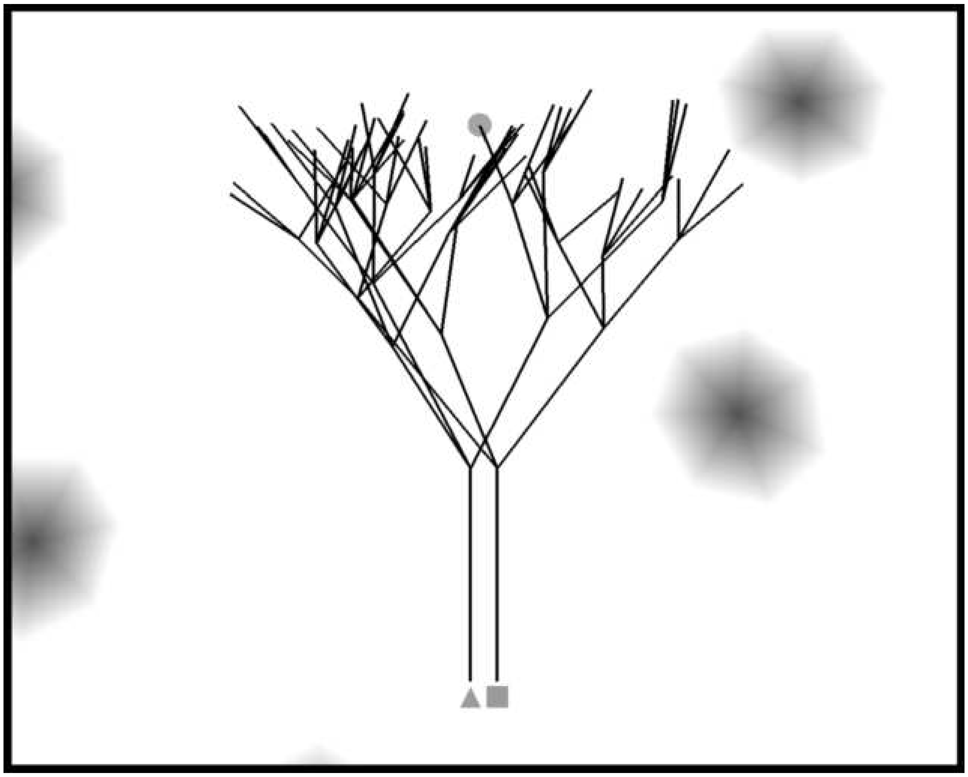

Figure 5: An example of the tree display used in Experiments 2 and 3. The colours and background have been modified for black and white reproduction.

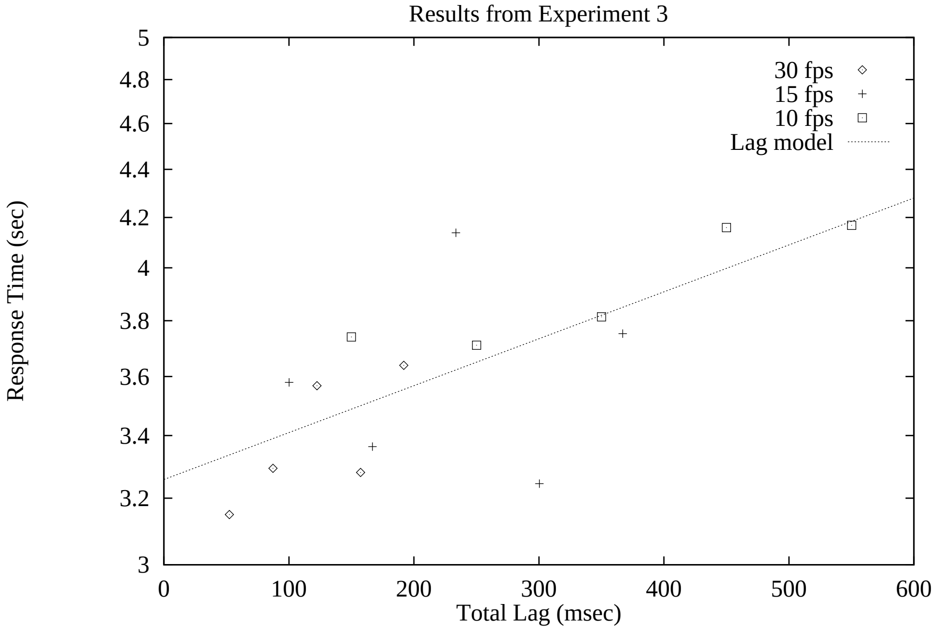

Figure 6: Plot of response time versus total lag for Experiment 3. Each point corresponds to an experimental condition with a particular lag and frame rate (see Table 3). The line is the best fit to the linear regression model involving total lag only.