Chapter 19

# SWIFT Explorations

Reinhold Kliegl and Ralf Engbert

SWIFT is a computational model of eye guidance in reading. It assumes (1) spatially distributed lexical processing, (2) a separation of saccade timing from saccade target selection, and (3) autonomous and parallel generation of saccades with inhibition by foveal targets. The model accounts for fixation probabilities as well as various measures of inspection time in their relation to lexical processing difficulty. We illustrate the dynamics associated with saccade generation and inhibition by foveal targets. In addition, we generate predictions for an experiment involving gaze-contingent display change.

## Introduction

Is it sufficient to assume that reading involves sequential shifts of attention from one word to the next or are several words within the perceptual span processed in parallel? There is still a very productive controversy surrounding this issue summarized recently by Starr and Rayner (2001). These authors concluded that

> One potential solution would be to abandon the serial framework of attention models of eye-movement control and replace it with a parallel mechanism. . . . Words would thus be processed in parallel, although the processing of information would be most accurate at the center of the attentional distribution. . . . However, such a model seems rather complicated and would be difficult to implement in a computational model. Thus a challenge for proponents of a parallel mechanism of attention during reading is to delineate the parameters of such a framework (Starr & Rayner, 2001, p. 162).

We proposed a model that fits this description (Engbert, Longtin & Kliegl, 2002). The model is based on three principles: spatially distributed lexical processing, a partial

separation of saccade timing from saccade target selection, and autonomous Saccade generation With Inhibition by Foveal Targets. From the last principle we also derived an acronym for the model (i.e., SWIFT). In the following we present a synopsis of the model components presenting in greater detail our assumptions about foveal inhibition and the dynamics of saccade generation and saccade cancellation. In addition, we evaluate the model with respect to predictions for an experiment with gaze-contingent display changes. With these explorations of the SWIFT model, we aim at a greater transparency of its core principles and demonstrate the utility of such a computational model for accounts of extant data as well as the prediction of novel aspects of eye guidance in reading.

## The SWIFT Model

### *Lexical Processing*

Figure 19.1 provides an overview of the model components. The "Lexical processing" box encapsulates how words are processed relative to the current eye position. We call this processing "foveal lexical activity." In contrast to other computational models such as E-Z Reader (Reichle, Pollatsek, Fisher & Rayner, 1998; Reichle, Rayner & Pollatsek, 1999) or the model proposed by Engbert and Kliegl (2001) we assume that the perceptual span encompasses four words, namely the word currently fixated as well as the one to the left and the two words to the right. The (normalized) processing rate depends on fixation location: It is largest for the fixated word [parameter estimate: $\lambda(0) = 0.798$] and considerably smaller for the left and right neighbor [$\lambda(1) = \lambda(-1) = 0.077$] and even smaller for the second word to the right [$\lambda(2) = 0.048$]. As the sum of $\lambda$s was fixed at 1.0, we used only two degrees of freedom for the three parameter estimates. Details about parameter estimation will be presented in the "Model evaluation" section; the model was estimated with a total of 11 free parameters (see Table 19.1).

A word is processed as soon as it is within the perceptual span leading to a dynamic change in what we call lexical activity associated with this word. We assume that there is a maximum of lexical activity with each word depending on its frequency and its predictability from the prior sentence context, using the specification proposed by Reichle *et al.* (1998), that is $L_n = (1-p_n)(\alpha - \beta \log f_n)$, where $p_n$ represents the predictability of *word_n* and $f_n$ the printed frequency of *word_n*, $\alpha$ and $\beta$ are model parameters which were estimated as 148.5 and 5.71, respectively. Thus, the maximum of lexical activity will range between 148.5 (=$\alpha$) for an unpredictable, very-low-frequency word and 0 for a perfectly predictable word. Over time lexical activity increases from zero for unprocessed words to the maximum lexical activity and then decreases back to zero. The processing time associated with the increase in lexical activity is called lexical preprocessing; the processing time required for the return to the zero-baseline is called lexical completion — again in rough analogy to a two-level processing introduced by Reichle *et al.* (1998). Obviously, preprocessing and completion could be conceived as two independent processes. Rather than estimating a single
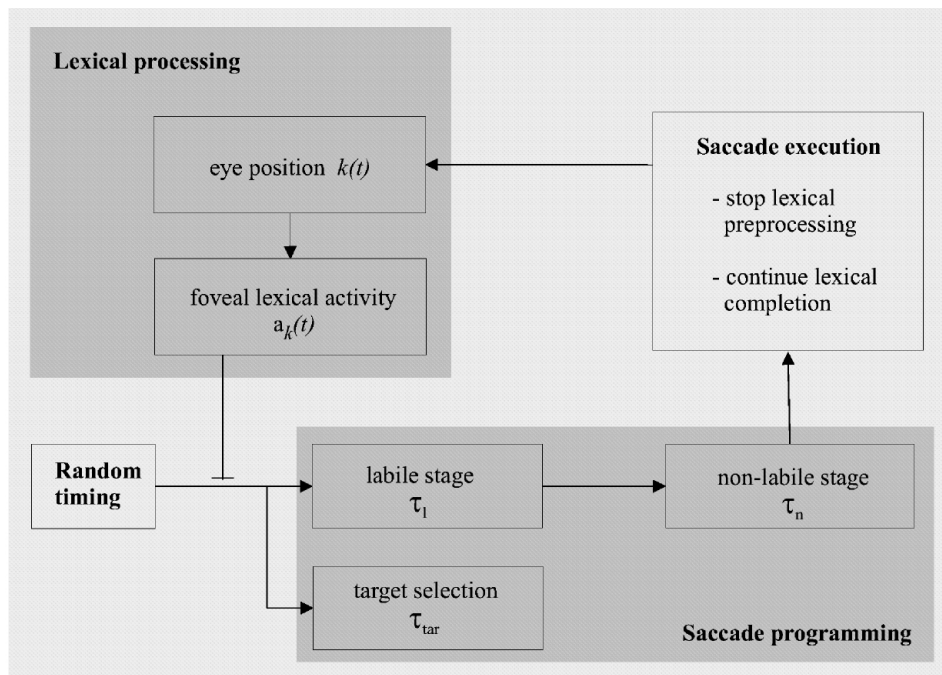
Figure 19.1: Schematic diagram of SWIFT. The main subsystems are saccade programming and lexical processing. These two subsystems are coupled via a foveally-inhibited random timing system and a saccade execution system which moves the eyes during saccades (from Engbert *et al.*, 2002).

processing rate and preprocessing factor, we could estimate two separate rates for these processes with the first process "running" from zero to maximum lexical activity and the second process from maximum to zero lexical activity. The remaining link between these two processes is that they use the same maximum of lexical activity for a given word as a stopping value for the process. As an analogy of this conceptualization it may be useful to interpret the maximum lexical activity of a word as an indicator of its "attractiveness" to the mind. In this sense, unpredictable or low-frequency words are more interesting to the mind and it takes longer for this interest to subside than in the case of predictable or high-frequency words. Note that the current formula based on printed frequency and predictability serves only as a descriptive interface to the general difficulty of the words in the perceptual span. Ideally, such lexical activity values should be provided by a theoretical account of sentence processing. Predictability and printed frequency are but convenient proxies of theoretically unspec-ified contributions of syntactic, semantic, and pragmatic sources of variance for the processing difficulty associated with a given word in its sentence context.

Processing rates for left and right parafoveal word could be constrained to the same estimate without loss of fit (i.e., 0.077). However, it is well-known that there is a

processing asymmetry in the direction of reading. Indeed, in our initial qualitative specification of parameters we assumed a higher processing rate for the right than the left parafoveal word. In the quantitative estimation it turned out that this asymmetry was already captured in the parameter that allows preprocessing and completion rates to differ by constant factor. This parameter ($f = 62.5$) suggests that lexical preprocessing is completed much faster than lexical completion. It turned out that lexical preprocessing usually occurs in the right parafovea and lexical completion in the right parafovea or fovea. If it occurred at all, processing in the left parafovea was restricted to lexical completion. Thus, lexical processing in the model reflects the well-known asymmetry of processing. Moreover, it allows us to interpret word-position dependent processing rates (i.e., the $\lambda$s) as indicators of retinal acuity which would be symmetric relative to the fixation location.

## *Saccade Programming*

The second major model component specifies "Saccade programming" which comprises (a) saccade initiation, (b) inhibition by foveal targets, (c) labile and non-labile stages of timing (When?), (d) target selection (Where?), and (e) saccade execution (see Figure 19.1). In the following we describe each of these aspects.

**Saccade initiation**   In the SWIFT model saccade initiation occurs autonomously (see "Random timing" box in Figure 19.1), an idea already implemented in our earlier sequential-attention shift model (Engbert & Kliegl, 2001) after a random interval generated by a timer ($t_S$). Assuming a gamma distribution, $t_S$ was estimated with a mean of 187.1 ms and a relative standard deviation of 0.239 of the mean (i.e., 44.7 ms). (A single value for relative standard deviations was estimated for saccade initition times and other saccade-related timing distributions, see next paragraph). The assumption of autonomous saccade initiation is fundamentally different from E-Z Reader and its predecessor models (e.g., Morrison, 1984) where saccade initiation is strictly coupled to some aspect of lexical processing.

**Inhibition by foveal targets**   The intuition guiding the assumption of autonomous saccade initiation is that during reading we initiate saccade programs according to some preferred mean rate. However, we want to allow for some influence of lexical processes. Specifically, we assume that high lexical activity delays the saccade initiation. In other words, if there is a chance that comprehension may lag behind the autonomous generation of saccades and, consequently, comprehension and eye position threaten to desynchronize, then saccade initiation can be postponed. In Figure 19.1, this intervention is represented by the inhibitory link from the foveal lexical activity to the link between "Random timing" and "Saccade programming". We assume an additive contribution of foveal lexical activity [$ha_K(t)$] to the random interval generated by the timer ($t_S$). Thus, a new saccade program is started after the time interval $t' = t_S + ha_K(t)$. For a word of maximum difficulty ($\alpha = 148.5$, $p = 0$), no parafoveal preprocessing and purely foveal processing the maximum "inhibition time"

$ha_K(t)$ amounts to 181 ms for the current set of parameter estimates [$\lambda(0) = 0.798$; $h = 50.3$] (see Appendix A). Thus, the next saccade will be initiated at the latest 181 ms after the value drawn from the distribution of saccade initiation times. (The calculation in the appendix also shows that the maximum inhibition time does not depend on the precise value of $h$ as long as $h$ is sufficiently large; for an infinite value of $h$, the maximum inhibition time would be 186 ms. Thus, in principle this free parameter might not be necessary in future model versions.) In general, however, the delay will be much smaller because of lower lexical activity, parafoveal preprocessing, asynchrony of maximum lexical activity, and determination of inhibition time during foveal processing. Indeed, the amount of fixation time due to foveal inhibition amounted to less than 15% for low-frequency words in the simulation. Nevertheless, foveal inhibition was necessary to explain the dependency of first fixation duration on word frequency. We will discuss a similar proposal by Yang and McConkie (2001; McConkie & Yang, this volume) in the final section of this chapter.

**Labile and nonlabile stages of saccade program ("when?")**  Once a saccade is initiated, that is once a saccade program is started, we assume two stages, a labile and a non-labile stage. Assuming gamma distributions, labile and non-labile times were estimated to last on average 128.6 (SD = 30.7 ms) and 41.6 ms (SD = 9.9 ms), respectively. The model assumes that a saccade can be cancelled and saccade targets can be modified during the labile phase. A distinction between labile and non-labile stage is implemented in E-Z Reader as well (Reichle *et al.*, 1998). However, E-Z Reader does not allow direct target modification. Rather a saccade program is always initiated to the next word at the completion of the first stage of lexical processing (i.e., the familiarity check); target modification can occur indirectly through the cancelation of a saccade during the labile stage. In the SWIFT model, a new saccade can be initiated during the preparation of an older one. Such interactions of a new saccade program with an older one will be the topic of the next section.

**Target selection ("where?")**  The distinction between "when?" and "where?" is motivated by neurophysiological results (e.g., Carpenter, 2000; Findlay & Walker, 1999; see Reilly & Radach, this volume, for another implementation of this distinction in computational model of reading). Saccadic target selection (where to move next) is specified as largely independent of saccade timing (when to move). Target selection was estimated to occur after 87% of the labile phase, that is on average after 112.1 ms in the current implementation. This dependence could be relaxed substantially in future versions of the model. This target of the next saccade, one of the words within the current perceptual window, is stochastically determined according to the values of current lexical activities. Thus, the word with the largest current lexical activity is the most likely target; perfectly predictable words with a lexical activity of zero will be skipped. The differences in processing rates associated with fixation position and the differences in lexical preprocessing and completion rates generate a "bow wave" of lexical activity pulling the eye in direction of reading across the sentence. The reason is that the currently fixated word is processed at the highest rate (see above) and is likely to be already in the stage of lexical completion with a continuous decrease

in lexical activity at the time of target selection. Words to the right of the fixated word are processed much slower than the fixated word and are likely to be still in the stage of lexical preprocessing with a continuous increase in lexical activity (or perhaps even "trail" the fixated word in the stage of lexical completion). Consequently, at the time of target selection the lexical activity is likely to be lower for the currently fixated words than its right neighbor. Therefore, the latter is more likely to be selected as the next saccade target.

**Consequences of target selection mechanism**  The stochastics of the selection process as well as differences between maximum lexical activities will lead to refixations of the current word and refixations to the previous word. In addition, there is the possibility that the eye moves on before a word (i.e., the word to the left of the fixated one) is completely processed with the residual lexical activity remaining at a constant level as long as the word is outside the perceptual span. These words will remain potential targets for saccades with a probability derived from their residual lexical activities. As the residual lexical activity is typically rather low, the selection is likely to occur late when the eye has moved towards the end of the sentence. We call such regressions to words to the left of the current perceptual span "long regressions". Their empirical plausibility (i.e., the accuracy of regressions to previous words in a text) has been demonstrated by Kennedy and Murray (1987).

**Saccade execution**  The end of the non-labile stage of saccade programming triggers "Saccade execution" (see Figure 19.1). At this time, lexical preprocessing is suspended but lexical completion continues. The rationale for this distinction is that lexical preprocessing requires perceptual input which is suppressed or strongly attenuated during saccade execution. Lexical completion should not be affected by saccadic suppression. In the simulation, times for saccade executions were fixed with a mean of 25 ms and a standard deviation of 8.3 ms assuming a gamma distribution of latencies. The saccade execution shifts the position of the eye which in turn leads to a change in the foveal and non-foveal lexical activities.

## Dynamics of Saccade Generation

Most of the model assumptions are fairly straightforward but there are some intricacies associated with the initiation of saccade programs. In particular, due to autonomous saccade generation it is possible, that a new saccade program is started while the previous saccade progam is still "under construction." The effects of such parallel saccade programs on the resulting fixation durations depend on the stage of completion of the first saccade program (completed, labile stage, non-labile stage, or execution stage). As described in the last section, lexical processing has an effect on the initiation of the next saccade program solely via the inhibition by foveal lexical activity. Lexical activity can only delay the initiation of a saccade program, it cannot start or cancel a saccade program. Thus, if the currently fixated word is very difficult

and if there is a high lexical activity when the initiation time for the next saccade is sampled, then the gamma-distributed latency will be extended accordingly. Therefore, for the following illustration the effects of lexical activity of foveal inhibition can be subsumed under the initialization time. Figure 19.2 illustrates a hypothetical sequence of saccade programs assuming for now a deterministic scheme with labile stages lasting 125 ms, non-labile stages 50 ms, and execution times 25 ms. We also assume that targets are selected 100 ms after saccade initiation. Now we turn to a description of the various cases of overlap between saccade programs.

### Non-overlapping Saccades

When a saccade program (SP) is started, the time for the initiation of the next saccade is determined as well. The initial saccade program SP0 is started with an initialization of $t = 0$ (see Figure 19.2).At the beginning of SP0 the initialization time for saccade program SP1 is sampled according to a gamma distribution , the dashed horizontal line indicates that SP1 is to start at $t = 250$ ms. According to the above specifications, for the first fixation duration F0 we simply compute the sum of labile and non-labile stages which amounts to 175 ms. The target for SP0 was specified after 100 ms during the labile stage. The first fixation (i.e., the nonlabile saccade program stage) is terminated with the execution of the saccade lasting 25 ms. Thus, the next fixation F1 starts at 200 ms. Note that there are still 50 ms of processing (250 ms — 200 ms) before SP1 is initiated. Thus, the inital saccade program does not overlap with the first one; SP1 is started during F1 at absolute time $t = 250$ ms. At this point in time a latency for SP2 is chosen with a value of 225 ms. Consequently, there is again enough time for SP1 to complete without interference from SP2: Its labile phase will end at absolute time 375 ms; the non-labile phase at absolute time 425 ms (amounting to a F1 duration of 225 ms = 50 ms + 175 ms); target selection occurred at absolute time 350 ms. The execution of the second saccade lasts from 425 ms to 450 ms. This illustration shows that there is no simple link between fixation duration and the time intervals at which saccade programs are initiated because the difference between the time for the initiation of the next saccade program and the time required for carrying out the current saccade program is allocated to the next fixation duration.

### Saccade Initiation During Labile Stage of Current Saccade Program

At absolute time 475 ms, saccade program SP2 is started and the next saccade program SP3 is to be initiated with a latency of 75 ms. The initiation of SP3 falls in the labile stage of SP2. In this case, SP2 is simply canceled and replaced with SP3 (i.e., the duration in the labile stage is reset to zero; any target would be de-selected as well). Obviously such a cancelation extends the duration of current fixation by the amount of time already spent in the labile stage of canceled saccade program (see F2). Note also that cancelation of a saccade program implies that only one saccade program is active.
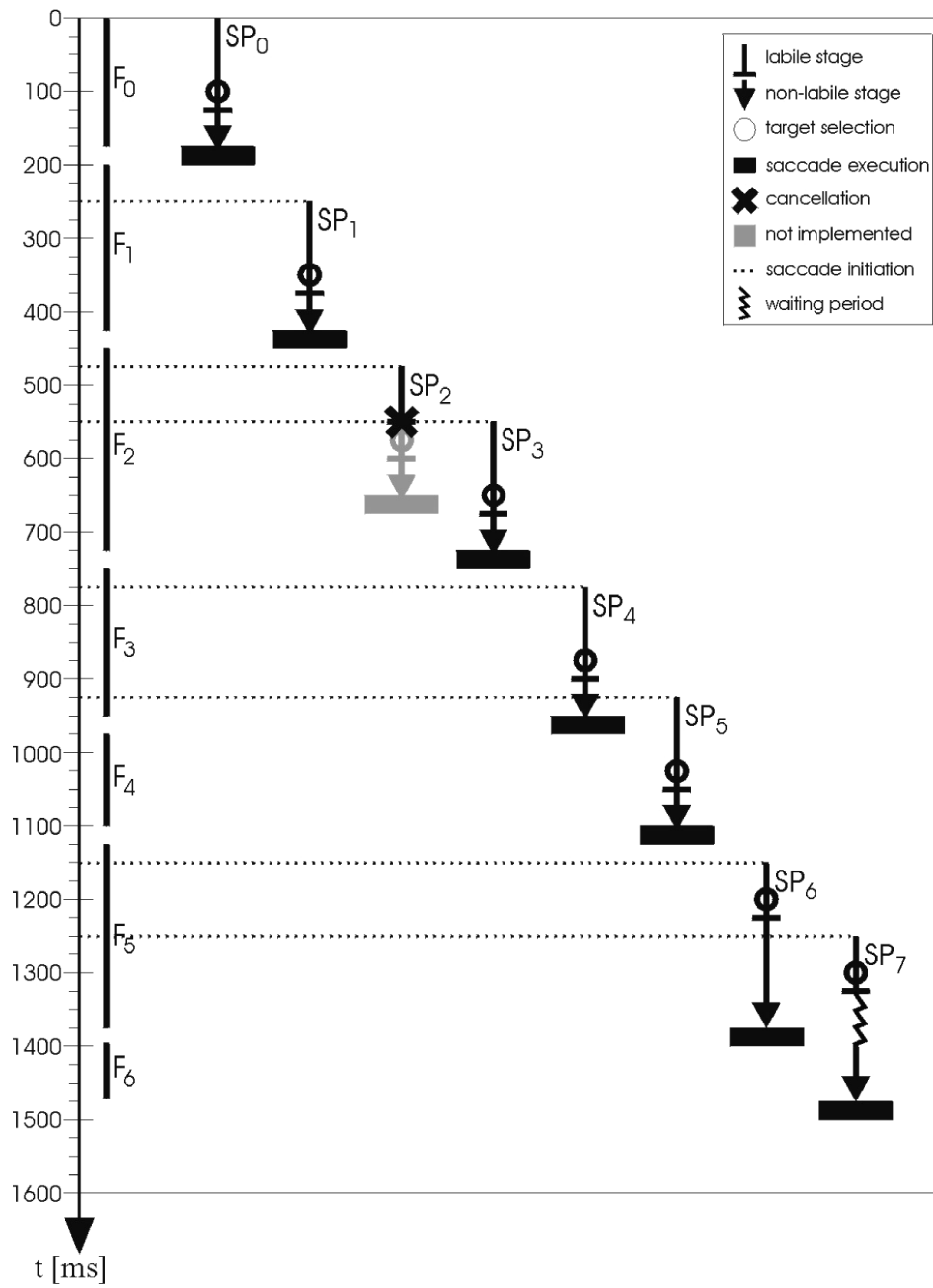
Figure 19.2: Dynamics of saccade generation in SWIFT. When a saccade program (SP) is started, the time for the initiation of the next saccade is determined as well. This will lead to non-overlapping (SP0 and SP1) and various cases of overlapping saccade

### *Saccade Initiation During Non-labile Stage of Current Saccade Program*

The case of a saccade initiation during the non-labile stage of a saccade program is illustrated for SP4 and SP5. As implied by the name of the stage, there is no consequence for a saccade program in the non-labile stage. The program will run its course and land at the target word selected earlier. There is also no consequence for the initiation of the new saccade program. Thus, in this case there are two saccade programs running in parallel without interference. The same logic applies if the initiation of a new saccade program occurs during the execution of the current saccade.

### *Two Saccade Programs in Non-labile or Execution Stages*

Durations of labile and non-labile stages were specified as stochastic variables. Consequently, the second saccade program might enter the execution stage while the first saccade is being executed. Obviously the eyes cannot go two targets at the same time. In this case the later saccade program simply has to wait for the current one to be completed. This case is illustrated with SP6 and SP7. Note that this case can lead to very short fixations because the eye is in rest only for the duration of the nonlabile stage of SP7.

## Model Evaluation

We used the corpus of Schilling, Rayner and Chumbley (1998) to fit the model and estimate parameters. The corpus comprises 48 sentences with a total of 536 words. Frequency and predictability values are available for each word. Moreover, Schilling *et al.* reported statistics of gaze, first, and single fixation duration as well as the probabilities of single, double, and zero fixations (i.e., skipping). Trials with regressions were discarded from the analysis. The same corpus was used by Reichle *et al.* (1998) and Engbert and Kliegl (2001). For each sentence we obtained 100 simulations. The model was fitted with 11 free parameters. These parameters, the best fitting values, and the associated standard deviations are listed in Table 19.1.

The panels of Figure 19.3 represent three simulations of the same sentence using the parameter estimates of Table 19.1. The sentence read was: "Mark told Jane that he would meet her after baseball practice". Time runs from left to right. The solid black line indicates the position of the eye at each point in time. Vertical gray lines indicated saccades; the vertical dashed line marks the end of reading. The thin lines

---

programs. Initiation of the second program during the labile stage of the first one leads to cancellation of the first program (SP2 and SP3). Initiation during the non-labile stage does not interfere (SP4 and SP5) but the two programs can not be in the non-labile stage at the same time (SP6 and SP7). In this case the second program has to wait.
Fixation durations are shown next to time axis (F0 to F6).

Table 19.1: Model parameter estimates.

|  | Parameter | Symbol | Value | SD |
|---|---|---|---|---|
| Lexical parameters | Difficulty, intercept | $\alpha$ | 148.5 | 3.6 |
|  | Difficulty, slope | $\beta$ | 5.71 | 0.29 |
| Processing rate | Foveal | $\lambda(0)$ | 0.798 | 0.017 |
|  | Parafoveal | $\lambda(1,-1)$ | 0.077 | 0.017 |
|  | Parafoveal | $\lambda(2)$ | 0.048 | 0.017 |
|  | Preprocessing | $f$ | 62.5 | 5.8 |
| Saccade parameters | Random timing (ms) | $t_S$ | 187.1 | 2.6 |
|  | Labile stage (ms) | $\tau_l$ | 128.6 | 3.2 |
|  | Nonlabile stage (ms) | $\tau_n$ | 41.6 | 4.7 |
|  | SD (relative to mean) | $\rho$ | 0.239 | 0.021 |
|  | Target selection (% labile stage) | $\tau_{tar}$ | 0.872 | 0.056 |
|  | Inhibition factor | $h$ | 50.3 | 14.1 |

*Notes:* Standard deviations (SD) were based on five runs of parameter estimation by genetic algorithm (see Engbert *et al.*, 2002, Appendix A). Processing rates were estimated with the constraint: $\lambda(0) + 2\lambda(1,-1) + \lambda(2) = 1$, yielding a total of 11 free parameters for the model.

above the solid eye-position line show the time course of lexical activity for each word. Finally, the solid black segments on the time axes indicate delays of saccade program initiation due to foveal inhibition. Obviously, aside from differences in total reading time between simulations, the three examples illustrate the high degree of complexity that follows from the theoretical principles and the inherent dynamics of the model and the large variance of eye movement traces that result from them. The same pinciples and dynamics, however, also strongly constrain the type of paths the eye can take through the sentence (Engbert, Longtin & Kliegl, in press). Indeed, such statistics may prove very useful for comparing computational models with each other and with human data.

The traces also illustrate links between lexical activity and various types of eye movements. The top panel represents a trace consisting solely of forward moves. At the beginning the eye is on the first word and the model starts processing the first three words as indicated in the increase and subsequent decrease of lexical activity visualized for each word. The rate of change is highest for the fixated word. After about 500 ms the eye moves to the second word. The decision point to select this word as the next target occured around 100 ms into the first fixation. Note that at this time lexical activity of the fixated word was probably already smaller than that of the second and the third word. The height of lexical activity indicates the probability with which a word is selected as the next target. Therefore, chances were high for the selection of the second or third word with the second word "winning" in this case. Note that in the second and third trace for this sentence (middle and bottom panels of Figure 19.3) the third word was selected as the target with the second word being skipped in the

process. Words are also skipped if they are already processed completely in the parafovea of earlier fixations. Examples of such skips are the words "that", "her" and "practice" in the first trace (top panel).

The second and third trace illustrate two different regressions to the skipped word "told". In the second trace (middle panel of Figure 19.3, "told" has the highest lexical activity among the four words in the perceptual span (i.e., "told Janet that he") during the second fixation and was selected as the taregt for the third saccade, leading to a regression back from "Janet". The same mechanism will also yield refixations of a word, indicated with a circle on the trace. For example, in the second trace (middle panel) "told" was fixated twice, that is it was selected again as target during the first fixation. Obviously, such refixations are more likely for low-frequency, unpredictable words.

We already encountered regressive movements of the eye to an earlier word within the perceptual span. A word will also be the target of a long regressive movement if its processing was not completed while it was in the perceptual span. For example, the word "told" in the bottom panel of Figure 19.3 was left in such an unfinished state because first "Janet" and then "he" were selected as targets causing "told" to fall to the left of the perceptual span. In the model the residual lexical activity of this word will stay at its last value in the perceptual span until it is selected again as a target. Any word with residual lexical activtiy will compete for target selection irrespective of whether or not it still is in the perceptual span. Typically such a residual lexical activity is low due to earlier processing and therefore the chances of being selected as a target are small as long as there are unprocessed words to the right. However, the predictability of words increases with serial word position and, consequently, maximum lexical activity will decrease across a sentence. Thus, as the eye approaches the end of the sentence chances increase again that words with residual lexical activity will be selected. In the third trace (bottom panel of Figure 19.3) "told" was selected as a target when the eye fixated "would." The following words had been processed completely; therefore reading resumes at the next word with lexical activity outside the perceptual span which is "after" in this case. (Incidentally, "after" was first processed during the fixation on "meet". Due to the subsequent regression, processing ceased while lexical activity was fairly high.)

There is good evidence that such long regressions are typically very precise, suggesting that reading sets up a spatial representation of word locations (Kennedy, 2000). Experimental evidence for long regressions due to residual lexical activity would constitute very strong support for the model because, traditionally, such long regressions have been attributed only to high-level sentence parsing problems, such as revisions of an interpretation in a garden-path sentence. Predictions relating to this distinction remain to be tested in experiments but at this point they illustrate the innovative potential of a dynamic model.

From such simulations we can compile summary statistics of various measures of inspection time as well as various measures of fixation probability and compare them with experimental results. These comparisons are displayed in the two panels of Figure 19.4 as a function of logarithmic word frequency. In general, the model reproduces the qualitative patterns very well. Most notable are the increase in skipping probability and the drop in single-fixation probability for high-frequency words. Also the decrease
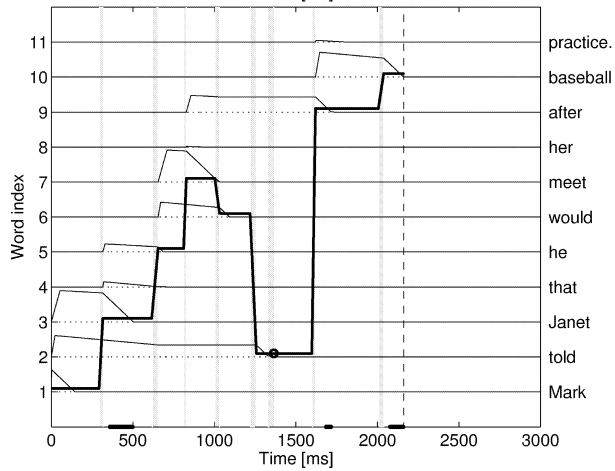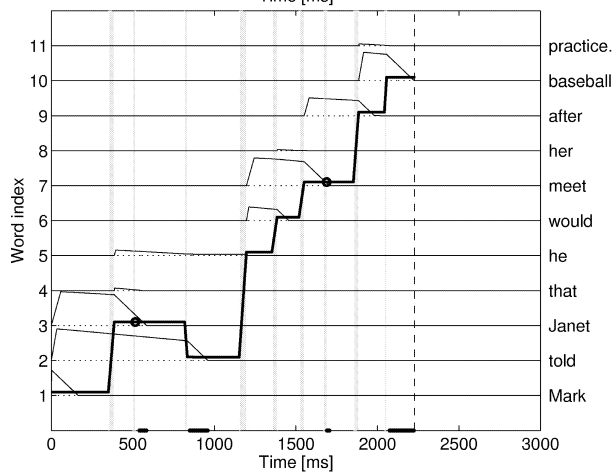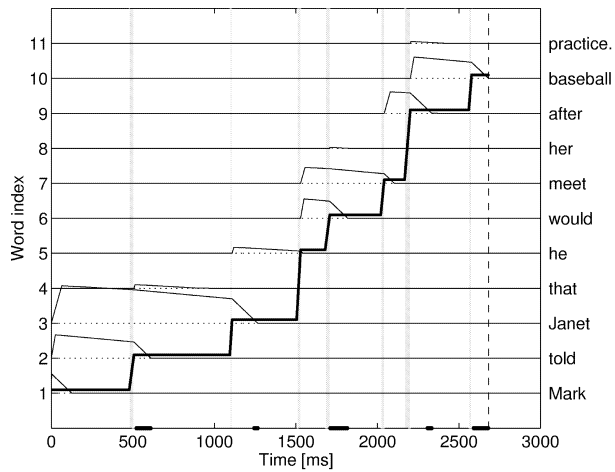
in inspection time measures as a function of logarithmic word frequency is reproduced very well. In the current implementation the height of lexical activity at the time of target selection is used as weight for determining the next saccadic goal. Explorations of the model showed that data could also be fitted if height determined the next target (i.e., a "winner-take-all" rule). Thus, the stochastic selection of target selection was not very critical. However, substantially lower parafoveal than foveal processing rates were critical for obtaining adequate model fits.

## Simulation of Eye-contingent Display Changes

In this section, we demonstrate that SWIFT can be used to predict several measures of a typical reading experiment. Properties of the perceptual span in reading are fundamental to the model. To derive the model we used psychologically plausible assumptions about the perceptual window. To test whether these assumptions lead to realistic behavioral patterns, we investigate predictions that can be obtained from numerical simulations with respect to an experimental manipulation of the display during reading (Binder, Pollatsek & Rayner, 1999; we restrict our analysis to the preview condition). In an eye-contingent display change experiment, a target word in a sentence was changed during preview. As soon as the eyes entered the target region, the preview word was replaced by the target word. The display change was performed during the saccade to the target region and the subjects tested were unaware of this manipulation.

Some results of the Binder *et al.* (1999) experiment are summarized in Table 19.2. If the preview was not changed, that is if an identical base word appeared at the target location, a skipping probability of 0.30 resulted. If a different word was used during preview, that is if the word was replaced with a different word during the approaching saccade, skipping probability decreased to 0.165. Also there was an increase of first fixation durations and regression probability for the changed word. These results highlight the importance of parafoveal lexical processing during reading.

In our numerical simulations, we used a sentence of the corpus by Schilling, Rayner and Chumble (1998) to demonstrate the effect of reduced preview in our model. Obviously, we only aimed for a qualitative reproduction of key results because of differences in the sentence material and experimental set-up. To this end, we reset the lexical activity of the target word to zero as soon as a saccade to the target region occurs (see Figure 19.5). As a result, we find a comparable reduction in skipping probability from 0.47 to 0.15. Furthermore, like Binder *et al.* (1999) we observe an

---

Figure 19.3: Trajectories for the same sentence from three simulation runs of SWIFT with parameter estimates of Table 19.1. Lexical activities (thin lines) are plotted over time together with the eye position (bold line). The execution of saccades is indicated by the shaded vertical regions. The beginning of a refixation is indicated by a circle. Foveal inhibition (delay of initiation of saccade program) is marked by the bold segments on the time axis. Sentence and data on word frequencies and predictability were taken from the Schilling *et al.* (1998) experiment.

increase in the first fixation durations in the changed compared to the identical preview condition (Table 19.2). Finally, to demonstrate the extraction of information to the left of the fixation point, we analyzed the probability of regressions to the target word. Our theoretical expectation was that the reset of lexical activity will lead to an increase of the number of regression. For the target word used here, the probability of regressing was 0.042 without preview manipulation. The increase of the regression probability to
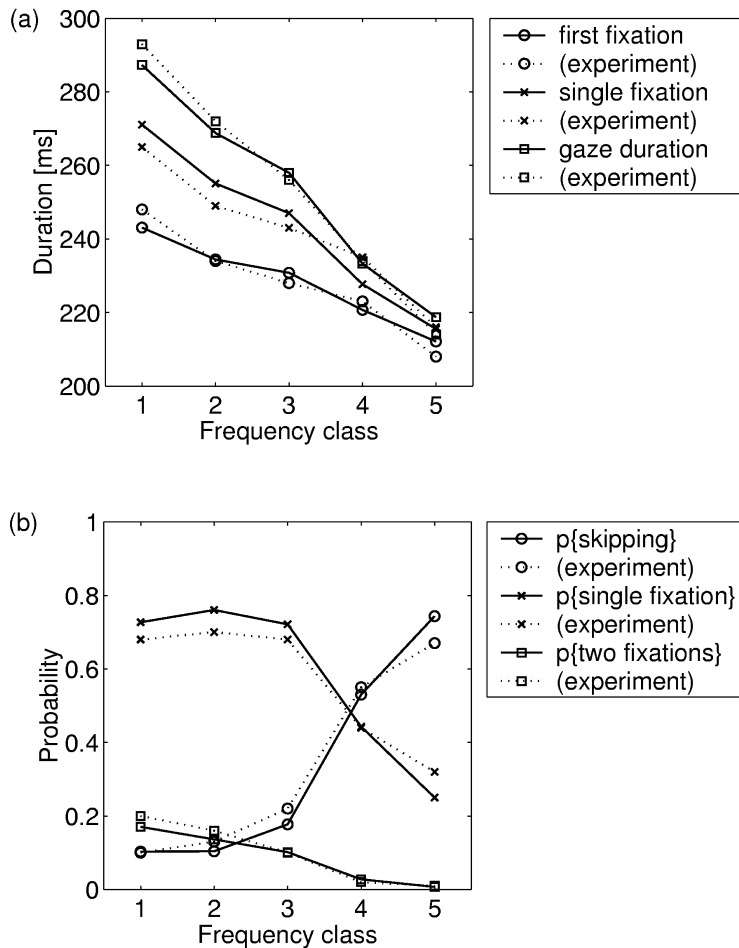


Figure 19.4:  Statistical evaluation of SWIFT performance; experimental data are from Schilling *et al.* (1998). (a) First fixation duration, gaze duration, and single fixation duration as a function of word frequency class (averaged over 1000 statistical realizations from SWIFT simulations, i.e. 1000 simulations of the model over the same corpus of sentences but with different pseudo-random numbers). (b) Probabilities for word skipping, performing a single fixation, and making two fixations (computed from the sampe runs as in (a)) as a function of word frequency class. (from Engbert *et al.*, 2002).

0.07 as a consequence of the reset of lexical activity is much smaller than the increase observed in the experiment by Binder *et al.* (1999). This discrepancy, however, may be explained by the fact that, in the model, regressions are completely caused by incomplete lexical processing. Different sources of regressions, as reviewed for example by Rayner (1998), are beyond the scope of the current version.

Table 19.2: Gaze-contingent display changes: A comparison of SWIFT and Binder *et al.* (1999).

| Reading measure / Preview | SWIFT | | Binder *et al.* (1999) | |
|---|---|---|---|---|
| | **Identical** | **Different** | **Identical** | **Different** |
| Probability of skipping preview | 0.470 | 0.150 | 0.300 | 0.165 |
| First fixation on target (ms) | 187 | 201 | 228 | 246 |
| Probability of regressing to target | 0.042 | 0.070 | 0.080 | 0.220 |

*Note*: Values for different preview of Binder *et al.* (1999) represent mean of related and unrelated preview changes.
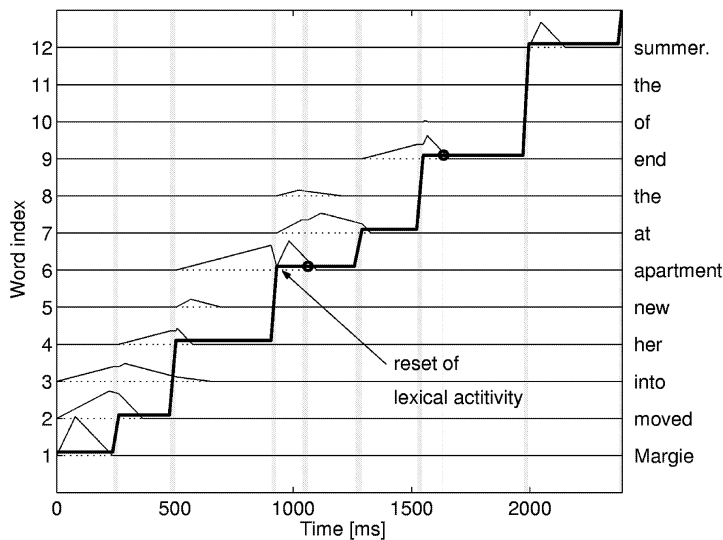


Figure 19.5: Numerical simulation of the preview experiment by Binder *et al.* (1999). A preview display was changed during the first saccade entering the target region (*word_6*). In our simulations, the lexical activity of the target word was reset to zero, when the first saccade to the target region was performed. In the example trajectory shown here, this occurs at time $t = 930$ ms. As a result, the fixation duration of $word_6$ increases. Also the word is skipped less often and is more likely to attract a regressive saccade.

In summary, the model can be used to predict typical measures used for the analysis of eye movement experiments with gaze-contingent display changes. The influence of preview manipulations and the mechanism of extraction of information to the left of the fixated word are qualitatively in good agreement with experimental results. These results underline the model's psychological plausibility and how properties of the perceptual window can be thought to influence the dynamics of eye movements.

## Comparisons and Perspectives

### *Comparison with E-Z Reader*

Computational model of eye movements in reading hold much promise for providing unifying accounts of rich and diverse sets of experimental results. Recent years have witnessed the emergence of a few attempts in this direction (Engbert & Kliegl, 2001; Reichle *et al.*, 1998, 1999; Reilly & Radach, this volume). The SWIFT model was developed as an alternative to the E-Z Reader (Reichle *et al.*, 1998), which at the time, in our opinion, was the most advanced computational model in the domain of attentional and ocular control during reading. Consequently, central assumptions guiding the design of E-Z Reader were adopted for SWIFT, such as two stages of lexical access (including the formula for combining word frequency and predictability) and the distinction between labile and nonlabile stages of saccade programs terminating with saccade execution. We also followed Reichle *et al.*'s (1998) lead to simulate reading at the level of words, not characters, which probably underestimates the role of perceptual factors given the correlation of word length and word frequency, but greatly reduced the complexity of model construction. Finally, as Reichle *et al.* (1998), we used the corpus of data from Schilling *et al.* (1998) to evaluate the model. In preliminary model comparisons, goodness of fit statistics were typically quite comparable. However, given the large architectural differences between the models, quantitative model fits may not be very meaningful. Rather we want to point to a few qualitative differences between the models which we considered to be critical in the design of the SWIFT model.

There are three pieces of empirical evidence that may prove problematic for the notion of sequential attention shifts (SAS) in reading and, consequently, also for computational models such as E-Z Reader (Reichle *et al.*, 1998) or our own earlier model of this type (Engbert & Kliegl, 2001) subscribing to this assumption (for references see Engbert *et al.*, 2002, p. 622f.): First, there is some evidence that processing of a fixated word is influenced by the difficulty of the next word. Second, there is evidence that information is picked up to the left of the fixated word. Third, there is little empirical support for longer fixation durations prior to skipped words — an implication of sequential attention shift models. The experimental evidence for the first problem is controversial. The second problem was handled at the cost of an increase in model complexity in more recent versions of the model (see Pollatsek, Reichle & Rayner, this volume).

However, in our opinion, there is no easy solution to the third problem: In SAS models, word skipping requires that a saccade targeting the next word is cancelled and re-programmed to the following word. Saccade cancellation necessarily increases the fixation duration prior to the saccade. Indeed both E-Z Reader and our own previous SAS model exhibited a strong effect of this sort (173 ms in E-Z Reader 5; 75 ms in Engbert & Kliegl, 2001). The three empirical studies cited by Reichle *et al.* (1998) in support of longer fixation durations prior to skipped reported the following effects: 3 to 7 ms (Hogaboam, 1982, Tables 18.1, 18.3), 21 ms (Pollatsek, Rayner & Balota, 1986, p. 126), and 38 ms (Reichle *et al.*, 1998, p. 147). Given that the effect can not be secured empirically in large data sets (e.g., McConkie, Kerr & Dyre, 1994; Radach & Heller, 2000), the SAS assumption should be reconsidered and SAS models minimally require architectural revision. One solution might be to postpone target selection, similar to the partial separation of saccade initiation and target selection in the SWIFT model. Accordingly, for SWIFT simulations we observed a range of 10 to 21-ms increases of fixation durations prior to skipped words which appears to be in agreement with the empirical data. Incidentally, in our opinion, it is to the credit of E-Z Reader and other SAS models that they can be seriously challenged with experimental evidence.

A general problem of E-Z Reader relates to the order-of-processing methodology. Even slight modifications of the model typically increase the number of states to be considered and seriously limit the complexity of the dynamics that can be covered in the model. Given that SWIFT was designed with at least some of these problems in mind, it is not surprising that it is not affected by these problems. And we should point out that SWIFT does not (yet) account for the full scope of the Schilling *et al.* (1998) corpus, most notably we did not fit distributions of fixation durations as function of logarithmic word frequency.

### *Comparison with Competition–Inhibition Theory*

Yang and McConkie (2001, also McConkie & Young, this volume) proposed a Competition–Inhibition theory which is very specific with respect to the timing of saccades, that is the "when?" component of saccade programs. Although the theory is not yet implemented as a computational model, there are two central assumptions of the SWIFT model that are conceptually very much in agreement with the Competition–Inhibition theory: autonomous timing of saccades and inhibition of foveal targets due to lexical factors. First, autonomous timing of saccades represents active search for new information as well as predictions and expectations about where relevant information is to be found. This is very different from models assuming that the completion of lexical or cognitive processes triggers the initiation of new saccades; the familiarity check in the E-Z Reader model is such an example (Reichle *et al.,* 1998). The proposal that the eye is acting on expectations that are corrected by sensory feedback if necessary (i.e., the "motor prediction" perspective of Wolpert & Flanagan, 2000), is also in line with more general theories of the relation of eye movements and complex actions such as driving or cricket batting (Land & Furneaux, 1997; Land & McLeod, 2000).

It seems plausible that eye movements in reading will eventually be understood as a special case of a more general theory of eye movements and action.

The second notion common to SWIFT and Competition–Inhibition theory is that lexical factors affect saccadic latency (and indirectly fixation duration) via an inhibition by foveal targets. The assumption is that the process of autonomous saccade generation can be delayed by lexical difficulty. In the current implementation of the SWIFT model, the sampled time interval for initiating the next saccade program increases with the lexical difficulty of the fixated word. Obviously, this prolongs the fixation on the current word — up to a maximum of an additional 181 ms. Yang and McConkie's (2001) inhibition process is very sophisticated. They distinguish between three types of saccades (early saccades initiated about 100 — 125 ms after fixation; normal saccades initiated after 175 — 200 ms; and late saccades initiated after 225 ms). Their most relevant result for the current discussion is that display changes of text content (word to nonword) affected only the initiation of late saccades. If there are qualitatively different saccade types and if lexical factors influence only the late ones, then SWIFT will have to be changed to accommodate this high degree of specificity, for example, by linking lexical processing with the labile phase of saccade generation conditional on some minimum amount of processing of the foveal word.

### The Next Version of SWIFT

We opted for some glaring simplifications in the evaluation and specification of the SWIFT model. They were motivated by keeping model complexity down and model comparability high. There are at least three necessary extensions. The first two extensions are to increase the scope of the data base that should be accounted for without major revisions of the current model. Specifically, as mentioned earlier, unlike E-Z Reader the SWIFT model does not account for distributions of fixation durations as function of logarithmic word frequency. Moreover, although SWIFT uses a single mechanism to generate all types of within-line eye movements (i.e., word-to-word, skippings, refixations, and regressions), we have not modeled regression probabilities, perhaps also based on a distinction between regressions within the perceptual span and long regressions. The reason for this omission was that the Schilling *et al.* (1998) corpus had removed sentences with regressions prior to the analysis. Fitting regression probabilities requires a new corpus of eye movement data.

The more serious extension of the SWIFT model concerns a switch from word-based to letter-based processing. The model should reproduce typical landing-position probabilities as a function of word length and saccadic launch distance (for a review of the relevant literature we refer to Radach and Heller, 2000). Moreover, the relation between fixation positions and fixation durations has been of lasting concern in eye movement research. Specifically, there is evidence for two independent effects of fixation position on fixation duration: (1) Fixation durations are longer for fixations in the center of words, (irrespective of whether they are single or first fixation durations (Vitu, McConkie, Kerr & O'Regan, 2001) and (2) fixation durations increase with the launch distance of the last saccade (Radach & Heller, 2000; Vitu *et al.*, 2001). Finally,

in this context it may be useful (or even necessary) to allow for a dynamic adjustment of the letter-based perceptual span in response to lexical difficulty. Obviously, such an extension requires a data base that includes information about landing positions of fixations at the letter level in addition to the statistics that were reported in Figure 19.4. Such an increase in data base is necessary to constrain the model parameter space. If successful, such an extension would provide a very desirable modeling framework for the joint consideration of oculomotor, perceptual and low-level cognitive control issues.

## Acknowledgments

## References

Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1162–1172.

Carpenter, R. H. S. (2000). The neural control of looking. *Current Biology, 10,* R291-R293.

Engbert, R., Longtin, A., & Kliegl, R. (in press). Complexity of eye movements in reading. *International Journal of Bifurcation and Chaos.*

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42,* 621–636.

Engbert, R., & Kliegl, R. (2001). Mathematical models of eye movements in reading: A possible role for autonomous saccades. *Biological Cybernetics, 85,* 77–87.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences, 22,* 661–721.

Hogaboam, T. W. (1983). Reading patterns in eye movements. In: K. Rayner (ed.), *Eye Movements in Reading.* New York: Academic Press.

Inhoff, A. W., Radach, R., Starr, M., & Greenberg, S. (2000). Attention and saccade programming. In: A. Kennedy, R. Radach, D. Heller and J. Pynte (eds), *Reading as a Perceptual Process.* Amsterdam: Elsevier.

Kennedy, A. (2000). Attention allocation in reading: Sequential or parallel? In: A. Kennedy, R. Radach, D. Heller and J. Pynte (eds), *Reading as a Perceptual Process.* Amsterdam: Elsevier.

Kennedy, A., & Murray, W. S. (1987). Spatial coding and reading: Some comments on Monk (1985). *Quarterly Journal of Experimental Psychology, 39A,* 649–718.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society London, B352*, 1231–1239.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3,* 1340–1345.

McConkie, G. W., Kerr, P. W., & Dyre, B. P. (1994). What are "normal" eye movents during reading: Toward a mathematical description. In: J. Ygge and G. Lennestrand (eds), *Eye Movements in Reading.* Oxford: Elsevier.

Morrison, R. E. (1984). Manipulations of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 667–682.

Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception & Psychophysics, 40,* 123–130.

Radach, R., & Heller, D. (2000). Spatial and temporal aspects of eye movement control. In: A. Kennedy, R. Radach, D. Heller and J. Pynte (eds), *Reading as a Perceptual Process* (pp. 165–191). Oxford: Elsevier.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157.

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research, 39,* 4403–4411.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26,* 1270–1281.

Starr, M.S., & Rayner, K. (2001). Eye movements during reading: some current controversies. *Trends in Cognitive Science, 5,* 156–163.

Vitu, F., McConkie, G.W., Kerr, P., & O'Regan, J.K. (2001). Fixation location effects on fixation durations during reading: an inverted optimal viewing position effect. *Vision Research, 41,* 3513–3533.

Yang, S.-N., & McConkie, G. W. (2001). Eye movements during reading: A theory of saccade initiation times. *Vision Research, 41,* 3567–3585.

Wolpert, D. M., & Flanagan, J. R. (2000). Motor prediction. *Current Biology, 11,* R729-R732.

## Appendix A:

## Analytical Calculation of the Theoretical Maximum of Inhibition Time

In the SWIFT model, the time between two subsequent decisions to start a saccade program is given by a random time interval $t_s$ and an additive contribution of foveal inhibition $h \cdot a_k(t)$. Let us denote the time of end of the last saccade (or, equivalently, the start of the current fixation) by $t'$. The next command to start a saccade program is generated at time

$$t' = t_s + h \cdot a_k(t) \tag{1}$$

The theoretical maximum of the contribution of the inhibition process, i.e. max $\{h \cdot a_k(t)\}$ can be calculated. For simplicity, we can choose $t = t'$. The inhibition mechanism reaches its maximum under three conditions:

1. The random component is — by chance — zero: $t_s = 0$.
2. There has been no preprocessing of the foveal word: $a_k(0) = 0$, i.e. the lexical activity of the foveal word is zero at the start of the fixation. Since lexical preprocessing time is short compared to lexical completion, however, we assume that $a_k(0) = L_k$ to further simplify our calculations.
3. The foveal word has a very low frequency, which implies that its lexical difficulty is $L_k = \alpha$.

Since the foveal word is lexically processed with rate $\lambda(0)$, its lexical activity decreases linearly according to the relation

$$a_k(t') = \alpha - \lambda(0) \cdot t' \tag{2}$$

Putting together this equation with Equation 1 with $t_s = 0$, i.e. $t' = h \cdot a_k(t')$, we obtain the relation $t'/h = \alpha - \lambda(0) \cdot t'$, which can be rearranged to the final equation for the maximum of the inhibition time

$$t' = \frac{\alpha}{\lambda(0) + \dfrac{1}{h}} \tag{3}$$

We interpret this result by discussing the two limiting cases of a vanishing or infinite inhibition parameter $h$:

$$h \rightarrow 0: \quad t' = 0$$

$$h \rightarrow +\infty: \quad t' = \frac{\alpha}{\lambda(0)} \tag{4}$$

In the first case, without inhibition, the maximum of $t'$ is obviously zero. In the second, and more interesting case, even an arbitrary large inhibition parameter leads to a finite contribution of $t' = \alpha/\lambda(0) = 186$ ms (for the estimated value $h = 50.3$, we obtain a slightly lower value $t' = 181$ ms). The small increase of 5 ms between the cases $h = 50.3$ and $h \rightarrow +\infty$ also explains the large standard deviation estimated for the inhibition factor $h$ (see Table 1).