

Scheduling Mental Operations in a Multiple-Response Sequence: Modeling the Effects of a Strategy to Minimize Variance in the Timing of Saccades

Roger W. Remington (roger.remington@jhuapl.edu)

The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road, Mail Stop 2-236, Laurel MD 20723-6099

Richard Lewis (rickl@umich.edu)

Department of Psychology, University of Michigan
1012 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043

Shu-Chieh Wu (scwu@mail.arc.nasa.gov)

San Jose State University and NASA Ames Research Center
Mail Stop 262-4, Moffett Field, CA 94035

Abstract

Common tasks in daily life are often accomplished by a sequence of actions. For highly practiced tasks, such as reading or piano playing, the eyes routinely move to the next item in sequence before the response to the current item is made (eye hand span), evidence of overlap in the mental processing of successive stimuli. We (Wu & Remington, 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 examined behavior under perceptual and central difficulty factors. Here we report a model of that task, which demonstrates how the observed timing of eye movements to successive stimuli could result from attempts to produce regular rhythmic output. The model makes two key assumptions about the generation of an eye movement to the next stimulus in a sequence. The first, dependence on perceptual encoding, asserts that the eyes remain fixated on the current stimulus until perceptual encoding is completed. The second, minimization of inter-movement time variability, asserts that people try to achieve a rhythmic regularity in motor performance, which includes the timing of eye movements. We discuss the conditions under which variance minimization might apply, and the implications for accounts that tie eye movements more directly to internal cognitive, perceptual, and motor operations.

Introduction

Psychological theories of task processing are derived largely, though not entirely, from studies using discrete-trial paradigms. Each trial typically requires a single response to a single stimulus or stimulus ensemble, the entire trial lasting about one to two seconds. Outside the laboratory such isolated stimulus-responses events are rare. Instead, we find overt sequences of looking, speaking, reaching, or other activities, and infer that these overt actions are coordinated with an intricate set of covert internal mental operations that

acquire information, build an understanding of events, and select appropriate actions. What are the rules that govern the coordinated scheduling of these overt and covert operations? While discrete-trial methods can uncover basic mechanisms that constitute the architecture of cognitive processing, they do not reveal how those mechanisms interact in continuous behavior. In the transition from discrete to continuous new behaviors emerge, not previously observed, such as coordination and overlapping among component mental operations. These phenomena require not just a set of resources, but rules that describe the constraints on scheduling them, and a description of the strategies people use in deploying their mental resources. We describe here a model that embodies these three elements -- resource architecture, constraints, and strategy -- in accounting for the coordination of manual responses, eye movements, and covert perceptual, cognitive, and motor operators presumed to underlie performance on a single discrete task.

One attempt to account for the coordination of internal resources in continuous tasks comes from a study of preview by Pashler (1994). In Pashler's preview paradigm, participants viewed a series of five letters and responded to each individually in sequence under different preview conditions. Pashler measured the reaction time (RT) to the first stimulus (RT1) and computed the inter-response intervals (IRIs) for subsequent responses. With no preview the next item appeared only upon the response to the current item. In that condition RT1 and subsequent IRIs were roughly equivalent and constant across the stimulus sequence. With preview the next item or the whole ensemble of five items were visible. In those conditions, RT1 was elevated, compare to no preview, while IRIs were faster and constant across successive items. The same effects were observed regardless of whether 1 or 4 preview items were presented. Pashler interpreted the results in terms of central bottleneck stage theory, in which stimulus processing is decomposed into three sequential stages: Stimulus Encoding (SE), Response Selection (RS), and Response Execution (RE). Consistent with central

bottleneck stage theory he posited that RS operations on the current item proceed concurrently with SE operations on subsequent items, and that RE on the current item can proceed in parallel with RS on the subsequent items. RS is the rate limiting operation, and the duration of IRI is a direct measure of the duration of the central RS stage. His model of the preview condition is shown in Figure 1. The fact that IRIs reflected the duration of RS is further supported by the findings that varying the duration of stimulus recognition and response production had little effect on the durations of IRIs.

Pashler's (1994) model showed how central bottleneck assumptions predicted the kind of overlap characteristic of skilled performance. No attempt was made to account for the placing of the stimulus encoding (SE) stages. Also,

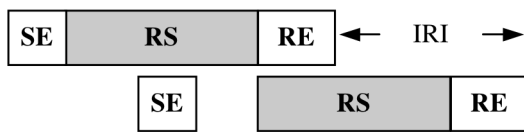


Figure 1: Overlapping of Stimulus Encoding (SE) and Response Execution (RE) with Response Selection (RS) for adjacent items showing Inter-Response Interval (IRI) following Pashler (1994).

because his stimuli were presented close to fixation it was assumed that eye movements played little or no role, and the model makes no attempt to incorporate them. Our goal was to extend Pashler's model to provide a quantitative account of the scheduling of resources in sequences of multiple choice response, which would include the coordination of eye movements with cognitive, perceptual, and motor components of task processing.

Understanding how eye movements are integrated into behavioral sequences is important for modeling reading, transcript typing, sight reading music, and the wide variety of other visually guided activities. The seemingly effortless and seamless integration of saccadic eye movements with other components of task performance provides possibly the best example of coordination and the most challenging task for human performance modelers. Yet existing implementations of eye movements (or gaze resources) rely on empirical findings from task conditions where eye movements are made to meet specifically instructions (e.g., Saltouse, 1983) rather than generated naturally in accord with task goals. Little work has addressed how task-driven eye movements are coordinated with the succession of stages and processes thought to characterize the underlying mental operations.

Eye Movements in Extended Tasks

The relationship of eye movements and other overt behaviors, such as hand movements, has been studied extensively in tasks ranging from golf putting (Vickers, 1992), driving (Land & Lee, 1994), to tea making (Land & Hayhoe, 2001). Eye movements have also been studied

using component tasks derived from everyday activities, such as block-copying (Pelz et al. 2001), reaching for and grasping objects (Johansson et al 2001; Terao et al 2002), and sequential looking and pointing (Epelboim et al., 1995). Although the specific use of gaze differs, in all of these activities the eyes tend to move in advance of the hands and head to fixate objects related to the immediate upcoming actions. For example, in block-copying a rhythmic pattern of movements was observed, in which the moving of each block began with the eyes fixating, followed by the head turning, and finally the hand (Pelz et al. 2001).

Similar patterns of proactive eye movements are also observed in activities such as transcription typing, which consist of series of repeated simple actions that, unlike the components of a reaching movement, are somewhat independent of each other. Yet, in skilled typing performance typists fixate from three (less skilled) to seven (more skilled) characters ahead of the ones being typed (Inhoff 1988; Inhoff and Wang 1992; Salthouse 1985, 1986). Similar pattern of eye movement preview is also found in sight reading of music (Furneaux and Land 1999; Kinsler and Carpenter 1995; Rayner and Pollatsek 1997; Truitt et al. 1997).

The existence of preview in extended task performance indicates that the eyes move to the next stimulus before the processing of the current stimulus is complete. *But, when can the eyes move and what determines it?* One possibility is that eye movements occur after the completion of some stage of processing. Some evidence suggests the eye can move after perceptual processing is complete (e.g., Sanders & van Duren, 1998). However, most accounts that link eye movements directly to internal processing stages posit that the eyes remain fixed until some amount of central processing as occurred. In part this is because typical fixation durations, which generally range from 200 to 400 ms, exceed the time needed for perceptual registration, which can be estimated at around 100 to 150 ms based on tachistoscopic experiments with backward masking (Salthouse & Ellis, 1980). Moreover, variables such as word frequency, which effect post-perceptual processing, also result in lengthened fixations that delay eye movements (e.g., Inhoff & Wang, 1992, 1998).

An alternative account has been offered for eye movement sequences in highly practiced continuous task such as reading or transcript typing. In these cases, eye movements tend to occur at regular intervals, although as noted above factors that increase central difficulty tend to delay the movements. A mixed model of eye movement control in reading has been suggested in which decisions on where and when to move the eyes are based on information from the current fixation as well as programs formed before the current fixation (Rayner & Pollatsek, 1981).

Here we describe a model of eye movement programming that shares with the mixed model above the assumption that eye movements are partly controlled by low-level programs, which attempt to produce saccades at regular intervals. The model demonstrates how the observed timing of eye

movements to successive stimuli could result from attempts to produce regular rhythmic output. It makes two key assumptions about the generation of an eye movement to the next stimulus in a sequence. The first, dependence on perceptual encoding, asserts that the eyes remain fixated on the current stimulus until perceptual encoding is completed. We do not specify here what the output of perceptual encoding is, but note that it is reasonable to assume fixation is required until the contents of a visual memory store have been transferred to cognitive mechanisms. The second, minimization of inter-movement time variability, asserts that people try to achieve a rhythmic regularity in motor performance, which includes the timing of eye movements. In highly practiced tasks, regular movements occur between 250-400 ms. It is possible that low-level mechanisms are restricted to times in this range. However, our model assumes that the timing of the movements is determined by the “just-in-time” strategic goal of providing perceptual information on the next stimulus at just the time central operations are available to process that information. Before describing the model in depth we describe two experiments using a multiple response paradigm, whose data will provide the empirical test of these assumptions.

Empirical Results

We model data from Wu & Remington (2004), which used a variant of Pashler’s (1994) preview experiment to examine the coordination of eye movement, manual responses and stimulus processing when executing a sequence of discrete choice response times. Pashler’s method was modified to force eye movements from one item to the next by reducing the size of the items and increasing the inter-item distance. 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 infra-red 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. The recording of eye movements began at the moment when the small square appeared, and ended after the participant had responded to the rightmost stimulus. A calibration procedure was administered before each block of trials to maintain accuracy of recordings.

Two experiments were conducted. Experiment 1 varied the duration of stimulus encoding by varying the luminance (Dim = 5.2 cd/m^2 ; Bright = 46.2 cd/m^2) of the target letters across blocks. Experiment 2 varied the duration of response selection by using two sets of stimuli to create two mapping conditions. One set consisted of four alphabetic characters T, D, Z, and Q mapped in this arbitrary order onto keys V, B, N, and M, and assigned to the four digits of the right hand; the other set consisted of the digits 1, 2, 3, and 4 mapped in this natural order to the same four keys and fingers.

Figure 2 shows the pattern of observable events in Experiment 1, along with the key dependent measures. The stimuli are listed in the order in which they were responded from top (leftmost) to bottom (rightmost). Horizontal bars reflect the time from fixation to response for each stimulus (S1-S5). RT1 refers to the response time to the first stimulus (S1). IRI (Inter-Response Interval) is the time between the overt manual responses for each pair of successive stimuli. EHS (Eye-Hand Span) is the total length of the bar, from fixation to response. Dwell time is the duration of the fixation on a stimulus, indicated by the shaded portion of

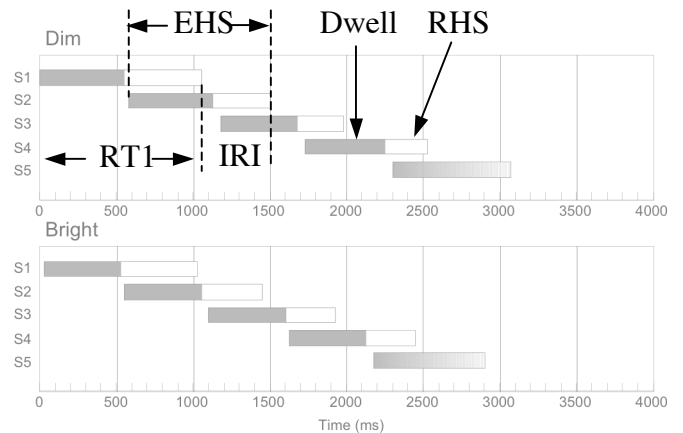


Figure 2: Mean times from fixation to response for each stimulus (S1-S5) in Experiment 1. Shaded portions are mean dwell times for each stimulus. EHS = Eye-Hand Span; RT1 = Response Time for S1; IRI = Inter-Response Interval; RHS = Release-Hand Span.

the bar. RHS (Release-Hand Span) is the time remaining after the eyes move until a response. EHS equals Dwell plus RHS. Dwell and RHS are meaningful only for S1-S4.

There are several key features of the data pattern (see Figures 4-6), which a computational model should capture.

- RT1 was slow, greater than 1 sec for this simple task.
- IRIs were fast and constant across stimuli (S2-S5).
- RT1 showed twice the luminance effect as IRI (30 ms vs 15 ms).
- RT1 showed four times the mapping effect as IRI (400 ms vs 100 ms)

- Dwell was relatively constant across stimuli, except for elevated Dwell on S1 in the hard mapping condition.
- Dwell across items showed same effect of luminance as RT1 and IRI.
- EHS decreased across stimuli, due to decrease in RHS.

Model

The results clearly indicate that some central process is rate limiting in this task. In earlier papers we had interpreted the increase in dwell time with difficult central mapping to indicate that eye movement initiation depended on partial completion of central operations. This was consistent with previous accounts of central difficulty leading to increased dwell times. The model we present here starts with the assumption that central processes are rate limiting without making the more detailed assumption they trigger the eye movement. Instead, we explore a model in which the timing of eye movement initiation is adjusted to achieve two related goals: completing perceptual processing on the next item as close as possible to completion of central operations on the current item, and maintaining a regular pace to minimize movement variance. The general case of these optimizations would be a complex tradeoff between the two criteria. However, because difficulty is blocked so that all items in a trial are of equal difficulty, it is possible to adopt a minimal variance eye movement regime that closely satisfies the “just-in-time” goal of fixating the next stimulus in time for its perceptual encoding to be complete when the central stage of the previous stimulus is complete. This timing would have the effect of minimizing delays in the initiation of central processing stages for successive stimuli.

Assumptions. Following Pashler we assume that mean IRI reflects the mean time for Response Selection (RS). We assume that the luminance manipulation affected Stimulus Encoding (SE), the mapping manipulation affected RS. Saccades and manual responses have two sequential components, an *Init* operator followed by a *Move* operator. However, we deviate somewhat from standard practice in assuming Init can proceed in parallel with central Response Selection operations. We assume periodic saccades that minimize movement time variance, whose period is determined by the “just-in-time” logic described above. To produce the regular period we introduce a variable called *Saccade Lag* that produces the constant delay required to achieve the just-in-time scheduling. Figure 3 shows the hypothesized sequence of events for successive stimuli (S2-S3).

Parameters. The model constructs a sequence of choice responses by scheduling component operations in accord with the assumptions listed above, illustrated in part in Figure 3. 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) operators were set to 50 ms, and Response Execution (RE) was set to 150 ms. Saccade movement time (E) was set to 30 ms. Response Selection (RS) was assigned the mean

IRI time for a given experimental condition. This directly reflects the rate-limiting assumption mentioned earlier. For the Dim condition of Experiment 1 SE was augmented by the average magnitude of the luminance effect.

The model does not have an explicit account of the RT1 elevation that would allow us to compute it from first

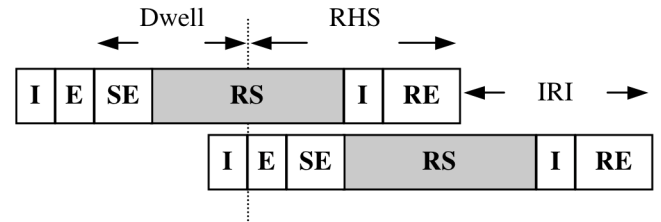


Figure 3: Overlapping task model . Processing of each stimulus consists of Stimulus Encoding (SE), Response Selection (RS), Response Execution (RE), preceded by an Eye Movement (E). E and RE are preceded by Init operators (I). Dwell, RHS, and IRI indicated (not to scale).

principles. Nor, is there data in the literature that would suggest a computation. 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. We estimate Prep by summing the mean values of the hypothesized components of RT1 then subtracting this from the observed RT1 as indicated here:

$$\text{Prep} = \text{RT1} - (\text{SE} + \text{RS} + \text{I} + \text{RE})$$

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 calculate *Dwell*. We have direct eye movement recordings that measure eye dwell times directly. However, for the model we estimate Dwell times by computing the total fixation duration for S1-S4, then dividing it by the number of fixations (4). Note there is no reliable fixation duration for the last stimulus, S5. Ideally, this would involve summing all the operators on the critical path. Because the model is stochastic we cannot be certain which elements will emerge on the critical path every run. For simplicity 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 assumed to be constant across items, dwell times could be estimated by dividing the total fixation time by four. By calculating total dwell time this way we make dwell time proportional on RS consistent with a just-in-time strategy.

$$\text{Dwell} = [\text{SE} + \text{Prep} + \text{RS} * 4 - (\text{SE} + \text{E})] / 4$$

Simulation. The model was implemented and Monte Carlo simulations run in the statistical package R. Each model parameter was assigned a standard deviation used by the model to draw times from a Gaussian distribution with the indicated mean and standard deviation. Comparison of model predictions with obtained data for Experiments 1 and 2 are shown in Figures 4-6. Model means are based on Monte Carlo simulations of 1000 trials.

The qualitative fits of the model are quite good, accurately reflecting the main trends in both the RT/IRI data and the eye movement data. The quantitative fit of the model is also quite good in most cases. Figure 4 shows that

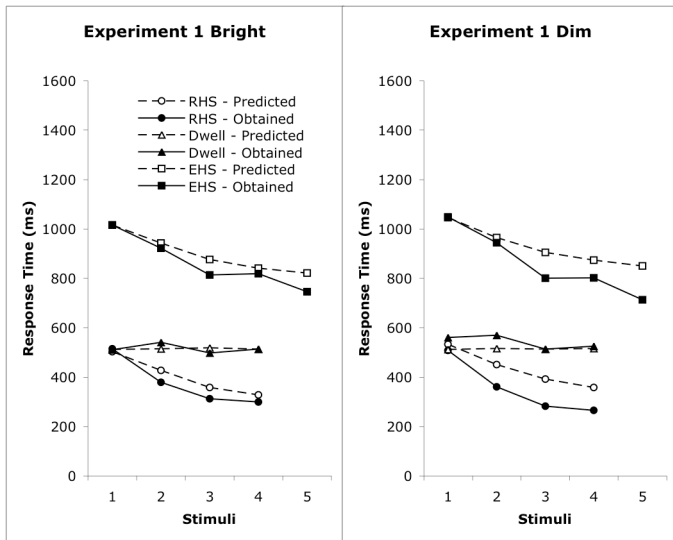


Figure 5: Predicted (open symbols, dashed lines) vs obtained (filled symbols, solid lines) eye movements in Experiment 1

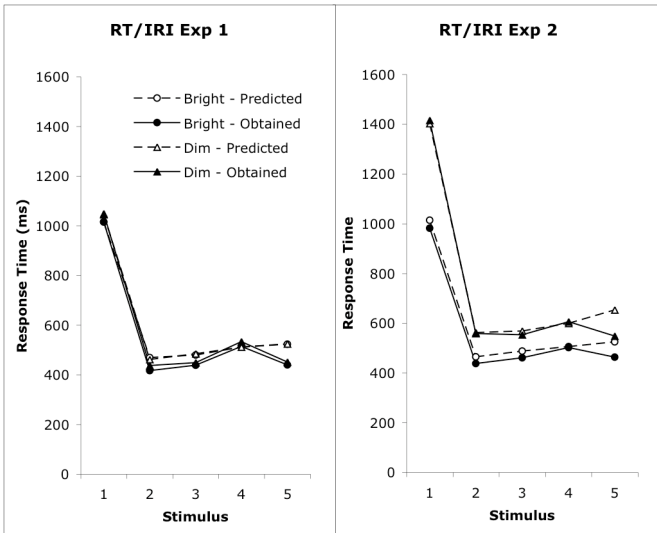


Figure 4: Predicted (open symbols, dashed lines) vs obtained (filled symbols, solid lines) for RT1 & IRI from Experiments 1 & 2.

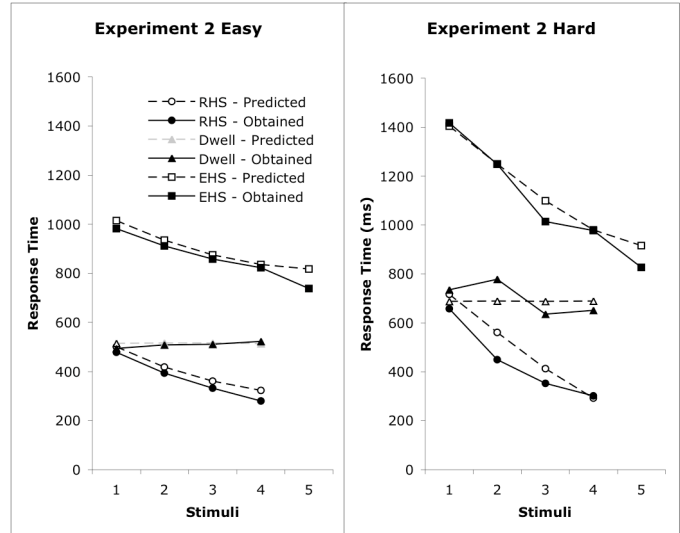


Figure 6: Predicted (open symbols, dashed lines) versus obtained (filled symbols, solid lines) eye movements in Experiment 2.

the model even predicts well the gradual increase in IRI from S2-S4. The only noteworthy deviation is the fit to the Release Hand Span in the Dim condition of Experiment 1. To assess how well the model follows the pattern of the data we ran correlations between predicted and obtained values for RT/IRI, EHS, and separately for RHS. To assess overall precision, Root Mean Squared Error was computed. For these analyses, mean times for all conditions of all experiments were pooled for each dependent measure. For RT/IRI, $R = 0.99$, $RMSE = 43$ ms; for EHS, $R = 0.977$, $RMSE = 59$ ms; for RHS, $R = .94$, $RMSE = 60$ ms.

General Discussion

We have described a simple quantitative model that accounts well both qualitatively and quantitatively for patterns of eye movement and manual responses in a task involving a sequence of choice responses. It shows that a rhythmic execution strategy can account for rate-limiting central operations without explicit dependencies on those operations. The model assumed regular movements whose period was tied to the desire not to be too late or too early in moving the eye to an upcoming stimulus. The success of these strategic assumptions suggests that one is not justified in attributing a direct link between saccade initiation and internal cognitive operation based only on evidence showing increases in eye dwell times with increases in central processing difficulty.

It is noteworthy that the simple assumptions of the model accurately reproduced the major trends in the data, and minor trends like the increase in IRI. Model assumptions dictated how to construct sequences and required only two parameters, RT1 and IRI from the data (aside from the luminance effect).

Is it possible to account for the data by assuming a direct link between internal processing and saccade initiation? We cannot rule out such an account with the evidence here. However, note that the success of the model in accounting for the decreasing RHS is due in large part to a decoupling of eye movement initiation from stimulus processing. Models that fail to decouple those tend to predict constant RHS.

The model in its present state is limited. More sophisticated optimization would be required for a general model of strategic scheduling. Also, the model has no mechanism for generating a regular, “just-in-time” schedule. We have not constructed a true computational model of just-in-time scheduling. For our purpose here it was sufficient to show what the effects of such a schedule would be.

Conclusions

We have demonstrated patterns of eye movements in routine multiple fixation tasks can be well fit by a model that assumes subjects are trying to minimize the variance of the movements. The model captures the rate-limiting effect of central processing stages by assuming that saccades are strategically timed to perceptual information just when central resources become available. Thus, it is not necessary to posit a direct dependence of saccades on central processing to account for observed correlations. Emergent properties in the scheduling of resources derived from strategic properties of behavior can produce very similar results. Indeed, a few assumptions were sufficient to reproduce the major effects.

Acknowledgements

This research was supported in part by funding from the Airspace Operations Systems (AOS) Project of NASA's Airspace Systems Program, and in part from the Johns Hopkins University Applied Physics Laboratory IR&D funding.

References

- Epelboim, J., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*, 3401-3422.
- Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight reading. *Proceeding of the Royal Society of London Series B*, *266*, 2435-2440.
- Inhoff, A. W., & Wang, J. (1992). Encoding of text, manual movement planning, and eye-hand coordination during copytyping. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 437-448.
- Inhoff, A. W. (1997). Eye movements and eye-hand coordination during typing. *Current Directions in Psychological Science*, *6*, 153-157.
- Johansson, R. S., Westling, G., Backstrom, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, *21*, 6917-6932.
- 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.
- Pashler, H. (1994). Overlapping task 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.
- Rayner, K., & Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology*, *33A*, 351-373.
- Rayner, K., & Pollatsek, A. (1997). Eye movements, the eye-hand span, and the perceptual span during sight-reading music. *Current Directions in Psychological Science*, *6*, 49-53.
- Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *American Journal of Psychology*, *93*, 207-234.
- Salthouse, T. A. (1985). Anticipatory processing in transcription typing. *Journal of Applied Psychology*, *70*, 264-271.
- Salthouse, T. A. (1986). Perceptual, cognitive, and motor aspects of transcription typing. *Psychological Bulletin*, *99*, 303-319.
- Sanders, A. F., & Van Duren, L. L. (1998). Stimulus control of visual fixation duration in a single saccade paradigm. *Acta Psychologica*, *99*, 163-176.
- Terao, Y., Andersson, N. E. M., Flanagan, J. R., & Johansson, R. S. (2002). Engagement of gaze in capturing targets for future sequential manual actions. *Journal of Neurophysiology*, *88*, 1716-1725.
- Truitt, F. E., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The perceptual span and the eye-hand span in sight reading music. *Visual Cognition*, *4*, 143-161.
- Vickers, J. N. (1992). Gaze control in putting. *Perception*, *21*, 117-132.
- Vera, A. H., John, B. E., Remington, R.W., Matessa, M., Freed, M. A. (2005). Automating human-performance modeling at the millisecond level. *Journal of 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.