

Quick Models for Saccade Amplitude Prediction

Oleg V. Komogortsev
Department of Computer Science
Texas State University-San Marcos
ok11@txstate.edu

Young Sam Ryu
Ingram School of Engineering
Texas State University -
San Marcos
yryu@txstate.edu

Do Hyong Koh
Department of Computer Science
Texas State University-
San Marcos
dk1132@txstate.edu

This paper presents a new saccade amplitude prediction model. The model is based on a Kalman filter and regression analysis. The aim of the model is to predict a saccade's amplitude extremely quickly, *i.e.*, within two eye position samples at the onset of a saccade. Specifically, the paper explores saccade amplitude prediction considering one or two samples at the onset of a saccade. The models' prediction performance was tested with 35 subjects. The amplitude accuracy results yielded approximately 5.26° prediction error, while the error for direction prediction was 5.3% for the first sample model and 1.5% for the two samples model. The practical use of the proposed model lays in the area of real-time gaze-contingent compression and extreme eye-gaze aware interaction applications. The paper provides theoretical evaluation of the benefits of saccade amplitude prediction to the gaze-contingent multimedia compression, estimating a 21% improvement in compression for short network delays.

Keywords: Saccade, prediction, Kalman filter, Human Computer Interaction

Introduction

A *Human Visual System* (HVS) exhibits a variety of eye movements: fixations, saccades, smooth pursuit, optokinetic reflex, vestibulo-ocular reflex, and vergence (Leigh and Zee, 2006). When a person is sitting in front of a computer screen, usually only fixations, pursuits, and saccades are present. Among those three eye movements, saccades are the fastest movements transitioning the eye between relatively stable fixation spots (Duchowski, 2007). The fixations provide the highest quality picture to the brain, while HVS is blind during saccades (Duchowski, 2007). Pursuits are rarely exhibited when a person is working in front of a computer screen, *i.e.* pursuits appear when a person looks at objects with translational motion.

Quality of vision during pursuit varies. Two areas inside the Human Computer Interaction (HCI) domain - gaze contingent compression (GCC) systems (Duchowski and Çöltekin, 2007; Komogortsev and Khan, 2008b; Parkhurst and Niebur, 2002; Reingold et al., 2003) and systems with direct eye gaze control (Huckauf and Urbina, 2008b; Jacob, 1990; Kumar and Winograd, 2007; Zhai et al., 1999) - employ the characteristics of eye movements to make HCI systems more efficient and responsive.

Real-time GCC systems exploit the properties of the HVS where the area of the highest visual acuity is approximately 2° of the visual angle, while the quality of vision in the periphery is severely degraded (Irwin, 1992). Real-time GCC systems have to accurately calculate the location of the fixation spot estimated by an area called a Region of Interest (ROI). A challenge for any

GCC is to minimize the ROI size, without letting a user see artifacts introduced by the GCC compression. Network transmission of the multimedia content while GCC is performed induces various transmission delays into the system. As a result of the delay, saccades can place a gaze on the low quality coded part of the image (Komogortsev and Khan, 2008b). Therefore, a real-time GCC system must have a saccade amplitude prediction algorithm that allows placing a high quality coded ROI on top of the future fixation/pursuit movement to compensate for the delay effects (Komogortsev and Khan, 2008b). A quick amplitude prediction model can reduce the delay, therefore improving the performance of a real-time GCC system. The discussion section of this paper provides a theoretical validation of this claim.

Today in the human computer interaction world, the mouse and the keyboard are the primary input devices. Recently eye-gaze aware interfaces, based on the eye tracker as an input device, have been gaining popularity in the HCI community (Huckauf and Urbina, 2008a; Huckauf and Urbina, 2008b; Istance *et al.*, 2008; Kumar *et al.*, 2007; MacKenzie and Zhang, 2008). The majority of the HCI systems use fixation duration (dwell time) as a trigger for interface actions (Kumar *et al.*, 2007; Sibert and Jacob, 2000; Zhai *et al.*, 1999). In such interfaces, the duration of the detected fixation triggers a “click”. Fixation-based selection necessitates data buffering, and therefore, introduces a delay in the system. Pursuit-based selection is an unexplored topic in the HCI community. Nevertheless, the definition of the pursuit implies that its detection will require a certain amount of data buffering. Due to its speed saccade selection would seem to be the most appropriate in the applications where the quickness of the target selection is of the utmost importance. Saccade’s characteristics would be employed to “click” a target even before the eye moves to the target’s location. Such a scheme would require that saccade’s amplitude and direction be predicted based on the first few or even one eye position samples belonging to the saccade trajectory.

We are aware of only two previously published works that discussed a prediction of the saccade amplitude. The first work was authored by Anliker (1976), where the author employed the fact that saccades in nature are ballistic (Leigh and Zee, 2006), i.e., once the peak velocity is detected, the remaining saccade trajectory resembles the trajectory before the peak. The second work authored

by Komogortsev and Khan (2007) employed a two state Kalman Filter (TSKF) for saccade amplitude prediction.

This paper’s aim is to create a quick saccade amplitude prediction model based on the first position samples at the beginning of a saccade and test this model for reliability. The model that is proposed in this paper employs both a Kalman filter and regression analysis and is tested by 35 subjects with a stimuli designed to evoke saccades of various amplitudes.

The structure of this paper is as follows: The following section, “Human Visual System Modeling by a Kalman Filter” provides a brief description of the Kalman filter related eye movement research, outlines the mathematical backbone of the Kalman filter framework, provides a Kalman filter-based model of the human visual system, and describes a chi-square test mechanism later employed for the amplitude prediction; the “Method” section describes the velocity-based amplitude prediction model proposed by Anliker (1976), provides the objective and the implementation details for the newly proposed saccade amplitude prediction model, and discusses two amplitude direction prediction models. It also provides the details of the experiment setup and the description of the evaluation metrics; the “Results” section provides the specific equation for the regression based prediction model and describes the performance details of every model; the “Discussion” section presents the discussion of the current challenges of amplitude prediction and provides theoretical evaluation of the amplitude prediction on the gaze-contingent compression field; and the “Conclusion” section summarizes the results presented in the paper.

Human Visual System Modeling by a Kalman Filter

The Kalman Filter has played an important role in eye movement related research. Sauter *et al.* (1991) has proposed a mechanism for the detection of the saccade onset/offset based on the innovations generated by a Kalman Filter. Rewari and Chi-Sang (1993) have applied a general likelihood approach to improve detection for the saccades of small amplitudes Abd-almageed *et al.* (2002) proposed parameters that allowed trackers to more accurately reconstruct a pursuit signal in cases when the signal was corrupted by noise. Komogortsev and Khan (2007) have applied a Kalman filter both in gaze contin-

gent compression systems and in systems with direct eye gaze input. In these systems a Kalman Filter was employed as a predictor of visual attention and as a filter for eye position samples not detected by an eye tracker. A Kalman Filter incorporated with the Oculomotor Plant Mechanical Model was employed as a predictor of the eye movement trajectories in cases when saccade amplitude was known (Komogortsev and Khan, 2008a; Komogortsev and Khan, 2009). The specific focus of the current work is quickness of prediction and evaluation of the accuracy of such prediction.

Kalman Filter

The Kalman filter is a recursive estimator that computes a future estimate of the dynamic system state from a series of incomplete and noisy measurements. The Kalman Filter minimizes the error between the estimation of the system's state and the actual system's state. Only the estimated state from the previous time step and the new measurements are needed to compute the new state estimate. Many real dynamic systems do not exactly fit this model; however, because the Kalman filter is designed to operate in the presence of noise, an approximate fit is often adequate for the filter to be quite useful (Brown and Hwang, 1997).

The Kalman Filter addresses the problem of trying to estimate the state $x \in \mathfrak{R}^n$ of a discrete-time controlled process that is governed by the linear stochastic difference equation (Brown and Hwang, 1997):

$$x_{k+1} = A_{k+1}x_k + B_{k+1}u_{k+1} + w_{k+1} \quad (1)$$

with the measurement

$$z_k = H_k x_k + v_k \quad (2)$$

The n-by-n state transition matrix A_{k+1} relates the state at the previous time step k to the state at the current step k+1 in the absence of either a driving function or process noise. B_{k+1} is an n-by-m control input matrix that relates m-by-1 control vector u_{k+1} to the state x_k . w_k is an n-by-1 system's noise vector with an n-by-n covariance matrix Q_k . $p(w_k) \sim N(0, Q_k)$. The measurement vector z_k contains state variables that are measured by the instruments. H_k is a j-by-n observation model matrix which maps the state x_k into the measurement vector z_k . v_k is a measurement noise j-by-1 vector with covariance R_k . $p(v_k) \sim N(0, R_k)$.

While Equations (1)-(7) provide the mathematical description of the process that is being modeled the actual state values x_{k+1} are unknown and have to be estimated. The estimation of x_{k+1} requires two distinct phases Predict and Update (Brown and Hwang, 1997).

Predict:

Equation (3) is used to predict the state vector ahead:

$$\hat{x}_{k+1}^- = A_{k+1}x_k + B_{k+1}u_{k+1} \quad (3)$$

The \hat{x}_{k+1}^- is a future estimation of the modeled state without a measurement from the measurement instrument. In the case of the eye movement prediction the value of \hat{x}_{k+1}^- can be employed as a predictor of the future gaze position.

One of the Kalman filter goals is to minimize the error between the actual state value x_{k+1} and the estimation of this value \hat{x}_{k+1} (Equation (6)). For these purposes, the first estimate of the error covariance matrix is computed following mathematical representation of the modeled process:

$$P_{k+1}^- = A_{k+1}P_k A_{k+1}^T + Q_{k+1} \quad (4)$$

Update:

The update phase improves the estimate of the modeled process by considering the measurement from the measurement device. The update phase can be broken down into three distinct steps.

Compute the Kalman gain:

$$K_{k+1} = P_{k+1}^- H_{k+1}^T (H_{k+1} P_{k+1}^- H_{k+1}^T + R_{k+1})^{-1} \quad (5)$$

Update the estimate of the state vector with a measurement z_{k+1} :

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}(z_{k+1} - H_{k+1}\hat{x}_{k+1}^-) \quad (6)$$

Update the error covariance matrix:

$$P_{k+1} = (I - K_{k+1}H_{k+1})P_{k+1}^- \quad (7)$$

The choice of Kalman filter gain K_k allows minimization of the estimate error covariance $P_k = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$. A more detailed description of Kalman filter mechanics is beyond the scope of this paper and can be found in (Brown and Hwang, 1997; Kalman, 1960).

Human Visual System

The approach that we use in this paper is to model an eye as a system with two states: position and velocity, with acceleration modeled as white noise with known maximum acceleration. Next we apply a Kalman filter framework to this eye representation creating a Kalman filter with two states, which we will call a *Two State Kalman Filter* (TSKF). To complete the description of this filter we describe our choice for the state vector x_k , control vector u_k , transition matrix A_k , and control matrix B_k . It is also necessary to derive a covariance matrix Q_k for the system's noise w_k and covariance matrix R_k defining the measurement noise v_k . Additionally, to map the actual system's state vector x_k to the measurement vector z_k , the observation matrix H_k is required. A detailed description is provided below.

Note that 2D eye movement parameters (position, velocity, acceleration) can be broken into the vertical and horizontal components, because essentially they are composed of superposition of their respective orthogonal components (Kohler, 1997). Therefore, we create two instances of the TSKF filter: the first is responsible for the horizontal component of movement and the second one is responsible for the vertical. As a result, an eye is represented as a system which has two state vectors x_k and y_k .

$$x_k = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \quad (8)$$

where $x_1(k)$ is the horizontal coordinate of the gaze position and $x_2(k)$ is the horizontal eye-velocity at time k .

$$y_k = \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} \quad (9)$$

where $y_1(k)$ is the vertical gaze position and $y_2(k)$ is the vertical eye-velocity at time k .

The state transition matrix for both horizontal and vertical states is:

$$A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \quad (10)$$

where Δt is the eye-tracker's eye-position sampling interval.

The observation model matrix for both state vectors is:

$$H = [1 \quad 0] \quad (11)$$

By definition, the covariance matrix for the measurement noise is $R_k = E[(v_k - E(v_k))(v_k - E(v_k))^T]$. Because only the eye position is measured, v_k is a scalar making $R_k = VAR[v_k] = \delta_v^2$, where δ_v is the standard deviation of the measurement noise. In this paper, it is assumed that the standard deviation of the measurement noise relates to the accuracy of the eye tracker and is bounded by one degree of the visual angle. Therefore δ_v was conservatively set to 1° . In cases when the eye tracker fails to detect eye position coordinates, the standard deviation of measurement noise is assigned to be $\delta_v = 120^\circ$. The value of 120° is chosen empirically, allowing the Kalman Filter to rely more on the predicted eye position coordinate \hat{x}_k^- .

The TSKF is initialized with zero valued initial vectors \hat{x}_0, \hat{y}_0 and an identity error covariance matrix P_0 .

By definition, the process noise covariance matrix is $Q_k = E[(w_k - E(w_k))(w_k - E(w_k))^T]$, where w_k is a 1x2 system's noise vector $w_k = [w_1(k) \quad w_2(k)]^T$. The TSKF simple model assumes that variables $w_i(k)$ are uncorrelated between each other (velocity is independent of eye position), i.e., $E[(w_m(k)w_n(k))] = E[(w_m(k))E[w_n(k)]]$ for all $n \neq m$ and $p(w_1(k)) \sim N(0, \delta_1^2)$, $p(w_2(k)) \sim N(0, \delta_2^2)$. These assumptions generate the following system noise covariance matrix: $Q_k = \begin{bmatrix} \delta_1^2 & 0 \\ 0 & \delta_2^2 \end{bmatrix}$. This simple model assumes that the standard deviation of the eye position noise $w_1(k)$ is connected to the characteristics of the eye-fixation movement. Each eye fixation consists of three basic eye-sub-movements: drifts, small involuntary saccades and tremors (Yarbus, 1967). Among those three movements, involuntary saccades have the highest amplitude—about a half degree of the visual angle; therefore, δ_1 is set conservatively to 1° . The standard deviation value for eye velocity was selected to be $\delta_2 = 1^\circ/s$.

Chi-square Test & Saccade Amplitude

The Chi-square test was originally employed by Sauter (Sauter et al., 1991) to detect the onset and the offset of a saccade. The Chi-square test monitors the difference between predicted and observed eye velocity:

$$\chi_i^2 = \sum_{i=1}^p \frac{(\hat{x}_2^-(i) - \hat{\theta}_i)^2}{\delta^2} \quad (12)$$

where $\hat{\theta}_i$ is the observed eye velocity and $\hat{\theta}_i^- = \hat{x}_2^-(i)$ is the predicted eye velocity computed by Equation (3). It is important to note that Equation (3) can be computed as a result of the Kalman filter framework presented by Equations (1)-(7) and specifically defined by equations (8)-(11). Komogortsev and Khan (2007) have suggested a function that connected the value of χ_i^2 to the amplitude of the corresponding saccade (2007). They suggested that the development of such a function is possible due to the fact that HVS uses phasic (fast) eye-muscle fibers with high motoneuronal firing rate for large saccades and tonic (slow) eye-muscle fibers with a lower motoneuronal firing rate for the saccades of lesser amplitude (Bahill, 1980). Such mechanism ensures different rate of rise of eye-muscle force for the saccades of various amplitudes providing higher acceleration to the eye globe during saccades of higher amplitude. Larger amplitudes produce larger eye velocity values therefore increasing the value for χ_i^2 . The next section provides a description of the chi-square based amplitude prediction model.

Method

Saccade Amplitude Prediction: Velocity Model

As a comparison model, we would like to employ a saccade prediction model proposed by Anliker (Anliker, 1976). Anliker's model uses the fact that saccadic movement is ballistic—i.e., a saccade trajectory is predetermined and cannot be altered once the movement starts, and the saccade trajectory resembles a bell curve. Once the peak velocity is reached, the rest of the saccade movement mirrors the movement prior to the peak. In our implementation of Anliker's model, the velocity peak is detected when a consecutive eye position point has a lesser absolute velocity value than the previous one. The saccade amplitude is made equal to double the distance of the path traveled, prior to the velocity peak.

Saccade Amplitude Prediction: Chi-square Regression Model

Objective & Evaluation

The aim of the quick saccade amplitude prediction model is to predict the saccade's amplitude as quickly as possible with amplitude predicted within the first two or even one eye position samples at the onset of a saccade. It is important to assess theoretically how much quicker

such models are, when they are, than the base Velocity Model. For simplicity sake, we call the First Sample model the model that requires just one eye position sample at the onset of a saccade for the amplitude prediction, and we call the Two Samples model the model that uses the first two eye position samples at the onset of a saccade for prediction. The time required for prediction for the First Sample model is constant $T_{predict} = 1/f$ (f is the frequency of the eye tracker equipment measured in Hz). For the Two Samples model the time of prediction is $T_{predict} = 2/f$. The Velocity Model requires peak velocity identification that occurs at the point representing the middle of the saccade plus an additional point required for velocity peak verification, making prediction time: $T_{predict} = T_{sac_dur}/2 + 1/f$ were saccade D_{sac_dur} represents a saccade duration measured in ms. Saccade duration can be estimated with the formula (Carpenter, 1977)

$$T_{sac_dur} = (2.2A_{sac_amp} + 21)/1000 \quad (13)$$

where A_{sac_amp} is an amplitude of a saccade measured in the degrees.

The amount of time saved by a prediction model when compared to the other model can be estimated by the formula: $T_{saved} = T_{model_1} - T_{model_2}$. Figure 1 provides the estimate of time saved for the First Sample model and the Two Samples model vs. the Velocity Model.

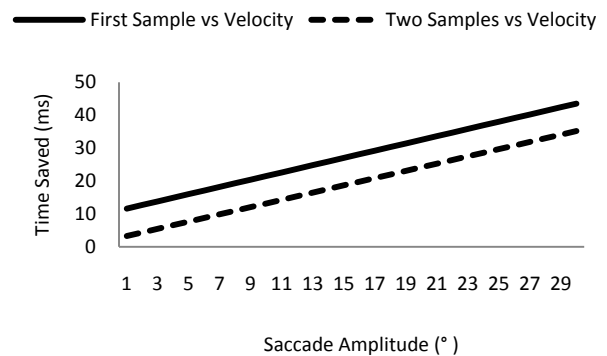


Figure 1. Amount of prediction time saved when First Sample and Two Samples models are compared to Velocity model

Figure 1 indicates that prediction time saved amounts to 12 ms for First Sample model and 3 ms for Two Sample model in case of 1° saccade. In case of 30° saccade the savings become 44 ms and 35 ms, respectively.

Implementation

We have selected a non-linear regression model with Gauss-Newton method to create the amplitude prediction model for the First Sample and Two Sample scenario. Nonlinear regression is widely employed as a powerful tool for analyzing scientific data, especially in physiology (Motulsky, 2001). Specifically, non-linear regression is more effective for curve-fitting than linear regression (Motulsky and Christopoulos, 2004). The proposed Chi-square Regression Model was created as a result of the analysis of the χ_i^2 data recorded with saccade evoking experimental setup. Chi-square test values were calculated by equation (12) for every eye gaze position sample with saccade parameters such as amplitude, onset, and offset detected by the Velocity Threshold model (I-VT). The I-VT model is considered a *de facto* standard by the eye tracking community, the description and parameters of the I-VT we employed can be found in (Komogortsev and Khan, 2009). Chi-square test values recorded at the onset of a saccade and the following eye position model created the data for the Chi-square Regression Model. The model was implemented using SAS software (SAS, 2009).

Saccade Direction Detection Model

This paper focused on only saccades with horizontal amplitude. Quick direction detection of this movement was not as easy as it might appear due to equipment noise. Quick direction detection schemes are prone to generate errors. In the next two paragraphs, we present two methods for the horizontal saccade's direction detection

First Sample Model

The direction of movement was connected to the sign of the velocity of the recorded signal. A positive sign of the velocity signal at the first eye position sample indicated a rightward direction of a saccade, and a negative velocity indicated a leftward direction of the saccade.

Two Samples Model

The first two velocity samples of the saccade trajectory were evaluated. The rightward saccade was predicted if both velocity samples had positive values. The leftward saccade was predicted if both velocity samples had negative values. In cases when velocity samples had different signs, the saccade direction was selected to be the same

as the sign of the velocity point with the highest absolute value.

Experimental Setup

Equipment

The experiments to collect data to generate saccade amplitude prediction and direction detection models were conducted with the Tobii x120 eye tracker, which is represented by a standalone unit connected to a 19 inch flat panel with a resolution of 1280x1024. The flat panel screen has its width as 37.8cm and its height as 30.2cm. In terms of degrees, the display has 30.22° as its width and 24.2° as its height. The distance between subjects and the screen is 70cm. This eye tracker performs binocular tracking with the following characteristics: accuracy 0.5°, spatial resolution 0.2°, and drift 0.3°. The sampling frequency for the eye tracker is 120Hz. Tobii x120 model allows 300x220x300 mm freedom of head movement. Nevertheless a chin rest was employed for higher accuracy.

Stimulus Presentations

A saccade inhibition stimulus was presented as a ramp stimulus (Leigh and Zee, 2006) where a dot appeared at the random horizontal location on the screen (vertical coordinate was fixed to the center of the screen). First, the dot flashed for 1000 ms.; then it disappeared; then, immediately a new dot appeared at the new, random location within the screen boundaries. The minimum distance between two consecutive dots was 2°. The maximum distance was 30° (approximately the width of the screen). Average presented distance amounted to 10°. Each subject was presented with a sequence of 30 dots.

Participants

Thirty five college students were recruited in undergraduate courses at Texas State University. Participants were compensated for their participation with extra credit in courses within the Departments of Psychology and Computer Science. All materials and procedures were approved by the Institutional Review Board at Texas State University, and informed consent was obtained from all participants prior to the testing session. On average, participants were 20.62 years of age [SD = 2 years; range = 18-25]. Of the 35 participants tested, 85% were male, 69% were of European-American descent.

Quality of the Recorded Data

Prior to the experiment, participants were screened for the actual accuracy and noise levels of the eye-tracker hardware using software developed in the Human Computer Interaction Laboratory at Texas State University (Komogortsev and Gowda, 2008). Participants with reported calibration error of more than 1° and a noise level of more than 16% were excluded from the analysis of eye movement data. The noise level is defined as the percentage of eye position samples for which the eye tracker failed to report the eye position coordinates. Some recording failures of the eye tracker occurred, due to conditions such as squinting and excessive moisture of the eye. In eye-tracking experiments, noise level parameters are rarely reported, but they serve as a major validation metric that should be specifically stated to verify the validity of the results.

Evaluation Metrics

The *Root Mean Squared Error* (RMSE) between the predicted A_{sac_pre} and the actual saccade amplitude A_{sac} determines the accuracy of the saccade prediction algorithm. $RMSE^M = \sqrt{\frac{\sum_{k=1}^N (A_{sac} - A_{sac_pre})^2}{N}}$. M is the model's name. The ideal saccade prediction model will have an RMSE of 0°.

Direction Prediction Error (DPE) represents the amount of erroneously detected saccade direction prediction—i.e., the rightward saccade was predicted as leftward saccade and vice-versa. The perfect scheme would have an error rate of 0.

Both metrics were computed in the following way. Out of 35 recordings, 25 were randomly selected to create functions connecting the Chi-square test value to the saccade amplitude according to the heuristic of each prediction model. The remaining 10 recordings were employed to compute the RMSE and DPE metrics.

Results

Amplitude Prediction Model

Equation (14) shows the chi-square regression model created for the Two Samples scenario, $p < 0.001$. The attempt to develop a reliable regression model for the First Sample scenario did not succeed due to the failure to provide significant terms in the model, $p > 0.7201$.

$$\begin{aligned}
 A_{sac_pre} = & -107 \cdot 10^{-13} \cdot (\chi_1^2)^5 + 1.848 \\
 & \cdot 10^{-8} \cdot (\chi_1^2)^4 \\
 & - 0.00001 \cdot (\chi_1^2)^3 \\
 & + 0.000266 \cdot (\chi_1^2)^2 \\
 & - 0.3008 \cdot (\chi_1^2) - 179 \\
 & \cdot 10^{-15} \cdot (\chi_2^2)^5 + 5.36 \\
 & \cdot 10^{-10} \cdot (\chi_2^2)^4 - 4.8 \\
 & \cdot 10^{-7} \cdot (\chi_2^2)^3 \\
 & + 0.000084 \cdot (\chi_2^2)^2 \\
 & + 0.0586 \cdot (\chi_2^2) \\
 & + 9.1502
 \end{aligned} \tag{14}$$

A_{sac_pre} is the amplitude of the predicted saccade, χ_1^2 represents the chi-square test value detected at the onset of a saccade, and χ_2^2 represents the chi-square test value recorded at the next eye gaze position after the onset of a saccade.

Amplitude Prediction Error

The Velocity Model yielded average RMSE of 3.38°. We hypothesize that these errors were due to high noise in the eye tracker. For example, spatial resolution of 0.2° results in a velocity noise of 24°/s without an eye effectively moving anywhere. For each subject in remaining 10 recordings, the average RMSE for the Two Sample chi-square regression model was 5.26°. However, RMSE for the Velocity Model was significantly lower than that for the chi-square regression model, $F(1,9)=15.07$, $p=0.0037$.

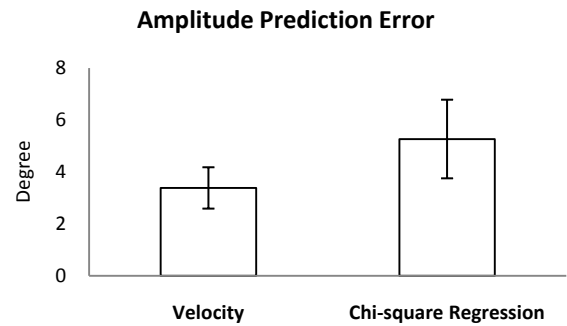


Figure 2. Saccade amplitude prediction errors by the two models based on average RMSE

Also, the scatter plots of actual saccade amplitude vs. predicted saccade amplitude by each of the two predic-

tion models are illustrated (Figures 3 and 4). These plots provide information on prediction error distribution according to the magnitude of saccade amplitude.

To quantify how close the distribution of predicted amplitude by each model is to the distribution of actual amplitude, Kullback-Leibler (KL) divergence was measured by assuming saccade amplitude as a discrete random variable. The results are:

$$D_{KL}(p||q_{velocity})=0.5483$$

$$D_{KL}(p||q_{chi-square\ regression})=0.5261$$

The values are not greatly different, thus it is hard to state which distribution is closer to the distribution of actual amplitude using KL divergence.

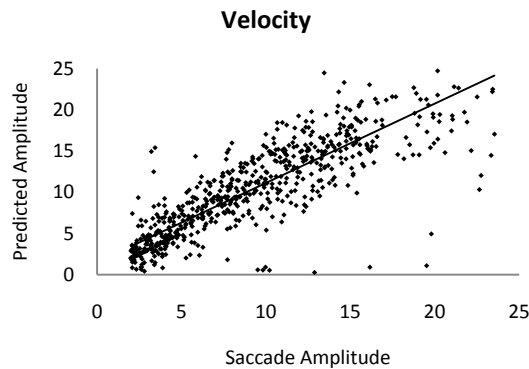


Figure 3. Scatter plots between actual saccade amplitude and predicted amplitude by the velocity model

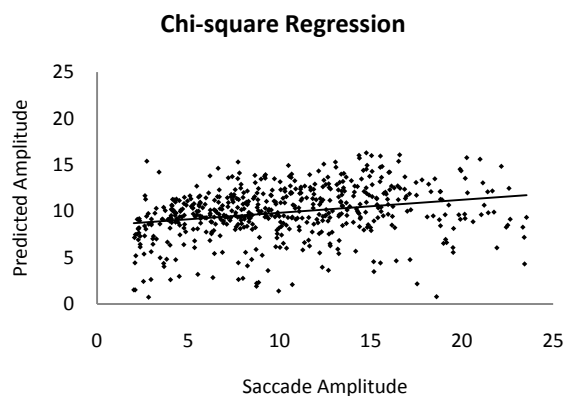


Figure 4. Scatter plots between actual saccade amplitude and predicted amplitude by the chi-square regression model

Direction Prediction Error

Average error rate for the First Sample model for all 35 recordings was 5.27%. The Two Samples model performed significantly better by reducing average error rate to 1.54%, $F(1,34)=25.87, p<0.01$.

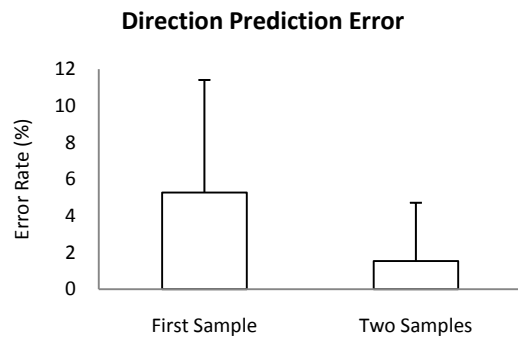


Figure 5. Direction prediction error rates for First Sample and Two Samples models.

To investigate the relationship between the magnitude of saccade amplitude and the frequency of direction error, the direction error distribution is provided (Figure 6). The distribution shows that direction error occurred more frequently for the saccades of smaller amplitudes. The probable reason for this type of behavior is that saccades of smaller amplitudes are distorted more by the eye tracking equipment noise than saccades of larger amplitudes.

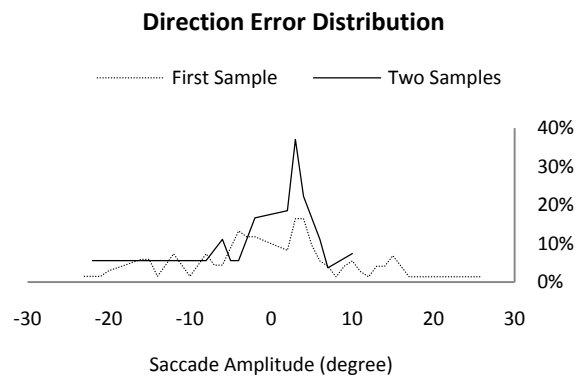


Figure 6. Direction error distribution according to saccade amplitude

Discussion

Prediction Accuracy Challenges

Average amplitude prediction error of 5.26° seems to be large for the average saccade amplitude of 10° . We hypothesize that the reason for this is the high amount of noise provided by the eye tracking equipment. Therefore, additional research is required to find noise reduction algorithms to reduce the effects of noise in the recorded data and thereby increase the accuracy of prediction.

It is worth mentioning that, in previous research, Komogortsev and Khan (Komogortsev and Khan, 2007) proposed the following equation for saccade amplitude prediction.

$$A_{sac_pre} = -24 \cdot 10^{-6} \cdot (\chi^2)^3 + 536 \cdot 10^{-4} \cdot (\chi^2)^2 + 1.5 \quad (15)$$

When this equation was applied to the data collected in the experiments described in this paper, the recorded average RMSE was more than 35° (larger than the monitor size). One possible reason for such low accuracy of prediction to the different frequency of eye position sampling recording in our experiments and in experiments presented in Komogortsev and Khan (2007). The frequency employed in their experiment was 50Hz and frequency employed in our experiments was 120Hz. The other possible issue is that the function presented in Equation (15) was derived empirically from only one recording of one subject.

Applications

Quick models for saccade amplitude prediction have potential to benefit the area of gaze-contingent-compression, by reducing the amount of lag created by sensing, processing, and transmission delays. Delay reduction allows achievement of higher levels of compression (Komogortsev, in press; Komogortsev and Khan, 2004). Specifically an amplitude prediction model would provide the location for the high quality ROI region that would be placed at the coordinates of the offset of a saccade. The quickness of the saccade's amplitude prediction would result in delay reduction. Below we provide a mathematical evaluation of the First Sample and Two Sample Chi-square regression models assuming that each of the models accurately predicts a saccade's amplitude (amplitude prediction error of 0°). In our analysis, we also fix eye position sampling interval to be 120 Hz, the

same as the sampling interval employed in our experiments.

The amount of delay reduction each model provides depends on saccade's duration and the amount of time each model requires to predict the saccade's amplitude.

$$T_{delay_reduc} = T_{sac_dur} - T_{predict} \quad (16)$$

Formulas for T_{sac_dur} and $T_{predict}$ were discussed in the "Method" section above.

Once delay reduction is estimated, it is possible to assess the amount of compression savings the models would provide for a GCC system. To provide such an estimate we have taken GCC performance results reported by Komogortsev (Komogortsev, in press). Table 1 provides a summary of the results.

Table 1. Average Perceptual Resolution Gain achieved with corresponding delay.

| Delay | 0 ms | 20 ms | 500 ms | 1000 ms | 2000 ms |
|-------|------|-------|--------|---------|---------|
| APRG | 2.6 | 2.2 | 1.6 | 1.5 | 1.4 |

Average Perceptual Resolution Gain (APRG) is a quantity that indicates the amount of compression a gaze-contingent approach provides given the value of the delay (Komogortsev and Khan, 2004). It is possible to provide a logarithmic approximation of APRG given the delay values

$$APRG_{orig} = -0.176 \ln(T_{delay}) + 1.5 \quad (17)$$

T_{delay} represents the amount of delay in the GCC system measured in seconds. In this approximation, lowest value for T_{delay} is assumed to be 0.001 ms. While Equation (17) provides the APRG estimation for the "Original" (no delay compensation) type of GCC system, the GCC system's performance aided by the amplitude prediction model can be presented as:

$$APRG_{mod} = -0.176 \ln(T_{delay} - T_{delay_reduc}) + 1.5 \quad (18)$$

In cases when expression $(T_{delay} - T_{delay_reduc}) \leq 0$ the value for $APRG_{mod}$ is computed with value of $(T_{delay} - T_{delay_reduc})=0.001$.

The resulting compression increase (CI) can be estimated as

$$CI = 100 \left(1 - \frac{APRG_{orig}}{APRG_{mod}}\right) \quad (19)$$

were $APRG_{mod}$ is the compression value provided as a result of delay compensation.

Figures 7 and 8 present comparison results created by two scenarios defined by the average amplitude of the exhibited saccades. The first scenario considers average saccade amplitude of 5° , and the second scenario considers average saccade amplitude of 10° .

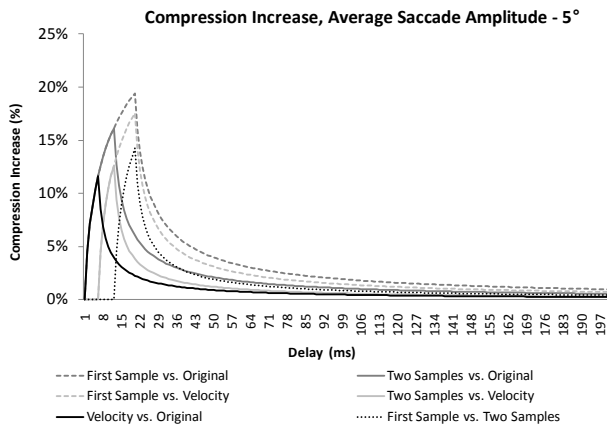


Figure 7. Compression increase provided by the amplitude prediction models. Average Saccade Amplitude is 5° .

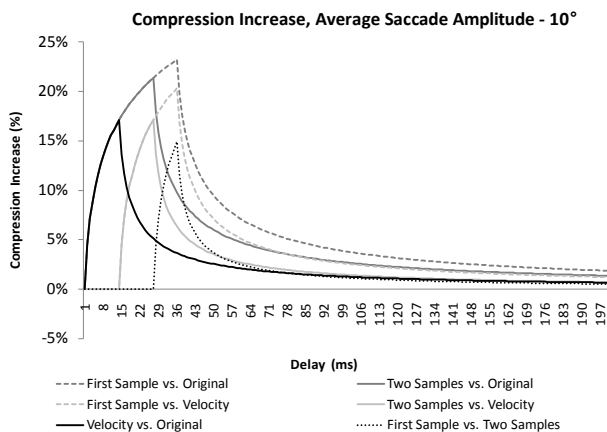


Figure 8. Compression increase provided by the amplitude prediction models. Average Saccade Amplitude is 10° .

Average saccade amplitude of 5° : the results indicate that the Velocity model would provide a 14% compression improvement for the delay of 9 ms. There is a steep reduction in performance after this value, e.g., delay of 50 ms achieves an improvement of performance of just 1%. The First Sample model increases the effective range of delay compensation, providing the highest compression

increase (21%) for the delays of up to 25 ms. There is a steep reduction in performance after that e.g., delay of 50 ms. reduces CI value down to 5% and delay of 100 ms. reduces the performance increase down to 2%. When performance of the First Sample model is compared to the performance of the Velocity model, the performance of both models is the same up to the 8 ms mark (the timing of the first interval); after this point we can see a constant rise of the performance until the point of 23-25 ms where performance reaches the value of 18%. There is a slow reduction of performance after this point.

The performance of the Two Samples model mimics the performance of the One Sample model when compared to the Original and Velocity aided compression. Two Sample model provides the peak of compression savings of 18% in the 15-16 ms delay range when compared to the Original compression and the improvement peak of 13% in case of the 15-16 ms delay when compared to the Velocity aided compression. The savings in compression are lower for both cases when the delay values are increased. The biggest gain in compression saving (15%) occurs at the 25 ms delay mark when the First Sample model is compared to the Two Samples model.

Average saccade amplitude of 10° : Larger saccades provide an opportunity to increase the compression improvement benefits and the effective delay compensation range produced as a result of saccade amplitude prediction. Considering average saccade amplitude of 10° the performance of the Velocity and One Sample models are the same up to 14 ms. After this point the Velocity model performance degrades, going down to 3% at the level of 50 ms., and the performance reaches the mark of 1% or less at the point of 85ms. The peak of performance improvement (23%) occurs in the 33-36 ms interval for the First Sample model, going down to the mark of 9% at the 50 ms. delay interval and 4% at the 100 ms delay interval. When the First Sample model is compared with the Velocity model, the peak of performance improvement occurs at the interval of 34-36 ms providing an improvement of compression by 20%, and the performance improvement decreases after this value.

The Two Samples model provides the peak of compression savings of 21% in the 24-27 ms delay range when compared to the Original compression and the improvement peak of 17% at 26-27 ms delay when compared to the Velocity model. The savings in compression are lower for both cases when the delay values are in-

creased. The biggest gain in compression saving (15%) occurs at the 36 ms delay mark when the First Sample model is compared to the Two Samples model.

We have provided a theoretical evaluation of the impact that quick saccade amplitude prediction models would have on the gaze-contingent compression domain. The actual implementation of such models in a real-time GCC system might change the estimated compression results. For example such factors as delay jitter, targeted amount of eye-gaze samples required to be contained by a GCC system (Komogortsev, in press), changes of the eye tracker's sampling frequency and prediction error compensation mechanisms were not considered in our evaluation. The detailed empirical evaluation of these factors is beyond the scope of this paper.

Conclusion

This paper has introduced a new saccade amplitude prediction model and compared it to the published work of Anliker (Anliker, 1976). The new model was based on a non-linear regression model using the chi-square test values computed by a two state Kalman filter implementation of the Human Visual System.

The results indicate that the Velocity model (Anliker's) was the most accurate producing a 3.46 degree error on average, but the model required that the middle of a saccade be reached, therefore reducing the potential benefit of saving time. The proposed regression model produced a higher saccade amplitude prediction error, but it predicted saccade amplitude in a constant amount of time. The regression model based on the data from just one eye position sample at the beginning of a saccade did not provide significant terms. The regression model aided by the data from the second position sample had significant terms with average amplitude prediction error of 5.26° (average amplitude of exhibited saccades was approximately 10°).

The saccade's direction prediction heuristic, presented in our paper, allows prediction of saccade direction with the success rate of 95% based on the data from just one eye position sample at the beginning of a saccade. Given the data from the second consecutive sample, the heuristic increases the success rate to 98.46%.

We have provided a theoretical evaluation of the benefits that saccade amplitude prediction can bring to the

realm of real-time gaze contingent compression. The result indicates that in the case using the two-eye position sample prediction model, the benefit of using the amplitude prediction model would be 21% of additional compression savings for the short network delays. For the case of the one eye-position sample model the compression savings would increase to 23%.

Reference

- Abd-almageed, W., Sami, M., and Bebis, F. G. (2002). A nonintrusive kalman filter-based tracker for pursuit eye movements. *American Control Conference*.
- Anliker, J. (1976). Eye Movements: On-Line Measurement, Analysis, and Control. Eye Movements and Psychological Processes, R. A. Monty and J. W. Senders, eds., Lawrence Erlbaum Associates, Hillsdale, 185–202.
- Bahill, A. T. (1980). Development, validation and sensitivity analyses of human eye movement models. *CRC Critical Reviews in Bioengineering*, 4, 311-355.
- Brown, R., and Hwang, P. (1997). *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley and Sons, New York.
- Carpenter, R. H. S. (1977). *Movements of the Eyes*, Pion, London.
- Duchowski, A., T., and Çöltekin, A. (2007). Foveated gaze-contingent displays for peripheral LOD management, 3D visualization, and stereo imaging. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 3(4).
- Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice* Springer.
- Huckauf, A., and Urbina, M. H. (2008a). Gazing with pEYEs: towards a universal input for various applications. Proceedings of the 2008 symposium on Eye tracking research & applications, ACM, Savannah, Georgia.
- Huckauf, A., and Urbina, M. H. (2008b). On object selection in gaze controlled environments. *Journal of Eye Movement Research*, 2(4), 1-7.
- Irwin, D. E. (1992). Visual Memory Within and Across Fixations. *Eye movements and Visual Cognition: Scene Preparation And Reading*, 146-165.

- Istance, H., Bates, R., Hyrskykari, A., and Vickers, S. (2008). Snap clutch, a model approach to solving the Midas touch problem. Proceedings of the 2008 symposium on Eye tracking research & applications, ACM, Savannah, Georgia.
- Jacob, R. J. K. (1990). What you look at is what you get: eye movement-based interaction techniques. Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people, ACM, Seattle, Washington, United States.
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME 6B "Journal of Basic Engineering"*(82 (Series D)), 35-45.
- Kohler, M. (1997). Using the Kalman Filter to track human interactive motion --- Modelling and initialization of the Kalman Filter for translational motion.
- Komogortsev, O. V. (in press). Gaze-Contingent Video Compression with Targeted Gaze Containment Performance. *Journal of Electronic Imaging*.
- Komogortsev, O. V., and Gowda, S. M. (2008). Eye Tracker Calibration Accuracy Evaluation Tool
- Komogortsev, O. V., and Khan, J. Predictive Perceptual Compression for Real Time Video Communication. *12th ACM International conference on Multimedia (ACM MM 04)*, New York, 220-227.
- Komogortsev, O. V., and Khan, J. Kalman Filtering in the Design of Eye-Gaze-Guided Computer Interfaces. *12th International Conference on Human-Computer Interaction (HCI 2007)*, Beijing, China, 1-10.
- Komogortsev, O. V., and Khan, J. Eye Movement Prediction by Kalman Filter with Integrated Linear Horizontal Oculomotor Plant Mechanical Model. *Eye Tracking Research & Applications Symposium 2008*, Savannah, GA, 229-236.
- Komogortsev, O. V., and Khan, J. (2009). Eye Movement Prediction by Oculomotor Plant Kalman Filter with Brainstem Control. *Journal of Control Theory and Applications*, 7(1), 14-22.
- Komogortsev, O. V., and Khan, J. I. (2008b). Predictive real-time perceptual compression based on eye-gaze-position analysis. *ACM Trans. Multimedia Comput. Commun. Appl.*, 4(3), 1-16.
- Kumar, M., Paepcke, A., and Winograd, T. (2007). Eye-Point: practical pointing and selection using gaze and keyboard. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, San Jose, California, USA.
- Kumar, M., and Winograd, T. (2007). Gaze-enhanced scrolling techniques. Proceedings of the 20th annual ACM symposium on User interface software and technology, ACM, Newport, Rhode Island, USA.
- Leigh, R. J., and Zee, D. S. (2006). *The Neurology of Eye Movements*, Oxford University Press.
- MacKenzie, I. S., and Zhang, X. (2008). Eye typing using word and letter prediction and a fixation algorithm. Proceedings of the 2008 symposium on Eye tracking research & applications, ACM, Savannah, Georgia.
- Motulsky, H. (2001). The GraphPad Guide to Nonlinear Regression. GraphPad Software.
- Motulsky, H., and Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*, NY Oxford University Press New York.
- Parkhurst, D., J., and Niebur, E. (2002). Variable resolution displays: A theoretical, practical, and behavioral evaluation. *Human Factors*, 44(4), 611-629.
- Rewari, G., and Chi-Sang, P. Saccadic eye movement detection using Kalman filtering and the generalized likelihood ratio approach. *Engineering in Medicine and Biology Society, 1993. Proceedings of the 15th Annual International Conference of the IEEE*, 418-419.
- Reingold, E., M., Loschky, L., C., McConkie, G., W., and Stampe, D., M. (2003). Gaze-Contingent multi-resolutional displays: An integrative review. *Human Factors*, 45(2), 307-328.
- SAS. (2009). The data analysis for this paper was generated using SAS software, Version [8] of the SAS System for Windows. Copyright © 2009 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
- Sauter, D., Martin, B. J., Di Renzo, N., and Vomscheid, C. (1991). Analysis of eye tracking movements using innovations generated by a Kalman filter. *Med. Biol. Eng. Comput.*, 63-69.
- Sibert, L. E., and Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, The Hague, The Netherlands.
- Yarbus, L. (1967). *Eye Movements and Vision*, Institute for Problems of Information Transmission Academy of Sciences of the USSR, Moscow.

Zhai, S., Morimoto, C., and Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, ACM, Pittsburgh, Pennsylvania, United States.