(OVP). This relationship is expressed by a U-shaped curve. As a similar pattern was found in text reading (e.g., McConkie et al.1989; Vitu et al. 1990), it was popularly assumed that a decision to refixate was made because of errors in the execution of saccades which do not land on the intended saccade target. To integrate these empirical observations, the E-Z Reader model admits that a proportion of refixations is planned because of mislocated initial positions.

We would like to address two questions to the authors. First, whether their model assumes two populations of refixations. Second, whether the presupposed factors that affect the decision to refixate also play a role in the computation of the metrics of refixation saccades. Indeed, even if this model addresses the question of the refixation probability, nothing is said about refixation saccade metrics – for example, what is the target for the refixation saccade?

An experiment was conducted in our lab (Doré-Mazars et al. 2003) to examine these questions further during reading of isolated long words. High- and low-frequency words of 8, 10, and 12 letters were displayed in parafoveal vision. With this procedure, the launch site (eccentricity) and the parafoveal preview were held constant. Critical aspects of early work about the refixation decision are replicated here: both length and frequency effects, and also the classical U-shaped curve describing the relation between the refixation probability and the initial landing position on the word. For each initial landing position, we found an effect of the length and the frequency of the word, the amplitude of the first being more important than the second one.

More interestingly, we provide arguments for the view that refixations do not result from saccadic error but are preplanned and sometimes canceled. We observed that the distribution of landing positions in refixation cases is clearly leftward-shifted relative to single fixation cases. In addition, the examination of the refixation saccade amplitude demonstrates that the saccade is planned on the basis of the word length with no effect of the initial landing position on the word. Indeed, the slope of the linear regression between first and second fixation position close to 1 indicates that the refixation saccade is computed as a fixed motor vector applied irrespective of the initial landing position on the word. We replicate here previous findings indicating that the refixation saccade is preplanned in parafovea relative to the word length integrated at this time (Vergilino & Beauvillain 2000). The absence of a target for refixation saccades stands against refixations as corrective saccades. In such a framework, we interpret the difference in initial landing position on the word between single- and refixation cases found in our experiment as the consequence and not as the cause of the planning of refixation saccades. Of course, because of the inherent variability of the text-reading situation (e.g., in launching sites), some refixations could be caused by mislocated landing positions, but their proportion and metrics remain to be assessed. Moreover, while refixation probability was affected by word frequency, no role of this factor in the computation of the refixation metrics was observed in our experiment. Indeed, we found a frequency effect neither on the mean refixation saccade amplitudes nor on the slope of the linear regression. This result is compatible with the notion that the lexical processing that progresses throughout the first fixation is likely to cancel a preplanned refixation saccade. However, since the frequency effect on refixation probability is around 10%, as usually observed in the literature, we assume that only a small proportion of refixations would be canceled by lexical processing. Word processing plays only the secondary role in refixating of long words.

In conclusion, one of the future challenges of the E-Z Reader model is to take into account not only the factors that influence the decision to make a refixation saccade, but also those that determine its metrics, to better explain refixations in reading.

The game of word skipping: Who are the competitors?

Ralf Engbert[a] and Reinhold Kliegl[b]

[a]*Department of Psychology and Center for Dynamics of Complex Systems,* [b]*Department of Psychology, University of Potsdam, D-14415 Potsdam, Germany.* **engbert@rz.uni-potsdam.de** **kliegl@rz.uni-potsdam.de** **www.agnld.uni-potsdam.de/~ralf/** **www.psych.uni-potsdam.de/people/kliegl/index-e.html**

Abstract: Computational models such as E-Z Reader and SWIFT are ideal theoretical tools to test quantitatively our current understanding of eye-movement control in reading. Here we present a mathematical analysis of word skipping in the E-Z Reader model by semianalytic methods, to highlight the differences in current modeling approaches. In E-Z Reader, the word identification system must outperform the oculomotor system to induce word skipping. In SWIFT, there is competition among words to be selected as a saccade target. We conclude that it is the question of competitors in the "game" of word skipping that must be solved in eye movement research.

In computational models based on the concept of sequential attention shifts (SAS), word skipping is a consequence of a competition between lexical processing and saccade programming (target article; cf. Engbert & Kliegl 2001; 2003; Reichle et al. 1998). This mechanism was proposed first by Morrison (1984). Such an explanation of word skipping is *qualitatively* different from the assumption underlying the SWIFT model (Engbert et al. 2002; 2004; Kliegl & Engbert 2003), that a field of lexical activities builds up during the eyes' random walk over the sentence. It is the relative strength of activity that determines the probability of selecting the next saccade target. The related theoretical framework of competition between targets for action is the dynamic field theory of movement preparation (Erlhagen & Schöner 2002). Consequently, the SWIFT model may be generalized as a model for eye-movement control in situations with many potential saccade targets such as visual search or general scene perception. To compare these differences between SAS models and SWIFT, we investigate the mechanism for word skipping using semianalytical techniques.

In E-Z Reader 7, currently the most advanced SAS model, a new saccade program is initiated at the end of stage 1 of the word identification system (Fig. 3 in the target article). Word skipping occurs if the saccade program is canceled by another saccade command during the labile stage. Such a cancellation will occur if the sum of the durations of L_2 (of the currently fixated word) and L_1 (of the skipped word) is smaller than the average duration of the labile saccade program M_1. To calculate the probability of skipping, we have to consider that saccade program stages are gamma-distributed[1] in E-Z Reader. As a consequence, the probability of skipping is given by an integral over the distribution $q^n_\tau(t)$ of durations of the labile saccade stage M_1,

$$p_{\text{E-Z Reader}} = \int_{L_1+\langle L_2\rangle}^{\infty} q^{\mathrm{n}}_{\tau}(t)\mathrm{d}t \qquad (1)$$

where the time constant τ is related to the mean of the labile saccade program by $\tau = M_1/9$. It is important to note that there are two oculomotor parameters, n and τ, in the probability. The integral in Equation 1 can be evaluated analytically. The probability for skipping a word, which needs an average processing time L_1 of the first stage of word identification, is given by

$$p_{\text{SAS}} = \left(\sum_{k=0}^{n} \frac{1}{k!}\left(\frac{L_1+\langle L_2\rangle}{\tau}\right)^k\right)\exp\left(-\frac{L_1+\langle L_2\rangle}{\tau}\right) \qquad (2)$$

Since stage L_1 refers to the skipped word, we have to estimate the average processing time during stage 1 by computing means over the five word-frequency classes for L_1. From low to high word frequency (classes 1 to 5) we computed the values 128.0 msec, 100.7

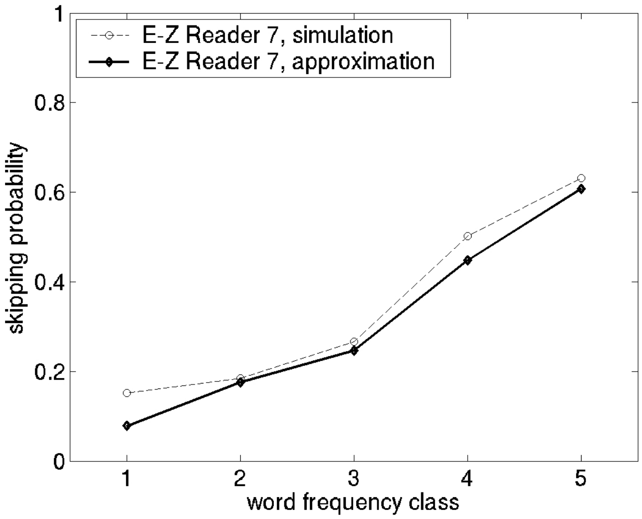

Figure 1 (Engbert & Kliegl). Skipping probability as a function of word frequency class.

msec, 90.8 msec, 60.7 msec, and 44.4 msec for L_1, using Equation 3 and corresponding parameter values given in Reichle et al. The average value of L_2 corresponds to an arbitrary word (the word left of the skipped word). Therefore, we used the ensemble average of L_2 over all words the corpus of sentences,[2] denoted by $\langle L_2 \rangle = 82.3$ msec. For a gamma distribution of order $n=8$ and a mean labile saccade duration $M_1=187$ msec, we obtained $\tau = 20.8$ msec. The resulting estimates for the skipping probability $p_{\text{E-Z Reader}}$ are in good agreement with simulated data from the target article (see Fig. 1).

In SWIFT, a field of lexical activities $a_n(t)$ evolves over time. The probability of target selection is given by the relative lexical activity. As a consequence, no additional assumptions must be made to produce forward saccades, refixations, and regressions. The probability of skipping word$_{n+1}$ is given by the probability to select word$_{n+2}$ as the next saccade target, which is computed by the fraction

$$p_{SWIFT} = \left. \frac{a_{n+2}(t)}{\sum_{k=1}^{n+2} a_k(t)} \right|_{t = \text{target selection}} \tag{3}$$

There is no oculomotor contribution to the skipping probability in Eq. (3) – an important difference to Equation (2) for E-Z Reader. Numerical estimates for p_{SWIFT} can be obtained by evaluating the set of lexical activities at the point in time where target selection occurs in SWIFT (for details see Engbert et al. 2002).

Diverging predictions can be derived from SAS and SWIFT models. In E-Z Reader, the probability of word skipping will depend on oculomotor parameters, because of the competition between saccade programming and word identification. In SWIFT, the competition between words for becoming selected as the next saccade target implies a structural stability of word skipping against oculomotor parameters. Therefore, dynamic models generate highly specific predictions, which might be most stimulating for future research: The current controversy on mechanisms of eye-movement control will still be resolvable by experimental results.

ACKNOWLEDGMENTS

We thank Erik D. Reichle for providing us with the simulated data for E-Z Reader 7. This work was supported by Deutsche Forschungsgemeinschaft (grants KL-955/3).

NOTES

1. The gamma distribution for saccade latencies can be written as $q_\tau^n = \frac{1}{\tau - n!}\left(\frac{t}{\tau}\right)^n \exp\left(-\frac{t}{\tau}\right)$, where τ is a time constant and n is the order of the distribution. Mean value and standard deviation are given by $\mu = (n+1)\tau$ and $\sigma = \sqrt{n+1}\tau$. For a relation of standard deviation to mean of one third (Reichle et al. 1998), we have to choose a gamma distribution of order $n = 8$.

2. This procedure may be interpreted as a *mean field approximation*, that is, using the average processing difficulty of the word left to the skipped word. To compute L_1 and $\langle L_2 \rangle$ according to Equation 3 in the target article, we used word frequencies, predictabilities, and the parameters β_1, β_2, and Δ.

Throwing the baby out with the bathwater: Problems in modeling aggregated eye-movement data

Gary Feng

Department of Psychology: Social and Health Sciences, Duke University, Durham, NC 27708. **garyfeng@duke.edu**
http://www.psych-shs.duke.edu/faculty/facultyfeng.html

Abstract: Parameters in E-Z Reader models are estimated on the basis of a simple data set consisting of 30 means. Because of heavy aggregation, the data have a severe problem of multicolinearity and are unable to adequately constrain parameter values. This could give the model more power than the empirical data warrant. Future models should exploit the richness of eye movement data and avoid excessive aggregation.

Because eye movement techniques produce an enormous amount of raw data, it is common practice to perform some sort of data reduction prior to modeling. However, there is a delicate balance between reducing computational complexity and preserving meaningful variance in the data. The E-Z Reader 7 model and its predecessors (Reichle et al. 1998; 1999) posit a comprehensive and elegant set of eye-movement control mechanisms, but the data set used to fit the models is too impoverished to adequately test these models.

The empirical data for E-Z Reader models (from Schilling et al. 1998) are averages of six eye movement variables (single fixation duration, first fixation duration, gaze duration, and the probability of skipping, making single fixations, and making two fixations) over five word frequency levels (Reichle et al. 1998, Table 1; see also Note 6 of the target article). Unfortunately, the structure of the empirical data is ill formed. The six variables are so highly correlated that the data space has far fewer than six independent dimensions, a problem known as *multicolinearity* in linear regression analysis.

Pair-wise correlation coefficients among the six variables range from 0.85 (between skipping rate and first fixation duration) to 0.998 (between first fixation duration and single fixation duration). Furthermore, all eye-movement measures are highly correlated with the logarithm of word frequency. A principal component analysis showed that the first component accounts for 94.6% of the total variance, the first two components account for 98.6%, and the first three components account for 99.999% of total variance. In short, with only 5% loss of information, the six eye-movement variables can be effectively reduced to a single variable, which in turn has an almost perfect linear relationship with log-transformed word frequency.

The consequences of multicolinearity in the dataset are profound. Free parameters (ranging from five in E-Z Reader 1 to at least seven in E-Z Reader 7 models) were effectively estimated on the basis of only five data points, creating a classic identification problem in parameter estimation, where some parameter values may be varied freely without affecting model fit. Moreover, flaws in the data threaten the internal validity of E-Z Reader as an empirical model. If we believe in the principle of parsimony, then the only model that will survive Occam's Razor would be something like "any eye-movement measure is a linear function of the log-transformed word frequency," which is both uninformative and wrong (see Kliegl et al. 1982).

https://doi.org/10.1017/S0140525X03270102 Published online by Cambridge University Press