DISCRIMINABILITY OF PREDICTION ARTIFACTS IN A TIME-DELAYED VIRTUAL ENVIRONMENT

Jae Y. Jung[1,2], Bernard D. Adelstein[1], and Stephen R. Ellis[1]
[1]NASA Ames Research Center, Moffett Field CA
[2]University of California, Berkeley CA

Overall latency remains an impediment to perceived image stability and consequently to human performance in virtual environment (VE) systems. Predictive compensators have been proposed as a means to mitigate these shortcomings, but they introduce rendering errors because of induced motion overshoot and heightened noise. Discriminability of these compensator artifacts was investigated by a protocol in which head tracked image stability for 35 ms baseline VE system latency was compared against artificially added (16.7 to 100 ms) latency compensated by a previously studied Kalman Filter (KF) predictor. A control study in which uncompensated 16.7 to 100 ms latencies were compared against the baseline was also performed. Results from 10 subjects in the main study and 8 in the control group indicate that predictive compensation artifacts are less discernible than the disruptions of uncompensated time delay for the shorter but not the longer added latencies. We propose that noise magnification and overshoot are contributory cues to the presence of predictive compensation.

INTRODUCTION

The negative consequences of latencies in interactive display systems have long been known for manual control (Sheridan & Ferrell, 1963) and visual-motor adaptation to spatial distortions (Held, Efsathiou, & Greene, 1966). More recent work has shown that latencies in virtual environments disrupt both objective measures of performance (Liu, Tharp, French, Lai & Stark, 1993; Ware & Balakrishnan, 1994; Ellis, Bréant, Menges, Jacoby, & Adelstein, 1997; Ellis, Adelstein, Baumeller, Jense, & Jacoby, 1999) as well as subjective sense of presence (Welch, Blackmon, Liu, Mellers, Stark, 1996; Ellis, Adelstein, Baumeller, Jense, & Jacoby, 1999)

End-to-end latency in a virtual environment (VE) is due to the sum of processing time internal to sensors, simulation computation, and graphics pipeline and rendering processes, as well as communication delays both between concurrent software processes and between computer(s) and attached sensors and displays. Thoughtful re-organization of VE system hardware and software architecture can reduce system latency, increase frame rates, and decrease frame rate variability (Jacoby, Adelstein, & Ellis, 1996). However, because computation, sensor, and display processing each take finite time to execute, there is a minimum latency which, even if approachable, cannot be circumvented. For example, in our VE system, baseline latency for the simple experiment application described below was measured with timing procedures from Jacoby et al. (1996) to be 35±5 ms (mean ± stdev) for Cartesian displacements and had a steady frame rate of 60 Hz. Quaternion rotation components are 5 ms less (Adelstein, Johnston & Ellis, 1995). Additional hardware and software "tweaking" can impose tighter synchrony in our UNIX system, reducing the displacement and rotation means by 8 ms, but does so at the expense of decreasing frame rate uniformity. The theoretical limit for our experiment application is about 23 ms for displacement and 18 ms for rotation. More complex VE simulations of course impose greater computational burdens and therefore can increase latencies drastically.

Psychophysical studies in a VE with a closed head mounted display (HMD) (Ellis, Young, Adelstein, & Ehrlich, 1999a, 1999b) indicate that subjects can discriminate latency differences at least as low as 16.7 (lowest value tested) up to 116.7 ms (highest tested). Furthermore, the latency increment detection curves plotted by Ellis et al. (1999a, 1999b) were invariant across all three (27, 97 and 194 ms) tested reference latencies. This suggests that the same detection curve and minimum detectable difference might apply just as well for latency discrimination with respect to a 0 ms reference condition and that, consequently, absolute (i.e., with respect to zero) latencies ≤16.7 ms may still be discernible to the VE user. This minimum detectable latency implication is expected to be even stronger for the more stringent dynamic image registration requirements of see-through augmented reality systems (Azuma & Bishop, 1994).

The only viable approach to eliminating—or at least mitigating the consequences of—the remaining latency once VE system hardware and software has been fully optimized and synchronized is predictive compensation. Such compensators has been demonstrated for tracked head and hand movement in VE's (Liang, Shaw, & Green, 1991; Friedmann, Starner, & Pentland, 1992; Azuma & Bishop, 1994; Wu & Ouhyoung, 1995; Mazuryk & Gervautz, 1995; So & Griffin, 1996; Kiruluta, Eizenman, & Pasupathy, 1997; Akatsuka & Bekey, 1998). All these prediction technique insert a mathematical algorithm to extrapolate to a future time ahead of the current position and orientation states obtained from motion sensors or trackers measurements. Though predictors may diminish overall latency, an unavoidable side-effect of the extrapolation algorithm is the introduction of undesirable artifacts such as overshoot and increased high frequency noise. Therefore a successful predictor implementation ultimately will diminish or nullify user awareness of *apparent* VE latency while at the same time not promote perceptually excessive compensation artifacts.

While only Wu and Ouhyoung (1995) have formally evaluated the effect of predictor implementations on visually mediated manual performance, none of the other cited work has examined prediction's perceptual impact. This work represents a first formal study of user assessment of predictive compensation for head tracking in an immersive VE. The remainder of this paper proceeds with the selection of a predictor structure and parameterization for this study, a description of a method for testing predictor artifact and latency discriminability, and conludes with presentation and discussion of the study's results.

PREDICTOR SELECTION

The majority of predictive compensation work for VE's has focused on Kalman Filter (KF) based techniques, either for their primary predictor formulation (Liang et al., 1991; Friedmann et al., 1992; Azuma & Bishop, 1994; Mazuryk & Gervautz, 1995; Kiruluta et al., 1997; Akatsuka & Bekey, 1998) or as a secondary implementation against which another technique is compared (Wu & Ouhyoung, 1995). As a primary or secondary KF design, most implementations use the same basic kinematic system model to propagate measured displacement states from time step to time step (Friedmann et al., 1992; Azuma & Bishop, 1994; Mazuryk & Gervautz, 1995; Wu & Ouhyoung, 1995; Kiruluta et al., 1997; Akatsuka & Bekey, 1998). This KF model is termed kinematic because it simply states that for either translational or rotational states velocity is the derivative of displacement and acceleration the derivative of velocity. This model also assumes that the acceleration state is constant, and therefore the derivative of acceleration (i.e., jerk) is not a function of the other states and can only be directly driven by plant noise. Explanations of the KF equation development for head motion prediction can be found in (Azuma & Bishop, 1994) and (Jung, Adelstein, & Ellis, 2000).

Because Azuma and Bishop (1994) contains the only predictor development that explicitly describes the orientation prediction problem, we adopt both its exemplary kinematic model formulation and KF noise component parametrization, but with one slight difference. We use a discrete-time state-transition matrix to update system states (Jung, Adelstein, & Ellis, 2000) that does not add the extraneous dynamics of Runge-Kutta integration (Azuma & Bishop, 1994), which results in apparently better performance as quantified by simple RMS error measures (Azuma & Bishop, 1994). The importance of the orientation problem in head tracking and prediction is predicated on the trigonometry of small rotations of the head potentially producing large translational shifts in the viewed VE images. Thus, the consequences of predictor induced jitter or overshoot are typically much more salient for head orientation than translation.

To illustrate the relative consequences of jitter and overshoot artifacts and motivate further this investigation of the potential effects of predictive compensation, Fig. 1 shows a sample section of the head motion equivalent to that from the experiment below. In this figure, the compensator predicts 50 ms ahead to compensate for a 50 ms latency in the system. It is noteworthy, especially in the elevation component, that the errors induced by prediction can be as obtrusive in magnitude as the tracking error of the delayed measurement.

METHODS

VE System Hardware and Software

The VE and KF predictor software for the experiment were run on a four CPU SGI Onyx workstation with dual-pipeline RealityEngine-2 graphics. The subjects viewed the VE in a Virtual Research V8 HMD. Position and orientation of subjects' head as well as a visually presented target object were measured by separate Polhemus FasTrak instruments (i.e., control boxes), each with a single receiver and single transmitter, and each interfaced to its own Onyx ASO 115 KBaud serial port.

The VE for the experiments consisted solely of a 10 cm diameter faceted virtual sphere (i.e., target) in a dark, empty space and lit as described in (Ellis et al., 1999b). Subjects were

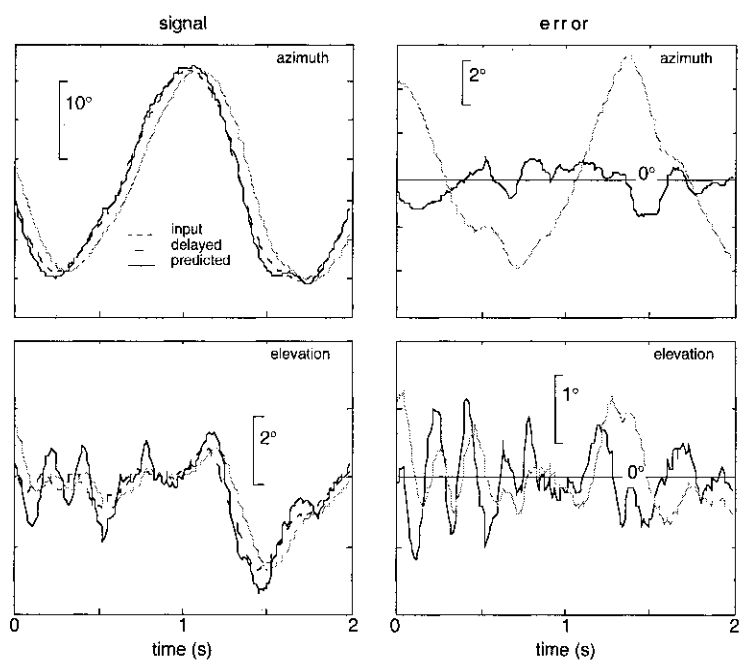

Fig. 1. Predicted head rotational components arising from a side-to-side head movement cycle (left). Input is artificially generated by shifting the acquired delayed measurements ahead by 50 ms. Errors for prediction and delayed measurement compared with input (right). The elevation components arise because the actual head motion is not a pure yaw with fixed vertical axis of rotation.

seated with the HMD's FasTrak receiver ~40 cm below the FasTrak transmitter. The virtual sphere, whose position in the VE was determined by the immobile second FasTrak receiver, was centered ~80 cm in front of the HMD. Ideally, with perfect measurement in the absence of any delay, the image of the sphere should move on the HMD LCD panels in a manner such that it appears to the observer to be fixed in space when her head moves. In the presence of inevitable delays or predictor imperfections, the virtual sphere will not be locked in space and may appear to move about its ideal fixed location.

The prediction procedure was written as a separate software process that could be interposed between the sensor data acquisition and VE simulation processes on the SGI workstation. Position data is transferred from sensor interface to predictor to VE processes via shared memory. A separate shared memory process enables experimentally controlled predictor parameters, such as prediction interval, to be revised in real time. The multi-processing, multi-processor architecture of our VE system allows the predictor to run without degradation to the other processes during our experiments. Predictor computation cycles (rotational and translational combined) rarely (< 0.05%) exceeded the 8.3 ms window required to maintain synchronization with the 120 Hz FasTrak sampling frequency.

Discrimination Experiment Protocol

The primary study aims to ascertain user awareness of any artifacts due to the presence of imperfect predictive compensation. The control examines user awareness of uncompensated end-to-end VE system latencies for the same underlying added latencies. The experimental approach used here is derived from a technique for assessing subjective detectability of changes in latency (Ellis, Young, Adelstein, & Ehrlich, 1999a, 1999b).

The procedure is based on a two alternative forced choice protocol. Seated subjects, paced by an 80 beat/min metronome (1.5 s per full back-and-forth cycle), yawed their heads through ~30° from side-to-side (See Fig. 1) while maintaining the virtual sphere in view. Using any perceivable quality in the appearance of the virtual sphere as they moved their heads, subjects were asked to judge whether sequentially presented VE conditions were the same or different and entered their automatically logged response through a hand-held push-button device. In the primary study, the VE could be running either Condition A, at the baseline 35 ms displacement latency without prediction, or Condition B, with artificial latency added to the baseline that was then matched by the predictor's compensation interval. In condition B, presumably, the underlying latency now matched that of Condition A with the only difference being the noise and overshoot artifacts induced by the predictor. In the control condition, the artificially added latency was *not* compensated. Prior to actual data collection, subjects were shown the effects of baseline minimum VE latency, and then, dependent on the study, the baseline plus 50 and 100 ms of added latency either with or without predictive compensation.

Each of six latency values (16.7 to 100 ms in 16.7 ms steps) was blocked in a randomly ordered set of 20 judgments such that each of the four possible A-B condition pairings was repeated five times. Ten subjects participated in the primary study of predictor artifact discrimination; eight were in the control study. The subjects, who were either lab members or paid recruits, all had normal or corrected to normal vision and no other known impairments.

RESULTS

Fig. 2 shows the percentage of correct discriminations between minimal VE latency and either compensated or uncompensated artificially delay grows monotonically with the increasing number of added 16.7 ms delay steps. Neither the mean proportions nor the standard errors computed for the binomial distribution of proportional data crossed the expected 50% level for random guessing given the balanced stimulus pair presentation. This implies that, on average, all conditions were discriminable from the VE system baseline latency.

A two-way ANOVA tested the effect of the added latency increment and the presence of predictive compensation on an arcsine transformation of the response proportions. The arcsine transformation converted the data to the normal distribution needed for the analysis (Sachs, 1984). The main effect of added latency on the proportion of correct responses was significant ($F = 24.902$; $df = 5,80$; $p < .001$), while the presence of predictive compensation alone was not ($F = 2.330$; $df = 1,16$; $p < 0.146$). Interaction between the two factors was significant ($F = 3.586$; $df = 5,80$; $p < .006$). This interaction result, in conjunction with Fig. 1 implies that artifacts introduced by predictive compensation may be less discernible than the disruptions attributable to uncompensated time delays for the shorter but not the longer added latencies. However, Scheffé contrasts performed on the arcsine transformed data did not reveal significant ($p < .10$) differences between compensated and uncompensated latencies—though the levels at 16.7 ms were marginal ($p < .103$).

DISCUSSION

The increasing proportion of correct judgments as latency is increased in the uncompensated control study is consistent with the latency detection levels reported by Ellis et al. (1999a, 1999b). The predictively compensated group's responses follow a similar pattern, indicating that subjects become more adept at discriminating predictor artifacts as the look-ahead interval was increased.

Subjects may rely on different cues to discriminate the presence of predictive compensation than they do for latency. In the control study, the difference between the delayed and baseline VE sphere's rendered displacement is simply the result of time lag and the consequent motion offset. In the main experiment, this difference arises not from lag, but from overshoot and noise artifacts induced by an imperfect predictor. The sample motion segments in Fig. 1 for a 50 ms latency added to the baseline, especially in the elevation plots, show substantial prediction overshoots that, in effect, trigger image instability (i.e., error) on par with those produced by an uncompensated delay. The assertion that noise and overshoot contribute to discriminability is supported by Fig. 3 showing growth in the power densities of higher frequency components that is commensurate with the growth in discriminability as the prediction interval is increased. The highlighted band corresponds to the highly oscillatory ~5 Hz activity apparent in the measured signal and that is exaggerated by the predictive compensator.

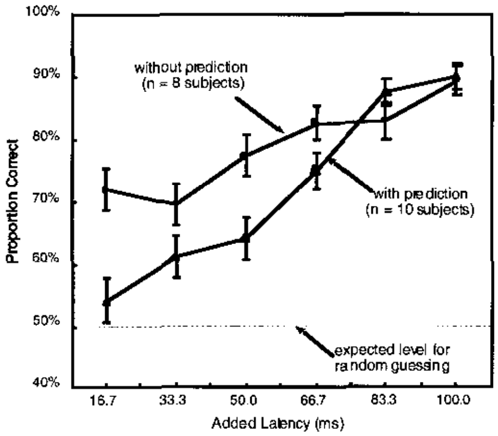

Fig. 2 Percent correct discrimination (mean ± binomial std error) as a function of latency added to the 35 ms VE system minimum.

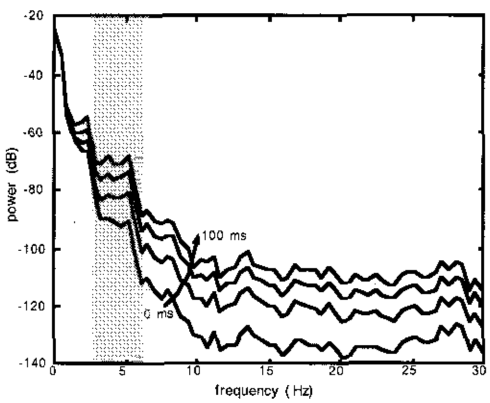

Fig. 3. Elevation component power spectra as prediction interval is increased from 0 to 100 ms in steps of 33 ms. Prediction is carried out off-line on a pre-recorded 20 s data set from which the short sample in Fig. 1 was drawn.

With the exception of one 16.7 ms step of predictive compensation, for which subjects' discrimination performance was consistent with random guessing, the predictor implementation used in this study did not offer dramatic improvement over the uncompensated latency condition. One reason might be that the KF parameterization applied in our system for these experiments was obtained in a different physical environment for a completely different VE head tracker technology (Azuma & Brown, 1994). However, when we parameterized the same KF predictor structure from optimizations for our own FasTrak sensors and the specific side-to-side head motion used in our experiments, no difference in discriminability were noted (Jung et al., 2000). Consideration of other predictor structures and parameterizations would be advisable. Psychophysical evaluations such as are presented here would be suitable in ascertaining the perceptual impact of new latency compensator designs.

REFERENCES

Akatsuka, Y., & Bekey, G.A. (1998) Compensation for end to end delays in a VR system. *Proceedings, IEEE VRAIS*

Adelstein, B.D., Johnston, E.R., and Ellis, S.R. (1996) Dynamic response of electromagnetic spatial displacement trackers. *Presence, 5*(3):302-318.

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

Brown, R.G., & Hwang, P.Y.C. (1997) *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley, New York.

Ellis, S.R., Bréant, F., Menges, B., Jacoby, R.H., & Adelstein, B.D. (1997) Operator interaction with virtual objects: effects of system latency. *Proceedings, HCI'97 International*, San Francisco, pp. 973-976.

Ellis, S.R., Adelstein, B.D., Baumeler, S., Jense, G.J., & Jacoby, R.H. (1999) Sensor Spatial Distortion, Visual Latency, and Update Rate Effects on 3D Tracking in Virtual Environments. *Proceedings, IEEE VR'99*, Houston TX.

Ellis, S., Young, M., Adelstein, B., & Ehrlich, S. (1999a) Discrimination of changes of latency during head movement . *Proceedings, HCI'99*, Munich, pp. 1129-1133.

Ellis, S., Young, M., Adelstein, B., & Ehrlich, S. (1999b) Discrimination of changes of latency during voluntary hand movement of virtual objects. *Proceedings, 43rd Ann. Meeting HFESy*, Houston, TX, pp. 1182-1186.

Friedmann, M., Starner, S., & Pentland, A. (1992) Synchronization in virtual realities. *Presence, 1*(1), 139-144.

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

Jacoby, R., Adelstein, B., & Ellis, S. (1996) Improved temporal response in virtual environment hardware and software. Proceedings, *SPIE Conf. on Stereoscopic Displays and Applications VII, Vol. 2653*, Bellingham WA, pp. 271-284.

Jung, J., Adelstein, B., & Ellis, S. (2000) Predictive Compensator Optimization for Head Tracking Lag in Virtual Environments. To appear in *Proceedings, IMAGE 2000*.

Kiruluta, A., Eizenman, M., & Pasupathy, S. (1997) Predictive head move movement tracking using a Kalman filter. *IEEE Trans. Sys., Man, Cybernetics—B, 27*(2), 326-331.

Liang, J., Shaw, C., & Green, M. (1991) On temporal-spatial realism in the virtual reality environment. *Proceedings, UIST'91) 1991*, Hilton Head, SC, pp. 19-25.

Liu, A., Tharp, G., French, L., Lai, S., & Stark, L. (1993) Some of what one needs to know about using head mounted displays to improve teleoperator performance. *IEEE Trans. Rob. Autom.*, 9(5), 638-648.

Mazuryk, T. & Gervautz, M. (1995) Two-step prediction and image deflection for exact head tracking in virtual environments. *Eurographics'95, 14*(3), C-29-42.

Sachs, L. (1984) *Applied Statistics: A Handbook of Techniques*. Springer-Verlag, New York.

Sheridan, T.B., & Ferrell, W.R. (1963). Remote manipulative control with transmission delay. *IEEE Trans. Hum. Fact. Elect., HFE-4*(1), 25-29.

So, R.H.Y., & Griffin, M.J. (1996) Experimental studies of the use of phase lead filters to compensate lags in head-coupled visual displays *IEEE Trans. Sys., Man, Cybernetics—A, Humans and Systems, 26*(4), 445-454.

Ware, C., & Balakrishnan, R. (1994) Reaching for Objects in VR Displays: Lag and Frame Rate, *ACM Trans. Comp. Hum. Interact., 1*(4) 331-356.

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, 5*(3), 263-273.

Wu, J.-R., & Ouhyoung, M. (1995) A 3D tracking experiment on latency and its compensation methods in virtual environments. *Proceedings, UIST'95*, Philadelphia, PA, pp. 41-49.