The Effects of Variation of System Responsiveness on User Performance in Virtual Environments

Watson, B., Walker, N., Ribarsky, W., & Spaulding, V.

Graphics, Visualization, and Usability Center

Georgia Institute of Technology

ABSTRACT

This paper reports on study of the effects on user performance of system responsiveness in VE systems. Responsiveness is a time-to-feedback measure, and includes the well-known system latency and frame time, as well as an additional delay between a user action and the next input sample used by the rendering process. After a detailed examination of the components of VE system responsiveness and a review of the methods by which this responsiveness can be measured and manipulated, three studies of the effects of mean responsiveness and responsiveness variation during task performance are presented. These studies used typical system responsiveness means and patterns of variation, and were performed on a immersive non-desktop VE system. Results indicated that variations in responsiveness can affect performance, but only at standard deviations above 82 ms. Effects were more detrimental when tasks required more feedback. This suggests that designers of VE systems implementing control of model complexity to manipulate system responsiveness need not tightly constrain variation in system responsiveness, and may wish to make their control sensitive to required task feedback.

INTRODUCTION

Designers of virtual environments (VEs) face a fundamental tradeoff. To make a VE more informative and useful, the level of visual detail (visual complexity) needs to be high. However, as visual complexity increases, system responsiveness decreases due to the time required to display the more complex model. Unfortunately, as system responsiveness drops so does both the sense of presence (e.g., Barfield & Hendrix, 1995) and performance (e.g., MacKenzie & Ware, 1993). Therefore a designer of a VE who is trying to optimize the usability of a VE is caught in a tradeoff between visual complexity and system responsiveness. This tradeoff has been identified as a critical issue facing the VE community (NSF, 1992; Van Dam, 1993).

One way to deal with this tradeoff is to manage the level of detail (Funkhouser & Séquin, 1993). Management involves reducing the level of detail (VE model or simulation complexity) whenever the time required to render the model at its current level of detail will result in a reduction in system responsiveness that impairs performance or reduces the sense of presence. However, unless the detail is precisely predicted and managed, the system will oscillate around the target frame rate as the user moves through the environment. In order to manage the level of detail appropriately, one must know the level beyond which system responsiveness (e.g., in terms of means and variation in frame rate) will negatively affect performance and the sense of presence.

Previous work that has attempted to define a minimum cutoff value in mean frame rate or latency has yielded differing values. Barfield and Hendrix (1995) found that 15 frames per second (Hz) seemed to be the minimum rate for maintaining a sense of presence. Bryson (1993) has argued that 10 Hz is the critical cutoff value. Ware and Balakrishnan's (1994) showed that frame rates above 10 Hz did not result in performance improvements for the tasks they investigated. However, Wickens and Baker (1995) have found that performance on a simulator can be impaired with lags as low as 50 ms.

The difficulty in establishing a consistent cutoff value for system responsiveness stems from two unanswered questions: 1) how does one measure system responsiveness? and 2) what measures should one use to determine performance? Each of these questions is addressed separately below.

System responsiveness is most often measured and reported as an average of frame rates or latencies throughout the specific task. In all of the studies cited above, the authors reported the mean frame rate. However, while mean frame rate is obviously an important measure of system responsiveness, it is not the only one. For example, one can have a mean frame rate of 15 Hz, but some portion of the task the frame rate may drop to 6 Hz due to the complexity of the visual scene and stay at that rate until the person looks at a different, less detailed area in the virtual environment.

A second reason for the inconsistent findings on the effects of frame rate on performance may be the types of tasks used in the experiments. Different types of tasks may be differentially affected by system responsiveness. Although there are many types of movement tasks and interactions that can be used in a VE (e.g., grasping an object, navigating through the environment, selecting from a menu with a ray pointer), we believe that they can be grouped into two distinct movement types: open- and closed-loop movements.

Movement tasks are often described according to whether or not a person can use feedback (visual or proprioceptic) to correct the movement (Wickens, 1992). Movements that do not allow feedback and correction are referred to as open-loop tasks. An example of an open-loop task is throwing a ball at a target. Once the movement has been planned and executed, no course corrections can be made. A closed-loop task is one in which a person makes an initial movement, obtains feedback about the accuracy of the movement, and then makes further movements to correct for any error. An example of a closed-loop task is using a mouse to position a cursor on an icon. In these experiments, we incorporated two tasks, open-loop and closed-loop, which varied in the degree to which the user could use visual feedback to guide and correct the movement. We will determine the effects that system responsiveness has on performance on each of these tasks. Before presenting the experiments, we must first review the concept of system responsiveness.

System Responsiveness and Its Effects

In any study of the effects of certain system parameters on user performance, care must be taken in the identification and manipulation of the parameters so that their effects can be properly understood. Responsiveness in real-time graphics and virtual environments (VE) systems is a deceptively complex phenomenon, and unfortunately many studies have not adequately identified the components of responsiveness, or at least not adequately described the manipulations made. In this section we present our analysis of the elements of system responsiveness, describe the methods by which responsiveness might experimentally be manipulated in typical VE systems, and review existing literature on the effects of system responsiveness in this context.

The components of system responsiveness**.** Wloka (1994) has provided an excellent analysis of many of the components of responsiveness and latency, and we shall refer to it here where appropriate. Our analysis has a stronger focus on user performance and an emphasis on single-processor VE systems, whereas Wloka focused on multiprocessor VEs.

Figure 1 about here

A commonly used measure of responsiveness in VE systems is *frame time*, the time elapsed from the beginning of one display frame until the beginning of the next. Frame time is the sum of three components (see Figure 1): *input collection time*, the time elapsed while collecting the input required for simulation and rendering (e.g. tracker and button input); *simulation time*, the time elapsed while calculating the behaviors of elements of the VE (e.g. motion, changes in model characteristics); and *rendering time*, the time elapsed while the graphics platform renders the current view of the VE into the back buffer, waits for the buffer swap to synchronize with the monitor refresh, and scans the buffer onto the display. Many advanced graphics architectures are pipelined to allow display of larger models through low level parallelism, the price of this reduced rendering time per model primitive is an increase in minimum overall rendering time. Often rendering time includes other display modes (e.g., auditory) besides the merely visual, resulting in multiple and different frame times; we consider here only visual display. Note that while much of the input collection and simulation processes may be parallelized, the output of these processes must eventually reach the rendering process for inclusion in display, and some time must be spent gathering the output of each of these processes. Frame time is often related to the alternative term *frame rate*, the number of frames displayed per second. Frame time effectively determines the rate at which the VE is sampled in time.

Unfortunately each displayed frame does not accurately represent the current state of input and simulation. Instead, the display of input and simulation presented in each frame is slightly aged. The measure of this age, and in effect of the currentness of each presented sample, is *system latency*. System latency is also the sum of three components (Figure 1): *input-sampling latency*, the time elapsed from a change in input device state until that change is reported to the rendering process (corresponding to Wloka's input device lag and certain components of synchronization lag); *sampling-rendering latency*, the time elapsed from the arrival of the input sample until rendering begins (Wloka's application lag); and *rendering time*, as defined above (Wloka's rendering lag). Note that sampling-rendering latency may or may not include simulation time, depending on whether or not simulation requires an input sample. If the VE system uses multiple input devices, there are different input-sampling and sampling-rendering latencies associated with each device, and thus multiple system latencies associated with each display frame.

Most researchers have focused on frame time and system latency as the crucial parameters for user performance in VEs. While these variables are certainly important, there is an additional variable of specific relevance to user performance. *System responsiveness* is the time that elapses from a user action until feedback is received from the system (system latency is a measure in the reverse direction, the time from display to the represented event). In VE systems, system responsiveness is the sum of three components (Figure 1): *event-input latency*, the time elapsed from the occurrence of a user action until the next following input sample of that event; the previously described *system latency*; and *rendering-perception latency*, the time elapsed between the display of the crucial display element and the user's perception of it. Since only one sample of each input device is used per frame, event-input latency varies between zero and roughly one frame time (subject to the temporal resolution of the input device). Rendering-perception latency might in fact be negative: it may be possible for the user to perceive the feedback sought before the display refresh cycle is complete. Rendering-perception latency includes elements of Wloka's rendering and frame-rate induced lags and varies with the location of the feedback in the displayed view, the complexity of the feedback event, and the perceptual abilities of the user.

As an example of system responsiveness and its components, consider a single processor VE system (see Figure 1). Even such a simple standard system contains parallel system components, in this case a 3D tracking device and the user. The user has just placed, to a first approximation, a virtual object. The user would now like feedback on the accuracy of the initial placement attempt. The user must first wait until the position of virtual object is sampled for input (event-input latency). The user must then wait until the input position is displayed (system latency). System latency is made up of the time elapsed from the moment the tracker samples the position of the object until the moment this information reaches the rendering process (input-sampling latency), the time elapsed from that moment until rendering begins (sampling-rendering latency), and the time that then elapses until rendering is complete (rendering time). The user must then perceive the displayed feedback (rendering-perception latency). Only then can the user improve the placement of the virtual object.

Relationships between and possible experimental manipulations of system responsiveness components. By now the complexity of system responsiveness should be quite clear. In this section, we discuss the relationships between system responsiveness, system latency, and frame time, and discuss possible manipulations of lower level components to affect experimental changes in these high level components.

Because frame time and system latency each include components not contained in the other, their relationship is not fixed. System latency can be longer than frame time, as often happens when input-sampling latency is large. In this case, VE samples will be displayed at a high rate, but each sample will be aged by one or more frame times. Though it occurs more rarely, frame time can also be longer than system latency, for example when both input-sampling and sampling-rendering latency are small. In this case, VE samples will be displayed infrequently, but each sample will be aged by less than one frame time.

Since system responsiveness includes two components not contained by system latency, system responsiveness will generally be longer than system latency. Wloka (1994) speculated that rendering-perception latency would 5 ms, while Ware and Balakrishnan (1994) posited an average delay of 0.75 frame times (though in our opinion it is difficult to see why this latency should be related to frame time). Considering event-input latency alone, system responsiveness will exceed system latency by an average of one half of frame time. In rare cases, if event-input latency happens to be extremely small and rendering-perception latency negative, system responsiveness may even be a few milliseconds shorter than system latency.

The relationship between system responsiveness and frame time is much looser. If system latency and rendering perception latency summed are smaller than frame time, then due to event-input latency, average system responsiveness will equal at least half of frame time, and may on rare occasions be much less than frame time. As system latency and rendering-perception latency grow, system responsiveness can exceed frame time quite dramatically.

There a great number of sources of variation in system responsiveness. Event-input latency varies probablistically between zero and one frame time (which in turn varies with simulation and rendering time). Input-sampling latency includes variation inherent to the input device, as well as possible synchronization variation between parallel input collection and rendering processes. Sampling-rendering latency, since it may include some simulation time, can vary with the complexity of the VE. Rendering time varies in part probablistically because of buffer swap-refresh synchronization, the maximum time for this variation is one monitor refresh cycle. In addition, rendering time can vary widely with the complexity of the VE. Rendering-perception latency, as outlined above, varies with the current view of the VE and with perceptual noise.

There are three software-based approaches for implementing experimental control of responsiveness in VE systems. The first, *frame-latency manipulation*, varies frame time, system latency and system responsiveness simultaneously by manipulating sampling-rendering latency. This mimics the effect of model-complexity-dependent variations in input-sensitive simulation time and rendering time. Typically this is done by adding delay to input collection time after the input is sampled. The second control approach, *frame-only manipulation*, varies frame time and system responsiveness. This can be done through addition of delay to input collection time before the input is sampled. Since input devices run in parallel to the graphics system, input-sampling latency is unaffected, and system latency unchanged. This mimics the effects of variations in pre-sample input collection time or input-independent simulation time. The third approach, *latency-only manipulation*, varies system latency and system responsiveness by varying input-sampling latency. This can be done by buffering tracker input in an input management process parallel to the rendering process, and mimics the effects of variations in the latency of input devices or off-renderer parallel input management processes.

<u>Existing experiments on the effects of system responsiveness in 3D virtual environments</u>**.** Before experimentation can begin, responsiveness in the experimental VE system must be characterized. Liang, Shaw and Green (1991) mounted a tracker on a pendulum and placed it next to the VE display. They recorded the motion of the real and the virtual pendulum with a single video camera. Since the velocity of the pendulum was known, they could with a single video field derive the time the tracker took to travel from its actual to its displayed position. Subject to the 60 Hz sampling error of a standard video camera, this measured the age (or system latency) of each displayed sample. Mine (1993) used a much more exacting method involving photodiodes and oscilliscopes, and instead of using the distance between the tracker and its displayed image, recorded the time between the moment a pendulum reached vertical and the moment its virtual image passed vertical. This was an accurate measure of system responsiveness. Ware and Balakrishnan (1994) used a method much like Liang, et al., but substituted a conveyor belt on a stepper motor for a pendulum. Again, they measured only system latency. For the research presented in this paper, we placed a tracker in front of a virtual display and recorded actual and virtual motion in a single video view. System responsiveness was measured by counting the number of 60 Hz video fields between the onset of actual motion and virtual motion, and averaging over multiple samples.

We are aware of only two studies on the effects of system responsiveness on users in 3D VEs. Tharp, Liu, French, Lai and Stark (1992) asked users to perform a highly demanding 3D tracking task. Users controlled a cursor with two table-mounted joysticks and viewed the VE with a head-mounted display (HMD) tracked only in two rotational degrees of freedom. In one experiment, frame-only manipulation was used to vary mean frame time. Tracking performance stopped improving when frame times fell below 100 ms. In a second experiment, latency-only manipulation was used to vary mean system latency. Even 50 ms latencies had a detrimental effect on performance. Ware and Balakrishnan (1994) asked users to perform 3D movement tasks. Users viewed the VE with a fishtank system (head-tracked stereoscopic desktop display), and controlled a cursor with a 3D tracker. The authors hypothesized a multiplicative relationship between the effects of task difficulty and responsiveness on performance. Two of their experiments are of interest here. In the first experiment, users moved the cursor until it was between two displayed planes. Latency-only manipulation was used to vary mean responsiveness to both head and hand input. Variation in head responsiveness was not significant, but this may be due to the nature of the fishtank system, which does not require much head motion. Hand responsiveness was significant. In their third experiment, users were asked to place the cursor inside a 3D box. In one condition, latency-only manipulation was used to vary mean responsiveness to hand input only. In a second condition, frame-only manipulation was used to vary mean responsiveness to hand input. In the final condition, frame-latency manipulation was used to vary mean hand input responsiveness. It should be noted that in the latter two conditions mean head responsiveness was effectively (due to event-input latency) also varied, with levels of 33, 50, 100, 167, 250, 500 and 750 ms in the second condition, and levels of 33, 50, 100, 167, and 250 ms in the third. Results did not show a strong multiplicative relationship. Ware and Balakrishnan speculated that this might be due to the difficulty of generalizing their Fitts' law based model from 2D to 3D. It may also be that the (perhaps unexpectedly) poor head responsiveness in the latter conditions of their third experiment was a confound in their results.

Both of these studies used desktop VE systems with limited ranges of motion and tracking. Users of many typical VE systems stand and have much more freedom of motion. In addition, neither study examined variation in responsiveness during the task itself. Since, as we have explained above, responsiveness is always varying, a study of in-task responsiveness variation is sorely needed. Finally, human factors researchers identify two ideal task types: *open-loop* and *closed-loop*. Open-loop tasks are accomplished without the use of feedback during the task (e.g. catching and jumping). In contrast, closed-loop tasks do make use of feedback, closing the feedback loop. Tasks of this type include driving and accurate placement of objects. As task type moves from closed to open-loop (as the amount of feedback required decreases), the importance of system responsiveness should decrease.

This paper describes three experiments in which we investigated the effects of system responsiveness on task performance in a typical VE system. In these studies we use a values of system responsiveness which bracket values of system responsiveness that are currently found in immersive virtual environments. We also systematically vary the amount of variation of system responsiveness, again in an attempt to mimic a more naturally occurring set of parameters for immersive environments. Across the three experiments, we used a wide range of mean frames rates and levels of variation around the means to establish the tradeoff functions between system responsiveness and open- and closed-loop movement performance in a typical immersive virtual environment.

EXPERIMENT 1

In the first experiment, we will use mean frame rates and variation around those rates at the low end of mean frame rates (9, 13 and 17 Hz) that are often found in current immersive environments. In the experiment, we have participants perform a grasping and a placement task, which vary in the degree to which the tasks allow the use of visual feedback. We assume performance on both the open-loop (grasping) and closed-loop (placement) tasks would be affected both by mean frame rate and variation around the mean frame rate.

Method

Participants. Eleven undergraduate students from the Georgia Institute of Technology participated in two 45 minute sessions for this study. They were inexperienced in virtual reality and head-mounted displays and their vision was normal or corrected-to-normal (via contact lenses). The subjects received course credit in an introductory course in psychology and were treated in accordance with APA guidelines. The subject with the best cumulative ranking at the end of the experiment received fifty dollars.

Apparatus. The experimental environment was displayed using a Virtual Research VR4 head-mounted display with Polhemous Isotrack 3D tracking hardware. The images were generated with a Silicon Graphics Crimson Reality Engine. The participants interacted with the environment using a plastic mouse, shaped like a pistol grip. During the experiment, they stood within a 1 m by 1 m railed platform. The platform was 15 cm high and the railing was 1.2 m high.

Stimuli. The subjects tracked a moving target object, grasped it, and placed it on a pedestal with a certain spatial accuracy tolerance. The target object was a white oblong box, measuring 31 cm in height and 15.5 cm in depth and width. A yellow cubic cursor, 9 cm across each side, represented the joystick/hand location within the virtual environment. Visual cueing guided the participants' grasp of the target object; the target object turned yellow and the cursor turned white when the subject successfully grasped the target object.

The virtual environment consisted of a black floor with a white grid superimposed on it, and a black background. The target object traveled at a constant velocity of .75 m/sec from left to right in a circular arc of 125 degrees and 1.5 m in length, at a constant radius of 69 cm from the center of the platform. The ends of the arc were marked by tall white posts (see Figure 2). After reaching the end of the arc and a 1.5 second pause, the target object reappeared at the left of the arc, effecting a *wraparound*. The target object moved up and down in an unchanging sinusoidal pattern. The amplitude of the sine wave measured 85 cm, and the target object described a single complete period of the sinusoid after traveling along the arc. The phase of the sinusoid was chosen randomly each time the target object appeared at the left end of the arc.

Figure 2 about here

The pedestal was white and located next to the base of the post marking the right end of the arc. It was an oblong box 1.5 m tall and 45 cm in depth and width. Success of the placement task was measured by testing the location of the target object: it had to be completely contained in a placement box. The placement box had the same depth and width as the pedestal and measured 55 cm in height. The placement box was blue and transparent and only appeared as feedback after the target object was incorrectly placed on the pedestal (see Figure 3).

A red and white bullseye was centrally positioned on a solid black background between trials. Subjects could not begin a trial until they had centered this bullseye in their view.

Figure 3 about here

A trial consisted of the subject orienting on the bullseye and squeezing the trigger button on the joystick to begin a trial. After a random delay (between 750 and 1750 ms) the target object appeared, and the bullseye disappeared. To grasp the target object, subjects had to squeeze the trigger button while the yellow cursor intersected the target object. Testing showed that the speed of the target object allowed only one grasp attempt. All subjects in these studies adopted an open-loop “predict motion and intercept” grasping strategy, rather than a closed-loop “track motion and click” grasping strategy. When the target object was successfully grasped, it would shift to a location underneath the cursor. This made placement difficulty independent of grasp location. To complete the trial, the participant transported the target object to the right side of the visual field and placed it on the white pedestal. For the placement to be correct, the target object box had to be placed completely inside the placement box described above when the trigger was released. Subjects required many corrective submovements to complete the task, and thus executed the placement task with closed-loop feedback.

Our focus in these studies was relevance to VE system designers attempting level of detail management. As pointed out above, only frame-latency manipulation mimics the effects of variations in model complexity (level of detail) on system responsiveness.

Therefore we implemented frame-latency manipulation of system responsiveness by ensuring that the virtual environment system would run well below target frame times, and adding delay after sampling time to reach the targeted frame time. Actual frame times were recorded to confirm experimental control. We measured system responsiveness in our system without delay or manipulation using the method outlined above and obtained a mean of 213 ms with a standard deviation of 30 ms. This then reflects minimal mean responsiveness and variation of responsiveness in our VE system. According to both Wloka (1994) and Ware and Balakrishnan (1994), this level of mean responsiveness is typical. Our VE system's variation in responsiveness also agrees well with Wloka's reported ranges of variation in responsiveness.

Control of mean system responsiveness through mean frame time and mean frame rate are equivalent. The same is not true of control of frame time variation. Each approach has its shortcomings. Posing variation control in frame time corresponds directly to variation in system responsiveness, and allows symmetric variation around the mean frame time. However, as frame times decrease, the standard deviations at which symmetric variation is possible decrease (e.g. symmetric variation in a ±40 ms range at a mean frame time of 33 ms is 73 ms maximum frame time, -7 ms minimum time). Posing variation control in terms of frame rate is inherently asymmetric, with a bias toward longer frame times (e.g. variation in a ± 5 Hz range at frame time of 50 ms is 67 ms maximum frame time, 40 ms minimum frame time). Furthermore, the range of this variation is dependent on the current frame time mean (e.g. variation in a ± 5 Hz range at a mean frame time of 50 ms gives a frame time range of 27 ms, at 67 ms a frame time range of 50 ms). Since we were interested in the effects of large amounts of variation even at high mean frame times, and believed that the effects of this variation would be primarily due to increases (not decreases) in frame time, we chose asymmetric frame time variation control based on frame rate. In a separate study of the effects of variation in system responsiveness on the same task by these authors (Watson, Spaulding, Walker, & Ribarsky, 1997), we used symmetric frame time variation control, and found effects less significant than those in these studies.

We considered controlling the pattern of this variation with sinusoids, however, we feared that their regular pattern would become a confound in our results. Instead, in correspondence with our emphasis on relevance for VE system designers, we recorded a typical 218 frame sample (see Figure 4) of the frame times from an existing, uncontrolled and unmanaged virtual environment application with approximately 7000 textured polygons (Hodges, Rothbaum, Kooper, Opdyke, Meyer, North, de Graff & Williford, 1995). Frame time in the experiment was set by looping over this sample. Mean frame time was changed by finding the difference between the mean frame time in the original sample and the desired mean, and adding this difference to each sample. Frame time standard deviation was changed by scaling the adjusted sample around its new mean, effectively changing the range of its variation. In the aforementioned study (Watson, et al., 1997), we implemented sinusoidal control of the pattern variation, and found a perceivable pattern of variation could indeed affect user performance.

Figure 4 about here

Design. This study utilized a 3 (average frame rate) X 3 (frame rate variation), within-subjects design. There were nine display conditions, determined by the two independent variables: the mean frame rate and the standard deviation from the mean frame rate. The mean frame rate variable consisted of three levels: 9, 13, and 17 Hz. Variation from this mean frame rate also had three levels: standard deviations of 0.5, 2.0, and 4.0 Hz. The corresponding values for system responsiveness are shown in Table 1.

Table 1 about here

There were four dependent measures, two for time and two for accuracy. The time measures were mean grasp time (average time to successfully grasp the target object), and mean placement time (average time to successfully place the target object on the pedestal). These mean times were calculated for the correct trials only. The measures of accuracy were the percentage of trials correctly performed and the mean number of attempts to grasp the target object.

Procedures. Each person participated in two sessions. Each session consisted of one block of 20 practice trials, followed by nine blocks of experimental trials. One display condition was presented in each experimental block. Three practice trials were presented at the onset of each display condition. Accurate placement of the target object within 30 seconds was defined as a correct trial, and there were five correct trials per block, per subject (i.e., 90 correct trials per subject over the two sessions). Thus, the subjects were required to complete five trials correctly in order to advance to the next display condition. Incorrect trials were discarded and subjects were required to complete all trials within each display condition before ending the session. The presentation order of the blocks was varied randomly between subjects and each order was used once.

Results

Five, two-way repeated measures analyses of variance (ANOVA) were performed on the four dependent measures. Bonferroni pair-wise comparisons were performed to follow-up significant main effects. Simple main effects tests were conducted when there was a significant interaction. An alpha level of .05 was maintained for all analyses. Cell means for the dependent measures are in Table 2.

Table 2 about here

Grasp Time. The ANOVA on grasp time yielded a significant main effect of mean frame rate [$F(2, 20) = 27.11$, $p < .001$] and a significant main effect of frame rate variation

[$F(2, 20) = 8.04$, $p < .01$]. The follow-up tests revealed a significant differences in mean grasp time between the 9.0 Hz level of average frame rate and each of the two higher levels (13.0 Hz, 17.0 Hz). In the same manner, the 4.0 Hz standard deviation level yielded significantly longer grasp times than the lower standard deviation levels (0.5 Hz, 2.0 Hz).

The ANOVA also yielded a significant interaction of frame rate mean and variation [$F(4, 40) = 6.27$, $p < .001$]. The follow-up examination of the simple main effects yielded a significant effect of mean frame rate at the 2.0 Hz and 4.0 Hz standard deviation levels. There was no significant effect at the 0.5 Hz standard deviation level. At the 2.0 Hz standard deviation level, grasp time at the 9.0 Hz mean frame rate was significantly longer than at the 17.0 Hz mean, but not the 13.0 Hz mean. At the 4.0 Hz standard deviation level, the 9.0 Hz mean frame rate was significantly different from each of the higher levels (13.0 Hz, 17.0 Hz).

The follow-up analysis of the simple main effects also resulted in a significant effect of variation at 9.0 Hz frame rate mean [$F(2, 20) = 13.62$, $p < 0.001$], but insignificant effects at the 13.0 Hz and the 17.0 Hz levels. At the 9.0 Hz frame rate, the 4.0 Hz standard deviation level was significantly different from each of the two lower standard deviation levels (0.5 Hz, 2.0 Hz).

Placement Time. The ANOVA on placement time yielded a significant main effect of mean frame rate [$F(2, 20) = 22.75$, $p < .001$]. Mean placement times for the different levels of mean frame rate were determined to be significantly different from one another (means of 3.84, 3.38, and 2.94 sec. for 9 Hz, 13 Hz, and 17 Hz respectively). Neither the main effect of the variation of frame rate nor the interaction of the variation and the mean frame rate were statistically significant.

Number of Grasps. The analysis of the average number of attempts to grasp the target object per trial yielded a significant main effect of average frame rate [$F(2, 20) = 13.15$, $p < .001$] and a significant main effect of variation in frame rate [$F(2, 20) = 4.67$, $p < .05$]. The average frame rate of 9.0 Hz (mean 2.8 grasps) had significantly more grasps than the two higher frame rates (means = 2.08 and 1.96 grasps for 13.0 Hz and 17.0 Hz). The two highest levels of standard deviation (means = 1.95 and 2.61 grasps for 2.0 Hz and 4.0 Hz) were significantly different from each other, although neither was significantly different from the 0.5 Hz level of standard deviation (mean = 2.28 grasps).

There was also a significant interaction of mean frame rate and variation, [$F(4, 40) = 3.71$, $p < .01$]. The effect of mean frame rate was significant at the 4.0 Hz standard deviation level [$F(2, 20) = 13.10$, $p < .001$], but not at the two lower levels (0.5 Hz, 2.0 Hz). The 9.0 Hz mean frame rate level was significantly different from each of the two higher levels (13.0 Hz, 17.0 Hz). Similarly, the effect of variation of frame rate was significant at the 9.0 Hz level of mean frame rate [$F(2, 20) = 9.04$, $p < .01$], but not at the two higher levels (13.0 Hz, 17.0 Hz). The 4.0 Hz standard deviation level was significantly different from each of the two lower levels (0.5 Hz, 2.0 Hz).

Accuracy. The analysis of percent correct trials identified a significant main effect of average frame rate [$F(2, 20) = 3.93$, $p < .05$]. The follow-up procedure indicated that the difference was between the 9.0 Hz (mean = 82.3%) and the 17.0 Hz (mean = 92.0%) levels. Neither the main effect of frame rate variation nor the interaction of mean frame rate and frame rate variation were significant.

Figure 5 about here

Discussion

These results clearly show that both variation of system responsiveness and mean responsiveness can affect performance. Variation in system responsiveness affected only the grasping task, and only when this variation was quite large (115 ms standard deviation). For both tasks, performance continued improving across the entire examined range of mean system responsiveness (259 ms - 337 ms). The interactions observed between mean frame rate and frame rate variation corresponded well with the resulting degrees of variation in system responsiveness introduced by the frame rate based control of variation.

EXPERIMENT 2

This study attempted to identify the effects of improved mean system responsiveness and higher levels of system responsiveness variation on performance of the same tasks examined in Experiment 1. Given the different effects of system responsiveness on performance of the two different tasks in the first experiment, we were particularly interested in identifying further possible differences of effect on these two tasks. By defining frame rate variation as a percentage of the frame rate mean, we largely eliminated the changes in the range of system responsiveness variation introduced by frame rate based control of variation. We expected that this would eliminate the interaction of frame rate mean and variation found in Experiment 1 and confirm the importance of system responsiveness as a predictor of human performance.

Method

Participants. Twelve undergraduate students from the Georgia Institute of Technology participated in two 45 minute sessions for this study. They had similar characteristics to the participants of Experiment 1 and were recruited and rewarded in the same manner.

Apparatus, Stimuli, Design, and Procedure. The apparatus and the stimuli used for Experiment 2, as well as the design and procedure, were the same as those of Experiment 1, with the exception of the values of mean frame rate and the frame rate variation. The mean frame rate variable had three levels: 17 Hz, 25 Hz, and 33 Hz. The three levels of the frame rate variation around the mean frame rate were: 5.60% of mean frame rate, 22.20%, and 44.40%. The corresponding values for system responsiveness are shown in Table 1.

Results

Repeated measures analyses of variance (ANOVA) were performed on the data and Bonferroni pairwise comparisons were performed on the significant main effects. An alpha level of .05 was maintained for both procedures. Due to the absence of any significant interactions, Table 3 provides the marginal cell means and illustrates the significant differences.

Grasp Time. The ANOVA on grasp time yielded no significant main effects or interaction. The mean grasp time was 2.498 seconds.

Table 3 about here

Placement Time. The overall ANOVA on placement time yielded a significant main effect of mean frame rate [$F(2, 22) = 7.49$, $p < 0.01$] and a significant main effect of variation in frame rate [$F(2, 22) = 8.43$, $p < 0.01$]. Follow-up test revealed that placement times were longer when the mean frame rate was 17 Hz than when the mean frame rate was 33.0 Hz. Follow-up analyses of the effect of variation revealed placement times were less when variation was 5.60% than in the 44.40% variation condition. The interaction of the mean frame rate and variation in frame rate was not significant.

Number of Grasps. The analysis of the average number of attempts to grasp the target object revealed no significant main effects or interaction. The grand mean number of attempts was 1.696.

Accuracy. An analysis of the percent correct trials revealed no significant main effects or interaction. The grand mean accuracy was 90.1%.

Figure 6 about here

Discussion

The differences in these results between the effects on the grasping (open-loop) and placement (closed-loop) tasks of our manipulations on system responsiveness is striking. For the grasping task, improving system responsiveness from 263 ms to 215 ms did not have a significant effect, indicating a performance threshold at a mean responsiveness of

233 ms. In addition, even the large ranges of variation in system responsiveness introduced here (33 ms to 186 ms standard deviation) had no effect. In the first experiment, standard deviation in system responsiveness of 115 ms did have an effect on the grasping task. Apparently mean responsiveness is the primary factor in performance for this task. It may also be that the grasping task is particularly sensitive to the high frame times used in the first experiment. In contrast, the placement task, which in the first experiment was affected only by mean system responsiveness, was in this experiment also affected by variation in system responsiveness. Clearly the placement task is more sensitive to system responsiveness than the grasping task. As expected, the interaction of system responsiveness mean and variation was not significant, indicating the value of system responsiveness as a predictor of performance.

EXPERIMENT 3

In the third experiment our system was pushed to its limits (optimal mean system responsiveness, variation in system responsiveness, and frame times) in an attempt to identify a performance threshold to the effects of system responsiveness on performance in the placement (closed-loop) task.

Methods

Participants. Ten undergraduate students from the Georgia Institute of Technology participated in two 45 minute sessions for this study. They had similar characteristics to the participants of Experiment 1 and were recruited and rewarded in the same manner.

Apparatus, Stimuli, Design, and Procedure. The apparatus and the stimuli used for Experiment 3, as well as the design and procedure, were the same as those of Experiments 1 and 2, with the exception of the mean frame rate levels and the frame rate variation. The mean frame rate variable consisted of three levels: 17 Hz, 33 Hz, and 41 Hz. The three levels of the standard deviation in frame rate were: 0.50 Hz, 3.77 Hz, and 7.80 Hz. The corresponding values for system responsiveness are shown in Table 1.

Results

As in the first two experiments, we ran four frame rate (3) by variation (3) repeated measures analyses of variance. The five dependent measures were grasp time, placement time, number of grasps, and accuracy. The means for the dependent measures are reported in Table 4.

Grasp Time. The ANOVA on grasp time yielded no significant main effects or interaction. The main effect of variation did approach significance ($F(2,18) = 3.07$, $p < .08$) but the null hypothesis could not be rejected.

Table 4 about here

Placement Time. The ANOVA on placement time yielded a significant main effect of mean frame rate [$F(2, 22) = 9.32$, $p < .01$]. Follow-up tests revealed that placement time was significantly longer with frame rate of 17 than for the two higher frame rates. The interaction of the mean frame rate and variation approached significance [$F(4,36) = 2.31$, $p < .08$], but again the null hypothesis could not be rejected.

Number of Grasps. The analysis of the average number of attempts to grasp the target object per trial yielded no significant main effects or interaction.

Accuracy. The analysis of the percent correct trials yielded no significant main effects or interaction.

Figure 7 about here

Discussion

Again, there was no significant effect on the grasping task, consistent with the results of experiment 2. While results for the placement task were consistent with the identification of a threshold of the effects of responsiveness, it may well be that further reductions in mean responsiveness beyond the capabilities of our system might have produced further improvements in task performance.

GENERAL DISCUSSION

The research presented here makes two important contributions. First, it provides a succinct outline of the complex phenomenon of responsiveness in immersive VE systems, and outlines the methods by which the responsiveness of existing VE systems can be characterized, and research on the performance effects of responsiveness performed. Second, it provides a meaningful example of this sort of research on a typical VE system, illustrating the effects of variation in responsiveness on different types of tasks.

Many researchers have noted that frame time and system latency have powerful effects on user performance in VE systems. We believe, however, that the the crucial factor in human performance, especially for closed-loop tasks, is system responsiveness, a time-to-feedback measure that includes both frame time and system latency. Calibration of any VE system should include not only measurements of frame time and system latency, but also system responsiveness. Furthermore, there are several important characteristics of VE systems that should be considered in any experimental examination of responsiveness and its effects. First, frame time has more of an effect on mean responsiveness than many may have realized: not only is frame time a component of system latency, but it also defines event-input latency. Second, variation in responsiveness is a natural and unavoidable part of any VE system. Because it comes from a number of different sources, this variation is highly complex and difficult to characterize. However, it should be noted this variation increases dramatically as mean frame time increases. Next, while reduction of mean responsiveness is always an admirable goal, it should be kept in mind that there will always be a floor (optimal) responsiveness in any VE system. Experimental examinations of the human performance effects of variables such as frame time and system latency will be most relevant to VE application designers if they examine the effects of these variables when responsiveness mean and variation are at levels typical in the field. Previous experiments that have attempted to identify frame time or frame rate thresholds beyond which human performance does not improve should be re-examined with the corresponding VE system's optimal responsiveness in mind; it may be that further reductions in that responsiveness could have resulted in further improvements in human performance.

The experiments presented here emphasized this relevance to designers of VE systems. This study of responsiveness, unlike the others of which are aware, examined effects on users in a fully head-tracked, non-desktop VE system. We examined effects of variation in system responsiveness at mean levels common in current VE systems, ranging from 213 ms to 337 ms. Responsiveness variations at standard deviations of 82 ms or less were never significant. For our grasping (open-loop) task, we found that variations in responsiveness was only significant when mean system responsiveness was above 298 ms, and the standard deviation of this responsiveness was 115 ms. For our placement (closed-loop) task, we found that variation in responsiveness with standard deviations of 128 ms or higher were significant across the entire examined range of mean responsiveness, although when responsiveness was 259 ms or higher, the effects of mean were dominant. As mean frame rate increased to 25 Hz and mean system responsiveness fell to 233 ms, we found a threshold of performance for the grasping task. We found a similar threshold at 33 Hz and 215 ms for the placement task, however, it is quite likely that further reductions in mean system responsiveness, not possible in our system, would have brought further improvements in user performance.

Clearly variation in system responsiveness can affect user performance in VE systems. However, even for the fairly difficult tasks used in this study, the range of this variation must be quite large to have any sort of effect. This study used a variation control methodology that placed more of the range of responsiveness variation above the mean responsiveness time than below it. In a similar study (Watson, et al, 1997) using symmetric variation around mean responsiveness, effects of variation were much less significant. This suggests that designers of VE systems implementing level of detail management of model complexity, rather than restricting frame times to a certain range, should be able to implement a ceiling-only management strategy that takes action only when frame times rise above a certain value. This should allow improvements in mean system responsiveness, and a looser form of detail control.

Obviously, the effects of responsiveness will vary with the task being performed. The results of this study suggest that tasks that require high levels of feedback (closed-loop) will be more sensitive to responsiveness than tasks that require low levels of feedback (open-loop). VE system designers may want to implement a system of detail control that is sensitive to task and the amount of feedback required by the task. Alternatively, they might reduce the demand on system responsiveness by lowering the amount of feedback a task requires.

This study has focused on the effects on user performance of changes in system responsiveness mimicking the changes introduced by variations in model and simulation complexity. Since the study demonstrated continued improvements of user performance even at optimal mean responsiveness, new research broadening the examined range of mean system responsiveness would be useful. Our results suggest that it is primarily the feedback provided in times longer than mean system responsiveness that are responsible for the effects of variation in system responsiveness. More explicit confirmation and description of this effect is needed. In addition, implicit in the idea of these studies is a tradeoff between visual and temporal fidelity (e.g. polygon count and system responsiveness). A direct experimental examination of this tradeoff is certainly in order. Finally, this research showed differential effects of system responsiveness depending on the amount of feedback required by a task. Reproducing these effects with different types of tasks would confirm the generalizability of this result.

REFERENCES

Barfield, W. & Hendrix, C. (1995). The effect of update rate on the sense of presence within virtual environments. Virtual Reality: Research, Development & Applications, 1, (3-15).

Bryson, S. (1993). Implementing virtual reality. SIGGRAPH 1993 Course #43 Notes, (pp. 1.1.1-1.6.6, 16.1-16.12).

Funkhouser, T. & Séquin, C. (1993). Adaptive display algorithm for interactive frame rates during visualization of complex virtual environments. Computer Graphics (SIGGRAPH 93 Conference Proceedings). (247-254).

Hodges, L.F., Rothbaum, B.O., Kooper, R., Opdyke, D., Meyer, T., North, M., de Graff, J.J., & Williford, J. (1995). Virtual environments for treating the fear of heights. IEEE Computer, 28, 27-34.

Lampton, D. R., Knerr, B. W., Goldberg, S. L., Bliss, J. P., Moshell, J. M., & Blau, B. S. (1994). The virtual environment performance assessment battery (VEPAB): Development and evaluation. Presence, 3, 145-157.

Liang, J., Shaw, C., & Green, M. (1991). On temporal-spatial realism in the virtual environment. Proceedings of the User Interface Software Technology, (19-25).

MacKenzie, I. S. & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. In Proceedings of the INTERCHI ’93 Conference on Human Factors in Computing Systems (488-493). New York: ACM.

Mine, M. (1993). Characterization of end-to-end delays in head-mounted display systems. University of North Carolina at Chapel Hill, Department of Computer Science. Technical report TR93-001.

National Science Foundation (NSF) (1992). Research directions in virtual environments. Report of an NSF Invitational Workshop, Computer Graphics, 26, (153-177).

Tharp, G., Liu, A., French, L., & Stark, L. (1992). Timing considerations of helmet mounted display performance. In Human Vision, Visual Processing, and Digital Displays III. SPIE Vol. 1666. (570-576).

Van Dam, A. (1993). VR as a forcing function: software implications of a new paradigm. IEEE Symposium on Research Frontiers in Virtual Reality. (5-8).

Ware, C. & Balakrishnan, R. (1994). Reaching for objects in VR displays: Lag and frame rate. ACM Transactions on Computer-Human Interaction, 1, (331-356).

Watson, B., Spaulding, V., Walker, N. & Ribarsky, W. (1997). Evaluation of the effects of frame time variation on VR task performance. In Proceedings of the Virtual Reality Annual International Symposium. (38-44). Albuquerque, NM. IEEE..

Wickens, C. D. (1992). Engineering Psychology and Human Performance (2nd ed.). *New York: HarperCollins.*

Wickens, C. D. & Baker, P.(1995). Cognitive issues in virtual reality. In *W. Barfield & T. Furness (Eds.),* Virtual Environments and Advance Interface Design (514-541). *New York: Oxford.*

Wloka, M. (1994). Lag in multi-processor VR tutorial: Virtual reality for visualization. In Proceedings of IEEE Visualization. (130-152).

Table 1

Means and Standard Deviations for Frame Time and System Responsiveness in All Three Experiments

Mean Frame Rate	Standard Deviation Frame Rate	Mean Frame Time	Standard Deviation Frame Time	Mean System Respons.	Std Dev System Respons.
9	0.5	111	6	337	60
9	2	117	26	345	82
9	4	133	55	370	115
13	0.5	77	3	285	47
13	2	79	12	288	57
13	4	85	30	298	77
17	0.5	59	2	259	41
17	2	60	7	260	46
17	4	62	15	263	55
17	5.6%	59	3	259	42
17	22.2%	62	14	263	54
17	44.4%	58	140	257	189
25	5.6%	40	2	230	36
25	22.2%	42	10	233	44
25	44.4%	39	95	229	128
33	5.6%	30	2	215	33
33	22.2%	32	7	218	38
33	44.4%	30	72	215	103
17	0.5	59	2	259	41
17	3.78	62	14	263	54
17	7.56	58	137	257	186
33	0.5	30	1	215	32
33	3.78	30	4	215	35
33	7.56	31	5	217	36
41	0.5	24	1	213	30
41	3.78	25	2	213	31
41	7.56	25	5	213	34

Table 2

Means for the Primary Dependent Measures by Frame Rate and Level of Variation in Experiment 1

Level of Variation (Hz)	Grasp Time	Placement Time	Number Grasps	Accuracy (percent)
		FR = 9.0 Hz		
0.5	3.58	3.72	2.61	0.84
2.0	3.30	3.72	2.10	0.85
4.0	6.32	4.08	3.68	0.78
		FR = 13.0 Hz		
0.5	2.90	3.36	2.18	0.87
2.0	2.37	3.40	1.86	0.87
4.0	2.82	3.38	2.21	0.93
		FR = 17.0 Hz		
0.5	2.46	2.99	2.05	0.85
2.0	2.05	2.77	1.89	0.95
4.0	2.26	3.04	1.95	0.96

Table 3

Time measures by frame rate and standard deviation for Experiment 2

Measure	Grasp Time	Placement Time	Total Time
Frame Rate (Hz)			
17.0	2.787	2.102**a**	4.888**a**
25.0	2.365	1.944	4.309
33.0	2.342	1.866**b**	4.208**b**
Level of Variation (%)			
5.60	2.235	1.837**b**	4.071**b**
22.20	2.664	1.986	4.651**a**
44.40	2.595	2.090**a**	4.684**a**

Note. Means in the same section of a column that do not share subscripts differ at $p < .05$ in the Bonferroni pairwise comparisons procedure.

Table 4

Means of the Primary Dependent Measures for Frame Rate and Level of Variation in Experiment 3

Level of Variation (Hz)	Grasp Time	Placement Time	Number Grasps	Accuracy (percent)
		FR = 17.0 Hz		
0.50	2.844	2.771	1.81	0.91
3.77	2.736	2.874	1.65	0.84
7.80	4.211	2.937	2.15	0.84
		FR = 33.0 Hz		
0.50	2.612	2.813	1.67	0.90
3.77	2.884	2.431	1.92	0.94
7.80	2.757	2.224	1.69	0.89
		FR = 41.0 Hz		
0.50	2.241	2.272	1.57	0.92
3.77	2.606	2.385	1.66	0.85
7.80	2.719	2.416	1.87	0.84

Figure 1: The components of system responsiveness in a simple single processor VE system, consisting of three parallel components: user, tracker and renderer. Here both system latency and system responsiveness exceed frame time.

Figure 2: A top down schematic of the experimental environment. Users on the platform begin by looking at the bullseye; the target object moves left to right across the visual field.

Figure 3: The target object and cursor after a trial with unsuccessful placement. The target object leans past the front edge of the pedestal, a common mistake. The placement box is visible.

Figure 4: The sample of frame times used to control the pattern of frame time variation (and thus system time variation).

Figure 5: Mean grasp and placement times from the nine conditions of experiment 1.

Figure 6: Mean grasp and placement times for experiment 2. Since there were no interactions of frame rate mean and variation, we do not show all nine conditions.

Figure 7: Mean grasp and placement times from the nine conditions of experiment 3.

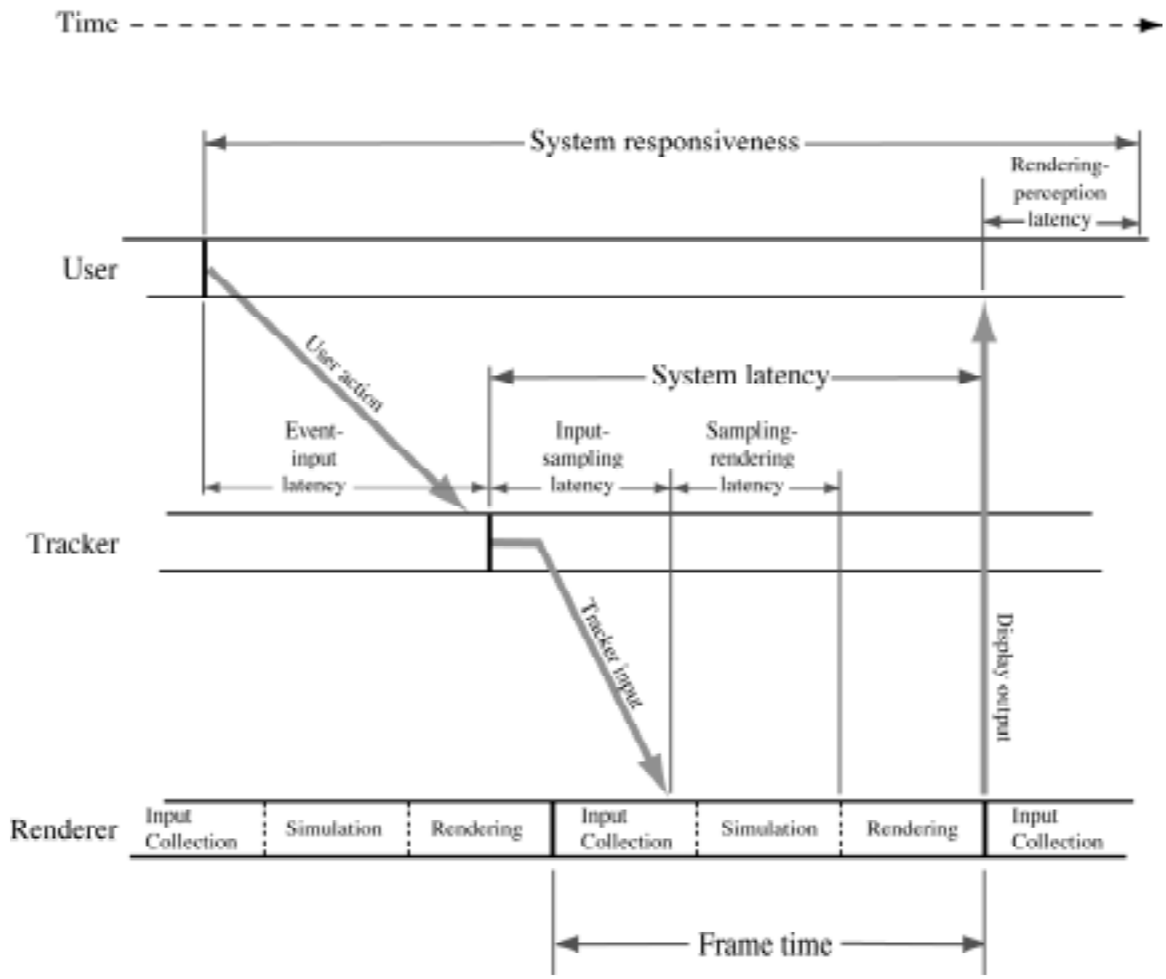
Time
System responsiveness
Rendering-perception latency
User
User action
System latency
Event-input latency
Input-sampling latency
Sampling-rendering latency
Tracker
Tracker input
Display output
Renderer
Input Collection
Simulation
Rendering
Input Collection
Simulation
Rendering
Input Collection
Frame time

Figure 1.

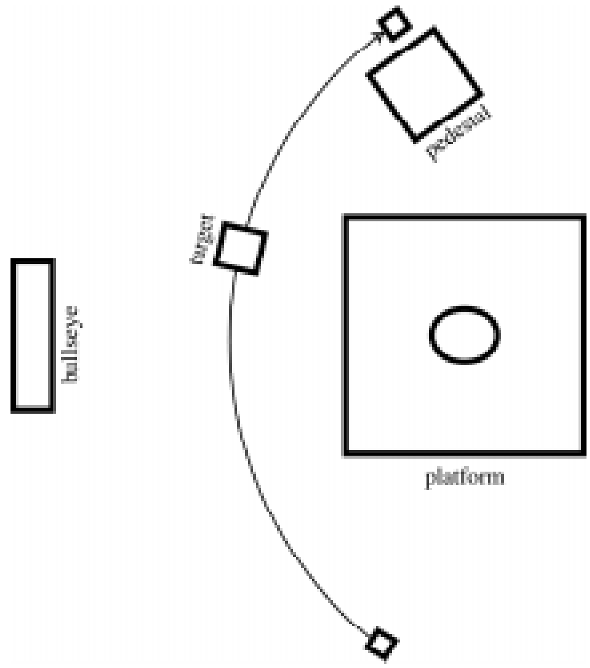

Figure 2.

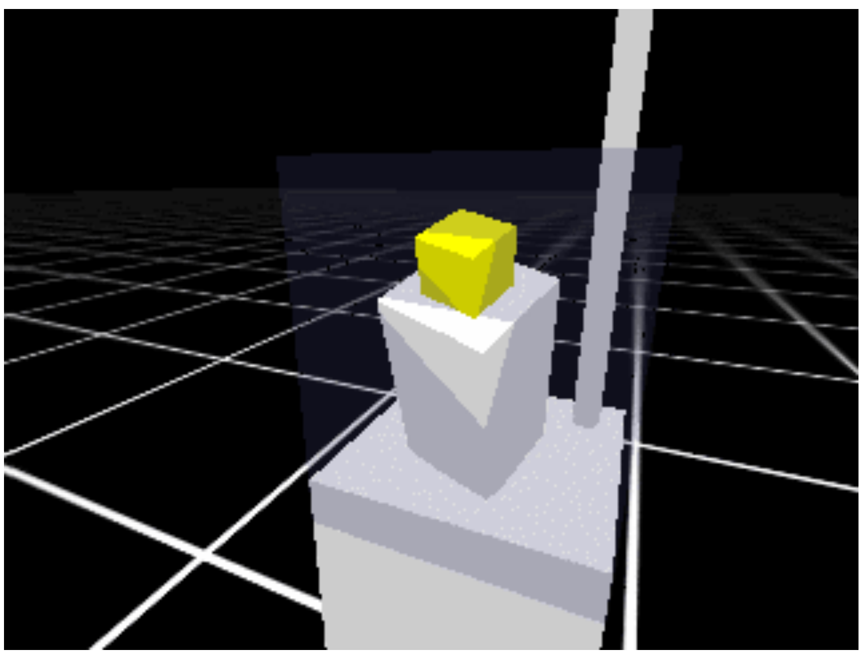

Figure 3.

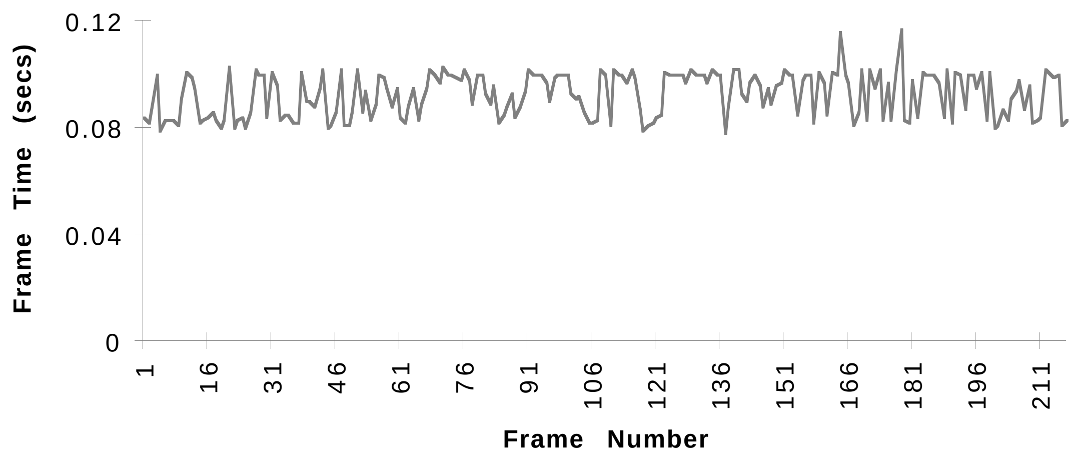

Figure 4.

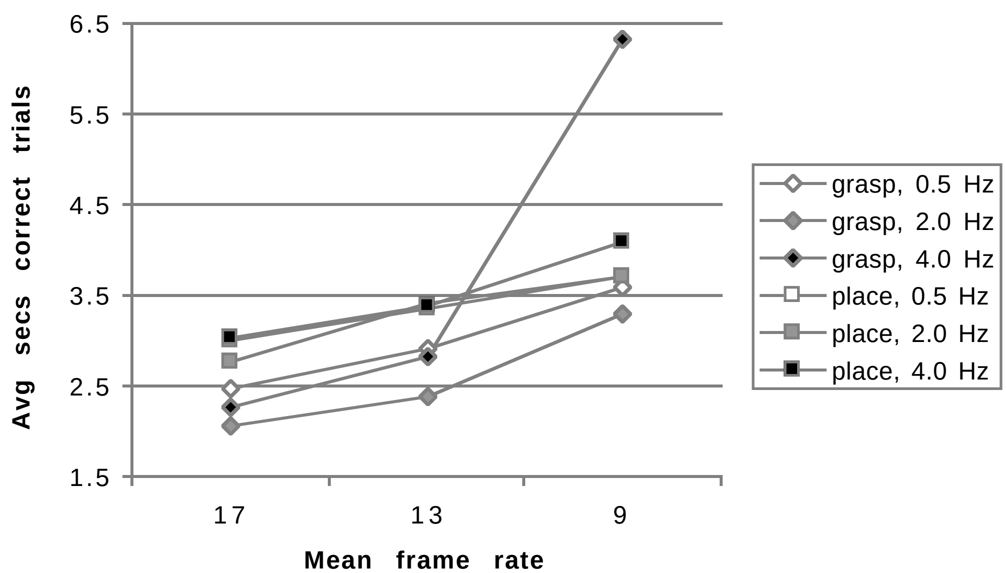

Figure 5.

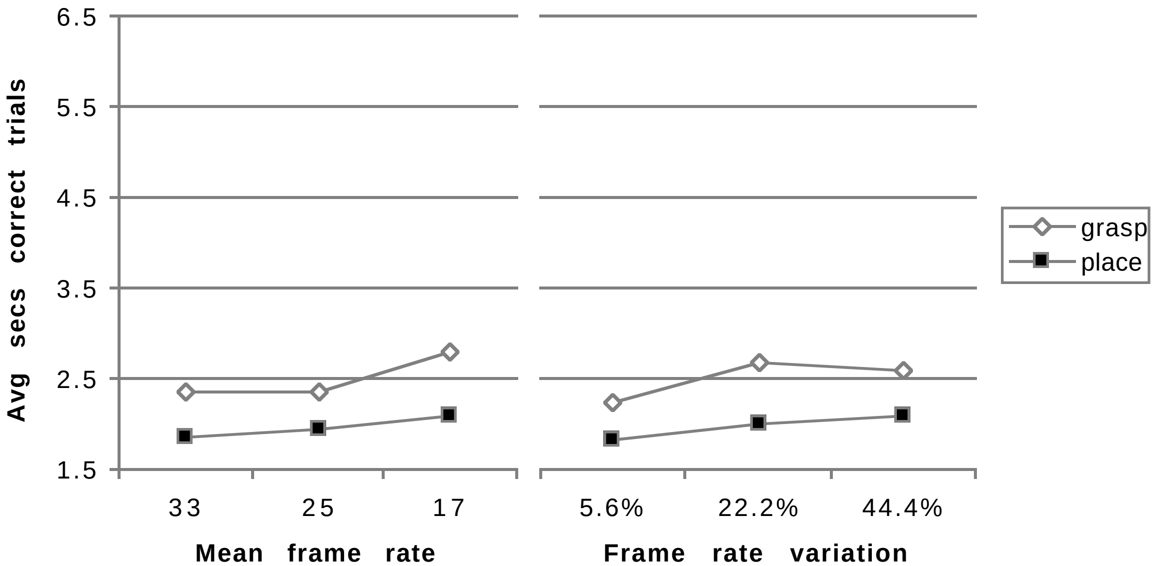

Figure 6.

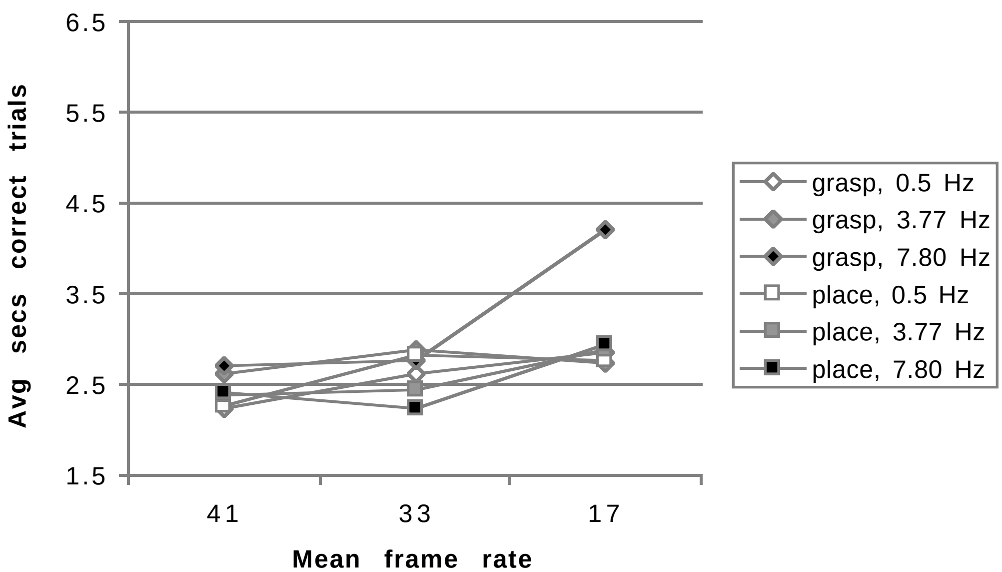

Figure 7.