A Model of Individual Differences in Gaze Control During Reading

Niels Landwehr[1] and **Sebastian Arzt**[1] and **Tobias Scheffer**[1] and **Reinhold Kliegl**[2]

[1] Department of Computer Science, Universität Potsdam
August-Bebel-Straße 89, 14482 Potsdam, Germany
{landwehr, sarzt, scheffer}@cs.uni-potsdam.de
[2] Department of Psychology, Universität Potsdam
Karl-Liebknecht-Straße 24/25, 14476 Potsdam OT/Golm
kliegl@uni-potsdam.de

Abstract

We develop a statistical model of saccadic eye movements during reading of isolated sentences. The model is focused on representing individual differences between readers and supports the inference of the most likely reader for a novel set of eye movement patterns. We empirically study the model for biometric reader identification using eye-tracking data collected from 20 individuals and observe that the model distinguishes between 20 readers with an accuracy of up to 98%.

1 Introduction

During skilled reading, the eyes of a reader do not move smoothly over a text. Instead, reading proceeds by alternating between brief fixations on individual words and short ballistic movements called *saccades* that move the point of fixation to a new location. Evidence in psychological research indicates that patterns of fixations and saccades are driven partly by low-level visual cues (*e.g.*, word length), and partly by linguistic and cognitive processing of the text (Kliegl et al., 2006; Rayner, 1998).

Eye-movement patterns are frequently studied in cognitive psychology as they provide a rich and detailed record of the visual, oculomotor, and linguistic processes involved in reading. Computational models of eye-movement control developed in psychology, such as SWIFT (Engbert et al., 2005; Schad and Engbert, 2012) or E-Z Reader (Reichle et al., 1998; Reichle et al., 2012), simulate the generation of eye movements based on physiological and psychological constraints related to attention, visual perception, and the oculomotor system. Recently, the problem of modeling eye movements has also been approached from a machine-learning perspective. Matties and Søgaard (2013) and Hara et al. (2012) study conditional random field models to predict which words in a text are fixated by a reader. Nilsson and Nivre (2009) use a transition-based log-linear model to predict a sequence of fixations for a text.

A central observation made by these studies, as well as by earlier psychological work (Erdmann and Dodge, 1898; Huey, 1908), is that eye movement patterns vary significantly between individuals. As one example of the strength of individual differences in reading eye movements, we cite Dixon (1951) who compared the reading processes of university professors and graduate students of physics, education, and history on reading material in their own and the two other fields. He did not find strong effects of his experimental variables (i.e., field of research, expertise in research) but "if there is one thing that this study has shown, it is that individual differences in reading skill existed among the subjects of all departments. Fast and slow readers were found in every department, and the overlapping of distributions from passage to passage was enormous" (p. 173). Even though it is possible to predict across a large base of readers with some accuracy whether specific words will be fixated (Matties and Søgaard, 2013), a strong variation between readers in attributes such as the fraction of skipped words and total number of saccades has been observed (Hara et al., 2012).

Some recent work has studied eye movement patterns as a biometric feature. Most studies are based on an artificial visual stimulus, such as a moving (Kasprowski and Ober, 2004; Komogortsev et al., 2010; Rigas et al., 2012b; Zhang and Juhola, 2012) or fixed (Bednarik et al., 2005) dot on a computer screen, or a specific image stimulus (Rigas et al., 2012a). In the most common use

case of biometric user identification, a decision on whether access should be granted has to be made after performing some test that requires the user's attention and therefore cannot take a long time. By contrast, our work is motivated by a less intrusive scenario in which the user is monitored continuously during access to, for instance, a device or document. When the accumulated evidence supports the conclusion that the user is not authorized, access can be terminated or additional credentials requested. In this use case, identification has to be based on saccadic eye movements that occur while a user is reading an arbitrary text—as opposed to movements that occur in response to a fixed, controlled visual stimulus. Holland and Komogortsev (2012) study reader recognition based on a set of aggregate features derived from eye movements, irrespective of the text being read; their work will serve as reference in our empirical study.

The paper is organized as follows. Section 2 details the problem setting. Section 3 introduces the statistical model and discusses parameter estimation and inference. Section 5 presents empirical results, Section 6 concludes.

2 Problem Setting and Notation

Let $\mathcal{R}$ denote a set of readers, and $\mathcal{X} = \{\mathbf{X}_1, ..., \mathbf{X}_n\}$ a set of texts. Each $r \in \mathcal{R}$ generates a set of eye movement patterns $\mathcal{S}^{(r)} = \{\mathbf{S}_1^{(r)}, \ldots, \mathbf{S}_n^{(r)}\}$ on the set of texts $\mathcal{X}$, by

$$\mathbf{S}_i^{(r)} \sim p(\mathbf{S}|\mathbf{X}_i, r)$$

where $p(\mathbf{S}|\mathbf{X}, r)$ is a reader-specific distribution over eye movement patterns given a text $\mathbf{X}$. A pattern is a sequence $\mathbf{S} = ((s_1, d_1), \ldots, (s_T, d_T))$ of fixations, consisting of a fixation position s_t (position in text that was fixated) and duration $d_t \in \mathbb{R}$ (length of fixation in milliseconds). In our experiments, individual sentences are presented in a single line on a screen, thus we only model a horizontal gaze position $s_t \in \mathbb{R}$.

At test time, we observe novel eye movement patterns $\bar{\mathcal{S}} = \{\bar{\mathbf{S}}_1, \ldots, \bar{\mathbf{S}}_m\}$ on a novel set of texts $\bar{\mathcal{X}} = \{\bar{\mathbf{X}}_1, ..., \bar{\mathbf{X}}_m\}$ generated by an unknown reader $r \in \mathcal{R}$. The goal is to infer

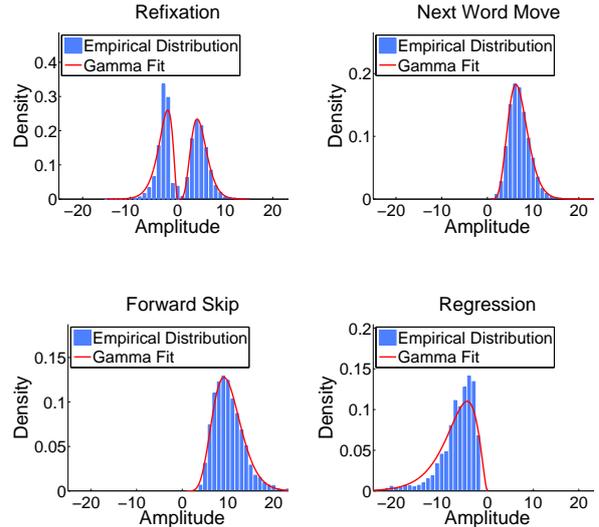$$r_* = \arg\max_{r \in \mathcal{R}} p(r|\bar{\mathcal{S}}, \bar{\mathcal{X}}). \quad (1)$$

Figure 1: Empirical distributions over amplitudes and Gamma fits for different saccade types.

3 Statistical Model of Eye Movements

We solve Problem 1 by estimating reader-specific models $p(\mathbf{S}|\mathbf{X}; \boldsymbol{\theta}_r)$ for $r \in \mathcal{R}$, and solving for

$$p(r|\bar{\mathcal{S}}, \bar{\mathcal{X}}; \boldsymbol{\Theta}) \propto \left(\prod_{i=1}^{m} p(\bar{\mathbf{S}}_i|\bar{\mathbf{X}}_i; \boldsymbol{\theta}_r) \right) p(r) \quad (2)$$

where all $\boldsymbol{\theta}_r$ are aggregated into a global model $\boldsymbol{\Theta}$. Assuming a uniform prior $p(r)$ over readers, this reduces to predicting the reader r that maximizes the likelihood $p(\bar{\mathcal{S}}|\bar{\mathcal{X}}; \boldsymbol{\theta}_r) = \prod_{i=1}^{m} p(\bar{\mathbf{S}}_i|\bar{\mathbf{X}}_i; \boldsymbol{\theta}_r)$.

We formulate a model $p(\mathbf{S}|\mathbf{X}; \boldsymbol{\theta})$ of a sequence $\mathbf{S}$ of fixations given a text $\mathbf{X}$. The model defines a dynamic probabilistic process that successively generates the fixation positions s_t and durations d_t in $\mathbf{S}$, reflecting how a reader generates a sequence of saccades in response to a text stimulus $\mathbf{X}$. The joint distribution over all fixation positions and durations is assumed to factorize as

$$p(s_1, \ldots, s_T, d_1, \ldots, d_T | \mathbf{X}; \boldsymbol{\theta})$$
$$= p(s_1, d_1 | \mathbf{X}; \boldsymbol{\theta}) \prod_{t=1}^{T-1} p(s_{t+1}, d_{t+1} | s_t, \mathbf{X}; \boldsymbol{\theta}).$$

The conditional $p(s_{t+1}, d_{t+1} | s_t, \mathbf{X}; \boldsymbol{\theta})$ models the generation of the next fixation position and duration given the current fixation position s_t. In the psychological literature, four different *saccadic types* are distinguished: a reader can refixate the current word (refixation), fixate the next word in the text (next word movement), move the

fixation to a word after the next word (forward skip), or regress to fixate a word occurring earlier than the currently fixated word in the text (regression) (Heister et al., 2012). We observe empirically, that modeling the amplitude as a mixture of four Gamma distributions matches the empirical distribution of amplitudes in our data well—see Figure 1. Modeling the amplitudes as a single distribution, instead of a mixture of four distributions, results in a substantially lower out-of-sampling likelihood of the model. Therefore, at each time t, the model first draws a saccadic type $u_{t+1} \in \{1, 2, 3, 4\}$ from a multinomial distribution and then generates a saccade amplitude a_{t+1} and fixation duration d_{t+1} from type-specific Gamma distributions. Formally, the generative process is given by

$$u_{t+1} \sim p(u|\boldsymbol{\pi}) = \text{Mult}(u|\boldsymbol{\pi}) \qquad (3)$$
$$a_{t+1} \sim p(a|u_{t+1}, s_t, \mathbf{X}; \boldsymbol{\eta}) \qquad (4)$$
$$d_{t+1} \sim p(d|u_{t+1}; \boldsymbol{\lambda}). \qquad (5)$$

Afterwards the model updates the fixation position according to $s_{t+1} = s_t + a_{t+1}$. The joint parameter vector $\boldsymbol{\theta}$ concatenates parameters of the individual distributions in Equations 3 to 5. Figure 2 shows a slice in the dynamical model.

Given the current fixation position s_t, the text $\mathbf{X}$, and the chosen saccadic type u_{t+1}, the amplitude is constrained to fall within a specific interval—for instance, within the characters of the currently fixated word for refixations. Therefore, we model the distribution over the saccade amplitude given the saccadic type (Equation 4) as truncated Gamma distributions, given by[1]

$$\mathcal{G}(x|[l, r]; \alpha, \beta) = \begin{cases} \frac{\mathcal{G}(x|\alpha, \beta)}{\int_l^r \mathcal{G}(\bar{x}|\alpha, \beta) \mathrm{d}\bar{x}} & \text{if } x \in [l, r] \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where} \quad \mathcal{G}(x|\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}}$$

is the Gamma distribution with shape parameter α and scale parameter β, and Γ is the Gamma function. For $x \sim \mathcal{G}(x|\alpha, \beta)$ it holds that $\mathcal{G}(x|[l, r]; \alpha, \beta)$ is the conditional distribution of x given that $x \in [l, r]$. The distribution over a sac-
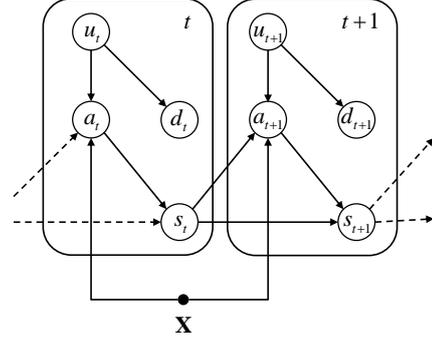

Figure 2: Graphical model notation of a slice in the dynamic model. Parameters are omitted to avoid notational clutter.

cade amplitude given the saccade type is given by

$$p(a|u_{t+1} = 1, s_t, \mathbf{X}; \boldsymbol{\eta}) =$$
$$\begin{cases} \mu \mathcal{G}(a|[0, r]; \alpha_1, \beta_1) & \text{if } a > 0 \\ (1 - \mu) \mathcal{G}(-a|[0, l]; \bar{\alpha}_1, \bar{\beta}_1) & \text{otherwise} \end{cases}$$
$$(6)$$

where the parameter μ reflects the probability for a forward saccade within a refixation, and

$$p(a|u_{t+1} = 2, s_t, \mathbf{X}; \boldsymbol{\eta}) = \mathcal{G}(a|[l^+, r^+]; \alpha_2, \beta_2)$$
$$p(a|u_{t+1} = 3, s_t, \mathbf{X}; \boldsymbol{\eta}) = \mathcal{G}(a|(r^+, \infty); \alpha_3, \beta_3)$$
$$p(a|u_{t+1} = 4, s_t, \mathbf{X}; \boldsymbol{\eta}) = \mathcal{G}(-a|(-l, \infty); \alpha_4, \beta_4).$$
$$(7)$$

Here, the truncation intervals reflect the constraints on the amplitude a_{t+1} given u_{t+1}, s_t and $\mathbf{X}$. Let w_l (w_r) denote the position of the left-most (right-most) character of the currently fixated word, and let w_l^+, w_r^+ denote these positions for the word following the currently fixated word. Then $l = w_l - s_t$, $r = w_r - s_t$, $l^+ = w_l^+ - s_t$, and $r^+ = w_r^+ - s_t$. The parameter vector $\boldsymbol{\eta}$ contains the parameters μ, $\bar{\alpha}_1$, $\bar{\beta}_1$ and α_i, β_i for $i \in \{2, 3, 4\}$.

The distribution over fixation durations given saccade type is modeled by a Gamma distribution

$$p(d|u_{t+1}; \boldsymbol{\lambda}) = \mathcal{G}(d|\gamma_{u_{t+1}}, \delta_{u_{t+1}})$$

with type-specific parameters γ_u, δ_u for $u \in \{1, 2, 3, 4\}$ that are concatenated into a parameter vector $\boldsymbol{\lambda}$.

It remains to specify the distribution over initial fixation positions and durations $p(s_1, d_1|\mathbf{X}; \boldsymbol{\theta})$, which is given by additional Gamma distributions

$$s_1 \sim \mathcal{G}(d|\alpha_0, \beta_0) \qquad d_1 \sim \mathcal{G}(d|\gamma_0, \delta_0)$$

---

[1]The definition is straightforwardly generalized to open truncation intervals.
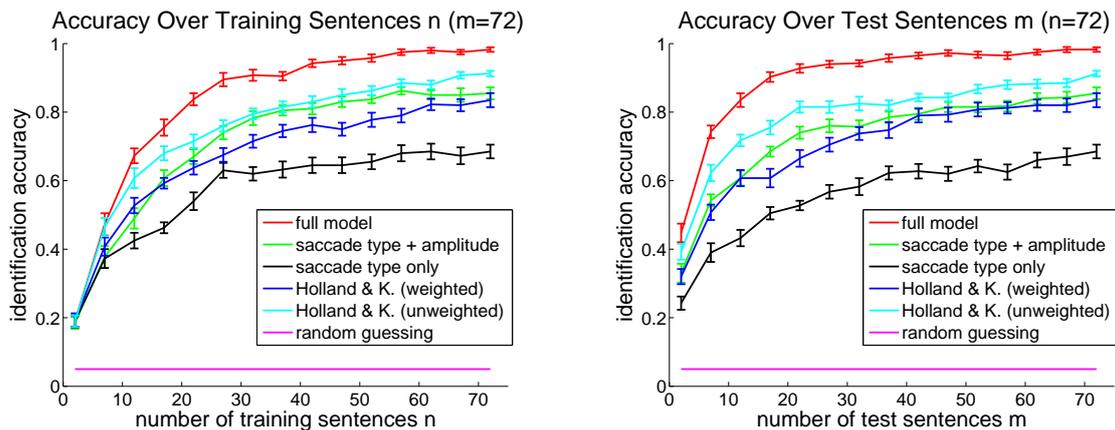
1812

Figure 3: Reader identification accuracy as a function of the number of training sentences (left) and test sentences (right) read for different model variants. Error bars indicate the standard error.

where the parameters $\alpha_0, \beta_0, \gamma_0, \delta_0$ are aggregated into the joint parameter vector $\boldsymbol{\theta}$.

4 Parameter Estimation and Inference

Given a set $\mathcal{S}^{(r)}$ of eye movement observations for reader $r \in \mathcal{R}$ on texts $\mathcal{X}$, the MAP estimate of the parameters is

$$\boldsymbol{\theta}_r = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathcal{S}^{(r)}, \mathcal{X})$$

$$= \arg\max_{\boldsymbol{\theta}} \left(\prod_{i=1}^{n} p(\mathbf{S}_i^{(r)}|\mathbf{X}_i; \boldsymbol{\theta}) \right) p(\boldsymbol{\theta}). \quad (8)$$

A Dirichlet distribution (add-one smoothing) is a natural, conjugate prior for the multinomial distribution; we use uninformative priors for all other distributions. The structure of the model implies that the posterior can be maximized by fitting the parameters $\boldsymbol{\pi}$ to the observed saccadic types under the Dirichlet prior, and independently fitting the distributions $p(a_t|u_t, \mathbf{X}, s_t; \boldsymbol{\eta})$ and $p(d_t|u_t; \boldsymbol{\lambda})$ by maximum likelihood to the saccade amplitudes and durations observed for each saccade type. The resulting maximum likelihood problems are slightly non-standard in that we have to fit Gamma distributions that are truncated differently for each data point, depending on the textual content at the position where the saccade occurs (see Equations 6 and 7). We solve the resulting optimization problems using a Quasi-Newton method. To avoid overfitting, we use a backoff-smoothing technique for $p(a_t|u_t, \mathbf{X}, s_t; \boldsymbol{\eta})$ and $p(d_t|u_t; \boldsymbol{\lambda})$: we replace reader-specific parameter estimates by estimates obtained from the corresponding data of all readers if the number of data points from which the distributions are estimated falls below a cutoff value.

The cutoff value is tuned by cross-validation on the training data.

At test time, we have to infer likelihoods $p(S_i|\mathbf{X}; \boldsymbol{\theta}_r)$ (Equation 2). This is done by evaluating the multinomial and (truncated) Gamma distributions in the model for the corresponding observations and model parameters.

5 Empirical Study

We empirically study the proposed model and several baselines using eye-movement records of 20 individuals (Heister et al., 2012). For each individual, eye movements are recorded while reading each of the 144 sentences in the *Potsdam Sentence Corpus* (Kliegl et al., 2006). The data set contains fixation positions and durations that have been obtained from raw eye movement data by appropriate preprocessing. Eye movements were recorded with an EyeLink II system with a 500-Hz sampling rate (SR Research, Osgoode, Ontario, Canada). All recordings and calibration were binocular. We randomly sample disjoint sets of n training sentences and m test sentences from the set of 144 sentences. Models are estimated on the eye movement records of individuals on the training sentences (Equation 8). The eye-movement records of one individual on all test sentences constitute a test example; the model infers the most likely individual to have generated these test observations (Equation 2). *Identification accuracy* is the fraction of times an individual is correctly inferred; random guessing yields an accuracy of 0.05. Results are averaged over 20 training and test sets.

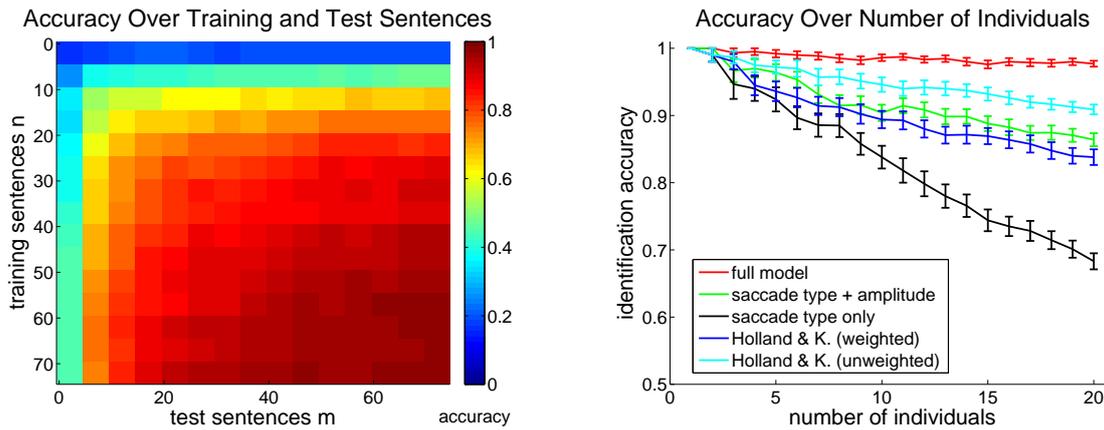We study the model introduced in Section 3

Figure 4: Identification accuracy as a function of the number of training and test sentences read for full model (left). Identification accuracy as a function of the number of individuals that have to be distinguished for different model variants (right). Error bars indicate the standard error.

(*full model*), a model variant in which the variable d_{t+1} and corresponding distribution is removed (*saccade type + amplitude*), and a simple model that only fits a multinomial distribution to saccade types (*saccade type only*). Additionally, we compare against the feature-based reader identification approach by Holland & Komogortsev (2012). Six of the 14 features used by Holland & Komogortsev depend on saccade velocities and vertical fixation positions. As this information was not available in the preprocessed data set that we used, we implemented the remaining features. There is extensive empirical evidence that saccade velocity scales with saccade amplitude. Specifically, the relationship between logarithmic peak saccade velocity and logarithmic saccade amplitude is linear over a wide range of amplitudes and velocities; this is known as the main sequence relationship (Bahill et al., 1975). Therefore, we do not expect that adding saccade velocities would dramatically affect performance of this baseline. Holland & Komogortsev employ a weighted combination of features; we report results for the method with and without feature weighting.

Figure 3 shows identification accuracy as a function of the number n of training sentences used to estimate model parameters (left) and as a function of the number m of test sentences on which inference of the most likely reader is based (right, cf. Equation 2). The full model achieves up to 98.25% accuracy, significantly outperforming the Holland & Komogortsev (2012) baseline (91.25%, without feature weighting) and simpler model variants. All methods perform much better

than random guessing. Figure 4 (left) shows identification accuracy as a function of both training size n and test size m for the full model.

We finally study how identification accuracy changes with the number of individuals that have to be distinguished. To this end, we perform the same study as above, but with randomly sampled subsets of the overall set of 20 individuals. In these experiments, we average over 50 random train-test splits. Figure 4 (right) shows identification accuracy as a function of the number of individuals. We observe that identification accuracy drops with the number of individuals for all methods; our model consistently outperforms the baselines.

6 Conclusions

We have developed a model of individual differences in eye movements during reading, and studied its application in a biometric task. At test time, individuals are identified based on eye movements on novel text. Our approach thus provides potentially unobtrusive biometric identification without requiring users to react to a specific stimulus. Empirical results show clear advantages over an existing approach for reader identification.

References

A. Terry Bahill, Michael R. Clark, and Lawrence Stark. 1975. The main sequence: a tool for studying human eye movements. *Mathematical Biosciences*, 24:191–204.

Roman Bednarik, Tomi Kinnunen, Andrei Mihaila, and Pasi Fränti. 2005. Eye-movements as a biometric. In *Proceedings of the 14th Scandinavian Conference on Image Analysis*.

W. Robert Dixon. 1951. Studies in the psychology of reading. In W. S. Morse, P. A. Ballantine, and W. R. Dixon, editors, *Univ. of Michigan Monographs in Education No. 4*. Univ. of Michigan Press.

Ralf Engbert, Antje Nuthmann, Eike M. Richter, and Reinhold Kliegl. 2005. SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4):777–813.

Bruno Erdmann and Raymond Dodge. 1898. *Psychologische Untersuchungen über das Lesen*. Halle: Max Niemeyer.

Tadayoshi Hara, Daichi Mochihashi, Yoshino Kano, and Akiko Aizawa. 2012. Predicting word fixations in text with a CRF model for capturing general reading strategies among readers. In *Proceedings of the First Workshop on Eye-Tracking and Natural Language Processing*.

Julian Heister, Kay-Michael Würzner, and Reinhold Kliegl. 2012. Analysing large datasets of eye movements during reading. In James S. Adelman, editor, *Visual word recognition. Vol. 2: Meaning and context, individuals and development*, pages 102–130.

Corey Holland and Oleg V. Komogortsev. 2012. Biometric identification via eye movement scanpaths in reading. In *Proceedings of the 2011 International Joint Conference on Biometrics*.

Edmund B. Huey. 1908. *The psychology and pedagogy of reading*. Cambridge, Mass.: MIT Press.

Pawel Kasprowski and Jozef Ober. 2004. Eye movements in biometrics. In *Proceedings of the 2004 International Biometric Authentication Workshop*.

Reinhold Kliegl, Antje Nuthmann, and Ralf Engbert. 2006. Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1):12–35.

Oleg V. Komogortsev, Sampath Jayarathna, Cecilia R. Aragon, and Mechehoul Mahmoud. 2010. Biometric identification via an oculomotor plant mathematical model. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*.

Franz Matties and Anders Søgaard. 2013. With blinkers on: robust prediction of eye movements across readers. In *Proceedings of the 2013 Conference on Empirical Natural Language Processing*.

Mattias Nilsson and Joakim Nivre. 2009. Learning where to look: Modeling eye movements in reading. In *Proceedings of the 13th Conference on Computational Natural Language Learning*.

Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422.

Erik D. Reichle, Tessa Warren, and Kerry McConnell. 2012. Using e-z reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16(1):1–21.

Ioannis Rigas, George Economou, and Spiros Fotopoulos. 2012a. Biometric identification based on the eye movements and graph matching techniques. *Pattern Recognition Letters*, 33(6).

Ioannis Rigas, George Economou, and Spiros Fotopoulos. 2012b. Human eye movements as a trait for biometrical identification. In *Proceedings of the IEEE 5th International Conference on Biometrics: Theory, Applications and Systems*.

Daniel Schad and Ralf Engbert. 2012. The zoom lens of attention: Simulating shuffled versus normal text reading using the swift model. *Visual Cognition*, 20(4-5):391–421.

Youming Zhang and Martti Juhola. 2012. On biometric verification of a user by means of eye movement data mining. In *Proceedings of the 2nd International Conference on Advances in Information Mining and Management*.