HEAD TRACKING LATENCY IN VIRTUAL ENVIRONMENTS: PSYCHOPHYSICS AND A MODEL

Bernard D. Adelstein[1], Thomas G. Lee[2], Stephen R. Ellis[1]

[1]NASA Ames Research Center
[2]San Jose State University Foundation
Moffett Field, CA 94035-1000

Quantification of perceptual sensitivity to latency in virtual environments (VEs) and elucidation of the mechanism by which latency is perceived is essential for development of countermeasures by VE designers. We test the hypothesis that observers use "image slip" (i.e., motion of the VE scene caused by system time lags) to detect the consequences of latency rather than explicitly detecting time delay. Our presumption is that forcing observers to change from constant rate to randomly paced head motion will disrupt their ability to discriminate latency based on perceived image slip. This study indicates that the disruption in motion pattern causes a shift in latency detection criteria and a minor degradation in discrimination ability. It is likely therefore that observers make at least some use of image slip in discriminating VE latency. It can also be inferred that when observers learn to discriminate latency, their Just Noticeable Difference (JND) remains below 17 ms.

INTRODUCTION

Latency, or time delay, from input action to visual display is an acknowledged shortcoming of current virtual environment (VE) and teleoperation (TO) technology. Excessive latency has long been known to hinder adaptation to spatial rearrangements (Held, Efsathiou, & Greene, 1966) and to degrade manual performance, forcing users to slow down to preserve manipulative stability (Sheridan & Ferrell, 1963). Both the quantification of perceptual sensitivity to latency and description of the mechanism by which VE latency is perceived will be essential to guide system designers in the development of countermeasures such as predictive compensation (e.g., Azuma & Bishop, 1994; Jung, Adelstein, & Ellis, 2000). Without these countermeasures, latency will remain an unavoidable consequence of the finite processing time of sensors, computation, and image rendering, as well as transmission delays and video display scan times inherent to VE and TO systems.

Several laboratories have investigated the perceptibility of latency within VEs. Allison, Harris, Jenkin, and Jasiobedzka (2001) observed with large virtual objects occupying the full head-mounted display (HMD) Field of View (FOV) that 50% thresholds for perceived image instability (oscillopsia) decreased from 320 to 180 ms as head velocity for 45° rotations increased from 22.5°/s to 90°/s. Regan, Miller, Rubin, and Kogelnik (1999), on the other hand, found 70.7% latency detection thresholds that were much lower, averaging 15 ms for a specialized single dimensional non-immersive CRT display.

Our previous studies of latency discrimination in VE hand (Ellis, Young, Ehrlich, & Adelstein, 1999a) and head tracking (Ellis, Young, Adelstein, & Ehrlich, 1999b) have shown observers are able to make reliable relative latency judgments. Using the Method of Constant Stimuli, false alarms fell below the rates expected for random guessing, remaining uniform across all added latencies for the three baseline levels (i.e., the references or pedestals) tested. Moreover, though not interpreted by Ellis et al. (1999a,b), the average Just Noticeable Difference (JND) and Point of Subjective Equality (PSE) for latency discrimination can be estimated from the plotted data as ~15-20 ms and ~50 ms respectively. These psychometric quantities appeared to be invariant across the three pedestals (33, 100, and 200 ms) in both studies. The apparent invariance of the detection function in Ellis et al. (1999a,b) demonstrated that the classic Weber's Law of psychophysics (JND $\propto$ pedestal) did not hold in this case. Were Weber's Law to hold, the slopes of these detection functions should have splayed in order to keep constant the ratio of JND to pedestal latency.

This invariance suggested that observers might have responded solely to "image slip" in the VE rather than the explicit time delay between input head motion and its displayed consequences (Ellis et al., 1999a). We define "image slip" as the virtual scene's artifactual concomitant motion with the observer's head resulting from time lag. In the absence of any lag, the virtual scene ideally would appear spatially fixed. Image slip therefore should be observable as displacement, velocity, and other kinematic quantities, either individually or in combination. Figure 1 illustrates image slip in terms of the displacements that would result from increasing time delay for a hypothetically sinusoidal, back-and-forth head motion of the type used both in Ellis et al. (1999a,b) and the present study. From the displacement curves in Figure 1, or from similar curves that might be plotted for other kinematic quantities, it could be postulated that observers discern latency differences by detecting changes in image slip features such as peak or RMS magnitude or the location of the peak in the motion cycle. It can be demonstrated mathematically that any of these slip features are dependent on the sinusoidal frequency of the head motion. This observation stands in contrast to the alternative that time delay is observed directly and is therefore independent of head motion frequency.

If image slip is considered in terms of displacement, as represented in Figure 1, different amounts of base latency (onto which incremental latency changes are added) would simply have the effect of shifting the slip location in front of the observer. Consequently, if latency discrimination were based solely on image slip, observers in Ellis et al. (1999a,b) would not be expected to exhibit Weber's Law performance. Interestingly, as shown in Figure 2, for certain sinusoidal motion frequencies, the amount of added incremental delay produces the same proportion of image slip in displacement, regardless of the latency pedestal.

The objective of the study is to determine whether image slip is the mechanism by which observers distinguish between differing latencies. In our prior studies, subjects were asked to compare two sequential observations while moving their head in a constant pattern determined by the same single pacing frequency. In those comparisons, subjects were simply looking for any difference between the presented conditions and we had no means to determine whether they might have relied on a directly observable change in delay time or on a change in image slip. In this study, a new head movement condition is added in which we randomly force subjects to move at different pacing frequencies for the two sequential observations. Were image slip (with its dependence on frequency) the sole proximate stimulus for latency, such a disruption of the constant pacing pattern should cause discrimination performance to deteriorate toward random levels.

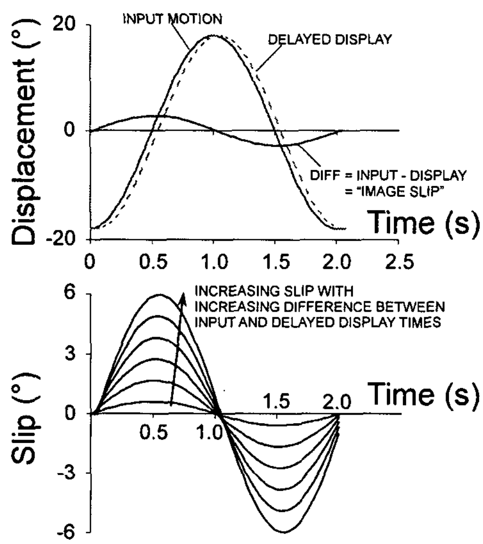

Figure 1. Image slip model for sinusoidal head motion for ±18° head yaw at 0.5 Hz with 33 ms base latency plus additional 16.7 ms delay increments. (Top) Because of inherent VE time lag and experimentally added latency, observer input head motion results in delayed motion in the display. The time difference between the input motion and the displayed consequence causes the image slip. (Bottom) The magnitude of the slip grows as the number of added delay steps increases from 0 to 5.

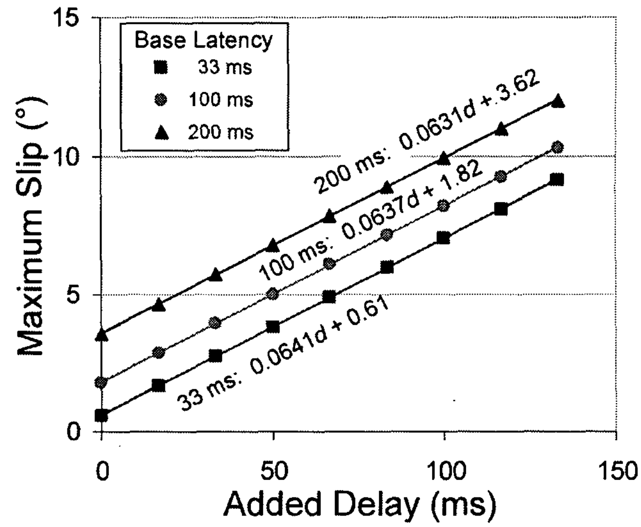

Figure 2. Image slip as a function of time delay for ±18° sinusoidal head yaw at 0.5 Hz. Each point marks the maximum slip displacement for 33 ms (lower line), 100 ms (middle line), or 200 ms (upper line) of base latency plus added 16.7 ms delay increments. The regression for each line is of the form $Ad + b$, where d is the added delay. The slope, A, is approximately equal for all three lines while the intercept, b, changes proportionally to the baseline.

MATERIALS AND METHODS

Subjects

Eleven observers (9 M and 2 F, age 20-29), either laboratory members or paid subjects, participated in this experiment. All had normal or corrected to normal vision and no known neuromotor impairment. With the exception of one subject (author TGL), all were naïve to the exact purpose of the experiment. One subject had extensive experience in a previous latency study. However, by the completion of this study, all participants were highly practiced, having spent an estimated 20-40 hours spread out over several weeks in the VE.

Apparatus

Participants viewed a very simple VE—an empty black environment containing only a blue octahedral frame (back-to-back right pyramids joined at their square base with apexes aligned vertically)—presented in a Virtual Research V8 HMD. Head motion was tracked at 120 Hz by a single-receiver Polhemus Fastrak. The position of the virtual octahedron coincided with a stationary second Fastrak receiver that was fixed to a bench top in the laboratory at eye-height ~80 cm in front of the seated viewer's head yaw axis. At this distance, the octahedron occupied a horizontal visual angle of 6°. The VE software, developed using Sense8's WTK API, ran on a 4-CPU dual-pipeline SGI Onyx computer with RE-2 graphics. Custom tracker drivers and a multi-processing,shared-memory architecture (Jacoby, Adelstein, & Ellis, 1996) ensured a 33±5 ms (mean ± stdev) base latency and constant 60 Hz update rate for the experiment VE.

Procedure

The subjects were instructed to yaw their head smoothly and sinusoidally from side-to-side (~36° end-to-end) so that the octahedron would remain fully visible while its motion spanned the HMD's entire 48° horizontal FOV. Subjects were paced by computer-generated beeps—listening to the interval between the first two beeps to establish the motion period and then moving during the remaining four intervals to complete two full back-and-forth cycles.

Latency conditions were presented in sequential pairs, one being a reference (R) and the other a probe level (P) composed of the reference level plus an added latency. R and P presentation order was pair-wise randomized. Using a 3-button hand controller, subjects advanced from one condition to the next and input their two-alternative forced-choice response as to whether the intervals were the same or different. No instructions were given concerning features to be used in making this judgment, so participants were free to form their own criteria.

The judgments advanced according to a staircase algorithm (Method of Limits) with uniform 16.7 ms increments. Because the increment size was limited by the HMD's and VE application's 60 Hz update rate, adaptive step sizes were not considered. The staircase method was chosen because our prior Constant Stimuli approach (Ellis et al., 1999a,b), which proved to be time consuming because of the large number of required comparisons, had demonstrated low false alarm rates.

Sessions comprised a single scripted set of 18 staircases, combining three ascending and three descending staircases for each of three (33, 100, and 200 ms) R levels. Descending staircases began with a randomly selected latency of 117, 133, or 150 ms added to R and ended when the subject responded that the paired stimuli were the same. Ascending staircases began with between one and three (randomly selected) repetitions of zero added latency during which P and R matched and ended when the subjects reported the paired stimuli to be different. Runs in which the initial zero augmentation was incorrectly judged to be different were not counted in the analysis. Participants completed two to four sessions per day, with never more than two hours VE exposure per day, including breaks between sessions and individual staircase runs.

For one-half of the study, subjects operated with a single 1 s beep interval (0.5 Hz yaw cycle). In the other half, three beep intervals of 0.5, 1, and 2 s were employed, corresponding respectively to back-and-forth yaw cycles of 1, 0.5, and 0.25 Hz. From Allison et al. (2001), these different pacing conditions would be expected to alter the discriminability for HMD image stability. To increase further the difficulty of our discrimination task under the latter condition, the beep interval was changed randomly following each individual stimulus such that comparisons of latency pairs were always made across unmatched head motion rates. Six subjects performed the single constantly paced half of the study first and then followed after by the randomly paced condition. A second group of five subjects performed in the opposite order.

Subjects continued under each pacing condition until they completed ten ascending and ten descending staircases at all three latency pedestals following stabilization of their response standard errors. The subjects' standard errors were computed separately for the two staircase directions at each pedestal from running windows of the ten previous judgments. In the work reported below, we consider only the stabilized staircases at the end of each group's first condition (epoch 1), and the next ten ascending and ten descending staircases, regardless of stabilization, immediately following the change in pacing condition (epoch 2).

Experiment Design

The experiment had a mixed design nesting pedestal latency (3 levels) and epoch (2 levels) within pacing order groups (2 levels). Dependent measures were the latency JNDs and PSEs extracted from the staircase data.

RESULTS

Response data from each subject were accumulated separately for each latency pedestal from the ten ascending and ten descending staircases making up the two epochs of interest. The accumulated data were compiled into detection rate versus added latency (the amount added above the pedestal) in each case and then fitted by Probit to a cumulative Gaussian distribution. The resultant Probit fits were then used to derive JNDs and PSEs for each subject and condition as illustrated by the sample data in Figure 3.

The JNDs and PSEs averaged across all subjects and conditions were 13.6 ± 0.6 ms (mean ± std err) and 58.8 ± 2.6 ms, respectively. The maximum JND and PSE observed under any of the conditions tested for any individual subject were 24.6 ms and 100.6 ms. Average JNDs and PSEs at each of the three latency pedestals in epochs 1 and 2 for the two subject groups are plotted in Figures 4 and 5. ANOVAs revealed a significant three-way interaction for JND (pedestal X epoch X pacing order: $F(2,18) = 4.744$, $p < .022$) and a two-way interaction for PSE (epoch X pacing order: $F(1,9) = 9.583$, $p < .013$). No significant main effects or any other interactions were observed.

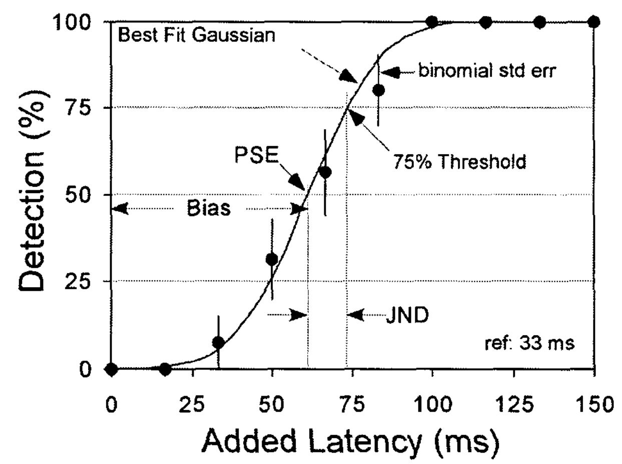

Figure 3. Sample psychometric function for one observer (IAD, epoch 1) for 33 ms latency pedestal. JND in this case is 12 ms; PSE (i.e., bias) is 61 ms.

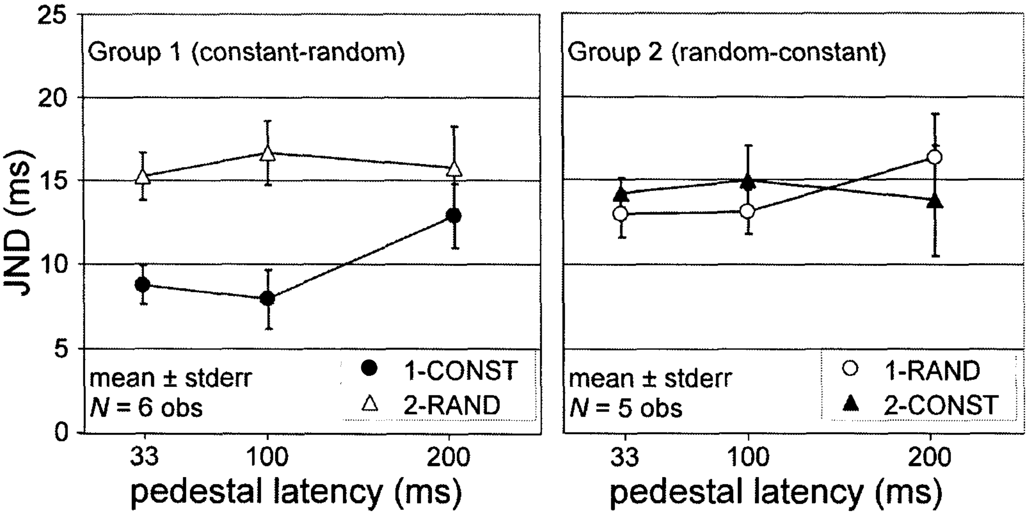

Figure 4. JNDs for epochs 1 and 2. Significant contrasts occur in Group 1 for 33 and 100 ms latency pedestals.

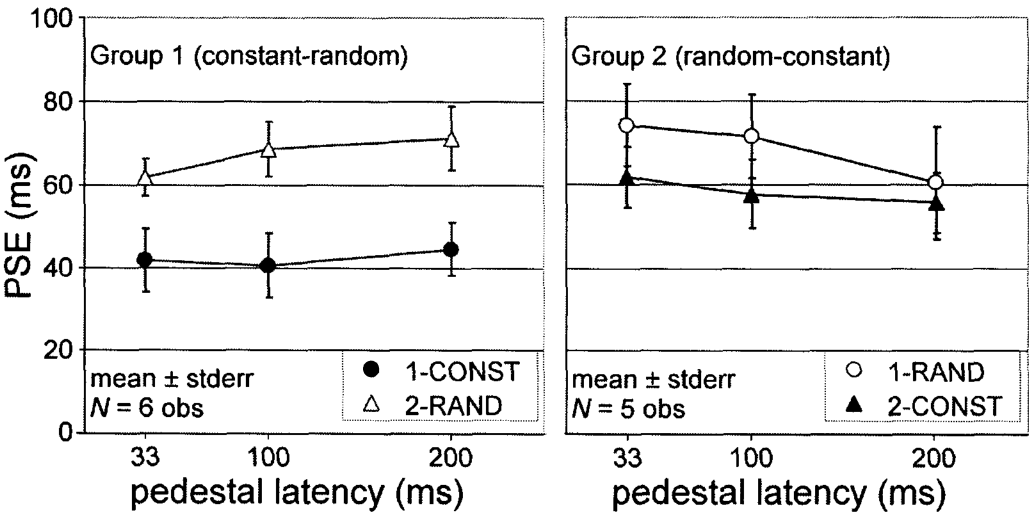

Figure 5. PSEs for epochs 1 and 2. Significant contrasts occur in Group 1 at all three latency pedestals.

Planned contrasts for Group 1, which had stabilized at constant pacing by epoch 1 and switched to random pacing in epoch 2, showed that statistically significant increases for JND occurred with the transition in pacing at the 33 ($t = 3.538$, $df = 10$, $p < .0054$) and 100 ms ($t = 3.334$, $df = 10$, $p < .0076$) pedestals, but not at 200 ms. For the 33 and 100 ms pedestals, Group 1's JNDs increased across the transition from an average of 8.4 to 16.0 ms. All PSE contrasts across the transition were significant for Group 1 (33 ms: $t = 2.244$, $df = 10$, $p < .049$; 100 ms: $t = 2.739$, $df = 10$, $p < .021$; and 100 ms: $t = 2.661$, $df = 10$, $p < .024$), with the average for the three latency conditions rising from 42 to 67 ms. Group 2, which had already stabilized for random pacing by epoch 1, did not exhibit significant contrasts for either JND or PSE at any of the latency pedestals upon changing to the constant pacing.

As expected from Ellis et al. (1999a,b), contrasts for JND and PSE between the three pedestal conditions were not significant within either subject group's two epochs (i.e., between pairs of points along any individual curve in Figures 4 and 5).

DISCUSSION

In general, the present results for JND and PSE overlap reasonably well with the respective 15 to 20 ms and ~50 ms values estimated from our previous constant pacing study (Ellis et al., 1999a,b), despite the difference in psychophysical method.

These latency sensitivities, however, cannot be compared fully with those of Regan et al. (1999) and Allison et al. (2001) because of the differences in data analysis methods. The 75% thresholds (by definition the sum of JND and PSE) for this study are notably higher than Regan et al.'s (1999) average 70.7% threshold of 15 ms. Potential causes for this difference cannot be assigned without the ability to separate Regan et al.'s data into PSE, which is controlled by the observers' response bias (e.g., Engen, 1971, p. 31), and JND.

Allison et al. (2001) reported 50% thresholds, equivalent to PSE, which at 180 to 320 ms, are significantly greater than those we obtained. JND values, however, were not provided by Allison et al. (2001). The higher PSEs (i.e., observer biases) may stem from the following differences in their experimental set-up. Allison et al.'s observers employed single interval judgments in which no direct comparison to a latency pedestal were made. Their virtual target object occupied the entire visual surround (i.e., field of regard) and thus was more a background than a target. Additionally, the head motion patterns in Allison et al.'s study might have been more abrupt because head travel was limited by mechanical bumpers rather than self-decelerated by the subject. More abrupt motion changes with a properly affixed HMD, however, would be expected to enhance latency sensitivity.

The present results showed statistically significant increases in JND and PSE between epochs 1 and 2 occurred for the subjects who transitioned from constant to random pacing (Group 1) but not for those who started in the random condition (Group 2). The asymmetric transfer in epoch 2 could be expected because Group 2's single pacing condition was a subset of the three pacing rhythms to which they had previously stabilized their responses. Group 1, on the other hand, encountered two additional randomly presented rhythms for which they were unpracticed, and that we hypothesized would degrade the acuity of their latency judgments.

While JNDs increased between epochs 1 and 2 for the 33 and 100 ms latency pedestals for Group 1, the absence of a significant JND interaction at 200 ms can be ascribed to two causes. First, if the sinusoidal displacement model depicted in Figures 1 and 2 is parameterized for the fastest (1 Hz) pacing condition and the 200 ms pedestal, incremental changes in image slip (which approximates JND) become less distinct from those at the lower pedestals. Second, at such long latencies, image slip exceeds 32° (2/3 of HMD horizontal FOV), making it difficult to view the target octahedron throughout the full extent of the motion.

There is evidence from the PSE shifts and JND degradation that the initial change in pacing for Group 1's subjects to a more difficult and irregular pattern did to some degree diminish the reliability of image slip as their cue for latency. Nevertheless, subjects in Group 1 could still perform the discrimination task immediately following the transition and do so at JND levels (mean ± stderr = 16.0 ± 1.2 ms; $N = 12$ observations) not much higher than those for Group 2 before or after the transition (14.8 ± 1.1 ms before, 13.2 ± 0.9 ms after; $N = 10$ observations). Thus it is likely that observers make at least some use of image slip in discriminating VE latency.

Our inability to completely suppress the capacity for latency discrimination may be due to the extent to which observers can still factor in head motion frequency when paced in an unpracticed random fashion. Furthermore, since the averaged JNDs reported for any of the conditions remain below 17 ms (see Figure 4), it can be therefore inferred that when observers learned to discriminate latency, their discrimination is robust to movement condition variations.

Finally, a design guideline for HMD-based VEs that emerges from this study's JNDs suggests that, to be imperceptible, system latency must be no higher than 17 ms. For some conditions or individual observers, this latency may have to be smaller still. For reference, in order to maintain a VE update rate that matches that of the standard 60 Hz video display, each simulation cycle's computation must be completed within 16.7 ms.

ACKNOWLEDGMENTS

Funding for this work was provided by the AOS Project of NASA's Airspace Systems Program, and by ONR's VIRTE program. We thank Jonathan Gips for development of the experiment software used in this study.

REFERENCES

Azuma, A., & Bishop, G. (1994) Improving static and dynamic registration in an optical see-through display. *Proceedings, SIGGRAPH*, pp. 197-204.

Allison, R.S., Harris, L.R., Jenkin, M., Jasiobedzka, U., & Zacher, J.E. (2001) Tolerance of temporal delay in virtual environments. *Proceedings, VR 2001*, pp. 247-253.

Ellis, S.R., Young, M.J., Ehrlich, S.M., & Adelstein, B.D. (1999a) Discrimination of changes of rendering latency during voluntary hand movement. *Proceedings, 43rd Annual Meeting HFES*, pp. 1182-1186.

Ellis, S.R., Young, M.J., Adelstein, B.D., & Ehrlich, S.M. (1999b) Discrimination of changes in latency during head movement. *Proceedings, Computer Human Interfaces*, pp. 1129-1133.

Engen, T. (1971). Psychophysics. In Kling, J.W., Riggs, L.A. (Eds.), *Woodworth and Schlosberg's Experimental Psychology*. (3rd edition), New York, Holt, Rinehart and Winston, pp. 11-46.

Held, R., Efsathiou, A., & Greene, M. (1966) Adaptation to displaced and delayed visual feedback from the hand. *J. Exp. Psychology*, *72*(6), 887-891.

Jacoby, R.H., Adelstein, B.D., & Ellis, S.R. (1996) Improved temporal response in virtual environments through system hardware and software reorganization. *Proceedings, SPIE Conference on Stereoscopic Displays and Applications VII, Vol. 2653*, pp. 271-284.

Jung, J.Y., Adelstein, B.D., & Ellis, S.R. (2000) Discriminability of prediction artifacts in a time delayed virtual environment. *Proceedings, 44th Annual Meeting HFES*, Vol. 1, pp. 499-502.

Regan, M., Miller, G., Rubin, S., & Kogelnik, C. (1999) A real-time low-latency hardware light-field renderer. Proceedings, SIGGRAPH, Los Angeles, CA, pp. 287-290.

Sheridan, T.B., & Ferrell, W.R. (1963) Remote manipulative control with transmission delay. *IEEE Transactions on Human Factors in Electronics, HFE-4*(1), 25-29.