Modeling the Scheduling of Eye Movements and Manual Responses in Performing A Sequence of Discrete Tasks

Shu-Chieh Wu (scwu@mail.arc.nasa.gov)
San Jose State University / NASA Ames Research Center, Mail Stop 262-4
Moffett Field, CA 94035 USA

Roger W. Remington (roger.remington@jhuapl.edu)
Johns Hopkins University Applied Physics Laboratory
Laurel, MD 20723-6099 USA
)
Richard Lewis (rickl@umich.edu)
Department of Psychology, University of Michigan
Ann Arbor, MI 48109-1043 USA

Abstract

Common tasks in daily life are often accomplished by a sequence of actions that interleave information acquisition through the eyes and action execution by the hands. How are eye movements coordinated with the release of manual responses and how may their coordination be represented at the level of component mental operations? We (Wu & Remington, 2004; Wu, Remington, & Pashler, 2004) have previously presented data from a typing-like task following Pashler (1994) requiring separate choice responses to a series of five stimuli. We found a consistent pattern of results in both motor and ocular timing, and hypothesized possible relationships among underlying components. Here we report a model of that task, which demonstrates how the observed timing of eye movements to successive stimuli could be accounted for by assuming two systems: an open-loop system generating saccades at a periodic rate, and a closed-loop system commanding a saccade based on stimulus processing. Future work will focus on testing the generality of this approach across a wide range of scanning conditions.

Introduction

Human performance modeling entails reconstructing human behavior through composition of elementary cognitive and motor operations. The common assumption derives from traditional theories of cognition, which decompose human performance into a succession of stages representing a limited number of component mental operations, such as perceptual, cognitive, and motor processes. This approach has achieved success in modeling highly practiced, perceptual-motor behaviors (e.g., Gray et al., 1993; John et al., 2002; Vera et al., 2005). However, these models rely on component estimates derived from discrete actions that often last less than a second. Real life tasks generally require the performance of a series of actions integrated into a fluid behavior sequence. Features not observed in discrete performances emerge, such as planning, scheduling, coordination, and optimization. It is an empirical as well as practical question in human performance modeling whether models of single-task performance, described at the level of elementary mental operations, are sufficient to characterize behavior in extended, continuous sequences with emergent features.

In a series of papers we have investigated one particular aspect of continuous behavior, the coordination between eye movements and manual responses. In human interactions with technologies, sequences of discrete actions are executed by hands through keyboards, mice, touch pads, or joysticks based on information taken in from the eyes with decisions made centrally. *How are eye movements coordinated with the release of manual responses in sequences of discrete actions that involve decisions?* In this paper, we present the results of one extended task experiment that examined the characteristics of eye movements in performing sequences of discrete actions and a computational model to account for the results. The empirical results have been reported previously (Wu, Remington, & Pashler, 2004).

Eye Movements in Extended Tasks

Eye movements occur naturally in almost all daily activities. Yet, to characterize patterns of eye movements, past researchers have focused on activities with a clear script. Examples of such activities range from golf putting (Vickers, 1992), driving (Land & Lee, 1994), to tea making (Land & Hayhoe, 2001), and block-copying (Pelz et al., 2001). A common finding in such observations is that the eyes move in anticipation of upcoming actions during activities that involve scripted behavior.

The existence of preview in extended task performance characterizes the proactive nature of eye movement control. In tasks that require mostly non-visually based decisions, it seems intuitive that the eyes could move away prior to the response as soon as information acquisition is completed. But, when can the eyes move and what determines it? Answers to these questions are critical to understanding the coordination between eye movements and other mental operations, as typical fixation durations generally range from 200 to 400 ms, exceeding the time needed for perceptual registration, which can be estimated at around 100 to 150 ms (Salthouse & Ellis, 1980). This observation suggests either that the initiation of an eye movement is based on post-perceptual processing, or alternatively, that

the timing of movements is driven by strategies of information acquisition.

To study when the eyes move we need a task that incorporate eye movements in a natural fashion while sensitive to changes in underlying processes. We adopted a typing like task developed by Pashler (1994). Participants viewed a row of five letters sequentially and responded to each individually in different preview conditions. By manipulating the amount of preview, whether information is made available prior to response, Pashler examined how the mental processing of two or more stimuli were overlapped in time. He measured the reaction time (RT) to the first stimulus (RT1) and computed the inter-response intervals (IRIs) for subsequent responses. With no preview, RT1 and subsequent IRIs were roughly equivalent and constant across the stimulus sequence. With preview, RT1 was elevated, compare to no preview, while IRIs were constantly low. The same effects were observed regardless of whether 1 or 4 preview items were presented. Pashler interpreted the constant IRIs as an indication of a bottleneck central processing stage of response selection, which would only allow the selection of one response at a time (see Figure 1 for a schematic diagram). The fact that IRIs reflected the duration of response selection is further supported by the findings that varying the duration of stimulus recognition and response production had little to modest effect on the durations of IRIs.
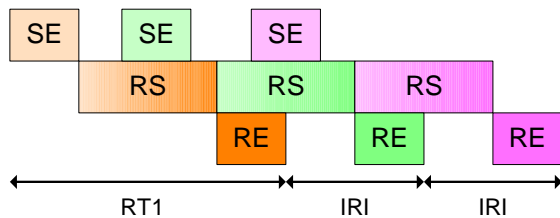

Figure 1. Pashler's (1994) model of the preview task

Pashler's (1994) provides a clear theoretical account of how the coordination of three critical mental components (perception, response selection, and response production) can account for the emergent effect of IRI in a sequence of choice responses. In this case, characterization of a single task was sufficient to account for the IRI results without further assumptions. The experimental paradigm represents a good compromise between the simplicity of typical discrete trial experiments, and real-world behavior.

To incorporate an eye movement component, we (Wu & Remington, 2004) adopted Pashler's complete preview condition and reduced the size of stimulus letters and increased the separation between them. Identification of stimulus letters thus required successive saccades and fixations. In two experiments we examined the effects of the level of difficulty in the perceptual and response selection

stages of manual responses on eye movement patterns. Perceptual difficulty was manipulated by having two stimulus luminance conditions (Bright and Dim). Response selection difficulty was manipulated by having two response mapping conditions, one natural (Easy) and one arbitrary (Difficult). In addition to RT and IRIs, we computed three eye movement related measures: 1) eye-hand spans (EH-Span), which represent the elapsed time between the initial fixation on a particular stimulus to the moment when the corresponding manual response is generated; 2) dwell time, which represents the duration for which fixation is maintain on a particular stimulus; and 3) release-hand spans (RH-Span), which represent the elapsed time between the end of fixation on a particular stimulus to the moment when the manual response is generated. In fact, dwell times and release-hand spans make up eye-hand spans.

In both experiments we (Wu & Remington, 2004) obtained patterns of manual RT results resembled what Pashler (1994) found in conditions with preview; specifically, the elevation of RT1 and constantly short IRIs of subsequent responses. The effect of perceptual difficulty was minimal on RT1/IRIs and appeared to be restricted to S1. Dwell time was lengthened in the Dim condition, though the amount of increase was small. Response selection difficulty had a strong effect on manual as well as oculomotor responses. Difficult response mapping resulted in increases in IRIs. It also significantly increased dwell times. Results from these two experiments suggest that fixations include perceptual as well as at least part of response selection related processes.

Present Research

In a follow-up study we (Wu, Remington, & Pashler, 2004) examined the effects of response selection difficulty within a trial on extended task performance using a Go/No-Go procedure. On each trial, only 2 or 3 positions contained target characters mapped with a key response. The rest were filled with dummy characters and participants were asked to skip them. We compare dwell time on Go and No-Go stimuli, and RT1 and IRIs on Go stimuli that are preceded and/or followed by No-Go stimuli to evaluate the impact of central difficulty. Our goal of using a Go/No-Go procedure was twofold. First, real-world tasks do not require actions to each and every stimulus, yet for those stimuli with no motor response requirements some amount of processing still takes place. By intermixing No-Go stimuli with Go ones we were able to more closely imitate real-world conditions. Second, in our previous research (Wu & Remington, 2004) with all Go stimuli we found substantial RT1 elevation in sequences of responses. With the inclusion of No-Go stimuli we could test whether RT1 elevation was due to delayed early responses in order to be coordinated with subsequent ones. In the following we recap the empirical results and access the validity of our underlying process assumptions using a computational model for the results.

Empirical Results

We model data from Wu, Remington, & Pashler (2004), which used a Go/No-Go version of Pashler's (1994) preview task. There were six experimental conditions, which differed in the number of required successive responses in a sequence (one, two, and three), and in the stimulus position on which these sequences occurred (S1 and S2). The six conditions can be represented as the following: TXXTT, TTXXT, TTTXX, XTXXT, XTTXX, and XTTTX, with T denoting letter stimuli that required a key response (Go stimuli) and X denoting letter stimuli that required no response (No-Go stimuli). Go stimuli were randomly drawn from the letter set T, D, and Z, with the constraint that no letter was repeated in two adjacent positions. This constraint however does not prevent repetition of responses; the same letter could occur in two positions interposed by Xs. Each participants received either 40 or 60 trials of each condition.

Each trial began with the presentation of a white fixation cross (0.3˚) in the center of the display. After the participant had maintained fixation within a 6˚ radius around the fixation for 500 ms, the fixation was erased and a small filled square (0.34˚) appeared at the leftmost stimulus position. Participants were instructed to move their eyes to fixate the small square when it appeared and maintain fixation at that location. The small square remained for 1 sec, followed by a blank interval of 500 ms. The five stimulus characters then appeared simultaneously, the leftmost appearing where the subject had been fixating. Subjects fixated and responded to each in turn from left to right. The characters were erased after the participant had responded to the rightmost character. The next trial began following an inter-trial-interval of 250 ms.

Eye movements were monitored and recorded using an infrared video-based eye tracking system (ISCAN), which outputs data at a temporal resolution of 120 Hz and a spatial resolution of approximately 0.5° visual angle. Fourteen undergraduate students recruited from local colleges near NASA Ames Research Center participated in the experiment for course credit.

Figure 2 (left panel) presents mean manual RTs as a function of stimulus. Cases where RT1 occurred to S1 (S1-RT1) are plotted separately from cases where RT1 occurred to S2 (S2-RT1). The general pattern of elevated RT1 followed by rapid, flat IRIs is apparent in conditions with a contiguous sequence of responses (TTTXX and XTTTX). More importantly, RTs of the Go responses followed by No-Go sequences (conditions TXXTT, TTXXT, XTXXT, and XTTXX) were closely aligned with the Go responses followed by Go sequences. The only significant effect of the arrangement of stimulus was that RT1 was significantly slower when made to S1 (S1-RT1) than to S2 (S2-RT1). The general elevation of RT1 for both S1-RT1 and S2-RT1 suggests that cost is incurred for the first response in a sequence, not just to the first possible stimulus position. These similarities in patterns and magnitudes strongly suggest that the RT1/IRIs patterns are related closely to the production of sequences of responses.
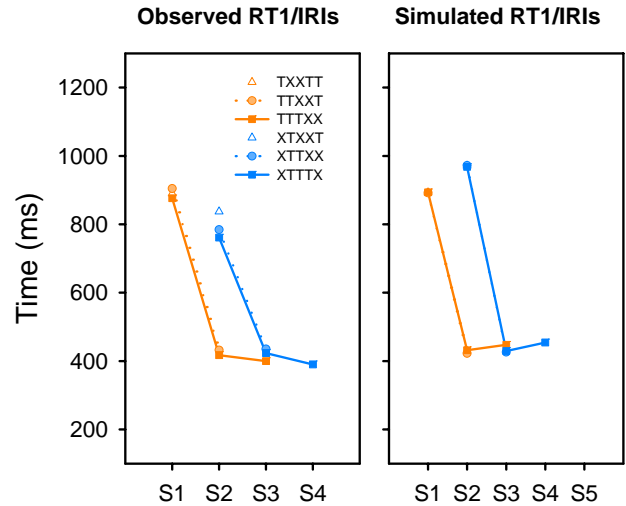

Figure 2. RT1/IRIs and eye-hand span results from Wu et al. (2004)

As in previous experiments, fixation durations remained relatively constant across stimuli. Not surprisingly, fixation durations on target (Go) stimuli were always longer than No-Go stimuli. More interesting comparisons arise when one regards fixation durations as a consequence of the previous stimulus (Figure 3). Here the dwell times suggest that the attempt to interleave the mental operations for successive stimuli pushes cost on to the subsequent stimulus. When the fixated stimulus is a target (a Go stimulus) dwell times were shorter by ~60 ms for targets that were preceded by dummy stimuli (i.e., XT) than by target stimuli (i.e., TT). When the fixated stimulus was a dummy stimulus this difference (TX compared to XX) was ~30 ms.

Condition	s1		s2		s3		s4	s5		T preceded by T or X
TXXTT	450.4	c	324.6	d	300.1	b	415.9	838.6	a	TT
TTXXT	467.6	a	478.8	c	315.1	d	307.9	768.9	b	XT
TTTXX	460.6	a	484.8	a	443.1	c	349.5	439.5		
XTXXT	339.9	b	394.4	c	344.0	d	329.9	776.8		X preceded by T or X
XTTXX	332.7	b	409.5	a	445.5	c	370.5	527.7		
XTTTX	336.6	b	395.7	a	433.3	a	473.3	581.8	c	TX
TXXXX	534.3		451.9		362.5		334.3	425.6	d	XX

Figure 3. Dwell times for S1-S5 in all conditions from Wu et al. (2004)

Model

Previously we (Remington, Lewis, & Wu, 2005) modeled a version of the preview task that consisted of trials with all "Go" stimuli. The model was developed under three key assumptions. First, it assumed that central processes are rate limiting (central bottleneck assumption). Second, an eye movement to the next stimulus in sequence can be initiated only after the completion of perceptual processing on the current stimulus. Third, the timing of the eye movement is strategically chosen so that perceptual processing of the next stimulus is completed at the same time as central processing on the current stimulus is completed. We refer to this as the

"just-in-time" assumption, since it attempts to minimize wait states in central processing by assuring that perceptual processing is complete as close as possible to when it is needed – when the central processor becomes free. In the present paper we modeled the Go/No-Go results based on the same three key assumptions. In the following sections we describe the structure of the model (see Figure 4 for a schematic diagram) and detail assumptions specific to the present model. The model was implemented in the statistical package R.

Assumptions

1. Following Pashler (1994) manual responses consist of three basic stages: stimulus encoding (SE), response selection (RS), and response execution (RE). Here we postulate that RS consists of two sub-stages, stimulus classification (RS.SC) and response preparation (RS.RP) to represent the differential processing need for Go and No-Go stimuli. RE is further broken down to two sequential components, an *Init* operator (I_M) followed by a *Motor Execution* operator (M).

2. We hypothesize that manual and oculomotor processes underlie each fixation. Specifically, each fixation begins with SE and ends with an Init component for the saccade (I_S). For fixations on No-Go stimuli, they include only the response clarification component (RS.SC). For fixations on Go stimuli, we introduce a variable called *Saccade Lag* that produces the constant delay required to achieve the just-in-time scheduling.

3. RS (including RS.SC and RS.RP) is the bottleneck process, occurring in strict sequence for all stimuli. Hence the task is rate-limited by RS, not SE or RE.

4. We assume that participants adopt an eye movement strategy that assures SE is completed just when RS is free to process the perceptual information. However with variability in processing times RS on a previous stimulus can be completed before or after SE. Stochastic durations in the model allow us to simulate both outcomes.
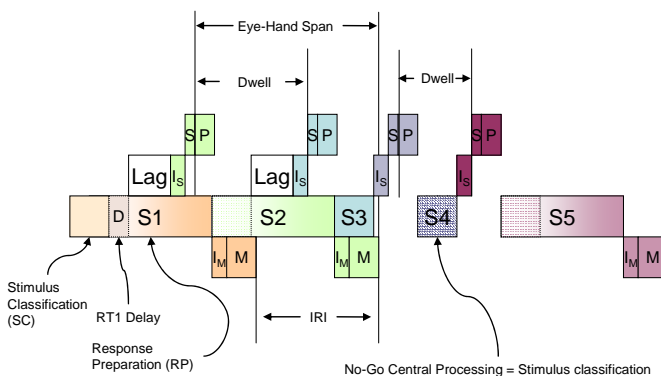
5. The model does not have an explicit account of the RT1 elevation that would allow us to compute it from first principles. Our conjecture is that the RT1 elevation results in part from preparation for the sequence of events to ensue. We insert an operation, *Prep*, into the sequence of RT1 after SE and prior to RS, and assume it is executed in series.

6. There is noise associated with the timing of all of these processes. In the model we assume that the noise patterns follow normal distributions.

Parameters

Though the model assumptions determine the order in which operations will be scheduled, durations must be assigned or computed. Parameters were estimated in three ways: 1) architectural conventions; 2) calculated from the data based on model assumptions; and 3) taken directly from the data.

Architectural Conventions Numerical parameter estimates for several necessary parameters were assigned values consistent with existing literature (e.g., Vera et al., 2005). Stimulus Encoding (SE) was set to 100 ms, both manual and saccade Init (I_M and I_S) operators were set to 50 ms, and Response Execution (RE) was set to 150 ms. Saccade movement time (S) was set to 30 ms.

Calculated from Data Although it would be desirable to set all parameters in advance, we have as yet no theoretical basis for assigning durations to some internal, unobservable states. For certain states the model provides guidance for computing estimates from observable measures. For example, to account for the elevated RT1 (on S1 and S2) we assume a response related preparation stage (Prep) is added to the normal model of a "Go" response. We estimate Prep.S1 by summing the mean values of the hypothesized components of RT1 then subtracting this from the observed RT1 as indicated here:

Prep.S1 = RT1 – (SE + RS + I + RE)

We estimated Prep.S2 by subtracting the observed difference in RT1 delay for S1 & S2. Thus,

Prep.S2 = Prep.S1 - 96

For Response Selection (RS), model assumptions dictate that for a sequence of "Go" trials, the IRI should ideally be a direct measure of RS. Thus, the entire RS stage was assigned the mean IRI time (416 ms) for pairs of adjacent "Go" trials. According to the model, "No-Go" trials have a smaller RS stage associated with Stimulus Classification (SC). The SC stage was estimated by constructing a model of final response time for the XTTTX sequence. This allowed RS.SC to be estimated from the data using the total response time and the delay of RT1 when it occurred on S2. Specifically,

Figure 4. A model of TTXXT and TXXTT conditions

$$RS.SC = Total.RT.XTTTX - SE*2 - I_S - S - Prep.S2 - RS*3 - I_M - M$$

A key assumption in the model is that eye movements are being delayed on "Go" trials in order to time the end of RS for the current stimulus with the end of SE on the subsequent stimulus. The final parameter needed to construct the sequence is an estimate of the Saccade Lag (Lag) that delays the saccade to produce the regular sequence of eye movements. To derive *Lag* we first note that the critical path for a total dwell time on a sequence of n successive "Go" trials starting from S1 is represented as:

$$Total.Dwell = SE_1 + Prep + RS*n - (SE_{n+1} + S)$$

The last term subtracts off that portion of the final RS where the eyes have moved to S_{n+1}. Here we use the TTTXX condition with n = 3. Because the model is stochastic we cannot be certain which elements will emerge on the critical path every run. For this estimate we assumed that in the majority of cases the critical path would be determined by the completion of RS stages (though it is possible the an SE stage will be delayed putting it on the critical path). Therefore the total time fixated would be equal to the total time minus the eye movement to the last stimulus. Because fixation duration is relatively constant across items, dwell times could be estimated by dividing the total fixation time by 3. We can compute the average Lag required to complete SE(n+1) at approximately the same time as RS(n) by

$$Saccade\ Lag = Total.Dwell / 3 - I_S - SE.$$

Simulation Results

Each model parameter was assigned a Gaussian standard deviation roughly equal to 25% of the mean. In simulations the model drew random durations from a Gaussian distribution with the indicated mean and standard deviation. Comparison of model predictions with obtained data for are shown in Figures x-x. Model means are based on Monte Carlo simulations of 1000 trials.

The qualitative fits of the model are moderately good, accurately reflecting the main trends in the RT/IRI data, the decreasing trends in the eye-hand span and release-hand span data, but failed to capture variances in dwell times. Although the model did yield different dwell times for Go and No-Go stimuli (465 ms versus 335 ms), it did not produce the pushback effects.

General Discussion

We (Remington et al., 2005) have previously presented a stochastic model that described how manual and ocular responses are generated in performing a sequence of linearly arrayed choice response time tasks. That model assumed that participants were trying to maintain a steady rate of eye movements (minimize eye fixation variability) and achieve
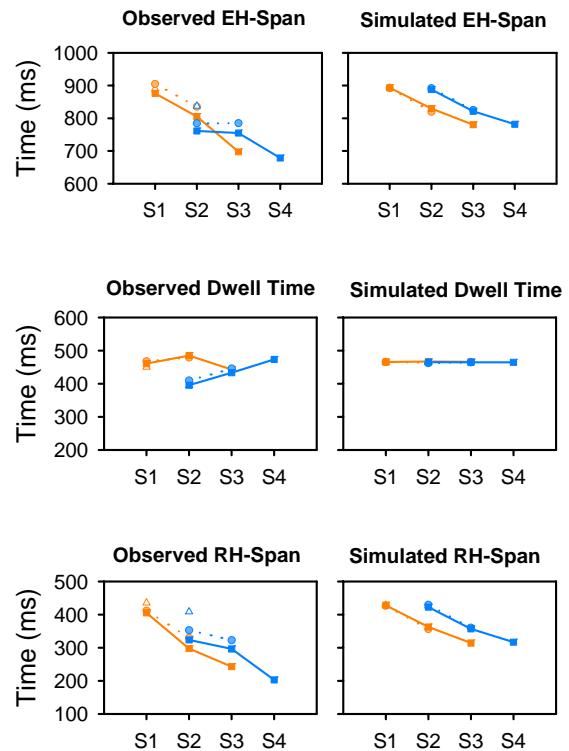

Figure 5. Observed and simulated eye-hand spans, dwell times, and release-hand spans data of Wu et al. (2004)

an efficient task completion rate (minimize waiting between central stages) by scheduling eye movements so that stimulus information would be available at the same time central processing resources became free ("just-in-time" scheduling). That model produced good approximations to data from trials where stimuli were of approximately equal difficulty. Here we extended that model to trials where stimuli were of unequal difficulty. Specifically, response selection demands were manipulated within a sequence of 5 characters by including "dummy" characters to which participants did not respond. Thus, some stimuli required a response ("Go"), some did not ("No-Go"). The model for heterogeneous trials retains the core logic of the homogenous case. Participants are assumed to establish a program that generates saccades at regular intervals. On "No-Go" trials an override is issued to the saccade execution system, which commands a saccade at that time. Thus, the model assumes the just-in-time goal remains in effect but can only be achieved on "Go" trials.

Moderately good qualitative fits to observed response time and oculomotor data were obtained. The model required four statistics from the empirical data to compute parameter estimates: RT1 for S1, RT1(S2), IRI, and total response time for XTTTX trials. Thus, the model fits rather than predicts the data. Nonetheless, the success of such a simple model shows that in principle it is possible to construct the extended sequence from assumptions about the scheduling of underlying component operations, at least for regular, ordered sequences such as text. The model also suggests dual-control approach where a succession of saccades is made. An open-loop pathway generates saccades

at a periodic rate, while a close-loop pathway intervenes to command a saccade on the basis of processed information.

This construction is similar to the SWIFT model of Engbert et al. (2005) developed to account for eye movements in reading, except their commanded signal is inhibitory, meant to delay the saccade for difficult words. They also derive movement parameters from assumptions about the underlying activation of words over space. Such a derivation has the desirable feature of providing a rigorous state representation from which decisions to move can be computed from a well-defined policy, such as speed-accuracy criteria. Yet, it is unclear whether this formulation provides any significant constraint on the outcomes possible.

Matessa (2004) developed a model of the homogeneous difficulty version also modeled by Remington et al (2005). Matessa's model required an extension of the ACT-RPM architecture to include an additional preparation stage. It produced qualitative fits to the data. It is unclear whether either the homogeneous or heterogeneous difficulty cases can be modeled in ACT-RPM without ancillary assumptions.

The use of central bottleneck theory provides a set of constraints that directly determine the scheduling and coordination of resources. However, unlike Engbert et al. (2005) we have yet to develop a representation of the underlying state of information processing that would allow us to generate a policy for optimal scheduling of eye movements. That is, we do not have a treatment that deals with factors that would affect tradeoffs between accuracy and duration of central processing, or allow us to estimate how information accrues or how accrual affects eye movement times. This is a direction for future work.

Conclusions

We have demonstrated that the patterns of manual responses and eye movements observed in a linear sequence of choice RTs can be modeled using a small set of assumptions derived from central bottleneck theory. While manual response emerged directly from bottleneck theory, the pattern of eye movements was accounted for by assuming two systems: an open-loop system generating saccades at a periodic rate, and a closed-loop system commanding a saccade based on stimulus processing. Future work will focus on testing the generality of this approach across a wide range of scanning conditions.

Acknowledgments

References

Engbert, R., Nuthmann, E. M., Richter, M., & Kliegl, R. (2005). SWIFT: A Dynamical model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8,* 237-309.

Land, M. F., & Hayhoe, M. 2001. In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559-3565.

Land, M. F., & Lee, D. N. 1994. Where we look when we steer. *Nature, 369,* 742-744.

Matessa, M. (2004). Anticipatory eye movements in interleaving templates of human behavior. In M. Lovett, C. D. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the Sixth International Conference on Cognitive Modeling.* Mahwah, NJ: LEA.

Pashler, H. (1994). Overlapping mental operations in serial performance with preview. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 47,* 161-191.

Pelz, J., Hayhoe, M., & Loeber, R. 2001. The coordination of eye, head, and hand movements in natural task. *Experimental Brain Research, 139,* 266-277.

Salthouse, T. A., & Ellis, C. L. 1980. Determinants of eye-fixation duration. *American Journal of Psychology, 93,* 207-234.

Vickers, J. N. 1992. Gaze control in putting. *Perception, 21,* 117-132.

Remington, R. W., Lewis, R., & Wu, S.-C. (2005). Scheduling mental operations in a multiple-response sequence: Modeling the effects of a strategy to minimize variance in the timing of saccades. Paper submitted to The Seventh International Conference on Cognitive Modeling.

Vera, A., John, B. E., Remington, R. W., Matessa, M., Freed, M. A. (2005). Automating human-performance modeling at the millisecond level. *Human-Computer Interaction, 20,* 225-265.

Wu, S.-C., & Remington, R. W. (2004). Coordination of component mental operations in a multiple-response task. In S.N. Spencer (Ed.), *Proceedings of the Eye Tracking Research and Applications Symposium 2004.* New York: ACM SIGGRAPH.

Wu, S.-C., Remington, R. W., & Pashler, H. (2004). Coordination of component mental operations in sequences of discrete responses. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Six Annual Conference of the Cognitive Science Society* (pp. 1476-1481). Mahwah, NJ: LEA.