JOURNAL OF COGNITIVE PSYCHOLOGY, 2011, 23 (5), 543–558

Psychology Press
Taylor & Francis Group

Temporal dynamics of the eye–voice span and eye movement control during oral reading

Albrecht W. Inhoff[1], Matthew Solomon[2], Ralph Radach[2], and Bradley A. Seymour[1]

[1]Psychology Department, Binghamton University, Binghamton, NY, USA
[2]Psychology Department, Florida State University, Tallahassee, FL, USA

The distance between eye movements and articulation during oral reading, commonly referred to as the eye–voice span, has been a classic issue of experimental reading research since Buswell (1921). To examine the influence of the span on eye movement control, synchronised recordings of eye position and speech production were obtained during fluent oral reading. The viewing of a word almost always preceded its articulation, and the interval between the onset of a word's fixation and the onset of its articulation was approximately 500 ms. The identification and articulation of a word were closely coupled, and the fixation–speech interval was regulated through immediate adjustments of word viewing duration, unless the interval was relatively long. In this case, the lag between identification and articulation was often reduced through a regression that moved the eyes back in the text. These results indicate that models of eye movement control during oral reading need to include a mechanism that maintains a close linkage between the identification and articulation of words through continuous oculomotor adjustments.

Keywords: Eye movement control; Eye–voice Span; Fixation–speech interval; Oral reading; Word identification.

Reading requires the acquisition of linguistic information from spatially ordered words. Since high acuity vision is confined to a relatively small spatial area, readers must execute eye movements (saccades) to view different segments of text in sequence. Detailed recordings of eye movements have shown that the eyes are at—or close to—a word from which linguistic information is sought during silent reading, and that visuospatial and lexical properties of individual words influence how they are viewed (see, for detailed reviews, Rayner, 1998, 2009). This has given rise to models of eye movement control according to which eye movement programming is primarily determined by the success of ongoing visual word recognition. Consequently, more time is spent viewing words that are difficult to recognise, e.g., long words or words with a low frequency of occurrence, than words that are easy to recognise. Two of these lexical control models, the E-Z Reader model (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003) and the SWIFT model (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005), have yielded remarkably accurate simulations of eye movements during silent reading.

Successful comprehension of sentence content also requires the execution of postlexical processes, and a recently developed extension of the E-Z Reader model seeks to accommodate instances in which the grammatical function and meaning of individual words is influenced by

---

Correspondence should be addressed to Albrecht Inhoff, Department of Psychology, Binghamton University, Binghamton, NY 13902-6000, USA. E-mail: inhoff@binghamton.edu

This research was support by NICHD Grant No. HD043405. We would like to thank Poujan Rahmani, Marika Inhoff, and Melissa Zeidler for their help with the analyses of speech data.

http://www.psypress.com/ecp
DOI: 10.1080/20445911.2011.546782

subsequent sentence context (Reichle, Warren, & McConnell, 2009; Warren, White, & Reichle, 2009). The lexical eye-guidance principle is maintained by assuming that a word is generally integrated while the next word is recognised. Generally, a word's integration occurs without difficulty, and its postlexical processing will not influence the progression of the eyes through a sentence. Postlexical processing influences eye movement programming only when words are difficult to integrate. The model postulates two types of integration failure: A "slow" failure that occurs when a word's integration duration exceeds the recognition duration of the following word, and a "rapid" failure that is detected immediately for a variety of reasons (e.g., in response to perceptual errors [cf. Reichle et al., 2009]). An integration failure will result in the shifting of attention to the source of the postlexical processing difficulty, and, depending on the state of eye movement programming, it can result in (1) a refixation, (2) longer viewing durations on the next word in the text, and/or (3) the programming of a regression that returns the eyes to the source of the integration difficulty (Inhoff, Greenberg, Solomon, & Wang, 2009; Pollatsek, Juhasz, Reichle, Machacek, & Rayner, 2008; Reichle et al., 2009; Warren et al., 2009).

Recent empirical work has also shown that identification of a word during silent reading can be followed by its subvocal articulation (Eiter & Inhoff, 2010; Inhoff, Connine, Eiter, Radach, & Heller, 2004). Similar to the hypothesised postlexical integration of an identified word, the subvocal articulation of an identified word could also occur while the next word in the text is identified. That is, eye movement programming could be controlled by the demands of visual word recognition, and the lexical guidance of the eyes could be ruptured only when subvocal articulation is impeded. When this occurs, readers could increase the time spent viewing the next word in the text and/or program a regression that returns the eyes to the source of the subvocal articulation difficulty (Eiter & Inhoff, 2010).

During oral reading, the viewing of a word and its articulation are also separated. Buswell's (1921) classic study showed that approximately 15–17 letter spaces, or about two to three words, intervene between the viewing and subsequent articulation of a word (this distance has been referred to as eye–voice span [EVS]). A similar-size EVS was recently obtained by Laubrock and Bohn (2008). Closely related measurements of the temporal interval between the onset of a word's fixation and its overt articulation (Laubrock & Bohn, 2008; Laubrock & Kliegl, 2010; Laubrock, Kliegl, & Engbert, 2007; Jarvilehto, Nurkkala, Koskela, Holappa, & Vierela, 2008; see also Jarvilehto, Nurkkala, & Kyosti, 2009) indicate that slightly more than half a second intervenes between the viewing and subsequent articulation of a word when skilled participants read grammatically correct text.

The results of other experiments that used report-based procedures to estimate the EVS during oral reading suggest an even longer lag between the viewing and articulation of a word. In these studies, the presentation of a sentence was terminated before its end was reached, and the number of words that could be articulated in the absence of visible text was used to index eye–voice distance. The most recent study that used this procedure (Stuart-Hamilton & Rabbitt, 1997) terminated the display when a specific trigger word was articulated, and this resulted in the reporting of approximately four subsequent words from the sentence. Reviews of earlier report-based EVS studies (Levin & Buckler-Addis, 1979) indicated that the span ranges from four to six words.

The results of Buswell's (1921) eye tracking study and of recent investigations are consistent with the view that relatively little buffering intervenes between the identification and articulation of a word during oral reading. When oral reading is fluent, the temporal interval between the viewing and articulation may encompass little more than the time required for the identification of a visual word and the programming of to-be-articulated speech. Although instantiation of overt speech is relatively time consuming (e.g., Balota & Chumbley, 1985), so that articulation of an identified word is likely to begin after the eyes have moved to a subsequent word in the text, eye movement control during fluent oral reading may be similar to eye movement control during silent reading. That is, the eyes are primarily guided by the time line of visual word identification, and the time line of overt articulation influences eye guidance only when lexical processing and speech production are not sufficiently synchronised.

Report-based EVS estimates of four to six words suggest, however, that a memory buffer intervenes between the recognition of individual words and their articulation. Through this buffering, lexical effects on eye movement control could be decoupled from speech-related effects. The ease with which a visual word is recognised could influence the time spent viewing it, but the success and rate of speech production may not influence the viewing of individual words. Instead, these effects could be buffered and distributed over the viewing of a sequence of words. The main goal of the current study was to determine how the temporal dynamics of speech production influence eye guidance during oral reading.

METHOD

Participants

Twenty-nine undergraduate students at Binghamton University volunteered to participate in exchange for course credit. All had normal or corrected to normal vision, and they considered themselves fluent readers.

Materials

The experimental materials consisted of 10 practice and 40 experimental sentences. None of the sentences contained difficult or anomalous syntactic constructions, and all were relatively easy to comprehend. A sample sentence is shown in Figure 1. Data analyses were confined to a sequence of five consecutive ''critical'' words that started four to six words after sentence onset so that FSI could be determined when sentence reading was fluent and after articulation had started (see Figure 1). That is, we excluded the beginning words of a sentence to avoid start up effects, and we excluded the final words of a sentence because the eyes could not move beyond the sentence ending.

Two indexes of lexical processing were obtained for each critical word: its frequency of occurrence (per million) and its length (in letter spaces, LS). The average log word frequency of critical words was 4.8 (*SD* = 3.1) (Brysbaert & New, 2009; words that were not included in the corpus were given a word frequency value of .001), and their mean length was 4.7 LS (*SD* = 2.0). Word frequency and word length are negatively correlated in natural text, and this was also the case in the current study, *r* = –.53.

Apparatus

A 17-inch LG Flatron L1720P monitor at a resolution of 1024 × 768 was used to display text. Individual sentences were shown in 12 point Courier black font on a grey background, with each letter occupying a constant horizontal area of 10 pixels. Although head movements were not constrained, the monitor was positioned to yield an average distance of 70 cm, and each character subtended approximately 0.3 degrees of visual angle.

A head mounted Eye-link II eye-tracking system was used to monitor eye movements so that the chin was free to move while reading out loud. The system also tracks head movements, so

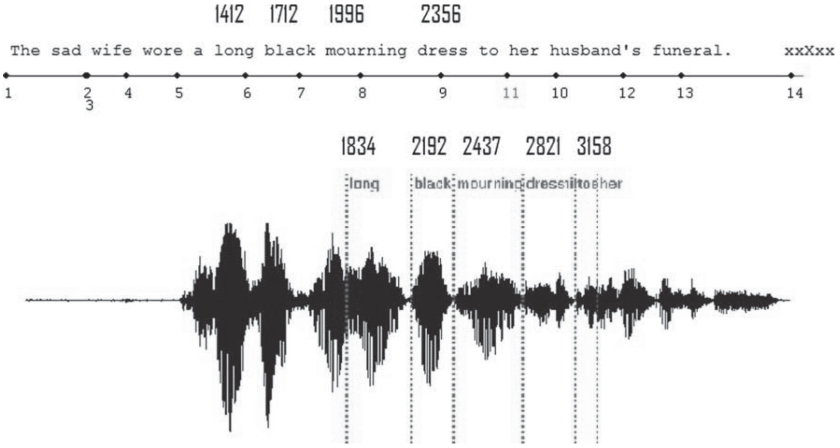

Figure 1. A typical sentence that was read without a speech production error. The numbered sequence of fixations is shown below the line of text. The interval between the onset of sentence reading and the onset of critical word viewing is shown above the line of text and the corresponding interval for the articulation of critical words is shown above the speech spectrogram that was produced during sentence reading. The fifth critical word, ''to'', was skipped. Fixation 11 moves the eyes back in the text while speech production progressed with word order.

that changes in head position can be discriminated from eye movements, but participants were asked to keep the head relatively stable and at a constant screen distance to maximise tracking accuracy. Fixation location was sampled at a rate of 500 Hz with a tracking error of 0.25 degrees of visual angle or less. Viewing was binocular but only movements of the right eye were recorded.

Oral reading was recorded with a small Super Micro Audio System, microphone model PA3, with a frequency range of 20–16,000 Hz. A Creative Sound Blaster X-Fi Sound card and short-latency ASIO driver were used to connect microphone input directly to the sound card so that the recording latency was 2–6 ms, with mean lag of approximately 3 ms. The entire experimental program, which included the presentation of individual sentences, the monitoring of eye position, and the recording of speech was run using SR Research ‘‘Experiment Builder’’ software (http://www.eyelinkinfo.com).

Procedure

Participants were tested individually in sessions lasting approximately 40 minutes. A horizontal three-point calibration of the eye-tracking system preceded the experiment. This required participants to fixate a sequence of four fixation markers as they appeared in random order at the left, centre, and right locations at the horizontal midline of the screen. This was followed by a validation routine that established an absolute tracking accuracy of 0.25 degrees of visual angle, or slightly better than a letter space (LS).

Participants were instructed to read individually presented sentences out loud with normal intonation. At the onset of an individual trial, a fixation marker was shown at the left side of the screen and the participant pushed a button on a gamepad to initiate the presentation of a sentence. It appeared subsequently 10 LS to the right of the marker. The point in time at which the eyes crossed the blank space preceding the first letter of the sentence was recorded and written into the speech- and eye-recording files to time-stamp the onset of sentence reading. The onset of critical word viewing and the corresponding onset of critical word articulation could thus be measured relative to the onset of sentence reading on a common time scale. Each trial was ended with a saccade to a sequence of Xs (xxXxx) that was shown 10 LS to the right of the sentence period. Readers were asked to fixate the upper-case centre position of the string and to depress the gamepad button once more. This fixation served to check tracking accuracy at the end of oral sentence reading, and the subsequent buttonpress erased the sentence and re-presented the fixation marker at the left side of the screen for the next trial.

Measurement

Entroware Communication software (http://www.entroware.com) was used to map saccades and fixations onto text. The program considered a word fixated when a fixation fell on one of its constituent letters or the preceding blank space, and it computed the point in time when a word was first viewed relative to the onset of sentence reading, the word viewing onset interval. The program also computed word-specific eye movement measures that have been used to index processing in a large number of reading studies (see Rayner, 1998; Inhoff & Radach, 1998). Four commonly used measures were analysed: (1) the size of a saccade that moved the eyes onto a critical word (all incoming saccades had to be forward directed), (2) the duration of the ensuing first fixation on a critical word (first fixation duration), (3) the cumulated critical word viewing time until the eyes moved to another word (gaze duration), and (4) outgoing saccade direction, i.e., whether the saccade out of a critical word was a progressive or regressive movement.

A sound file was obtained for each spoken sentence, and each file was listened to by a native speaker of English to distinguish sentences with and without speech production errors. When sentence reading was error free, Soundforge software (www.sonycreativesoftware.com) was used to determine the speech onset of each critical word. Spectral properties of the spoken sentence were used in a first pass to determine the onset of words with clear visual separations, such as the words ‘‘black’’ and ‘‘mourning’’ in Figure 1. Prosodic cues and knowledge of the to-be-articulated sentence content were used to determine word boundaries when visual separation was not possible. These transition segments were listened to selectively and incrementally until a transition point could be determined that separated the two words from each other, although complete and precise separation was not always possible due to coarticulation. To minimise word segmentation inaccuracies, all onset markings were checked and approved by an expert with training in psychoacoustics. The specification of critical word onsets was used to determine a critical word's speech onset interval, i.e., the interval between the onset of sentence reading and the onset of the critical word's articulation. The fixation–speech interval of a word (FSI) was then defined as the numeric difference between its speech and viewing onset times. Instances in which viewing of a word preceded its articulation were given positive FSI values, and instances in which articulation preceded viewing were given negative FSI values. Following Jarvilehto et al. (2008), we are using the term fixation–speech interval (FSI) in our work to specifically denote the temporal aspect of eye–voice coordination in natural, continuous reading. This is an important special case of the more general eye–voice span concept.

Data selection and data analyses

No speech onset measure was obtained for 22% of the trials because of various types of speech errors, and no FSI measure was available for 23% of the remaining 4529 critical words because they were skipped and did not yield a fixation onset measure. In addition, we excluded outliers that were defined as trials (1) with a FSI of less than –1500 ms or more than 3000 ms, (2) with first fixations of less than 50 ms or more than 800 ms, and (3) with incoming saccades of less than 1 LS or more than 15 LS.[1] Together, these selection criteria removed approximately 1% of the trials which left 3422 instances of critical word viewing.

All data were analysed using linear mixed models (LMM), as implemented in the lme4 package of the R system for statistical computing (Bates, Maechler, & Dai, 2008; R-version 2.10.1, R Development Core Team, 2008). Trial-based data are entered which makes computation relatively immune to inequalities in the number of available data. The mixed model also accommodates predictor correlations that are present when continuous text is read (Baayen, 2008), and multiple random factors can be entered simultaneously.

Four primary oculomotor measures were specified for each critical word, the size of the incoming saccade, the ensuing first fixation and gaze duration, and the direction of the outgoing saccade. To determine whether and how the predictor FSI influenced these four components of critical word viewing, each LMM model included linear and quadratic effects of the FSI at the onset of critical word viewing. In addition, each model included the properties of individual critical words as predictors, i.e., their log word frequency, the number of their constituent letters (word length), and the interaction of word frequency with word length as fixed factors. Participants and items were entered as crossed random factors.

Other oculomotor predictors were added as the viewing of a critical word unfolded in time so that the influence of an earlier eye movement component on subsequent oculomotor activity could be taken into account. The basic predictors FSI, log word frequency, word length, and the interaction of word length with word frequency, were thus used to analyse incoming saccade size. An additional predictor, incoming saccade size, was added to the analysis of first fixation duration and gaze duration, and yet another predictor, gaze duration, was added to the analysis of outgoing saccade direction. An identity link function was used to analyse incoming saccade size and the two viewing duration measures; a binomial link function was used to analyse saccade direction.

The coordination of speech production with critical word viewing could also be spread over the viewing of spatially adjacent words. If this were the case, then the saccade out of a critical word and the following fixation duration could also be influenced by the FSI at the onset of critical word viewing. A total of 3091 saccades moved forwards after the viewing of a critical word, and two supplementary LMMs were applied to this data set to examine potential FSI spillover. Specifically, the analyses of *outgoing* saccade size and of the *following* fixation duration were examined as a function of the FSI at the onset of critical word viewing, and also as a function of the critical word's lexical properties and of the oculomotor activity during critical word viewing, so that FSI-contingent spillover could be determined in combination with other potential spillover effects.

The LMM analyses yielded intercepts, regression coefficients relative to the intercept (b), standard errors (SE), and t- or z-values, and these statistics are reported in data tables. All t- and z-values of 2.0 or larger were significant at the

---

[1] Outgoing saccades of less than 1 LS and more than 15 LS were also removed from the analysis of outgoing saccade sizes.

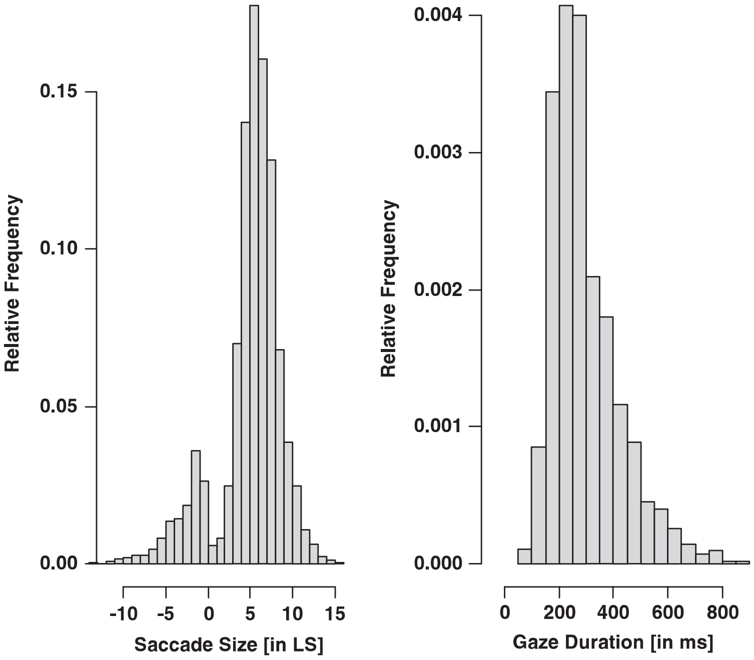

Figure 2. The left panel shows the relative frequency of outgoing saccade sizes; these outgoing saccades could either move forwards in the text or regress to a prior text segment. The right panel shows the distribution of critical word gaze duration.

$p < .05$ level, and the statistical significance of saccade size and fixation duration effects was confirmed through Markov Chain Monte Carlo (MCMC) sampling via the mcmcsamp program of the lme4 package (for which we specified 1000 samplings). Pertinent significance levels are reported in the body of text.

RESULTS

Slightly more than one fifth of the critical words were not fixated. The skipping frequency of critical words was a function of their length with a higher skipping rate for short words, 58%, 35%, and 18%, and 10% for two- to five-letter words, respectively, than for long words, 4%, 3%, 3%, and 3% for six- to nine-letter words. Movements of the eyes out of critical words did not always progress with word order (see Figure 2, left panel). The distribution of outgoing saccade sizes was distinctly bimodal, with one mode for progressive saccades (89%) and another for regressions (11%). The mean size of forward-directed outgoing saccades and of regressions was 6.6 LS and 4.0 LS, respectively. Mean gaze (first pass word viewing) duration was 299 ms, and the distribution, shown in the right panel of Figure 2, was skewed towards longer durations.

The FSI mean was 486 ms, and the relative frequency of FSIs is shown in Figure 3. Although 11% of saccades moved the eyes to a prior word in the sentence, negative FSIs were exceedingly rare (less than 2%). That is, the onset of a word's fixation almost never occurred at a later point in time than the onset of its articulation.

Overall, this pattern of eye movements is similar to oculomotor activity during silent reading, except that gaze durations were somewhat longer and saccades were somewhat smaller than during silent reading.[2] In addition, the results show that readers keep the eyes relatively close to an articulated word. More detailed analyses sought to elucidate the influence of the FSI at the onset of critical word viewing on eye movement control.

Incoming saccade size

The left panel of Figure 4 shows incoming saccade size as a function of critical word frequency and FSI. The right panel shows the corresponding

---

[2] According to Rayner and Pollatsek (1989), gaze durations during silent reading are typically between 250 ms and 275 ms, and saccade sizes range between 6 and 9 LS. The distribution of gaze durations in the current study is similar in shape to the distribution of gaze durations for silent reading (Rayner, 1998).

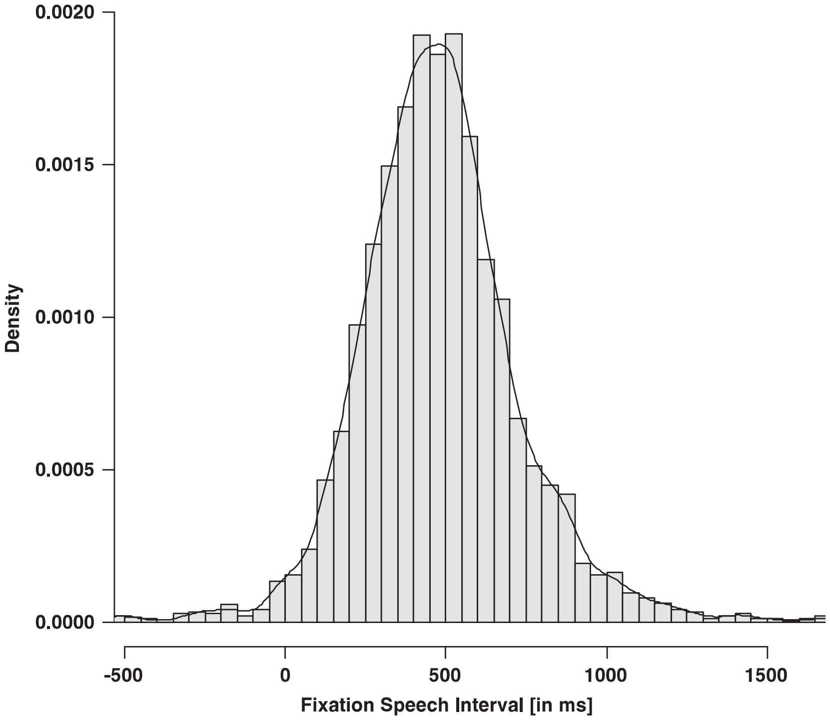

Figure 3. The relative frequencies of the fixation–speech interval.

effects of word length and FSI. LMM coefficients for each predictor, standard errors, and *t*-values are shown in Table 1.

A larger saccade to a critical word increased the temporal gap between the viewing and articulation of critical words, and longer FSIs at the onset of critical word viewing were associated with larger incoming saccade sizes ($p < .01$). The analysis also revealed a significant quadratic FSI component ($p < .01$), as FSIs of more than 900 ms were associated with smaller rather than larger incoming saccades.

A relatively long FSI after a small incoming saccade suggests that the preceding FSI was also relatively long. A small incoming saccade could thus be part of an oculomotor correction strategy. To examine whether the quadratic FSI effects was a byproduct of an oculomotor correction strategy, we removed critical words with outgoing regressions from the data set and then reanalysed these data. The results revealed a highly significant linear FSI component ($p < .01$), and the quadratic component was no longer reliable ($p > .08$).

As can be seen in Figure 4, incoming saccades also increased with the length of the critical word ($p < .01$). The effect of word frequency and the interaction of word frequency with word length were not significant ($p > .1$ and $p > .5$, respectively).

Critical word viewing durations

Figures 5 and 6 show first fixation duration and gaze duration as a function of FSI, critical word frequency, and word length. Table 1 shows the LMM statistics for the two measures.

As can be seen in Figures 5 and 6, the duration of the first fixation on a critical word and its gaze duration increased with FSI. The trend was linear for FSIs between 250 ms and 750 ms, and the linear FSI component was highly reliable for first fixation duration and gaze duration (both *p*s $< .01$). Large FSIs at the onset of critical word viewing were thus routinely down-regulated through increases in word viewing duration.

First fixation duration and gaze duration also yielded a negative quadratic FSI component ($p < .01$ and $p < .05$, respectively), as both viewing durations decreased when FSIs increased over 750 ms. Inspection of the data indicated that the unexpected decrease in viewing durations for relatively long FSIs was due to relatively short viewing durations when critical word viewing was terminated with a regression. Trials with outgoing

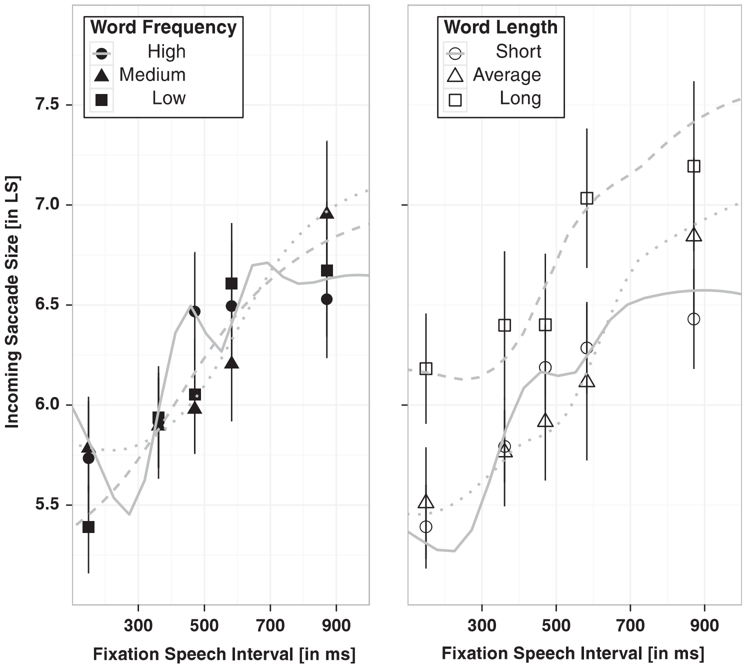

Figure 4. The left panel shows incoming saccade size as a function of the fixation–speech interval (FSI) and critical word frequency. The right panel shows the corresponding effects of word length. Bin cuts were applied to FSI to create five equal-size groupings (very short, short, medium, long, very long), and the means of these bins are shown with error bars that mark the 95% confidence interval. Least square regression lines, applied to the full set of ungrouped FSI data, are also shown.

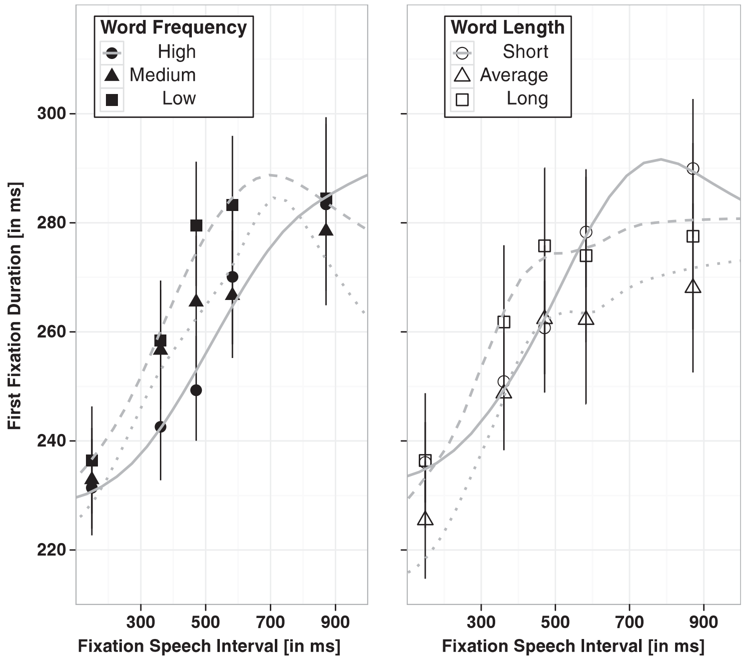

Figure 5. The left panel shows the duration of the first fixation on critical words as a function of the fixation–speech interval (FSI) and critical word frequency. The right panel shows the corresponding effects of word length. Bin cuts were applied to FSI to create five equal-size groupings (very short, short, medium, long, very long), and the means of these bins are shown with error bars that mark the 95% confidence interval. Least square regression lines, applied to the full set of ungrouped FSI data, are also shown.

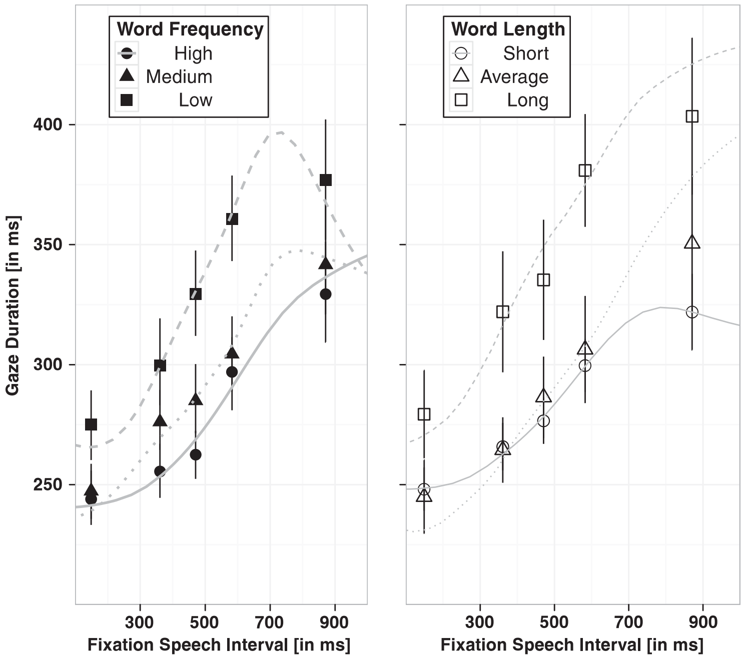

Figure 6. The left panel shows the gaze duration (the cumulated viewing duration) of critical words as a function of the fixation–speech interval (FSI) and critical word frequency. The right panel shows the corresponding effects of word length. Bin cuts were applied to the FSI to create five equal-size groupings (very short, short, medium, long, very long), and the means of these bins are shown with error bars that mark the 95% confidence interval. Least square regression lines, applied to the full set of ungrouped FSI data, are also shown.

TABLE 1
The LMM models used for the analyses of forward-moving (incoming) saccade size and the duration of the following first critical word fixation and critical word gaze duration

Measure and predictor	*Effect size*	*SE*	*t-value*
Incoming saccade size			
Intercept	4.82		
Word frequency (log WF)	7.4E–02	4.8E–02	1.57
Word length	2.03E–01	4.4E–02	4.56
WF × Word length	5.0E–03	1.0E–02	0.57
Linear FSI	1.3E–01	2.1E+00	6.18
Quadratic FSI	−6.85E+00	2.1E+00	−3.31
First fixation duration			
Intercept	238		
Incoming saccade size	5.5E+00	7.7E–01	7.23
Word frequency (log WF)	−2.0E–01	1.7E+00	−0.11
Word length	−3.48E–01	1.7E+00	−0.21
WF × Word length	−4.43E–01	3.6E–01	−1.23
Linear FSI	7.96E+02	9.7E+01	8.21
Quadratic FSI	−2.71E+02	9.3E+01	−2.91
Gaze duration			
Intercept	158		
Incoming saccade size	7.2E+00	1.1E+01	6.69
Word frequency (log WF)	9.05E+00	2.7E+00	3.4
Word length	2.24E+01	2.5E+00	8.87
WF × Word length	−2.68E+00	5.5E–01	−4.92
Linear FSI	1.62E+03	1.4E+01	11.89
Quadratic FSI	−2.83E+02	1.3E+02	−2.17

Significance levels are reported in the text. FSI = fixation–speech interval.

regressions were removed in a supplementary analysis to determine whether the quadratic component was due to a confounding of the FSI effect with outgoing saccade direction. The results of this supplementary analysis revealed a highly significant linear component for first fixation duration and gaze duration (both ps $<.01$), and the quadratic FSI components of the two measures were no longer reliable (both ps $>.17$).

The viewing of a critical word was also influenced by prior oculomotor activity and by critical word properties. First fixation duration and gaze duration were shorter when the incoming saccade was small than when it was large (both ps $<.01$), and gaze duration increased with decreases in word frequency and with increases in word length (both ps $<.01$). The effects of word frequency and word length interacted, with larger word frequency effects for long words ($p < .01$). The first fixation duration of critical words was not influenced by word-specific properties (all ps $>.2$).

The relationship between FSI and critical word gaze duration is shown for each reader in Figure 7. As can be seen, FSI effects on word viewing duration were quite stable across readers. Except for four participants (P6, P10, P11, and P26), readers increased critical word viewing duration in response to a large FSI at the onset of critical word viewing.

Outgoing saccade direction

The relationship between the FSI and the rate of outgoing regressions is shown in Figure 8, and LMM statistics are shown in Table 2.

The FSI at the onset of critical word viewing had a profound effect on the direction of outgoing eye movements. The rate of outgoing regressions

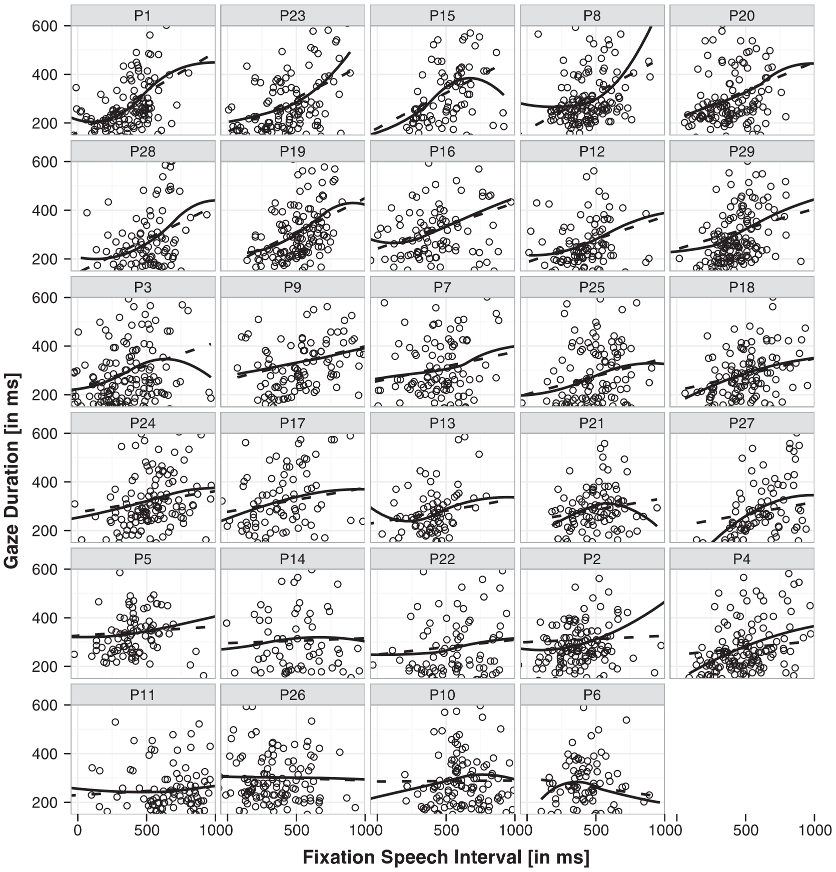

Figure 7. The relationship between gaze duration and the fixation–speech interval (FSI) for the 29 readers. Readers are ordered according to the slope of their linear FSI effect (dashed lines). Full lines show least square regression lines. P = participant.

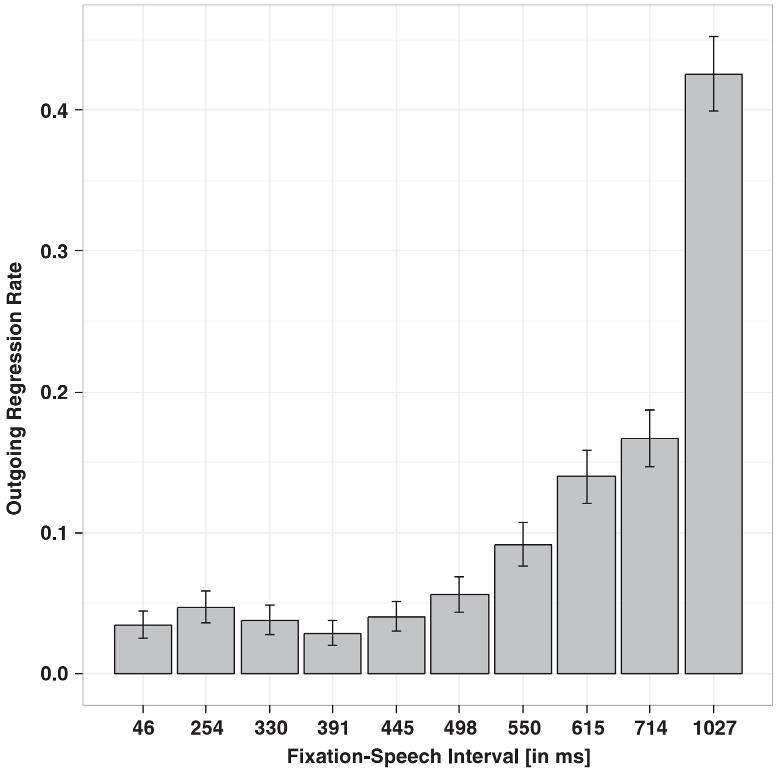

Figure 8. The relative frequency of outgoing regressions as a function of the fixation–speech interval (FSI) at the onset of critical word viewing. Bin cuts were used to create 10 FSI bins each of which contained an equal number of FSIs. The x-axis shows the mean FSI value for each bin, and the error bars indicate the SE for individual bins computed over subjects.

was less than 5% for FSIs of 500 ms and less, and it exceeded 40% for FSIs of more than 794 ms (the lowest FSI value for the 1027 ms FSI interval shown in Figure 9). The linear and quadratic FSI components were reliable (both *p*s $<.01$). The linear effect was due to an increase in regression rate with increases in FSI, and the quadratic effect appears to be due to a dip in regression rate for FSIs of approximately 390 ms (see Figure 8). Regressions rate also increased with critical word length ($p < .01$), especially when the word had a low frequency of occurrence ($p < .05$). Increases in regression rate were also associated with shorter critical word gaze durations ($p < .01$), as readers terminated critical word viewing “prematurely” when a long FSI was corrected with a regression after a critical word was fixated.

Together, the four analyses of critical word viewing indicate that readers responded immediately to the FSI at the onset of a word’s fixation with two complementary FSI regulation strategies. Almost all readers responded to increases in the FSI at the onset of critical word viewing with immediate increases in critical word viewing duration. However, when the FSI at the onset of critical word viewing was relatively long, readers often terminated critical word viewing prematurely with a regression that moved the eyes closer to the spoken word.

Spillover

Figure 9 shows the size of progressive saccades out of critical words and the duration of the following fixation as a function the critical word’s FSI and word length. LMM statistics are shown in Table 2.

The analysis of outgoing saccade sizes yielded a negligible linear component ($p > .3$) and a significant quadratic FSI component ($p < .01$). The quadratic component could be part of a FSI regulation strategy that spilled over into the viewing of the following word, but its precise source is not clear. The size of forward-directed saccades out of critical words was also influenced by critical word properties, with larger outgoing saccades for

TABLE 2
The LMM models used to for the analyses of outgoing saccade size and outgoing saccade direction

Measure and predictor	*Effect size*	*SE*	*z-value*
Outgoing saccade direction (regression rate)			
Intercept	−1.6		
Incoming saccade size	−2.39E–02	3.16E–02	−0.76
Word frequency (log WF)	4.48E–02	6.12E–02	0.73
Word length	2.32E–01	6.08E–02	3.83
WF × Word length	−2.37E–02	1.25E–02	−1.89
Incoming saccade size	2.39E–02	3.16E–02	−0.75
Linear FSI	9.03E+01	5.63E+00	16.04
Quadratic FSI	−2.87E+01	4.63E+00	−6.19
Gaze duration	−7.50E–03	7.05E–04	−10.63
			t-value
Outgoing forward-directed saccade size			
Intercept	4.15		
Incoming saccade size	−2.26E–02	1.69–02	−1.57
Word frequency (log WF)	1.07E–01	4.01E–02	2.67
Word length	4.52E–01	3.90E–02	11.59
WF × Word length	−7.46E–03	8.24E–03	−1.89
Linear FSI	−2.67E+00	2.39E+00	−1.11
Quadratic FSI	−7.82E+00	2.19E+00	3.57
Gaze duration	6.16E–04	2.69E–04	2.29
Fixation duration after critical word viewing			
Intercept	251		
Incoming saccade size	8.27E–01	9.16E–01	0.90
Word frequency (log WF)	−2.23E–00	1.96E–00	–1.19
Word length	4.92E–00	1.95E–00	2.52
WF × Word length	2.92E–01	4.05E–01	0.72
Linear FSI	−4.38E+02	1.12E+02	3.90
Quadratic FSI	−9.07E+01	1.03E+02	–0.87
Gaze duration	−2.94E–02	1.47E–02	2.01

Significance levels are reported in the text. FSI = fixation–speech interval.

high frequency words and long words than for low frequency words and short words ($p < .01$, for both main effects). In addition, outgoing saccade size increased with critical word gaze duration ($p < .05$).

The duration of the fixation following a progressive outgoing saccade increased with the FSI at the onset of (prior) critical word viewing. The linear FSI component was reliable ($p < .01$), the quadratic component was not ($p > .4$). The fixation duration following critical word viewing also increased with critical word length ($p < .025$), and it was longer when critical word gaze duration had been long ($p < .05$). Together, the two analyses indicate that the FSI at the onset of critical word viewing influenced oculomotor activity even after the critical word was viewed and that other aspects of critical word processing also spilled over into the viewing of a subsequent word.

DISCUSSION

The current study showed that the eyes were kept relatively close to an articulated word during oral reading and that eye movement control was immediately influenced by ongoing speech production and visual word recognition. Specifically, increases in the FSI at the onset of critical word viewing were immediately responded to with increases in first fixation duration and gaze duration, and a very long FSI was often corrected with a regression that moved the eyes closer to an articulated word. FSI effects were not completely resolved during the viewing of individual words, as the FSI at the onset of word viewing influenced the size of progressive outgoing saccades and the following fixation duration.

Mean FSI at the onset of critical word viewing was approximately 500 ms, and critical word gaze duration was slightly less than 300 ms. This

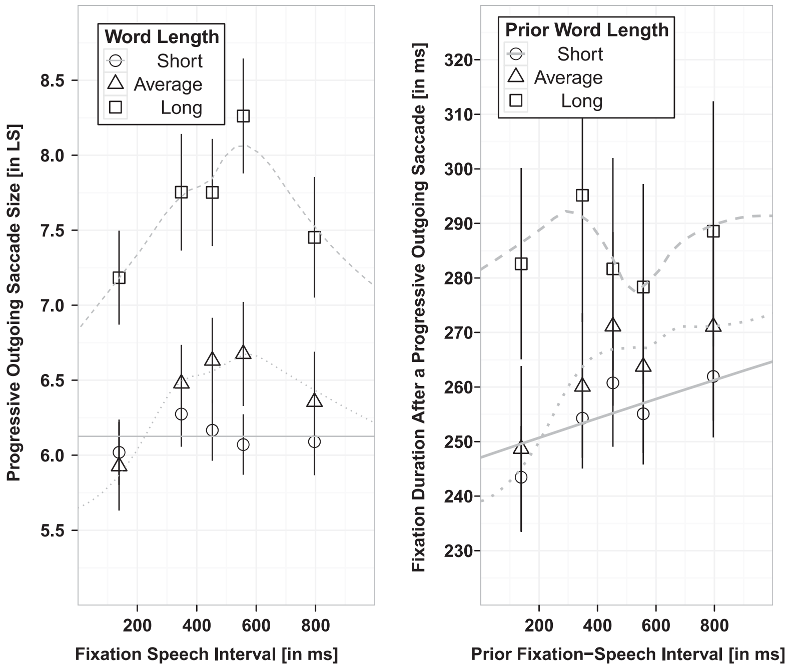

Figure 9. The left panel shows the size of outgoing progressive saccades and the right panel shows the following fixation duration as a function of the fixation–speech interval at the onset of (prior) critical word viewing and of the length of the previously fixated critical word.

implies that one word typically intervened between the viewing and articulation of a word after a forward-directed saccade was executed, and this corresponds to an eye–voice distance of approximately two words. This differs considerably from earlier work with recall-based measures that yielded EVS estimates between four and six words. It is exceedingly unlikely that the much smaller EVS in the current study is due to peculiarities of our sentence materials or of the tested population. Our sentences were easy to comprehend, and no member of the relatively large group of participants reported a reading-related disability. It is also unlikely that the mean FSI of approximately 500 ms was a byproduct of an overly cautious reading strategy. Mean first fixation duration and gaze duration when reading out loud, approximately 260 ms and 300 ms, respectively, were only 50–60 ms longer than the corresponding viewing durations during silent reading (see, for reviews, Rayner, 1998, 2009; Rayner & Pollatsek, 1989). Furthermore, similar to silent reading, word skipping was relatively common, especially when a critical word was short (Brysbaert, Drieghe, & Vitu, 2005; Rayner & McConkie, 1976), small saccades to critical words were followed by shorter first fixation and gaze durations (Inhoff & Rayner, 1986), and the frequency and length of a critical word influenced its gaze duration (Klieg, Grabner, Rolfs, & Engbert, 2004; Pollatsek et al., 2008). The relatively short FSI thus cannot be attributed to reading impediments, reading anomalies, oculomotor anomalies, or to atypical reading strategies. The current FSI estimate is also quite similar to other recent FSI estimates for skilled readers (Jarvilehto et al., 2008; Laubrock & Bohn, 2008; Laubrock et al., 2007), and it appears safe to conclude that report-based EVS estimates were inflated. Readers could have looked ahead and/or guessed to improve performance with this procedure.

A FSI of approximately 500 ms indicates that relatively little time intervened between the identification of a word and its overt articulation. In naming tasks that require identification and articulation of individual words, articulation onset latencies typically range between 500–600 ms, and they are often less than 500 ms when a to-be-named word has a high frequency of occurrence (e.g., Balota & Chumbley, 1985). A mean FSI of slightly less than 500 ms in the current study thus suggests that readers started to program the articulation of a word immediately after it was identified. Furthermore, immediate and robust FSI effects on first fixation duration and gaze duration indicate that the progression of the eyes through the text was regulated so that the identification and articulation of a word were closely coupled during oral reading. In the current study, a typical eye–voice span of approximately two words could thus have emerged when a reader identified a fixated word (n), specified the speech parameters for the prior word (n – 1), and started to articulate the word before that (n – 2).

Regulation of the FSI occurred primarily through adjustments of critical word viewing duration, as FSI increases were responded to with increases in gaze duration for FSIs up to approximately 750 ms. A yet longer FSI at the onset of critical word viewing was often responded to with a regression. When the FSI was relatively long, two or more words could intervene between the fixation and articulation of a word, and this could have ruptured the contiguity between a critical word's identification, speech programming, and articulation, and this was corrected with a regression.

The pattern of FSI effects on word viewing durations during fluent oral reading suggest that processes that occur after visual word identification, in particular the monitoring of speech production, have a pervasive and immediate effect on eye movement programming. The most recent version of the E-Z Reader model includes a mechanism through which readers can immediately respond to postlexical integration failures, and one may speculate that a similar—though more pervasive—mechanism regulates eye–speech coordination during oral reading. Specifically, an articulation delay could be detected and responded to routinely, i.e., whenever the identification of a word cannot be followed by the programming and/or production of corresponding speech because the viewing and articulation of a word are too close to each other or too far apart. The contiguity of word identification and speech production could occur through adjustments of word viewing duration, unless the FSI at the onset of word viewing was relatively long.

Alternatively, it could be assumed that long FSIs will occur naturally as part of a continuous distribution, and the response of the eye movement control system could be equally graded. Such a response mode can be envisioned in an interactive activation (IA) framework of word processing and oculomotor control in reading. In one IA model of eye movement control during reading model, Glenmore (Reilly & Radach, 2006), words are represented as potential saccade targets on a saliency map, where saliency values are determined as a combination of bottom up visual and top down cognitive activation. Critically, a fixation centre receives feedback on the global amount of excitation in the saliency network and triggers an eye movement to the most salient target word when activation goes down below a threshold. In such a framework, a saccade out of the currently fixated word could be delayed when a secondary saliency peak is maintained at the location corresponding to the ensuing articulation. A very long FSI would increase the chance for the articulated word to have a higher saliency than other potential targets in the current visual configuration at the moment of saccade triggering, and this would cause the observed regressive interword saccades towards the current locus of articulation. There are several current models in the domain of oral word reading that have an IA architecture and/or use dynamic activation functions (see Biedermann, Coltheart, Nickels, & Saunders, 2009, for a discussion). Within such a tandem of dynamic activation models, the FSI could be expressed as a distance function between concurrent peaks representing ongoing visual word processing and articulation.

The supplementary analysis of progressive outgoing saccades and of subsequent fixation durations also showed that FSI effects and lexical effects were distributed over the viewing of more than one word. A long FSI at the onset of critical word viewing not only increased the time spent viewing the critical word, it also increased the fixation duration immediately following critical word viewing. The precise source of these spillover effects cannot be determined with the current data. According to the E-Z Reader model, the FSI at the onset of critical viewing could influence the acquisition of useful information from the next (parafoveally visible) word, possibly because a long FSI delayed forward-directed shifts of attention. Similarly, word-length contingent spillover may have occurred because readers obtained less useful information from a parafoveally visible word when the fixated word was long than when it was short. FSI-contingent spillover could also have occurred because of some temporal buffering between visual word identification and speech programming and/or between speech programming and speech production. Further work is needed to distinguish between these alternatives.

Original manuscript received April 2010
Revised manuscript received November 2010
First published online April 2011

REFERENCES

Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, *24*, 89–106.

Bates, D., Maechler, M., & Dai, B. (2008). Linear mixed-effects models using S4 classes. R package version 0.999375-27. [Software]. Vienna, Austria: R Foundation for Statistical Computing.

Biedermann, B., Coltheart, M., Nickels, L., & Saunders, S. (2009). Effects of homophony on reading aloud: Implications for models of speech production. *Language and Cognitive Processes*, *24*, 804–842.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53–77). Oxford, UK: Oxford University Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Buswell, G. T. (1921). The relationship between eye-perception and voice response in reading. *Journal of Educational Psychology*, *12*(4), 217–227.

Eiter, B., & Inhoff, A. W. (2010). Visual word recognition during reading is followed by subvocal articulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 457–470.

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., Nuthmann, A., Richter, E. M., & Kiegl, R. (2005). SWIFT: A dynamic model of saccade generation in reading. *Psychological Review*, *112*, 777–813.

Inhoff, A. W., Connine, C., Eiter, B., Radach, R., & Heller, D. (2004). The phonological representation of words in working memory during sentence reading. *Psychonomic Bulletin and Review*, *11*, 320–325.

Inhoff, A. W., Greenberg, S. N., Solomon, M., & Wang, C. A. (2009). Word integration and regression programming during reading: A test of the E-Z Reader 10 model. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1571–1584.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance and scene perception* (pp. 29–54). Amsterdam, The Netherlands: Elsevier, North Holland.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, *27*, 457–464.

Jarvilehto, T., Nurkkala, V.-M., Koskela, K., Holappa, E., & Vierela, H. (2008). Reading as anticipatory formation of meaning: Eye movement characteristics and fixation-speech intervals when articulating different types of text. *Journal of Transfigural Mathematics*, *1*, 73–81.

Jarvilehto, T., Nurkkala, V.-M., & Kyosti, K. (2009). The role of anticipation in reading. *Pragmatics and Cognition*, *17*, 509–526.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262–284.

Laubrock, J., & Bohn, C. (2008). Broadening the scope of eye-movement research in reading: Oral reading and proof reading. *International Journal of Psychology*, *43*, 190.

Laubrock, J., & Kliegl, R. (2010). The eye-voice span during reading aloud. *Manuscript submitted for publication*.

Laubrock, J., Kliegl, R., & Engbert, R. (2007, August). On the eye–voice span during oral reading. Paper presented at the European conference on Eye Movements, Potsdam, Germany.

Levin, H., & Buckler-Addis, A. (1979). *The eye–voice span*. Cambridge, MA: MIT Press.

Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: A reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 726–750.

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829–837.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

R Development Core Team (2008). *R: A language environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org

Reichle, E., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–526.

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., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin and Review*, *16*, 1–21.

Reilley, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*, 34–55.

Stuart-Hamilton, I., & Rabbitt, P. (1997). The decline of eye-voice span in elderly readers. *Educational Gerontology*, *23*, 389–400.

Warren, T., White, S., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and the E-Z Reader. *Cognition*, *111*, 132–137.