Perceptual Sensitivity to Head Tracking Latency in Virtual Environments with Varying Degrees of Scene Complexity

Katerina Mania
Department of Informatics
University of Sussex, UK

Bernard D. Adelstein, Stephen R. Ellis, Michael I. Hill
Human Factors Research and Technology Division
NASA Ames Research Center, USA

Abstract

System latency (time delay) and its visible consequences are fundamental virtual environment (VE) deficiencies that can hamper user perception and performance. The aim of this research is to quantify the role of VE scene content and resultant relative object motion on perceptual sensitivity to VE latency. Latency detection was examined by presenting observers in a head-tracked, stereoscopic head mounted display with environments having differing levels of complexity ranging from simple geometrical objects to a radiosity-rendered scene of two interconnected rooms. Latency discrimination was compared with results from a previous study in which only simple geometrical objects, without radiosity rendering or a 'real-world' setting, were used. From the results of these two studies, it can be inferred that the Just Noticeable Difference (JND) for latency discrimination by trained observers averages ~15 ms or less, independent of scene complexity and real-world meaning. Such knowledge will help elucidate latency perception mechanisms and, in turn, guide VE designers in the development of latency countermeasures.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism, Virtual Reality

Keywords

Latency, Sensitivity thresholds, Simulations.

1 Introduction

Virtual Environment (VE) latency is the time lag between a user's action in a VE and the system's response to this action. This lag typically takes the form of a transport delay and arises from the sum of times associated with measurement processes of the various input devices, computation of the VE contents and interaction dynamics, graphics rendering, and finite data transmission intervals between these various components (Figure 1).

Excessive latency has long been known to degrade manual performance, forcing users to slow down to preserve manipulative stability, ultimately driving them to adopt 'move and wait' strategies [Sheridan and Ferrell 1963; Sheridan 1992; Smith and Smith 1962; Smith et al. 1962]. While users can exhibit sensorimotor adaptation that might improve manual performance to time delays in situations where task preview is available [Cunningham et al. 2001a; Cunningham et al. 2001b], the presence of delay has been shown to hinder operator adaptation to other display distortions such as static displacement offset [Held et al. 1966].

The literature has also established that delays in immersing VEs have a significant impact on user performance [Ellis et al. 1997; Ellis et al. 2002] and user impressions of simulation fidelity [Ellis et al. 1999a; Ellis et al. 1999b; Jung et al. 2000; McCandless et al. 2000; Adelstein et al. 2003; Mania et al. 2003]. Latency negatively affects user performance in 3D object placement tasks [Liu et al. 1993; Watson et al. 2003].

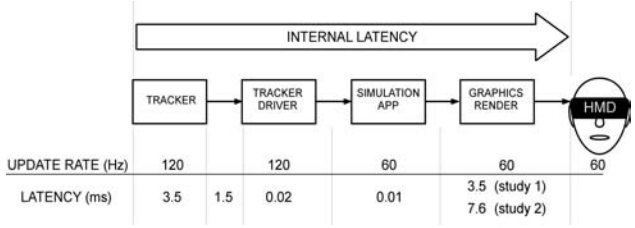

Figure 1. Internal VE latency stems from processing times within as well as the communication time between each of the system components shown. Tabulated are the measured update rate and latency of each component. The graphics rendering time is reported separately for [Ellis et al. 2004] (study 1) and the work presented here (study 2).

Interest has more recently been directed toward the subjective impact of system latency relevant to virtual reality simulations. Latency as well as update rate have been considered as factors affecting the operator's sense of presence in the environment [Welch et al. 1996; Uno and Slater 1997]. In a recent study, lower latencies were associated with a higher self-reported sense of presence and a statistically higher change in heart rate for users while in a stress-inducing (fear of heights), photorealistic environment involving walking around a narrow pit [Meehan et al. 2003].

Since the combination of sensing, computation, rendering, and transmission delay is unavoidable in most VE, tele-operation, and augmented reality applications, interest naturally is directed to how detectable differing levels of latency might be. Both the quantification of perceptual sensitivity to latency and description of the mechanism by which latency is perceived are essential to the development of countermeasures such as predictive compensation [Azuma and Bishop 1994; Jung et al. 2000], necessary for future VE system design.

Previous research has also examined the precision, stability, efficiency, and complexity of operation interaction and performance with latency-plagued systems [McCandless et al. 2000]. Additionally, the first measures of human operators' discrimination of the consequences of latency during head- or hand-tracked movements have been provided [Ellis et al. 1999a; Ellis et al. 1999b]. Related investigations have explored the hypothesis that observers do not explicitly detect time delay, but rather detect the consequences of latency–i.e., they use the artifactual displacement and motion of the VE scene away from its normally expected spatially stable location caused by system time lags [Adelstein et al. 2003]. Relevant perceptual thresholds (i.e., Just Noticeable Difference or JND) were identified to average 8-17 ms, depending on viewing condition. This psychometric quantity appeared to be invariant across different pedestals (33, 100 and 200 ms, standard stimuli). The apparent invariance of the detection function in Ellis et al. [1999ab] and Adelstein et al. [2003] demonstrated that the classic Weber's Law of psychophysics (that JND is linearly proportional to the magnitude of the standard stimulus) did not hold for latency. In other words, observers of long latency VEs will be as sensitive to changes in latency as those who use prompter, more advanced systems. It can also be inferred that the same sensitivity would also apply for comparisons against zero latency pedestal.

Regan et al. [1999] found 70.7% latency thresholds averaging 15 ms for a specialized non-immersing CRT display. Assuming Gaussian psychometric functions and zero response bias for two-interval forced-choice judgments with balanced presentation order, the 70.7% threshold from Regan et al. [1999] can be equated with a JND of 18.6 ms.

Allison et al. [2001] observed on the other hand that for large virtual objects occupying the full Head Mounted Display (HMD) Field-of-View (FOV), 50% thresholds for perceived image instability (oscillopsia) were found to be 180-320 ms depending on head motion velocity. This threshold indicates the latency level at which observers were equally likely as not to say the image was unstable and represents their average response bias or preference. Such response biases may be attributable to, among other things, the amount of observer training before the data was collected and the form of judgment task required. In the case of Allison et al. [2001], participants performed single interval judgments—i.e., they did not compare each presentation against a standard stimulus but relied on their own internal notion of when an image was no longer stable. Data from Ellis et al. [1999ab] and Adelstein et al. [2003] show their participants' response bias ranged between 40 and 70 ms for a two-interval judgment of whether the stimulus was the same as or different than the pedestal standard. In contrast, the participants in [Regan et al. 1999] were forced to choose which of the two stimulus intervals was actually the one with added latency, which though not reported, leads to a presumption of zero bias.

The much higher threshold reported by Allison et al. [2001] might also be attributable to the fact that their participants viewed a textured virtual background (the inside surface of red and white faceted sphere) that completely enveloped their head and thus always occupied their entire FOV. Surrounding observers with such a geometrically structured environment contributes to the phenomenon of visual capture. The term 'visual capture' implies that when concurrent multisensory spatial information is available, the observer will weight the visual channel more heavily in constructing a percept. It has been demonstrated that, even with very simple VE graphics in an HMD, visually discrepant information will bias proprioceptive and vestibular feedback of static head pitch angle [Nemire et al. 1994]. Since awareness of VE image instability relies on visual, vestibular, and proprioceptive information, the full structured background viewed in Allison et al. [2001] may have diminished their observers' sensitivity to latency-induced oscillopsia. Furthermore, without the inclusion of nearer objects in their environment, participant head movement does not trigger motion of scene contents relative to the background and thus does not provide cues through internal image shear.

One aim of our ongoing research on latency perception has been to quantify the latency that a VE system can exhibit without being perceptible to the user. In our prior studies, we employed very sparse environments containing only a single simple object such as a faceted sphere [Ellis et al. 1999a; Ellis et al. 1999b] or a hollow-framed octahedron [Adelstein et al. 2003] against an empty black background. In the studies reported here, we employ synthetic environments with differing levels of graphical complexity with the goal of extending the generality of our results for participant sensitivity latency in VEs.

In particular, the focus of this paper is quantification of observer sensitivity to latency differences during head movements in a realistic, immersing VE (e.g., room, building, etc.)—sensitivity that has not been measured in previous research. On the one hand, because there could be an inherent association with how the real world is perceived, we might expect observers to be more sensitive to the visual consequences of latency when viewing a scene representing what could be a real-world space rather than a sparse, simplified scene with only one or two artificial objects. On the other hand, an enveloping structured scene could promote visual capture, thereby degrading observers' sensitivity to VE latency.

During an earlier study more fully reported in [Ellis et al. 2004], a simple white-red checker sphere surrounding the observer, such

as that used in [Allison et al. 2001] and/or a hollow-frame octahedron in front of the observer, as in [Adelstein et al. 2003] served as the VE's visual content. Participants viewed two sequential stimulus presentations with experimentally manipulated VE latency while moving their head in a rhythmic pattern. They reported whether the stimuli differed in appearance. The presented study here employed the same experimental methodology. Instead, the visual scene was a pre-computed radiosity rendering of two interconnected rooms that include real-world objects. Here, we also compare sensitivity results derived from Ellis et al. [2004] and the study presented in this paper. Both studies also explore whether relative motion shear between more than one artificial object in the VE could be a mechanism contributing to observer perception of head tracking latency.

2 Psychophysics

Psychophysics is an area of perceptual psychology that employs specific behavioral methods to study the relation between physical stimulus intensity and sensation reported by a human (or animal) observer. [Lederman, 2002]. The amount of change in a stimulus required to produce a just noticeable difference in sensation is defined as the *difference threshold*. For example, if the intensity of a stimulus is 10 units and the stimulus must be increased to 12 units to produce a just noticeable increment in sensation, then the difference threshold would be 2 units [Gescheider 1997]. In recent times, the term *difference threshold (DL)* has been used interchangeably with the term *Just Noticeable Difference* (JND). The basic procedure for measuring thresholds is to present a stimulus to observers and asking them to report whether they perceive the stimulus. Biological systems (e.g., humans), however, are not fully deterministic in their reactions. Therefore, such thresholds are defined and studied in a statistical manner detailed below.

In the studies discussed here, participants' psychophysical functions for the discrimination of latency were measured with a two-interval, two-alternative forced-choice (2AFC) technique. The two intervals are the reference stimulus (R) (i.e., the standard) and the probe stimulus (P), which may or may not be different from the standard. The standard stimulus, R, in this and the related study [Ellis et al. 2004] was held to a particular constant, with the order of presentation of reference and probe stimuli randomized. The standard stimulus for the particular VE used in this study was based on the system's 12.5 ms minimum latency setting. In these two studies, the observers were forced to choose between whether the two viewed intervals were 'different' or the 'same' (i.e., no difference was observed).

3 Materials and Methods

3.1 Apparatus

The VE system employed in this work included a single receiver Polhemus FasTrak for motion sensing running at 120 Hz and a Virtual Research V8 HMD. Separate software applications to interface to the FasTrak (a customized version of AuTrak by AuSIM, Inc., Mountain View CA) and to model and render the experiment VE were written in Visual C++ for use under Windows 2000. All software ran on a Dell Precision 530 workstation equipped with dual 2.4 GHz Xeon processors and an NVidia 3D graphics card. For the simple (~100 polygon) environments in [Ellis et al. 2004], a graphics card based on the GeForce4 MX-440 graphics processing unit (GPU) was used. The more complex (~35,000 polygons) radiosity environment in this study was run on a card with an NVidia GeForce FX-5900 GPU. All environments were displayed in stereo at VGA resolution. For the HMD's specified 48°H X 36°V Field of View (FOV) at 100% binocular overlap, this resolution corresponds to 4.5 arcmin/pixel.

The internal VE system latency was controlled and measured for both the simple and complex environments [Hill et al. 2004]. We define the internal latency as the portion of the end-to-end interval ending at the top of the video frame when the first colour channel activity is detected on the VGA input to the HMD. The internal latency is the time interval depicted in Figure 1 that includes the transduction of a mechanical event by the FasTrak [Adelstein et al. 1996] through serial transmission of the data to the host computer until the visual consequence of that event is ready to be rendered in this display [Jacoby et al. 1996]. As defined, internal latency excludes temporal components that are dependent on the specific display technology—e.g., the time to scan out the image to the bottom of the frame and the physical dynamics of the display elements themselves (e.g., TFT-LCD in the V8). Thus, internal latency has a magnitude resulting solely from the sum of processing and transport times along the single data path (i.e., pipeline), beginning with each input measurement up until the last instant before any of that input's consequences will be visible in the HMD. Moreover, as implemented in our system [Hill et al. 2004], the internal latency is fully independent of the update rates of the components that make up the pipeline. For the simple environments on the MX-440 GPU, the baseline internal latency was controlled to be ~8.5 ms [Hill et al. 2004]. For the complex environment on the FX-5900 GPU, a baseline internal latency of ~12.5 ms was maintained [Hill et al. 2004].

The components contributing to the total internal latency are shown in Figure 1. The baseline value is the minimum latency that can be sustained irrespective of the contents of the visual scene without dropping video frames. Increasing the computation or rendering load beyond the worst-case level for which the baseline latency was set would result in dropped frames that, in turn, would cause the VE's instantaneous effective update rate to fall. We define the effective update rate as that at which the image content itself is actually redrawn and not necessarily that at which the physical display device is refreshed. Thus, the effective update rate is due to the instantaneous refresh frequency of the slowest component in the pipeline between input motion and displayed output.

For either the simple or complex environment, the VE system operated at a constant 60 Hz effective rate—a limitation that is imposed by the 60 Hz refresh frequency of the V8 HMD's electronics. The statistical characteristics of the end-to-end latency were initially confirmed with an automated version of the swing-arm measurement procedure described by [Jacoby et al. 1996]. Since the control of latency and update rate was demonstrated to

be stable and without significant variance, reliable single measurements could be made directly from the lapsed time between FasTrak transmitter fields and the contents of the VGA color channel signals [Hill et al 2004].

3.2 Visual Content

During the previous study [Ellis et al. 2004], participants performed the same latency discrimination task across three visual conditions that were distributed within subjects according to a Latin-square experiment design [Coolican 1999]. The three environments, shown in Figure 2, included a 2 m diameter faceted sphere viewed from the inside duplicating the viewing condition used by [Allison et al. 2001], a hollow octahedral frame, and an environment that included both the sphere and octahedron. The scene involved in this study was a pre-computed radiosity rendering of two interconnected rooms (4 X 4 m each) with objects placed at various heights. The two rooms were connected by a wall with a large opening (Figure 3).

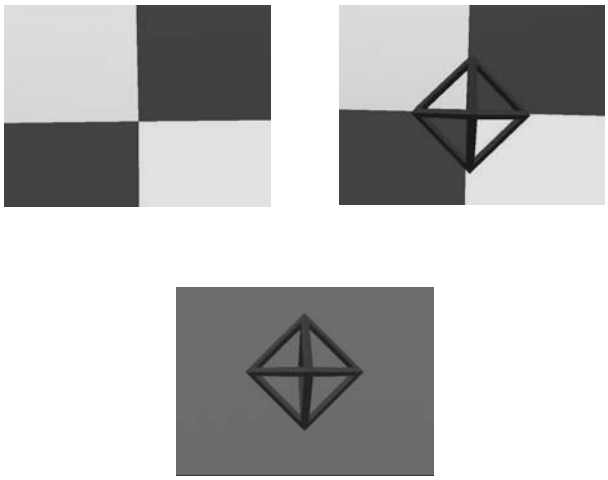

Figure 2. Experimental visual conditions from the previous study [Ellis et al. 2004].

Figure 3. Experimental visual condition.

Radiosity algorithms display view-independent diffuse interreflections in a scene assuming the conservation of light energy in a closed environment. The surfaces of the scene are broken up into a finite number of *n* discrete patches, each of which is assumed to be of finite size, emitting and reflecting light uniformly over its entire area. The result of the radiosity solution employed here was an interactive three-dimensional representation of light energy in the environment allowing for soft shadows and colour bleeding that contributed towards a photorealistic image but without any specular reflections. The final ~35,000 polygon scene was rendered with one incandescent light source in each room using the same spherical photometric web (.ies format) without any texturing or further optimization. A photometric web is a diagram that represents the three-dimensional flow of light energy from a light source. This study's scene presented a much more complex VE than we had used previously, and included a much greater variety and number of contours, textures and depth planes.

The room was fixed at eye height, with the far wall ~6 m from the seated participant's head yaw axis. The room position was automatically adjusted for each participant's eye height.

3.3 Participants

Ten participants (6M-4F, ages 25-45) were recruited for this study. All but one of these had participated in the previous study [Ellis et al. 2004]. All participants had normal or corrected to normal vision and no reported neuromotor impairment. With the exception of two participants (authors KM - participant #5, and MH - participant #4), all were naïve to the exact purpose of the experiment.

3.4 Procedure

The participants were instructed to yaw their head smoothly from side-to-side (with ~30° motion extent). The end-to-end motion was sized so that each visual scene would span nearly the entire 48° horizontal FOV of the HMD, while still maintaining the scene within view. If they turned too far, the scene darkened by 58% to signal excessive rotation. Participants were advised not to activate this cue after getting accustomed to the yaw motion amplitude so that their average rotation actually ranged between 20-30°

Participants were paced by computer-generated beeps at 1 second intervals. They used the first beep interval without moving to establish the side-to-side motion period and the remaining four intervals to complete two full back-and-forth yaw cycles at this rhythm.

As discussed in Section 2, reference (R) and probe (P) stimuli were presented sequentially in randomised-order pairs. Participants used a 3-button hand controller (Figure 4) both to signal their 2AFC response , as well as to advance to the next stimulus pair.

Figure 4. Experimental set-up.

In the interest of shortening the duration of each participant's involvement, we employed an adaptive staircase method rather than the lengthy method of constant stimuli approaches that we used previously [Ellis et al. 1999ab]. In the present studies, we employed a Two-Down, One-Up (2D-1U) stepping procedure [Levitt 1970]—i.e., two consecutive 'different' responses diminishes the added latency, while a single 'same' response augments the latency. In an adaptive method, the initial step size is diminished at each reversal (when the observer's changes from response of 'same' to two consecutive 'different' or vice versa) until the final latency step size is achieved. Once the final step size is achieved, the procedure is continued until a sufficient number of response transitions have been recorded. The adaptive staircase method is an efficient means of threshold measurement because it allows the experiment to focus most of the stimuli in a region of interest that is near the final staircase settling level. The 2D-1U staircase has a theoretical settling level that corresponds to the 70.7% threshold, which is helpful for exploring the region of the underlying psychometric function between the 50% (the Point of Subjective Equality or PSE) and 75% detection rates (one JND higher).

Preliminary observations were used for setting the approximate range of latency values between the lowest at which it is seldom perceived and the highest at which it is almost always perceived. Based on data from our prior studies [Ellis et al. 1999ab; Adelstein et al. 2003], staircases either started with the probe, P, set equal to the reference, R, and ascended, or with P set to 133 ms above R and typically descended. The adapting procedure began with a latency step size of 66.7 ms that was halved at each reversal until reaching a final step size of 8.3 ms (resolution of our VE tracker). Once the final step size was achieved, the procedure continued through seven more reversals. Finally, pairs of ascending and descending staircases were concurrently interleaved to minimize the possibility that participants could track their progress through an individual staircase. Throughout the experiment, a record of the 'different' or 'same' responses was kept.

The present study comprised a single scripted block of three paired staircases. The reference stimulus for this study was set to a stable internal latency of 13.5 ms—greater than the 12.5 ms baseline for the more complex radiosity environment in order to ensure running at the VE system's maximum 60 Hz effective frame rate.

Prior to the experiment, participants were given a broad explanation of latency and its visual consequences. Participants were instructed to report any apparent differences between stimulus pairs that could relate to delay, visual lag, oscillation, or visual instability of the scene. It was suggested they should associate these artifacts with how they perceive real-world surroundings and objects as being stable during head motion. This discussion was followed by a brief demonstration by the experimenter of the head motion required to complete the task. The instructions and the task were the same as in [Ellis et al. 2004].

All participants were extensively trained before commencing the actual experiment, to the point they could easily detect the visual consequences of ~120 ms latency. At that point, the experiment started. The duration of these training sessions varied according to the individual. In some cases, warm-up training was utilized for participants returning to the experiment after more than a week's absence. Also, in certain cases, participants were asked to repeat a section when they appeared tired or not focused on the task.

Participants completed as many sections (or paired staircases) as they could handle per session, taking rest breaks every 10-15 minutes. By the completion of the prior study [Ellis et al. 2004], all participants were highly practiced, having spent an estimated 16-20 hours each, sometimes spread over a few weeks, in the laboratory. Participants required an additional 3-4 hours each to complete the present study, sometimes spread over a few days.

4 Results

For each participant, the proportion of 'different' responses was computed from the total number of presentations at each probe stimulus level (i.e., the amount of latency added to the reference stimulus, R) accumulated from all the staircases. The resulting proportion of 'different' responses out of the total number of presented pairs at each probe level (P) can be plotted as a function (ordinate) of stimulus intensity (abscissa), as demonstrated in Figure 5 for one participant. The curve fitted to the data set, is called the psychometric function. The quality of the fit in Figure 5 is typical of all participants in this study and indicates that the estimated parameters underlying the fit provide a good representation of the data.

The psychometric function provides a statistical estimate, derived from the experiment data, of the detection rate expected for different stimulus levels. The function follows a monotonically increasing S-shape, termed an ogive, which corresponds to the diminished detection rate expected for small stimuli that grows to assured detection for large stimuli. Although more general functional forms are possible, we employ the best-fit cumulative Gaussian resulting from a Probit procedure to estimate the individual ogive for each participant and visual condition in these studies.

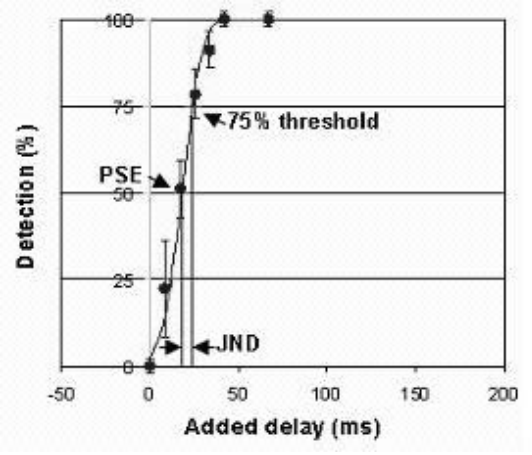

Figure 5. Typical psychometric function fitted to one subject's (#7) accumulated detection rate data from the present study. JND in this case is 5.8 ms and PSE is 17.8 ms.

The resultant Probit fits were then used to derive JNDs and PSEs for each participant as illustrated by the sample data in Figure 5. The PSE (point of subjective equality) is defined as the 50% detection level on the psychometric function. The PSE signifies the stimulus level in this study that the observer will judge with equal likelihood as being 'same' or 'different'.

The difference between PSE and the actual reference stimulus, R, represents the bias in the observer's response. Note in Figure 5 that 'Added delay,' by definition, subtracts out the reference, R, from all stimulus levels. Therefore, here, bias is equal to PSE. By convention, JND is the amount of additional stimulus needed to increase a participant's detection rate from 50% (PSE) to 75% as estimated by the fitted Gaussian psychometric function. Thus, the steeper the slope of the fitted psychometric function in the region of PSE (as in Figure 5), the smaller will be JND.

It is important to emphasize that because the fitted psychometric function is a statistical model of the observer's stochastic perceptual process, JND does not simply impose a hard threshold above or below which detection is turned on or off. JND represents a gradation over which the likelihood of detection will change from one defined level to another. In this case, because the psychometric function is represented by a cumulative Gaussian distribution, the increment of additional stimulus represented by JND is proportional to that distribution's standard deviation.

Average JNDs and PSEs for the three visual conditions in [Ellis et al. 2004] are reported in Table 1. The statistical significance of any differences between visual conditions was investigated by ANalysis Of Variance (ANOVA) [Coolican 1999]. Significance decisions involve rejecting or retaining the null hypothesis (which claims that groups are identical). The null hypothesis is typically rejected when the probability that a result occurring under it is less than 0.05. The ANOVA for all thirteen participants in [Ellis et al. 2004] did not reveal a statistically significant effect of visual condition on either JND ($F_{2,24} = 0.459$; $p < 0.642$; n.s.) or PSE ($F_{2,24} = 0.536$; $p < 0.592$; n.s.). Additional detail and discussion on the prior study can be found in [Ellis et al. 2004].

N =13	**Both**	**Object Only**	**Back-ground Only**	**Room**
JND	12.6 (2.0)	15.0 (2.7)	12.5 (2.5)	
PSE	32.3 (5.2)	29.0 (4.9)	33.3 (4.5)	
N = 9				
JND	11.1 (2.3)	11.5 (1.9)	12.2 (3.6)	8.0 (1.3)
PSE	25.4 (5.6)	28.0 (6.2)	25.9 (4.0)	13.9 (3.0)

Table 1. JNDs and PSEs for each visual conditions for all thirteen participants in [Ellis et al. 2004] and separately for the nine participants who took part in both studies. Standard error of means in parentheses. *N* = number of participants.

Subject	**JND**	**PSE**
1	17.3	30.8
2	6.1	7.7
3	4.4	8.4
4*	19.3	18.0
5	8.7	16.1
6	6.1	7.7
7	5.8	17.8
8	7.5	17.5
9	5.7	-0.5
10	10.4	19.2

Table 2: JNDs and PSEs derived from the psychometric function fitted to each subject's accumulated detection data. (*) indicates subject that did not participate in the prior study [Ellis et al. 2004].

The JNDs and PSEs for the 10 participants in this study were respectively 9.1±1.6 ms and 14.3±2.7 ms (mean ± standard error). The quality of the individual psychometric function fits (e.g., Figure 5) from which the JNDs and PSEs in Table 2 were obtained was uniformly very high ($0.964 < r_{pearson} < 0.998$), as determined by correlations between detection rate data and the corresponding fitted ogive levels.

Differences between the two studies were examined for the nine participants who participated in both. For this subgroup of nine participants, ANOVAs again did not show statistically significant differences between the visual conditions in [Ellis et al. 2004] for either JND ($F_{2,16} = 0.049$; $p < 0.952$) or PSE ($F_{2,16} = 0.235$; $p < 0.794$). Because of this absence of significant differences, participants' average JND and PSE from the three viewing conditions were employed for comparisons with the results from the second study. Pairwise contrasts for the nine participants showed that averaged JNDs in [Ellis et al. 2004] (11.6±1.6 ms) did not differ significantly ($t = 1.970$; $df = 8$; $p_{two\text{-}tail} < 0.084$) from those of the present study (8.0±1.3 ms), an observation confirmed by a nonparametric sign test ($p < 0.090$). On the other hand, while a paired contrast failed to demonstrate significance ($t = 1.952$; $df = 8$; $p_{two\text{-}tail} < 0.087$) for the change in PSE from the averages of [Ellis et al., 2004] (25.7±4.9 ms) to the radiosity-rendered room (13.9±3.0 ms), the reduction seen in eight of the nine subjects was significant ($p < 0.020$) by a sign test. Finally, this significance pattern for JND and PSE was unchanged by removal of author KM's data from the t and sign test analyses.

5 Discussion

In general, the results for JND from the present study overlaps well with the 8 to 17 ms JNDs measured in earlier investigations with an immersing HMD [Adelstein et al. 2003] as well as with the 19 ms JND we estimate for [Regan et al. 1999]'s non-immersing desktop CRT system. Both of these studies included only simple objects on a plain background. The consistency of the present JNDs derived from synthetic environments of differing visual complexity suggests this range may be a fundamental attribute for human perception of latency. However, this supposition is based on observations from systems with a 60 Hz update rate.

JND did not show statistically significant differences between [Ellis et at. 2004] and this study for the subgroup of nine participants involved in both studies, just as no differences were detected between the three visual conditions of [Ellis et al. 2004]. This indicates that the presence of multiple objects with relative image shear (versus single objects) did not significantly enhance the detection of VE latency differences. Backgrounds such as the sphere in the previous study and the rooms here, which would be expected to promote visual capture, were not shown to significantly affect latency detection. Most importantly, the radiosity-rendered scene depicting a meaningful real world setting also did not have a statistically significant impact. It is likely therefore that observers make use of features that are common to all four of the VEs studied, and that these latency detection cues are present regardless of whether the environment contains a single discrete object in the near field, an FOV-filling background, a combination of near and distant objects and backgrounds, or a real-world setting.

The PSE in the present two studies is much lower than the 180 to 320 ms reported by Allison et al. [2001] for the same scene content and HMD, and for similar head motion. The *background_only* condition in [Ellis et al. 2004], which matched the viewing conditions of [Allison et al. 2001], shows that the expected heightened visual capture of a fully encompassing scene versus that of a single object on an empty background does not explain the difference with Allison et al.'s result. The principle protocol differences that remain with present studies are Allison et al.'s use of single interval judgments that rely on the observer's internal reference for stability as well as their participants' training and experience over a much briefer exposure to the added latency conditions. Equipment differences include the use of mechanical bumpers rather than a visual signal to limit the amplitude of motion and the added encumbrance of a mechanical linkage for head tracking, both of which may hamper natural oculo-vestibular cues that we may rely upon for detecting visual instability in the real world.

The PSE levels reported here are also lower than the 40-60 ms PSEs previously seen for the identical 2AFC 'same'/'different' judgment, constant pacing frequency, and head yaw motion in [Adelstein et al. 2003]. This lower PSE or bias may be attributable to other differences in the experiment protocol such as the adaptive staircase technique, which shortened the overall duration of each subject's participation, or perhaps to subtle changes in the image content such as color, or the underlying improvements in the temporal stability of VE system graphics and computation hardware.

PSE levels in this study are also marginally (i.e., significant by sign test, but not by paired t-test) lower than those in [Ellis et al. 2004], which is suggestive of a criterion shift for psychophysical judgments that were made under the same staircase procedure. Since this study was carried out after completion of the experiments in [Ellis et al. 2004], an order confound prevents ascertaining whether the improvement in PSE may be due to training or some other feature of the experiment. Moreover, because the head motions and psychophysical tasks were performed without explicitly providing error-correcting feedback to the participants, and because latencies were continually varied according to the staircase procedure between probe and reference stimulus exposures, it is doubtful that sensorimotor adaptation to augmented latency would occur in our experimental situation and account for the PSE difference between the two studies.

Additionally, potential basement effects are a consideration in attempting to quantify changes and measure the significance of differences in JND and PSE between the various visual conditions in [Ellis et al. 2004] and the present study, as can be seen from the already low values for individual participants in Table 2. First, JND cannot be less than zero. Additionally, the stimulus resolution imposed by the minimum 8.3 ms of the latency step size, along with computational limits of the Probit procedure could prevent the fitted psychometric function from capturing the instantaneous transition from zero to full detection. Moreover, in practice, for real observer data, there will always remain some variability, which will result in positive-valued JND. For PSE, the

nature of the question posed can affect the outcome. In the present latency discrimination study and in [Ellis et al. 1999ab; Adelstein et al. 2003; Ellis et al. 2004], rather than identify which interval has the shorter latency, subjects simply answered whether the interval pairs appeared to be the same or different. Unlike psychophysical responses requiring definitively correct answers (e.g., identification), which theoretically have zero bias for properly counterbalanced stimulus presentation, our studies' judgments can also have positive biases that, in the limit, approach zero. (Note that the one slightly negative PSE in Table 2 is an artifact of the Probit fitting procedure, and is also essentially zero in magnitude.)

In summary, it can be inferred from this study that when unburdened with any other performance tasks, well-practiced subjects learn to discriminate latency in VEs with average JND below 15 ms. This observation appears to hold regardless of scene complexity, the relative location of objects, the 'meaningfulness' of the scene context, and possibly the degree of photorealism. These results provide guidelines to help decide when the implementation of latency management strategies such as predictive compensation (e.g., [Jung et al. 2000]) is necessary.

6 Acknowledgments

This work was supported by the NASA-AOS program, and the ONR-VIRTE program.

7 References

AZUMA, A. & BISHOP, G. 1994. Improving static and dynamic registration in an optical see-through display. In *Proc. of ACM SIGGRAPH '94*, 197-204.

ALLISON, R.S., HARRIS, L.R., JENKIN, M., JASIOBEDZKA, U. & ZACHER, J.E. 2001. Tolerance of temporal delay in virtual environments. In *Proc. of IEEE Virtual Reality 2001*, 247-253.

ADELSTEIN, B.D., JOHNSTON, E.R. & ELLIS, S.R. 1996. Dynamic response of electromagnetic spatial displacement trackers. *Presence: Telepresence and Virtual Environments*, *5*, 3, 302-318.

ADELSTEIN, B.D., LEE, T.G., ELLIS, S.R. 2003. Head tracking latency in Virtual Environments: Psychophysics and a model. In *Proc. of the 47th Annual Meeting Human Factors and Ergonomics Society*, 2083-2087.

COOLICAN, H. 1999. *Research Methods and Statistics in Psychology*, 3rd edition. Hodder & Stoughton.

CUNNINGHAM, D.W., BILLOCK, V.A., & TSOU, B.H. 2001a. Sensorimotor adaptation to violations in temporal contiguity. *Psychological Science,* 12, 6, 532-535.

CUNNINGHAM, D.W., CHATZIASTROS, A., VON DER HEYDE, M. & BULTHOFF, H.H. 2001b. Driving in the future: Temporal visuomotor adaptation and generalization. *Journal of Vision,* 1, 88-98.

ELLIS, S.R. BRÉANT, K., MENGES, B.M., & ADELSTEIN, B.D. 1997. Operator interaction with virtual objects: effects of system latency. *Proc. HCI International*, 973-976.

ELLIS, S.R. MANIA, K., ADELSTEIN, B.D., & HILL, M.I. 2004. Generalizability of latency detection in a variety of virtual environments. To appear in *Proc. of the 48th Annual Meeting Human Factors and Ergonomics Society.*

ELLIS, S.R., YOUNG, M.J., EHRLICH, S.M., & ADELSTEIN, B.D. 1999a. Discrimination of changes of rendering latency during voluntary hand movement. In *Proc. of the 43th Annual Meeting Human Factors and Ergonomics Society*, 1182-1186.

ELLIS, S.R., YOUNG, M.J., ADELSTEIN, B.D. & EHRLICH, S.M. 1999b. Discrimination of changes in latency during head movement. In *Proc. Computer Human Interfaces*, 1129-1133.

ELLIS, S.R. WOLFRAM, A. & ADELSTEIN, B.D. 2002. Large amplitude three-dimensional tracking in augmented environments: a human performance trade-off between system latency and update rate. In *Proc. of* the 46th *Annual Meeting Human Factors and Ergonomics Society*, 2149-2154.

GESCHEIDER, G.A. 1997. *Psychophysics: The Fundamentals*, 3rd edition. Laurence Erlbaum Associates, London.

HELD, R., EFSATHIOU, A. & GREENE, M. 1966. Adaptation to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology* 72, *6*, 887-891.

HILL, M.I., ADELSTEIN, B.D., & ELLIS, S.R. 2004. Achieving Minimum Latency in Virtual Environment Applications. In *Proc. of IMAGE Society Annual Conference*, Scottsdale AZ.

JACOBY, R., ADELSTEIN, B.D., ELLIS, S.R. 1996. Improved temporal response in virtual environment hardware and software. In *Proc. SPIE Conference on Stereoscopic Displays and Applications VII, Vol. 2653*, 271-284.

JUNG, J.Y., ADELSTEIN, B.D., ELLIS, S.R. 2000. Discriminability of prediction artifacts in a time delayed virtual environment. In *Proc. of the 44th Annual Human Factors and Ergonomics Society Meeting*, 499-502.

LEDERMAN, S.J. 2002. *Haptic Research: Psychophysics*, http://haptics.mech.nwu.edu/Psychophysics.html.

LEVITT, H. 1970. Transformed up-down methods in psychoacoustics. *J. Acoustical Soc. of America* 49, *2*, 467-477.

LIU, A., G. THARP, S. LAI, L. FRENCH, and Stark, L.W. 1993. Some of what one needs to know about using head-mounted displays to improve teleoperator performance. *IEEE Transactions on Robotics and Automation* 9, *5*, 638-48.

MANIA, K., TROSCIANKO, T., HAWKES, R., CHALMERS, A. 2003. Fidelity Metrics for Virtual Environment Simulations based on Human Judgments of Spatial Memory Awareness States. *Presence: Teleoperators and Virtual Environments* 12, 3, 296-310.

MCCANDLESS, J.W., ELLIS, S.R., & ADELSTEIN, B.D. 2000. Localization of a time-delayed monocular virtual object superimposed on a real environment. *Presence: Teleoperators and Virtual Environments* 9, 1, 15-24.

MEEHAN, M., RAZZAQUE, S., WHITTON, M., BROOKS, F. 2003. Effect of Latency on Presence in Stressful Virtual Environments. In *Proc. of IEEE Virtual Reality '03*, 141-148.

NEMIRE, K., JACOBY, R.H. & ELLIS, S.R. 1994. Simulation fidelity of a virtual environment display. *Human Factors* 36, *1*, 79-93.

REGAN, M., MILLER, G., RUBIN, S. & KOGELNIK, C. 1999. A real-time low-latency hardware light-field renderer. In *Proc. of ACM SIGGRAPH'99*, 287-290.

SHERIDAN, T.B. 1999. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments* 1, 1, 212-227.

SHERIDAN, T.B. & FERRELL, W.R. 1963. Remote manipulative control with transmission delay. *IEEE Transactions on Human Factors in Electronics* 4, 1, 25-29.

SMITH, K.U. & SMITH, W.M. 1962. *Perception and Action.* Saunders, Philadelphia, 247-277.

SMITH, W.M., MCCRARY, J.R. & SMITH, K.U. 1962. Delayed visual feedback and behavior. *Science,* 132, 103-1014.

UNO, M. & SLATER, M. 1997. The sensitivity of presence to collision response. In *Proc. of IEEE VRAIS 1997*, 95-101.

WATSON, B., WALKER, N., WOYTIUK, P., RIBARSKY, W. 2003. Maintaining Usability during 3D Placement despite delay. In *Proc. of IEEE Virtual Reality 2003*, 133-140.

WELCH, R.B., BLACKMON, T.T., LIU, A., MELLERS, B.A., STARK, L.W. 1996. The effects of pictorial realism, delay of visual feedback and observer interactivity on the subjective sense of presence. *Presence: Teleoperators and Virtual Environments* 5, 3, 263-273.