PREDICTIVE COMPENSATOR OPTIMIZATION FOR HEAD TRACKING LAG IN VIRTUAL ENVIRONMENTS

Jae Y. Jung
Graduate Student Researcher
University of California, Berkeley/NASA Ames Research Center

Bernard D. Adelstein, Ph.D.*
Engineer
&
Stephen R. Ellis, Ph.D.
Psychologist
NASA Ames Research Center

*Corresponding author

Abstract

We examined the perceptual impact of plant noise parameterization for Kalman Filter predictive compensation of time delays intrinsic to head tracked virtual environments (VEs). Subjects were tested in their ability to discriminate between the VE system's minimum latency and conditions in which artificially added latency was then predictively compensated back to the system minimum. Two head tracking predictors were parameterized off-line according to cost functions that minimized prediction errors in (1) rotation, and (2) rotation projected into translational displacement with emphasis on higher frequency human operator noise. These predictors were compared with a parameterization obtained from the VE literature for cost function (1). Results from 12 subjects showed that both parameterization type and amount of compensated latency affected discrimination. Analysis of the head motion used in the parameterizations and the subsequent discriminability results suggest that higher frequency predictor artifacts are contributory cues for discriminating the presence of predictive compensation.

Introduction

Predictive compensation has been widely considered as a means for mitigating the consequences of the time delays inherent to sensors, computation, and display rendering in virtual environment (VE) systems (Liang, Shaw, & Green, 1991; Friedmann, Starner, & Pentland, 1992; Azuma & Bishop, 1994; Mazuryk & Gervautz, 1995; Nelson, Hettinger, Haas, Russell, Warm, Dember, & Stoffregen, 1995; Wu & Ouhyoung, 1995; Zikan, Curtis, Sowizral, & Janin, 1995; So & Griffin, 1996; Kiruluta, Eizenman, & Pasupathy, 1997; Akatsuka & Bekey, 1998). While visually mediated manual tracking experiments in VEs have demonstrated the potential benefit of predictive compensation for human performance (Wu & Ouhyoung, 1995; Nelson et al., 1995), the direct perceptual impact of such compensation has never been examined.

Presented at the IMAGE 2000 Conference
Scottsdale, Arizona 10-14 July 2000

Predictive compensators for VEs operate on current position and orientation measurements to extrapolate future positions and orientations based on kinematic and dynamic models of motion via Kalman filtering or other techniques. Perfect prediction, in principle, should remove all effects of sensor-to-display latency and introduce no additional artifacts (e.g., noise or overshoot) to the VE system. Notwithstanding, we presume that a *practical* predictor need neither remove all latency nor avoid all artifacts, it need only avoid making the user aware of those compensator imperfections.

An important element of Kalman Filter (KF) predictor implementation is parameterization of the KF components. One approach has been numerical optimization to find parameter sets that minimize RMS error between the input body part motion and predicted VE output (Liang et al., 1991; Azuma & Bishop, 1994; Mazuryk & Gervautz, 1995; Kiruluta et al., 1997). Parameters can also be chosen from estimates or models of sensor and human motion characteristics (Liang et al., 1991; Friedmann et al., 1992; Kiruluta et al., 1997).

Predictor performance has also been evaluated on the basis of simple RMS error metrics (Azuma & Bishop, 1994; Kiruluta et al., 1997; Akatsuka & Bekey, 1998) as well as less rigorously from the cursory appearance of co-plotted motion and prediction traces (e.g., Liang et al., 1991; Friedmann et al., 1992). However, as can be observed from sample plotted measurement and prediction records in the cited prior studies, these parameterization and evaluation methods may imply good performance while still failing to capture fully the undesirable noise and overshoot artifacts introduced by prediction. Consequently, we propose that these prior simple metrics are incomplete indicators of the user's perceptual experience and are therefore insufficient for ascertaining the impact of predictive compensation on human performance in VEs.

This paper presents new work in two areas for VE predictive compensator design. One is an investigation of more sophisticated cost functions for optimizing KF predictor parameters. The second is an experimental method for subjective evaluation of candidate predictor parameterizations. The remainder of this paper begins with an overview of the Kalman Filter predictor formulation and a discussion of parameters available for optimizing VE head motion prediction. Observations of

predictor output from pre-recorded motion data leads to candidate parameter optimization strategies. A perceptual experiment to compare the subjective discriminability of different KF predictor parameterizations is described next. Finally, implications of the predictor optimization analysis and the discriminability experiment results are discussed.

Kalman Filtering and Prediction

Kalman Filter Estimation

A general linear continuous-time dynamic process to describe head (or other body part) motion is

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{w}(t) \qquad \text{(Eq. 1)}$$

where $\mathbf{x}(t)$ is the vector of system state variables, $\mathbf{A}(t)$ is the system matrix describing the dynamic relationships between the state variables, and $\mathbf{w}(t)$ is the driving plant noise (assumed zero-mean Gaussian and uncorrelated between states). Plant noise comprises random inputs that drive displacements and their higher order time derivatives. For a head tracked VE system, these inputs include low-frequency volitional commands as well as involuntary, higher frequency signals such as those that might be driving head tremors.

When sampled at uniform time interval h, Eq. (1) is rewritten as the discrete system

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \mathbf{w}(k) \qquad \text{(Eq. 2)}$$

in which k represents the sampled values at time $t = t_k = kh$. The state transition matrix, $\Phi(k)$, is the matrix exponential of the scalar sampling interval, h, multiplied by the system matrix, $\mathbf{A}(t)$, evaluated at the beginning of the time interval, t_k:

$$\Phi(k) = e^{h\mathbf{A}(t_k)} \qquad \text{(Eq. 3)}$$

Measurements reported by the sensor are given by

$$\mathbf{y}(k) = \mathbf{C}(k)\,\mathbf{x}(k) + \mathbf{v}(k) \qquad \text{(Eq. 4)}$$

$\mathbf{C}(k)$ combines the states contributing to the measurement vector $\mathbf{y}(k)$. $\mathbf{v}(k)$ is a noise vector, again assumed to be zero-mean, Gaussian, and uncorrelated between measurement channels, that contributes to random variations in sensor output. One typical source of sensor noise in VEs is the electromagnetic interference associated with tracker induced image jitter.

Given the discrete-time expressions for the head tracking and measurement processes of Eqs. (2) and (4), the equation

$$\begin{aligned}\hat{\mathbf{x}}(k+1) &= \Phi(k)\hat{\mathbf{x}}(k) \\ &\quad + \mathbf{K}(k+1)\left[\mathbf{y}(k+1) - \mathbf{C}(k+1)\Phi(k)\hat{\mathbf{x}}(k)\right]\end{aligned} \qquad \text{(Eq. 5)}$$

employs an observer, $\mathbf{K}(k+1)$, to update an estimate, $\hat{\mathbf{x}}(k)$, of the state vector $\mathbf{x}(t)$.

The *optimal* observer, i.e., the Kalman Filter gain, is calculated from

$$\begin{aligned}\mathbf{K}(k+1) &= \mathbf{P}(k+1)\mathbf{C}(k+1)^T \times \\ &\quad \left[\mathbf{C}(k+1)\mathbf{P}(k+1)\mathbf{C}(k+1)^T + \mathbf{V}(k+1)\right]^{-1}\end{aligned} \qquad \text{(Eq. 6)}$$

where $\mathbf{V}(k+1)$ is the covariance matrix of the sensor noise, $\mathbf{v}(k+1)$, and $\mathbf{P}(k+1)$ is the covariance matrix of the state estimator error, $\hat{\mathbf{x}}(k+1) - \mathbf{x}(k+1)$. ($\times$ denotes matrix multiplication.).

The error covariance, $\mathbf{P}(k+1)$, is propagated from the previous time step according to

$$\begin{aligned}\mathbf{P}(k+1) = &\Phi(k)\mathbf{P}(k)\Phi(k)^T + \mathbf{W}(k) \\ &-\Phi(k)\mathbf{P}(k)\mathbf{C}(k)^T \\ &\times\left[\mathbf{C}(k)\mathbf{P}(k)\mathbf{C}(k)^T + \mathbf{V}(k)\right]^{-1} \\ &\times\mathbf{C}(k)\mathbf{P}(k)\Phi(k)^T\end{aligned} \qquad \text{(Eq. 7)}$$

where $W(k)$ is the covariance matrix of $\mathbf{w}(k)$. After substituting for $\mathbf{K}(k)$

$$\begin{aligned}&\mathbf{P}(k+1) = \\ &\quad \Phi(k)\left[\mathbf{I} - \mathbf{K}(k)\,\mathbf{C}(k)\right]\mathbf{P}(k)\Phi(k)^T + \mathbf{W}(k)\end{aligned} \qquad \text{(Eq. 8)}$$

Either of Eqs. (7) and (8)—or one of several other equivalent forms—may be selected based on computational performance (Brown & Hwang, 1997, p. 219).

For the work described below, both plant and sensor noise covariance, as well as the measurement matrix will be considered time invariant—i.e., $\mathbf{V}(k) = \mathbf{V}$ and $\mathbf{W}(k) = \mathbf{W}$, and $\mathbf{C}(k) = \mathbf{C}$. This leaves the KF gain matrix, $\mathbf{K}(k)$, dependent only on updates of the error covariance, $\mathbf{P}(k)$, and the state transition matrix, $\Phi(k)$.

Thus the KF procedure of Eqs. (5) to (8) represents an algorithm to compute an estimate of the state vector from sensor measurements, $\mathbf{y}(k)$, such that the expected value of the error, $\hat{\mathbf{x}}(k) - \mathbf{x}(k)$, between estimated and actual states is minimized in the least squares sense.

Prediction

From the solution over a single time step for the differential equation in Eq. (1) provided by of Eq. (3), and given that the expected value of the noise $\mathbf{w}(k)$ is zero, the extrapolation from time $t = t_k$ out to $t = t_k + \tau$ of the current state vector estimate, $\hat{\mathbf{x}}(k)$, is expressed likewise by

$$\hat{\mathbf{x}}(t_k + \tau) = e^{\tau\mathbf{A}(t_k)}\hat{\mathbf{x}}(t_k) \qquad \text{(Eq. 9)}$$

In the special case that the extrapolation is across an integral number of sample steps $\tau = Nh$, Eq. (3) shows that

$$e^{\tau\mathbf{A}(t_k)} = e^{Nh\mathbf{A}(t_k)} = \Phi^N(k) \qquad \text{(Eq. 10)}$$

Therefore, when predicted N steps ahead, the state estimate is

$$\hat{\mathbf{x}}(k+N) = \Phi^N(k)\hat{\mathbf{x}}(k) \qquad \text{(Eq. 11)}$$

State Space Model

Following Friedmann et al. (1992), Azuma and Bishop (1994), Mazuryk and Gervautz (1995), Kiruluta et al. (1997), and others, a simple, purely kinematic model is selected to describe the interrelation of motion states. Translational motion is described independently in each Cartesian component, (x, y, z), by its position, p, velocity, v, and acceleration, a. Beginning with simple kinematic derivatives in the x component,

$$\frac{d}{dt}\begin{bmatrix} p_x \\ v_x \\ a_x \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} p_x \\ v_x \\ a_x \end{bmatrix} + \begin{bmatrix} w_{px} \\ w_{vx} \\ w_{ax} \end{bmatrix} = \mathbf{A}_{3\times 3}\begin{bmatrix} p_x \\ v_x \\ a_x \end{bmatrix} + \mathbf{w} \qquad \text{(Eq. 12)}$$

we note that only the noise component w_{ax} can affect the acceleration state a_x. The state vector for all three translational components, $\mathbf{x}_t = (p_x, v_x, a_x, p_y, v_y, a_y, p_z, v_z, a_z)^T$, has system matrix,

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{A}_{3\times 3} & 0 & 0 \\ 0 & \mathbf{A}_{3\times 3} & 0 \\ 0 & 0 & \mathbf{A}_{3\times 3} \end{bmatrix} \qquad \text{(Eq. 13)}$$

such that

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{A}_t\mathbf{x}_t + \mathbf{w}_t \qquad \text{(Eq. 14)}$$

where $\mathbf{w}_t$ is the vector of translational noise components.

Rotational displacements in our system are described in terms of the unit quaternion (e.g., Kuipers, 1999)

$$\mathbf{q}_r = \begin{bmatrix} q_w \\ q_x\mathbf{i} \\ q_y\mathbf{j} \\ q_y\mathbf{k} \end{bmatrix} = \begin{bmatrix} \cos(\theta/2) \\ u_x\sin(\theta/2)\mathbf{i} \\ u_y\sin(\theta/2)\mathbf{j} \\ u_z\sin(\theta/2)\mathbf{k} \end{bmatrix} \qquad \text{(Eq. 15)}$$

where θ is the twist angle about the instantaneous rotation (i.e., Euler) axis $\mathbf{u}_r = (u_x\mathbf{i}, u_y\mathbf{j}, u_z\mathbf{k})^T$.

The rotational velocities are $\omega_r = (\omega_x\mathbf{i}, \omega_y\mathbf{j}, \omega_z\mathbf{k})^T$ and have as their respective time derivatives the accelerations $\alpha_r = (\alpha_x\mathbf{i}, \alpha_y\mathbf{j}, \alpha_z\mathbf{k})^T$. From Chou (1992), the quaternion rate and the rotational velocities are related by

$$\frac{d\mathbf{q}_r}{dt} = 0.5\mathbf{q}_r \quad \omega_r \qquad \text{(Eq. 16)}$$

After carrying out the quaternion multiplication denoted by , Eq. (16) can be restated as either of the matrix multiplications

$$\frac{d\mathbf{q}_r}{dt} = 0.5\mathbf{Q}_r\omega_r = 0.5\Omega_r\mathbf{q}_r \qquad \text{(Eq. 17)}$$

in which

$$\mathbf{Q}_r = \begin{bmatrix} -q_x & -q_y & -q_z \\ q_w & -q_z & q_y \\ q_z & q_w & -q_x \\ -q_y & q_x & q_w \end{bmatrix} \qquad \text{(Eq. 18)}$$

and

$$\Omega_r = \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix} \qquad \text{(Eq. 19)}$$

Once the nonlinear product terms in Eq. (17) are linearized locally about the states' instantaneous value at $t = t_k$, the full rotational system model for the 10 element state vector, $\mathbf{x}_r$, can be stated as

$$\frac{d}{dt}\begin{bmatrix} \mathbf{q}_r \\ \omega_r \\ \alpha_r \end{bmatrix} = \begin{bmatrix} 0.5\Omega_r & 0.5\mathbf{Q}_r & 0_{4\times 3} \\ 0_{3\times 4} & 0_{3\times 3} & I_{3\times 3} \\ 0_{3\times 4} & 0_{3\times 3} & 0_{3\times 3} \end{bmatrix}\begin{bmatrix} \mathbf{q}_r \\ \omega_r \\ \alpha_r \end{bmatrix} + \mathbf{w}_r = \mathbf{A}_r\mathbf{x}_r + \mathbf{w}_r \qquad \text{(Eq. 20)}$$

Since the dynamic systems in Eqs. (14) and (20) are independent of each other, the translational and rotational components are treated individually as separate predictors. One advantage in separating the translational and rotational problems is that the system matrices to be manipulated are smaller, with fewer zero entries. Another advantage is that the different dynamic system structures of Eqs. (14) and (20) require wholly different estimation and prediction procedures. On one hand, the time-invariant translational system in Eq. (14) produces a time-invariant state-transition matrix, $\Phi_t(k) = \Phi_t$, which yields a simpler steady-state formulation for the Kalman Filter of Eqs. (5) to (7) and the extrapolation of Eq. (11). On the other hand, because Eq. (17) is a nonlinear function of instantaneous state values, the linearized rotational system matrix $\mathbf{A}_r$ does vary and therefore must be updated regularly to compute the instantaneous matrix exponential $\Phi_r(k)$. This repeated model re-parameterization makes the rotational component estimator a so-called "Extended Kalman Filter" (e.g., Gelb, 1979).

Noise Parameterization for Head Tracking

The system model dynamics of the two KF estimator-predictors, Φ_t and $\Phi_r(k)$, are built solely from expressions defining the time derivatives of translational

and rotational displacement. As such, Eqs. (12) and (20) are purely kinematic models; they cannot be re-parameterized to reflect dynamic properties such as damping, bandwidth, or more complex neuromotor control elements. This leaves only the sensor and plant noise covariance matrices $\mathbf{V}$ and $\mathbf{W}$ as design parameters that might be tuned to adjust KF predictor performance.

$\mathbf{V}$ and $\mathbf{W}$ may be identified in two general ways. The covariances can be predefined from prior calibration or from validated analytic models. Otherwise, if these driving noises cannot be well characterized by measurement or analysis, the covariance matrices (like any other tunable KF model quantity) can be parameterized through numerical optimization of estimator or predictor performance against a cost criterion. Because VE sensor noise parameters can be measured by standard engineering techniques, we predefined $\mathbf{V}$ and chose to investigate more generally the parameterization of plant noise, $\mathbf{W}$, which at a fundamental level arises from physiological muscle activity or neural signals, and is therefore not easy to measure or model analytically.

Following Azuma and Bishop (1994), sensor noise covariance matrices for the translational and rotational predictors, $\mathbf{V}_t$ and $\mathbf{V}_r$, were respectively set to be diagonal with 0.01 m^2 for the three Cartesian and 0.0001 (dimensionless units) for the four quaternion components. Despite Azuma and Bishop's (1994) sensors being based on a completely different technology, off-line tests for our system's sensors and environment showed little KF predictor sensitivity to changes in $\mathbf{V}_t$ and $\mathbf{V}_r$ parameterization

As in Azuma and Bishop (1994) and Liang et al. (1992), we used Powell's method, a multi-dimensional direction set optimization algorithm (Press, Flannery, Teukolsky, & Vetterling, 1986) to parameterize $\mathbf{W}_t$ and $\mathbf{W}_r$. The procedure optimizes parameters of a system of equations with respect to the numerical performance of a functional cost criterion selected by the designer for a given input data collection. In our case, the input data (with the specific exception noted below) were position and quaternion measurements recorded from a single subject performing the same head and body movements that would subsequently be required during the discrimination study described below. The algorithm output throughout was constrained to produce diagonal (i.e., uncorrelated between states), positive semi-definite covariance matrices—negative auto-covariance has no physical meaning and could cause unstable KF behavior.

The translational plant covariance $\mathbf{W}_t$ was optimized using the same cost function as Azuma and Bishop (1994): the simple RMS difference over the sample interval between predicted and actual measured position in each of the three Cartesian coordinates. Informal analytic and subjective observation showed limited predictor sensitivity to $\mathbf{W}_t$ parameterization. We therefore focused our investigation on parameterization of $\mathbf{W}_r$.

One significant difference between our and Azuma and Bishop's implementations (and therefore also the parameter search) is that we employ discrete-time state transition matrices for updating both our translational and rotational KF predictor designs; Azuma and Bishop used 4th order Runge-Kutta numerical integration to propagate their system. Informal observation from simulations with Azuma and Bishop's data sets indicate that our state transition matrix approach yielded smaller RMS errors—possibly due to the additional dynamics introduced by the Runge-Kutta technique.

Twist Optimization

The first of three candidate $\mathbf{W}_r$ parameterization cost functions considers RMS magnitude of the twist angle, $\Delta\theta_i$, over the sampled movement history $(i = 1,\ldots,N)$ between actual and predicted head orientations as in (Azuma & Bishop, 1994).

$$\mathbf{q}_i^{\Delta} = \mathbf{q}_i^{a} \quad \left(\mathbf{q}_i^{p}\right)^{-1} \qquad \text{(Eq. 21)}$$

gives the quaternion difference between actual ($\mathbf{q}_i^a$) and predicted ($\mathbf{q}_i^p$) head orientation quaternions. (Kuipers, 1999) The twist angle is calculated by

$$\Delta\theta_i = 2\cos^{-1}\left(q_{w_i}^{\Delta}\right) \qquad \text{(Eq. 22)}$$

in which $q_{w_i}^{\Delta}$ is the scalar component of $\mathbf{q}_i^{\Delta}$. The *twist* cost function is simply

$$J_{twist} = \sqrt{\sum_{i=1}^{N}\left(\Delta\theta_i\right)^2} \qquad \text{(Eq. 23)}$$

VE Object Displacement Optimization

While angular error would be expected from the projective geometry of rotations in a head tracked VE to be an important indicator for prediction accuracy, Eq. (22) does not account for the twist error's direction. The cost function of Eq. (23) consequently does not permit selective weighting of errors in preferred directions. Thus, we consider the translational displacement components projected by the prediction error, $\mathbf{q}_i^{\Delta}$, of Eq. (21).

Defining $\mathbf{r} = r_x\mathbf{i} + r_y\mathbf{j} + r_z\mathbf{k}$ to represent the vector from a point on the head mounted display (HMD) to some spatial location or "object of interest" in the VE, the translational displacement of this object due to error in predicting head orientation is given by the quaternion version of rotational transformation (Kuipers, 1999)

$$\mathbf{r}_i^{\Delta} = \mathbf{q}_i^{\Delta} \quad \mathbf{r} \quad \left(\mathbf{q}_i^{\Delta}\right)^{-1} \qquad \text{(Eq. 24)}$$

where the inverse is identical to the conjugate for the unit quaternion $\mathbf{q}_i^{\Delta}$. $\mathbf{r}_i^{\Delta} = \Delta x_i\mathbf{i} + \Delta y_i\mathbf{j} + \Delta z_i\mathbf{k}$, in effect, gives the difference between the predicted and actual location of a rendered object in the VE due to $\mathbf{q}_i^{\Delta}$ as seen through the HMD. We note that once an "object of interest" in the VE is selected, the length scaling provided by Eq. (24) allows direct comparison between translation and orientation induced predictor errors. These comparisons generally support the observation for the dominance of orientation over translation effects in head tracking predictor accuracy (Zikan et al., 1995), especially for the experimental task described below.

The cost function minimizing errors according to the components of $\mathbf{r}_i^{\Delta}$ is

$$J_{disp} = \sqrt{\sum_{i=1}^{N} \left[\alpha \left(\Delta x_i \right)^2 + \beta \left(\Delta y_i \right)^2 + \gamma \left(\Delta z_i \right)^2 \right]} \quad \text{(Eq. 25)}$$

where α, β, and γ represent weighting factors to selectively penalize error components in a Cartesian coordinate frame fixed to the HMD. For the case in our VE system where the y and z components lie in the plane of the HMD, β and γ are weighted equally. From trigonometry, small twist angle errors produce negligible x variation, making α less important. The "object of interest" for the parameterization is the virtual target sphere in the experiment described below, situated at $\mathbf{r}$, 0.8 m in the x direction.

Noise Frequency Content Optimization

Predictive compensation, particularly when based on the purely kinematic model described above, exacerbates higher frequency image jitter in VEs, regardless of whether the source is sensor or plant noise. This exacerbation is due to the differentiator action needed to generate velocity and acceleration state estimates from displacements. Fig. 1 shows samples of increased jitter for J_{twist} optimized predictor output in the 5 Hz band and beyond as the look-ahead time (and therefore the amount of differentiator action) is increased for a subject yawing his head side-to-side while wearing the HMD. This 5+ Hz spectral band remains present when the subject stops volitional movement and sits still but vanishes when the HMD and sensor are removed from the subject and supported instead on an inanimate fixed base. Because of its frequency content, we believe this human generated activity is associated with involuntary normal (so-called "physiological") tremor of the head and body (e.g., Desmedt, 1978). Under usual real environment conditions (or in an ideal VE free of sensor noise), the vestibulo-ocular consequences of one's own physiological tremor are subthreshold and not visually detectable by the human. However, under KF prediction, we propose that differentiation elevates tremor induced image jitter to the point where it is clearly visible and interferes with perceived image stability.

To penalize high frequency jitter amplification through parameterization of $\mathbf{W}_r$, an optimization can be carried out on a high-pass filtered version of a pre-recorded motion history consisting primarily of measured components that we want the KF predictor response to suppress. The cost function to minimize jitter amplification in the HMD, similar to Eq. (25), is

$$J_{noise} = \sqrt{\sum_{i=1}^{N} \left[\alpha \left(\delta x_i \right)^2 + \beta \left(\delta y_i \right)^2 + \gamma \left(\delta z_i \right)^2 \right]} \quad \text{(Eq. 26)}$$

where δx_i, δy_i, and δz_i are the high frequency KF predictor orientation output projections into Cartesian coordinates from Eq. (24) that should be driven to zero. α, β, and γ represent the same directional weighting factors as in Eq. (25). Although not examined in this work, another reasonable alternative is a re-structuring of the state matrix $\Phi_r(k)$ to include low-pass filtering which could attenuate high frequency jitter.

Optimization applied to Eq. (26) alone yields a predictor that has good noise attenuation characteristics but poor tracking performance because of very low gains in KF matrix $\mathbf{K}(k)$. To allow better response for low frequency volitional activity, the image displacement cost of Eq. (25) is reintroduced into a hybrid function that combines overall tracking and noise attenuation criteria. Because tracking accuracy can come at the expense of increased KF predictor noise, the hybrid cost function, formulated as

$$J_{hyb} = a \frac{J_{disp}}{J_{disp}\big|_{default}} + b \frac{J_{noise}}{J_{noise}\big|_{default}} \quad \text{(Eq. 27)}$$

allows trade-offs between a and b—respectively the relative importance attached to tracking accuracy, J_{disp}, and noise reduction, J_{noise}. The denominators of each term are scale factors—the respective cost functions evaluated at the default parameter values reported in Azuma and Bishop (1994)—serving to make the weighting factors a and b meaningful.

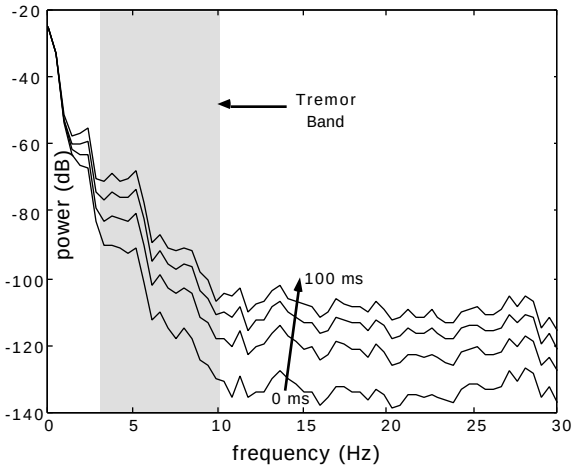

Fig. 1. Increasing power spectral densities for vertical displacement projected from head rotation as prediction interval is increased from 0 to 100 ms in steps of 33 ms.

Predictor Analyses

Based on extensive off-line numerical simulations, three different predictor parameterizations were ultimately selected for experimental study. The *default* and *twist* parameterizations are both based on the cost function of Eq. (23). The *hybrid* parameterization is from Eq. (27). The default parameterization uses Azuma and Bishop's (1994) exact $\mathbf{W}_t$ and $\mathbf{W}_r$ values. $\mathbf{W}_t$ for the twist and hybrid parameterizations and $\mathbf{W}_r$ for the twist and the displacement portion (J_{disp}) of the hybrid parameterizations were optimized for a prediction look-ahead of 50 ms from a pre-recorded sample set (~20 s of data recorded at 120 Hz) of the side-to-side human head motion that would later be required of the subjects in the experiment described below. The noise portion, J_{noise},

of the hybrid design, however, was based separately on a high pass filtered (2nd order, breakpoints at 0.5 and 5 Hz) record of a subject's head motion while sitting still and looking straight ahead. The hybrid parameterization sets $\beta = \gamma$ and $\alpha = 0$ for both its J_{disp} and J_{noise} portions (i.e., emphasizing horizontal and vertical image displacement errors) and $a = b$ to equally weight tracking accuracy and noise reduction.

Table 1 lists a variety of metrics describing the three parameterizations' performance at 50 ms of prediction for the same 20 s sample movement history used in the optimizations. The tabulated values are each ratios. The denominator of each ratio—the RMS difference based on the particular metric between the undelayed displacement input to the sensor and its uncompensated (delayed) measurement—expresses the error introduced by delay. For the noise metric, the sensor displacement input is set to zero. The numerator is the RMS difference for the particular metric between the displacement input to the sensor and the output of the predictive compensator—a measure of the error introduced by imperfect prediction. Thus, the smaller the tabulated quality, the better a particular predictor's performance is with regard to that specified metric. The twist RMS error is defined by Eq. (23). The horizontal RMS displacement error and noise arise from Eqs. (25) and (26) respectively when $\alpha = \gamma = 0$; the vertical components arise when $\alpha = \beta = 0$. Distance (or depth) errors (i.e., $\beta = \gamma = 0$), perpendicular to the plane of the HMD, were omitted from the table because of their minimal impact on the rendered VE in the discrimination experiment.

From Table 1, the horizontal displacement errors appear to be 10% to 15% of the vertical, implying that horizontal tracking is better for any of the predictor parameterizations. A more likely explanation for this imbalance is that the side-to-side motion record used to parameterize the predictors and to calculate these scores contributes mainly to the horizontal metric's denominator, and therefore dominates it but not the vertical. Conversely, the horizontal and vertical noise scores only use the record of the subject sitting still (rather than moving side-to-side) from which essentially all voluntary activity was removed by high pass pre-filtering, making their denominators nearly identical for $J_{default}$. In comparing the performance of the three predictor parameterizations, the vertical displacement error and horizontal noise offer no specific insight. Vertical noise suggests that the J_{hyb} design is best, while the horizontal displacement error shows that the J_{twist} and J_{hyb} parameterizations are equally good. Interestingly, the twist error metric indicates that J_{twist} is best, but that J_{hyb} is worse than even the default parameterization. However, it is important to caution that the observations in Table 1 are calculated from the same specific (i.e., side-to-side or stationary) motion records used to form the parameterizations.

Fig. 2 presents time domain tracking excerpts for the three predictor parameterizations from the optimization data sets. These plots show horizontal and vertical displacements arising from head rotations per Eq. (24). The horizontal trace is dominated by 0.67 Hz side-to-side volitional head motion. The same 0.67 Hz rhythm is not obvious in the vertical plots. The vertical plots appear more oscillatory in the 5 Hz range, which, consistent with the performance metrics' scores in Table 1, may simply be due to the absence of the larger scale horizontal motion. Examination of the magnified plots shows that none of the three predictors track the 5 Hz input component particularly well—all three outputs show 90 ° or more of phase lag. In fact, the hybrid predictor output remains in phase with the delayed measurement, evidently not exhibiting the other designs' overshoot simply because these higher frequency components are not being predicted.

Power spectral density plots in Fig. 3 summarize the frequency domain attributes of the three parameterized predictors averaged over the entire 20 s length of the side-to-side motion optimization data set for the calculated horizontal and vertical components. In general, all predictors match the amplitude of the actual input in the range of voluntary head motion up to ~1.5 Hz. The horizontal component has greater power density in this region again because the motion is predominantly side-to-side. The vertical component spectrum shows a prominent bulge in the vicinity of 5 Hz—the oscillatory activity associated with head and body tremor—that is not discernible in the horizontal plot. In the horizontal plots, beginning at ~1.5 Hz, the predictors' outputs initially all rise together above the actual input, eventually spreading to between 20 and 40 dB higher, indicative of predictor action at higher frequencies. This behavior is seen in the vertical spectrum, but only for the default and twist optimized predictors. The hybrid predictor follows the input spectrum magnitude very closely (up only ~4 dB at the 5 Hz bump), confirming the lack of prediction at higher frequencies noted in the vertical plots of Fig. 2. The comparably heightened noise gain for the default and twist predictors' spectra but not the hybrid's (which had essentially unmagnified

Metric	Twist Error	Displ Error	Displ Error	Noise	Noise
Component	—	Horizontal	Vertical	Horizontal	Vertical
Input Data	Raw	Raw	Raw	High-Pass	High-Pass
$J_{default}$	0.181	0.161	1.047	2.250	2.251
J_{twist}	0.135	0.104	0.934	2.703	1.993
J_{hyb}	0.457	0.108	1.041	2.449	1.043

Table 1. Performance metrics for default, twist, and hybrid optimized parameterization with 50 ms prediction.

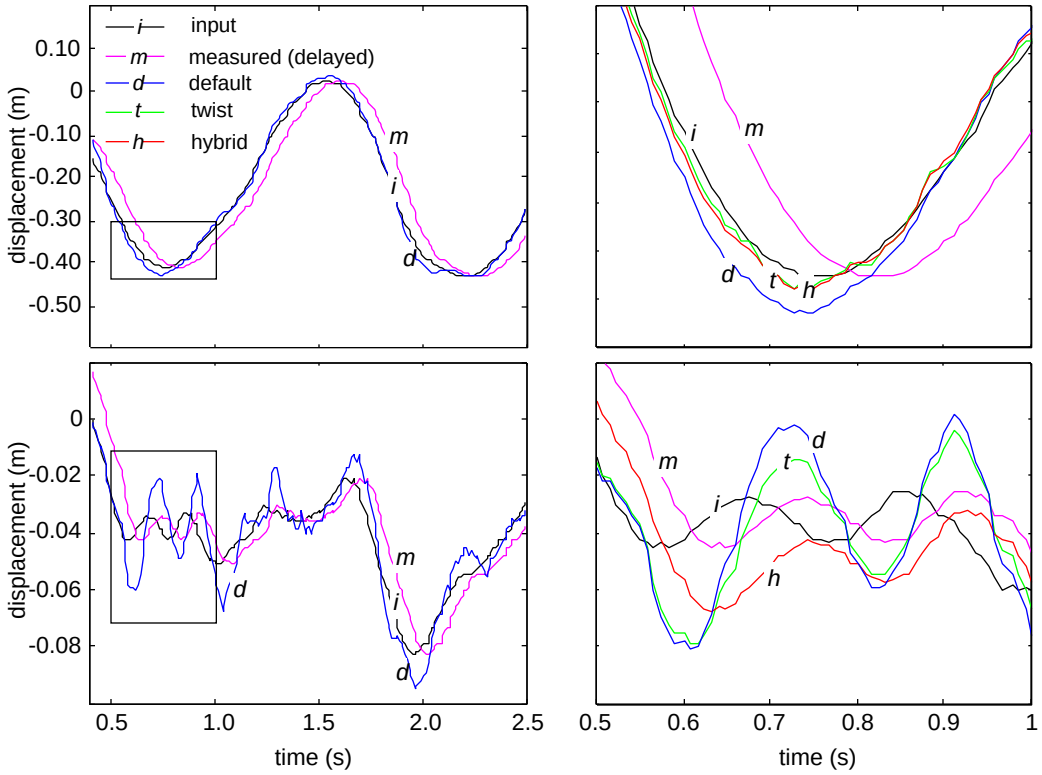

Fig. 2. Predictor performance for 50 ms of compensation for horizontal displacement (top) and vertical displacement (bottom) components. Note the different displacement scales for all plots. (Left) Default parameterization with actual and measured (delayed) displacements converted from head rotation. (Right) Regions magnified from corresponding boxes in left plots also include twist and hybrid predictor output.

vertical noise) concurs with the noise values reported in Table 1, and thereby supports use of the hybrid design to prevent magnification of high frequency vertical noise components.

It is important to note that while the time and frequency domain plots and performance metrics represent data for 50 ms of predictive compensation, results are similar for the longer 67, 83, and 100 ms latencies covered in the discrimination experiments.

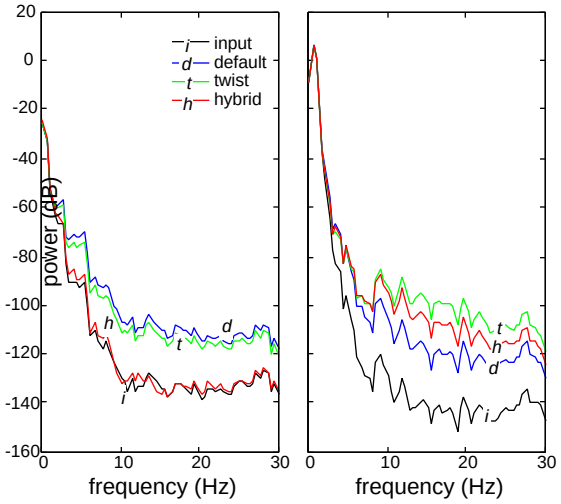

Fig. 3. Predictor power spectra at 50 ms look-ahead. Vertical (left) and horizontal displacement (right).

Experiment

VE System Hardware and Software

The experiment VE and KF predictor software ran on an SGI Onyx workstation with four R4400 CPUs and dual-pipeline RealityEngine-2 graphics. The subjects viewed the VE in a Virtual Research V8 HMD. The position and orientation of subjects' head as well as those of 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 115.2 KBaud serial port. A variety of software techniques, including the streaming of sensor input to its own stand-alone software process that then transfers data to other simulation processes on the SGI computer via shared memory (Jacoby, Adelstein, & Ellis, 1996), enable us to produce fast VEs with low latency, high update rates, and reduced temporal variability. The sensor-to-display "internal" latency for the VE used in our experiment was measured by the methods of Jacoby et al. (1996) to be 35±5 ms (mean ± stdev) for Cartesian displacement at a steady frame rate of 60 Hz. Concurrently reported quaternion rotations averaged 5 ms less (Adelstein, Johnston, & Ellis, 1995). A number of additional software and hardware procedures that further reduce translation and rotation latencies respectively to 30±5 and 25±5 ms or better were not invoked because of potential degradation to frame rate uniformity. This low internal VE system latency and uniform frame rate make

possible the controlled addition of the time delays required for our experimental study.

The VE for the experiments consisted solely of a virtual faceted sphere (i.e., target) in a dark, empty space, lit as described by Ellis, Young, Adelstein, and Ehrlich (1999a). Subjects were seated with the HMD's FasTrak receiver 0.4 m below the FasTrak transmitter. The virtual sphere, whose position in the VE was governed by an immobile second FasTrak receiver, was centered 0.8 m 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. In the presence of the inevitable delays, the virtual sphere will not be locked perfectly in space and may appear to move about its ideal fixed location.

Software embodying the estimation and prediction procedures was implemented as a separate process interposed between the sensor and VE processes on the SGI workstation. The predictor process receives raw data from the sensor process in one shared memory location and deposits the compensated results into another location. A separate shared memory process serves to revise predictive compensator parameters in real time. The predictors are coded in C and, at present, use Matlab's C/C++ Library for computing the matrix exponential function of Eq. (3). Predictor parameterization sets are typically developed beforehand in off-line optimizations coded in the Matlab language.

The multi-processing, multi-processor architecture of our VE system allowed 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 ultimate goal of our parameterization effort is to produce predictors that remove VE system while at the same time not introduce compensation artifacts. These experiments aim to study user awareness of any artifacts due to the presence of imperfect predictive compensation. Our experimental approach for determining the effect on user perception of different compensator parameterizations is derived from a technique to assess subjective detectability of changes in VE latency (Ellis, Young, Adelstein, & Ehrlich, 1999a, 1999b).

The experimental procedure is based on the following two alternative forced choice protocol. The seated subjects were required to yaw their heads from side-to-side in time to a 80 beat/min metronome (0.67 Hz or 1.5 s per full back-and-forth cycle) 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. 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.

Each of six latency values (16.7 to 100 ms in 16.7 ms steps) was blocked into its own randomly ordered set of 20 judgments such that each of the four possible A-B condition pairings was repeated five times. The order of blocks of individual latencies was also randomized. Predictor type was blocked so that subjects completed all tests with one predictor before proceeding to the next. The six possible presentation orders for the three predictors (standard, twist, and hybrid) were balanced between the 12 subjects. Each subject's proportion of correct discriminations for a particular condition was calculated from its set of 20 responses. The subjects, who were either lab members or paid naive recruits, all had normal or corrected to normal vision and no other known impairments.

Discrimination Results

Fig. 4 shows the percentage of correct discriminations between the compensated delay and minimal delay conditions averaged across all twelve subjects. A 50% correct response proportion would be expected if subjects were guessing randomly at the balanced presentation of stimulus pairs. Discriminability of any of the predictors grows monotonically with the number of steps of added latency. The separations of the curves and standard error bars for in Fig. 4, calculated from the binomial distribution for the proportional data, suggests that for compensated latencies $\geq$ 33 ms the hybrid predictor's presence is less discriminable than the other designs.

A three-way (latency X predictor type X predictor order) ANOVA was carried out on the proportional responses following the arcsine square root transformation to convert the data to a normal distribution (Sachs, 1984, p. 339). The main effects of added latency (F = 56.347; df = 5,30; p < .001) and predictor type (F = 11.239; df = 2,12; p < .002) on the proportion of correct responses were significant. Predictor order was not significant. Neither were any of the interactions between the main factors.

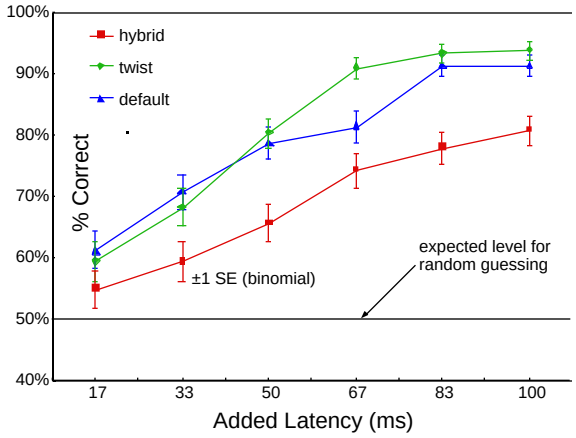

Fig. 4 Percent correct discrimination averaged across all 12 subjects (mean ± std error) as a function of predictor type and latency added to the 35 ms VE system minimum.

Discussion

The experiment tested human performance in a specific, stereotyped head motion meant to elicit both the transient VE image drifts commonly associated with system latency and the overshoots of prediction. Head yaw motion was chosen because it is a principal component of gaze direction—the action of visually locating objects in both real and virtual environments. Furthermore, because head rotation rather than translation by the body is generally expected to cause larger shifts of the visual scene (Zikan et al., 1995), VE images will be more sensitive to response imperfections in the rotational components of spatial motion.

In the experiment, artificial latency was deliberately added and then compensated back toward the VE system baseline delay. This avoided prediction all the way down to zero absolute latency for which we could not then produce a prediction-free VE control condition. It also enabled us in a separate study (Jung, Adelstein, & Ellis, 2000) to compare the subjective effects specifically of the artifacts introduced by prediction against the degradation caused by uncompensated, artificially added VE latencies.

The experiment results show the hybrid predictor parameterization to have better performance (lower discriminability) than the two other designs tested, and that this improvement appears to increase as the amount of additional compensated latency rises. The twist and default predictors did not exhibit significant differences in discriminability. While the twist and default predictors' parameterizations were based on the same cost function, this result is still somewhat surprising since the default parameter quantities (Azuma & Bishop, 1994) were developed for completely different VE system components and input human motion. It is noteworthy that all the predictor designs still remained above 50 percent discriminability (including the standard error ranges)—the level associated with random guessing. If the subjects did no better than random guessing this would have implied that any artifacts introduced by prediction were, on average, not perceptible to the subjects.

The better performance of the hybrid parameterization is attributable to its ability to predict in the horizontal direction while not predict in the vertical. Thus, while the slow (and therefore less overshoot prone) side-to-side motion was acted upon, amplification (and consequent increased visibility) of the higher frequency noise components dominant in the vertical was avoided. The behavior of this parameterization was dictated by a cost function that weighted equally the displacement magnitude of higher-frequency noise (such as tremor) and errors from tracking low-frequency volitional inputs—all as projected into the plane of the HMD. None of the cost criteria examined explicitly weighted time lag. Also, the hybrid cost function as weighted did not itself differentiate between horizontal and vertical response: the difference arose because of the directional anisotropy of the motion history upon which the optimization was carried out. As such, the resulting predictor parameters may only be suited for the specific type of motion from which they were optimized.

A number of metrics based on the cost criteria for the different optimization schemes were examined as indicators of predictor performance. Two of the four metrics related to the hybrid optimization indicated that the hybrid predictor would have the lowest vertical noise amplification and tie for best tracking accuracy. The other two of the four metrics demonstrated no preferred parameterization. The twist error metric picked the twist design—more or less as expected since this metric was the cost function for optimizing twist. Interestingly, the same criterion also suggested that the hybrid design would fare worse than the default parameterization. It might be claimed that this is because the simple twist cost function is unable to capture all the nuances that the directional metrics can. At issue, however, is whether these metrics are ultimately suitable only for scoring optimizations obtained with the same cost function as the metric. This question warrants a more detailed analysis of these performance metrics, using data from many different motion records and from a variety of subjects rather than the same single data set used in the optimizations for this study.

Though the optimization cost criteria introduced in the course of this work may have general utility, the specific numerical parameterizations were developed for the exact same subject movements used in the experiment. Hence, the predictors studied might not achieve the same performance for other head or body segment motions—recall that the hybrid parameterization did not predict vertical components. One possible approach to this limitation would be to develop banks of appropriate predictors whose relative weightings could be modulated depending on the type of human operator activity.

Acknowledgments

Thanks are due to Dr. Mark Young for development of the VE experiment application and to Tony Wolfram for assistance with subject testing. This research was supported by NASA HQ-Code UL (RTOP 131-20-30).

References

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

Akatsuka, Y., & Bekey, G.A. (1998). Compensation for end to end delays in a VR system. *Proceedings, IEEE Virtual Reality Annual International Symposium*, (pp. 156-159). Piscataway NJ: IEEE.

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

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

Chou, J.C.K. (1992). Quaternion kinematic and dynamic differential equations. *IEEE Trans. Robotics and Automation, 8*(1), pp. 53-64.

Desmedt, J.E, ed. (1978). *Physiological Tremor, Pathological Tremors and Clonus. Progress in Clinical Neurophysiology, Vol. 5*. Basel: Karger.

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

Ellis, S., Young, M., Adelstein, B., & Ehrlich, S. (1999b). Discrimination of changes of latency during voluntary hand movement of virtual objects. *Proceedings, 43rd Annual Meeting Human Factors Ergonomics Society*, (pp. 1182-1186), Houston TX: HFES.

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

Gelb, A., ed. (1979). *Applied Optimal Estimation*, 5th ed, Cambridge, MA: MIT Press.

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

Jung, J., Adelstein, B., & Ellis, S. (2000). Discriminability of Prediction Artifacts in a Time-Delayed Virtual Environment. *Proceedings, 44th Annual Meeting Human Factors Ergonomics Society*, San Diego CA: HFES.

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

Liang, J., Shaw, C., & Green, M. (1991). On temporal-spatial realism in the virtual reality environment. *Proceedings, Symposium on User Interface Software and Technology (UIST'91) 1991*, (pp. 19-25). Hilton Head SC: ACM.

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

Nelson, W.T., Hettinger, L.J., Haas, M.W., Russell, C.A., Warm, J.S., Dember, W.N., & Stoffregen, T.A. (1995). Compensation for the effects of time delay in a helmet-mounted display: perceptual adaptation versus algorithmic prediction. *Proceedings, Conference Helmet- and Head-Mounted Displays and Symbology Design Requirements II, 2465*, (pp. 154-164). Bellingham WA: SPIE.

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. (1986). *Numerical Recipes: The Art of Scientific Computing*, New York: Cambridge University Press.

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

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. Systems, Man, Cybernetics—A, 26*(4), pp. 445-454.

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

Zikan, K., Curtis, W.D., Sowizral, H.A., & Janin, A.L. (1995). A note on dynamics of human head motions and on predictive filtering of head-set orientations. *Proceedings, SPIE Conference Telemanipulator and Telepresence Technologies, 2351*, (pp. 328-336). Bellingham WA: SPIE.