Journal of Experimental Psychology: General
2011, Vol. 140, No. 4, 552–572

© 2011 American Psychological Association
0096-3445/11/$12.00 DOI: 10.1037/a0023885

Coregistration of Eye Movements and EEG in Natural Reading: Analyses and Review

Olaf Dimigen
Universität Potsdam and Humboldt-Universität zu Berlin

Werner Sommer and Annette Hohlfeld
Humboldt-Universität zu Berlin

Arthur M. Jacobs
Freie Universität Berlin

Reinhold Kliegl
Universität Potsdam, Germany

Brain-electric correlates of reading have traditionally been studied with word-by-word presentation, a condition that eliminates important aspects of the normal reading process and precludes direct comparisons between neural activity and oculomotor behavior. In the present study, we investigated effects of word predictability on eye movements (EM) and fixation-related brain potentials (FRPs) during natural sentence reading. Electroencephalogram (EEG) and EM (via video-based eye tracking) were recorded simultaneously while subjects read heterogeneous German sentences, moving their eyes freely over the text. FRPs were time-locked to first-pass reading fixations and analyzed according to the cloze probability of the currently fixated word. We replicated robust effects of word predictability on EMs and the N400 component in FRPs. The data were then used to model the relation among fixation duration, gaze duration, and N400 amplitude, and to trace the time course of EEG effects relative to effects in EM behavior. In an extended Methodological Discussion section, we review 4 technical and data-analytical problems that need to be addressed when FRPs are recorded in free-viewing situations (such as reading, visual search, or scene perception) and propose solutions. Results suggest that EEG recordings during normal vision are feasible and useful to consolidate findings from EEG and eye-tracking studies.

Keywords: EEG, eye tracking, fixation-related potentials, artifact correction, natural viewing

Reading is a complex cognitive task, unfolding at the same time at visual, attentional, lexicosemantic, and oculomotor levels. Comprehension requires the processing of visual input across a complex series of brief fixation pauses and saccadic eye movements as well as retrieving, updating, and integrating contents of memory. Current research on reading makes heavy use of two methods: recording eye movement (EMs) and event-related brain potentials (ERPs). Traditionally, these research methods have used different experimental protocols: In EM studies, subjects read sentences or paragraphs of text while their fixation position is monitored with an eye tracker. The durations, positions, and sequences of fixations are then used to make inferences about the underlying cognitive processes (Rayner, 1998). Procedures in these studies often resemble everyday reading without unusual task demands. In contrast, in ERP studies of reading, serial visual presentation (SVP) has typically been used to avoid saccade-related measurement artifacts in the electroencephalogram (EEG). In SVP, readers fixate the center of the screen while sentences are presented word by word at a predefined pace. ERPs are then time-locked to stimulus presentations.

In the present study, we demonstrate effects of a critical variable—the predictability of a word from the prior sentence context—in EMs and ERPs that were recorded simultaneously during left-to-right sentence reading. The predictability effect has figured prominently in both the EM- and the ERP-research traditions. Coregistration of EM and EEG may grant new perspectives on the relation between fixation time and single-trial EEG amplitude, as well as on the time course of predictability effects in both measures. Their simultaneous recording also raises several methodological problems to which we propose solutions. We argue that methodological advances in coregistration, as exemplified for reading in the present article, will also apply to other free viewing situations. In the following, we summarize (a) the rationale for focusing on word predictability effects, (b) the potential benefits of simultaneous recordings, (c) previous EEG studies in which some

This article was published Online First July 11, 2011.

Olaf Dimigen, Department of Psychology, Universität Potsdam, Germany, and Department of Psychology, Humboldt-Universität zu Berlin, Germany; Werner Sommer and Annette Hohlfeld, Department of Psychology, Humboldt-Universität zu Berlin, Germany; Arthur M. Jacobs, Department of Psychology, Freie Universität Berlin, Germany; Reinhold Kliegl, Department of Psychology, Universität Potsdam, Germany.

Funding was provided by Deutsche Forschungsgemeinschaft Grants KL 655/6 and FOR 868. Results were first presented at the International Conference on Cognitive Neuroscience held September 2005 in Havana, Cuba, and at the 13th annual meeting of the Cognitive Neuroscience Society held April 2006 in San Francisco, California.

The authors thank Petra Grüttner for assistance in data collection, Rainer Kniesche for technical support, and Mark Minnes for proofreading.

Correspondence concerning this article should be addressed to Olaf Dimigen, Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24-25, Potsdam 14476, Germany. E-mail: dimigen@uni-potsdam.de

form of EM coregistration has been used, and (d) the methodological challenges that have limited the use of this technique.

Word Predictability in Reading

A word's predictability in the context of a given sentence is known to modulate both oculomotor behavior (e.g., Balota, Pollatsek, & Rayner, 1985; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Ashby, Pollatsek, & Reichle, 2004) and N400 amplitude (e.g., Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Kutas & Hillyard, 1984). The well-established N400 component describes a negative-going ERP deflection, which is most pronounced around 400 ms after stimulus onset at centroparietal recording sites (Kutas & Hillyard, 1980). N400 amplitude is largest when a word violates the semantic context of a preceding sentence fragment, but is also larger for semantically correct words that are less predictable from the context. Because of its context sensitivity, N400 amplitude is thought to reflect the difficulty in retrieving conceptual knowledge associated with a word from memory or in integrating it into the context of the sentence or discourse (Kutas, Van Petten, & Kluender, 2006). However, it remains controversial whether N400 effects reflect facilitated access to lexicosemantic features (Lau, Phillips, & Poeppel, 2008), a late postlexical process of semantic context integration (Brown & Hagoort, 1993; Holcomb, 1993), or semantic inhibition (Debruille, 2007), and it is possible that multiple mechanisms contribute to the N400. Regardless of the theoretical viewpoint, the N400 provides information about the time course of semantic processing, and its onset can be interpreted as an upper time limit for the initial access to word meaning. One aim in the present study was therefore to test for the existence of an N400 in a normal reading situation and describe its properties.

Predictability also figures prominently in current conceptualizations of reading from the perspective of eye movement control: Highly predictable words are skipped more frequently (e.g., Balota et al., 1985; Drieghe, Rayner, & Pollatsek, 2005; Vitu, 1991); fixations on them are shorter (e.g., Altarriba, Kroll, Sholl, & Rayner, 1996; Balota et al., 1985; Rayner et al., 2004); and high predictability of an upcoming word is associated with a longer fixation on the previous word (Kliegl, Nuthmann, & Engbert, 2006). Understanding the role of predictability in reading is also part of the broader question of whether lexical processing is spatially distributed over several adjacent words and whether lexicosemantic information—in addition to low-level visual and orthographic properties—is extracted from not-yet-fixated words in the parafovea (Kennedy, Pynte, & Ducrot, 2002; Kliegl, 2007; Kliegl et al., 2006; Rayner, Pollatsek, Drieghe, Slattery, & Reichle, 2007).

Potential Benefits of Simultaneous Recordings

From the perspective of ERP research, there is no doubt that SVP has proven itself to be extremely successful method of studying the electrophysiological correlates of word recognition (Kutas et al., 2006). At the same time, it presents a strong simplification of the normal reading process, which differs in several ways from SVP: In normal reading, readers determine how long each word is fixated and which word to fixate next. Words are therefore not inspected in a strictly serial fashion but frequently are skipped or fixated several times, and regressive saccades toward earlier words are common. Unlike SVP, normal reading allows for the preprocessing of upcoming words in parafoveal vision. At the same time, words are not always fixated at their center (as in SVP) but are often processed from nonoptimal viewing positions near the word boundaries (Nuthmann, Engbert, & Kliegl, 2005). Another major difference concerns speed: While most ERP studies present words at stimulus-onset asynchronies of between 400 and 1,000 ms (i.e., 60–150 words per minute), average reading fixations last only 200–250 ms, and reading speeds of 250 words per minute are common. Accordingly, new visual input is obtained at much higher rates than in most SVP experiments. Finally, SVP imposes secondary-task demands—to maintain fixation and refrain from blinking—absent in normal reading.

As a result of these differences, the extent to which SVP results apply to normal reading is unknown. Several approaches to improve the ecological validity of SVP have been introduced. One is to present words at speeds that are very fast (Kutas, 1987), reading-like (Dambacher, Rolfs, Göllner, Kliegl, & Jacobs, 2009), or under the control of the reader by pushing a button (Ditman, Holcomb, & Kuperberg, 2007). Another proposal is to grant a parafoveal preview on the upcoming word during SVP without eye movements (Barber, Donamayor, & Kutas, 2010). Finally, in several studies, the same sentences have been used in separate EM and SVP experiments with different subjects (Dambacher & Kliegl, 2007; Raney & Rayner, 1993; Sereno, Rayner, & Posner, 1998).

All these techniques preclude direct comparisons between EEG measures and oculomotor behavior. As a consequence, the basic relationship and temporal contingency between the dependent variables in EM research (e.g., fixation duration) and ERP research (e.g., component amplitude) is unresolved (Sereno & Rayner, 2003). On the topic of predictability, one interesting question concerns the apparent discrepancy in the timing of effects in EMs and ERPs (Rayner & Clifton, 2009; Sereno & Rayner, 2003): In normal reading, predictability acts early enough to influence the initial fixation on a word (Rayner, Binder, Ashby, & Pollatsek, 2001; Rayner et al., 2004), which lasts less than 250 ms on average. In contrast, N400 effects from SVP studies only begin to arise at 200–250 ms and reach their maximum at 400 ms or later. This raises the question of whether predictability effects in both methods reflect a common underlying process, or not.

A potential alternative is to record EMs and ERPs from the same reader simultaneously. Because little or no useful information is acquired during the saccade (Matin, 1974; Ross, Morrone, Goldberg, & Burr, 2001), fixation onsets provide natural EEG time-locking points to study information processing in normal vision. Averaged potentials aligned to fixation onsets are called fixation-related potentials (FRPs), while those aligned to saccade onsets are called saccade-related potentials (SRPs).[1]

Existing Research Integrating EM and EEG

Recording SRPs and FRPs is not a new technique. Large and single saccades, measured via electro-oculogram (EOG) electrodes

[1] With regard to the family of visually evoked components that follow saccade on- and offset, SRPs have also been referred to as *lambda waves* in the literature. Likewise FRPs have also been called *eye-fixation-related potentials* (EFRPs). We use SRP and FRP for averaged saccade- and fixation-aligned data, because these abbreviations are short and symmetric, and their meaning is not confined to early visual processing.

near the eyes, have frequently been used in basic EEG research on postsaccadic visual processing, oculomotor preparation, and decision making (e.g., Everling, Krappmann, & Flohr, 1997; see Methodological Discussion for additional references). In contrast, only a handful of studies on visual word recognition have allowed for saccades. In several early studies, researchers recorded SRPs following a single saccade toward a word presented in the periphery (e.g., Marton, Szirtes, & Breuer, 1985). In two recent studies with eye tracking (Baccino & Manunta, 2005; Simola, Holmqvist, & Lindgren, 2009), pairs of words have been presented in order to investigate whether the lexicality of a parafoveal word and its semantic relation to the foveal word influence the ERP while subjects still fixate the foveal word. Baccino and Manunta (2005) have reported an effect of semantic relatedness before the saccade to the parafoveal word and as early as 110 ms after stimulus onset. Simola and colleagues (2009) found a lexicality effect for parafoveal words in the right hemifield, but no evidence for parafoveal semantic access. To avoid saccade-related measuring artifacts, investigators in both studies restricted data analysis to a short segment of EEG before the first saccade.

In a study by Hutzler et al. (2007), subjects read an array of five unrelated words and had to judge whether the final word had been presented as part of the array or not. The old/new effect—a late positivity for correctly recognized old words—was observed in FRPs and also during SVP of the same words. As part of their pioneering studies on SRPs (Marton, 1991). Marton and colleagues even allowed their subjects to read full sentences from left to right (Marton & Szirtes, 1988a, 1988b). However, due to various technical constraints, the final word of the sentence was displaced 20° to the left or right. After time-locking the SRP to the saccade onset, the authors observed an N400-like effect when the final word violated the sentence context. As in the study of Hutzler et al., the critical saccade was last in the sequence, so it was not possible to compare neural activity with fixation durations.

Finally, several investigators have recorded SRPs during largely unconstrained scanning behavior such as reading (Barlow, 1971; Kurtzberg & Vaughan, 1979; Takeda, Sugai, & Yagi, 2001; see also Burdette, Walrath, Gross, & Stern, 1986), rapid-eye-movement (REM) sleep (Abe, Ogawa, Nittono, & Hori, 2004), or picture viewing (Kurtzberg & Vaughan, 1979). However, without concurrent eye tracking, SRPs could not be related to fixation durations or the fixated item, but were instead aggregated across all saccades or compared globally for different stimuli or saccade types (e.g. reading vs. picture scanning; Kurtzberg & Vaughan, 1979). Finally, a study by Graupner, Velichkovsky, Pannasch, and Marx (2007) allowed for free EM behavior during picture viewing. Subjects scanned a scene, and distractor stimuli were occasionally flashed near the current fixation. Different distractor conditions were then compared in terms of their effect on fixation duration and the visual potential evoked by distractor onset.

Methodological Challenges

To our knowledge, there have been no previous attempts to co-register EM and EEG in unconstrained viewing situations in order to directly compare brain activity to oculomotor behavior as a function of the properties of the currently fixated item. This is likely due to at least four methodological problems associated with such recordings: (a) the need to co-register precise gaze position without technical interference, (b) corneoretinal and myogenic eye movement artifacts, (c) varying degrees of overlap between brain responses elicited by successive fixations, and (d) low-level, visuomotor influences on cortical activity before and after fixation onset. In the Methodological Discussion section, we review the relevant background information on each of these problems and propose solutions.

The Present Study

Given the important role of predictability in reading research, we deemed it a suitable proving ground for an attempt to co-register EMs and ERPs in natural vision with in- and outgoing saccades. Subjects read sentences at their own pace, moving their eyes freely, with no task other than comprehension. We assumed that well-known effects could be recovered and that we could reap benefits from coregistration that went beyond what usually can be inferred from separate recordings. Data analyses are structured as follows: First, we describe basic properties of the artifact-corrected FRP and SRP in multisaccadic vision. Second, we demonstrate that standard word predictability effects are recovered under coregistration. Third, we establish the basic relationship between EM behavior and N400 in the same set of fixations. We model this relation at the level of individual fixations and trace the time course of semantic processing relative to fixation onset.

Method

Subjects

Thirty students (22 women and 8 men; age range 17–37 years; mean age 23.0 years) participated in the 2.5-hr session. All were native speakers of German with a mean of 15 years of education and no history of reading difficulties or neurological or psychiatric disorders. They were paid 25 euros or received course credit. All subjects had normal or corrected-to-normal visual acuity (Bach, 1996). Twenty-five subjects were right-handed, one left-handed, and four ambidextrous (Oldfield, 1971). Data from four additional subjects were recorded but not analyzed because of EEG voltage drifts.

Apparatus

Subjects were seated in a dimly lit, sound-attenuated room at a distance of 85 cm from a 17-in. (43.18-cm) monitor (SyncMaster 171T TFT [Samsung Group, Seoul, South Korea], resolution 800 × 600 pixel, 60 Hz vertical refresh). The screen of the monitor was framed with a light gray cardboard mask that subtended 60° × 75°. The mask served to homogenize the characteristics of the visual field across different on-screen fixation locations and to reduce any resulting influences on the morphology of postsaccadic visually evoked lambda waves (see Methodological Discussion).

Materials

Subjects read the Potsdam Sentence Corpus (Kliegl et al., 2004), which contains 144 unrelated German sentences (1,138 words) with a large variety of grammatical structures and semantic contents. All sentences are semantically and syntactically legal. Sen-

tence length ranges from 5 to 11 words, with a mean of 7.9 words. The corpus previously has been used to study predictability effects on EMs (Kliegl et al., 2006) and stimulus-locked ERPs (Dambacher et al., 2006). Thirty-two samples of these sentences are provided in Kliegl et al. (2004). For the present analyses, we considered only open-class words of the corpus (nouns, adjectives, verbs, adverbs; $n = 813$ words) and excluded words at the beginning (Word Positions 1 and 2) or end (final word) of the sentence. Sentence-initial words were excluded to avoid influences of the trial-initial fixation check on fixation behavior. Sentence-final words were excluded because they tend to be fixated longer (Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000) and elicited more positive-going ERPs (Friedman, Simson, Ritter, & Rapin, 1975; Hagoort, 2003) than words at intermediate positions ("sentence wrap-up" effects).

To study predictability effects, we used only the normal range of cloze probabilities in the sentences of the Potsdam Corpus. In order to do so, we categorized the remaining 499 words (henceforth called *target words*) according to cloze probability. The cloze probability of a word in a given sentence context is defined as the probability that a reader will correctly guess it as the upcoming word after knowing all preceding words of the sentence. Cloze probabilities for every word in the corpus were collected in a norming study with 282 German native speakers (for details, see Kliegl et al., 2004). Each subject generated predictions for a subset of the sentences, yielding 83 complete protocols for the entire corpus.

While predictability is always defined as cloze probability in the present study, it is important to note that cloze probability is typically correlated to, but not identical with, the amount of contextual constraint imposed by the preceding sentence. For example, both a weakly and a strongly constraining sentence frame can be completed by an equally unpredictable final word (e.g., Federmeier, 2007; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). We therefore also computed sentence constraint at the position of the target word, which was operationalized as the number of different predictions generated during the norming study.[2] The theoretical range of this variable was therefore from 1 (perfectly constraining sentence frame; allows only one completion) to 83 (uninformative sentence frame; every rater guesses a different upcoming word). As it is typical for a corpus of normal sentences, cloze probability correlated not only with constraint ($r = -.43$) but also with word length ($r = -.23$), word position ($r = .18$), and CELEX-based word frequency ($r = .33$; Baayen, Piepenbrock, & van Rijn, 1995). To control the influences of these covariates, we included them as predictors in a linear mixed model of N400 amplitude.

For most SRP/FRP analyses, we used a three-level categorization of *low-predictability* words (cloze $p \leq .01$, $n = 187$ words), *medium-predictability* words ($.01 <$ cloze $p \leq .25$, $n = 235$), and *high-predictability* words (cloze $p > .25$, $n = 83$). Mean cloze probabilities for these categories were .00, .07, and .55, respectively (Table 1). To compare effects on EMs and FRPs, we also used a finer categorization into five bins, where unpredictable words (cloze $p = 0$) were assigned to the first bin, and the remaining words were assigned to four additional bins, each containing the same number of words.

Procedure

The experimental procedure was designed to approximate a natural reading flow, including leftwards return saccades at the end of each trial. At the beginning of the trial, a fixation point appeared on the left side of the center line of the screen (Figure 1A). Five-hundred milliseconds after fixation point onset, the eye tracker started to poll the subjects' eye position. Once it registered a stable (>150 ms) fixation, a full sentence was presented as one line of text on the center line of the monitor, thereby replacing the fixation point. Text was displayed in black on a white background in a monospaced font (9-point Courier) at a visual angle of 0.26° per character. The horizontal position of the sentence was set so that the initial fixation was always located slightly left of the center of the first word (the optimal viewing position; O'Regan & Lévy-Schoen, 1987). Subjects then read the sentence at their individual pace, moving their eyes freely over the words. After they finished reading, subjects looked for 500 ms at a second small point near the right margin of the screen. This fixation initiated a new trial: The sentence and right fixation point disappeared and were replaced by the left fixation point, the fixation of which triggered presentation of the next sentence.

The subjects' task was to read the sentences and to answer simple three-alternative multiple-choice questions presented after 20% of the sentences. Questions pertained to the content of the preceding sentence and were answered by a mouse click (mean accuracy: 96%). There was no instruction to suppress eye blinks. Subjects read 10 warm-up sentences before the experiment.

EM Recording

EM were recorded from the right eye with a table-mounted IView-X Hi-Speed eye tracker (SensoMotoric Instruments, Teltow, Germany) at a sampling rate of 240 Hz. Viewing was binocular. The infrared video-based system has an instrument spatial resolution of less than 0.025° and an absolute gaze position accuracy of up to 0.2°. Thus, calibrated eye position was recorded accurately at the level of letters. Head movements were minimized by the eye tracker's built-in chin and forehead rests. Proper calibration of the eye tracker was automatically assessed at the onset of every trial: If gaze was not detected within an invisible 0.5° × 0.5° box around the left fixation point within 2 s, the system was recalibrated with a 13-point grid.

EEG Recording

The EEG and EOG were recorded from 32 Ag/AgCl electrodes on the scalp and around the eyes. Twenty-eight electrodes were mounted in an elastic electrode cap (Easycap GmbH, Herrsching, Germany) at positions FP1, FP2, Fz, F3, F4, F7, F8, FC5, FC6, FT9, FT10, Cz, C3, C4, T7, T8, A2, CP5, CP6, Pz, P3, P4, P7, P8, PO9, PO10, O1, and O2 of the international 10/10 system. Four EOG electrodes were affixed to the outer canthi and infraorbital ridges of both eyes. Foam cushions were fitted to the subjects' forehead to preclude pressure artifacts from contact between fron-

[2] The same results pattern was obtained when we defined contextual constraint not as the number of different words expected in the cloze procedure, but as the cloze probability of the most expected word.

Table 1
Target Word Properties and Resulting Effects on Eye Movement Behavior and Amplitude of Fixation-Related Brain Potentials

			Predictability						
	All targets		Low		Medium		High		
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*
Word and sentence properties[a]									
Cloze probability	0.12	0.21	0.01	0.00	0.07	0.07	0.54	0.21	.000
Word length (no. of characters)	5.8	2.7	6.6	2.7	5.4	2.5	4.9	2.3	.000
CELEX frequency (log per million)	4.7	1.3	4.0	1.2	5.1	1.2	5.3	1.2	.000
Word position in sentence	5.0	1.7	4.6	1.6	5.1	1.7	5.8	1.6	.000
Sentence constraint (no. of guessed words)	25.1	11.0	27.8	10.8	26.3	10.7	15.4	6.3	.000
Sentence length (no. of words)	8.2	1.4	8.2	1.5	8.2	1.3	8.4	1.3	*ns*
Eye-movement behavior[b]									
First-fixation duration (in milliseconds)	224	25	235	28	219	25	213	21	.000
Gaze duration (in milliseconds)	278	41	304	54	268	38	247	34	.000
Duration previous fixation $n - 1$ (in milliseconds)	213	25	214	26	210	24	217	29	.046
Duration next fixation $n + 1$ (in milliseconds)	212	32	216	29	210	34	207	38	.025
Refixation probability	0.24	0.07	0.28	0.10	0.22	0.08	0.18	0.09	.000
Incoming saccade amplitude (in degrees)	2.0	0.5	2.0	0.5	1.9	0.5	2.0	0.5	.037
Outgoing saccade amplitude (in degrees)	1.4	0.6	1.4	0.6	1.4	0.6	1.5	0.7	*ns*
Fixated after sentence onset (in milliseconds)	1,301	198	1,244	177	1,300	208	1,431	253	.000
Sentence reading duration (in milliseconds)	2,490	713	2,566	737	2,465	704	2,400	687	.000
Fixation-related potentials[b]									
Amplitude at Pz, 300–500 ms (in microvolts)	−0.77	0.62	−1.27	0.90	−0.69	0.74	0.03	1.31	.000

Note. CELEX = CELEX lexical database.
[a] Statistics are based on words. [b] Statistics are based on fixations.

tal electrodes and the eye tracker's forehead rest. Seating position and head position in the eye tracker were carefully adapted to avoid myogenic interspersion from neck and temple muscles. Impedances were kept below 5 kΩ. An additional electrode at FPz served as ground. Signals were amplified with a Brainamp AC amplifier (Brain Products GmbH, Gilching, Germany) at a band-pass of 0.01–70 Hz and digitized at a rate of 250 Hz. All electrodes were initially referenced to left mastoid (A1) but were converted to average reference offline, thereby recovering A1 as a recording electrode. Thus, the data of 33 electrodes entered the analyses. For use in artifact correction, three-dimensional electrode locations were determined with a CMS20 digitizer (Zebris Medical GmbH, Isny, Germany). The EM and EEG records were synchronized via a common transistor–transistor–logic (TTL) trigger sent at the beginning and end of each trial from the stimulus presentation PC (running Presentation Software, Neurobehavioral Systems, Albany, CA) and looped through to two additional PCs recording EMs and EEG.

EM Analysis

The EM record was screened for loss of measurement and eye blinks. If a sentence contained only a single blink very early (<200 ms) or late (>2 s) after sentence onset (12.8 % of trials), the remaining data was used for fixation detection. Otherwise, or if multiple blinks occurred, the trial was discarded (5.1 % of trials). Saccades were detected as outliers in two-dimensional-velocity space with the monocular variant of the algorithm detailed in Engbert and Mergenthaler (2006). Saccade detection led to a total pool of 38,538 reading fixations. In a first level of screening, we discarded 2,775 fixations that occurred during intervals in which the EEG contained nonocular artifacts. In a second step, the pool was constrained to 22,321 fixations that occurred more than 700 ms after sentence onset. Earlier fixations were excluded to avoid temporal overlap between FRPs and the ERP evoked by the screen onset of the sentence. In line with previous experiments with the Potsdam Corpus (Kliegl et al., 2006), we eliminated extremely short (<50 ms, $n = 1{,}157$) or long (>750 ms, $n = 64$) fixations. In a final step, the pool was restricted to *first* fixations on *target words* in *first-pass reading*: 12,607 of the remaining fixations were on targets, 9,237 fixations of these were first fixations rather than refixations, and 7,113 occurred in first-pass reading. All EM and EEG analyses were based on this final pool of 7,113 fixations. Because fixations with a bad concurrent EEG record were removed, EM and FRP analyses were always conducted on the exact same set of fixations.

Dependent variables for behavioral analyses were first-fixation duration (FFD) and gaze duration (GD). Gaze duration is defined as FFD plus the duration of all immediate refixations. FFD and GD were submitted to repeated-measures analyses of variance (ANOVAs) on the factor predictability. Results are reported with *p* values corrected for violations of sphericity according to Huynh & Feldt (1976), the original degrees of freedom, and the epsilon (ε) value.

EEG Ocular Artifact Correction

To correct for corneoretinal eye movement artifacts (see the Methodological Discussion), we applied surrogate multiple-source eye correction (MSEC; Berg & Scherg, 1994; Ille, Berg, & Scherg, 2002) as implemented in BESA (Version. 5.1; MEGIS Software, Gräfelfing, Germany). The method combines the recording of

calibration eye movements, principal component analysis (PCA), and dipole modeling to separate artifact and brain activity. In surrogate MSEC, calibration eye movements are averaged to derive characteristic scalp topographies for different types of ocular artifacts for each subject. In addition to these empirically derived artifact topographies, a set of brain signal topographies is defined by a generic dipole model of the brain, which is identical for all subjects. This "surrogate" brain model is not used to directly model the artifact-free EEG, but its purpose is to preclude the subtraction of genuine brain activity that is spatially correlated to the artifact. On the bases of these spatial definitions for artifact and brain activity, a linear inverse operator is computed in which the experimental data are decomposed into linear combinations of brain and artifact activity; that is, the activation time courses for the artifact topographies are determined in the presence of the brain model. In a final step, this estimated artifact activity is subtracted from the raw EEG.

Technical details were as follows: In a 15-min session before the experiment, subjects performed 120 calibration saccades (15° amplitude) in the four cardinal directions. Saccades were aimed at targets on the mask surrounding the monitor. Saccade direction was indicated by an arrow, which appeared in the screen center every 3 s. In addition, 40 spontaneous eye blinks were recorded during fixation. Short EEG segments following each of the three movement types (vertical, horizontal, blink) were then averaged and subjected to three separate PCAs. The first PCA factor (typically explaining $> 97\%$ variance) was used to define the topography for each type of artifact. Note that PCA was used here as an optional preprocessing step (see Berg & Scherg, 1994) to extract the most characteristic artifact topographies from calibration data; PCA was not applied to the experimental data. Brain signal topographies were defined by BESA model RS4.par. This model contains 12 dipoles with fixed location and orientation, placed at spatially distributed, strategic positions of the brain. After defining artifact and brain topographies, the activity time course for each topography was determined in the experimental data using the spatial filter operator detailed in Ille et al. (2002, p. 123). For correction, the activity assigned to the artifact topographies was subtracted. After MSEC, the corrected continuous data were high pass-filtered at 0.25 Hz (48 dB/octave) and EOG channels were treated as regular EEG channels. Application of surrogate MSEC is detailed in Scherg (2003). Recommendations for recording clean calibration movements are given by Ruchkin (in Berg, 2002, p. 7-2). A comparison with other correction methods is provided by Ille et al. (2002).

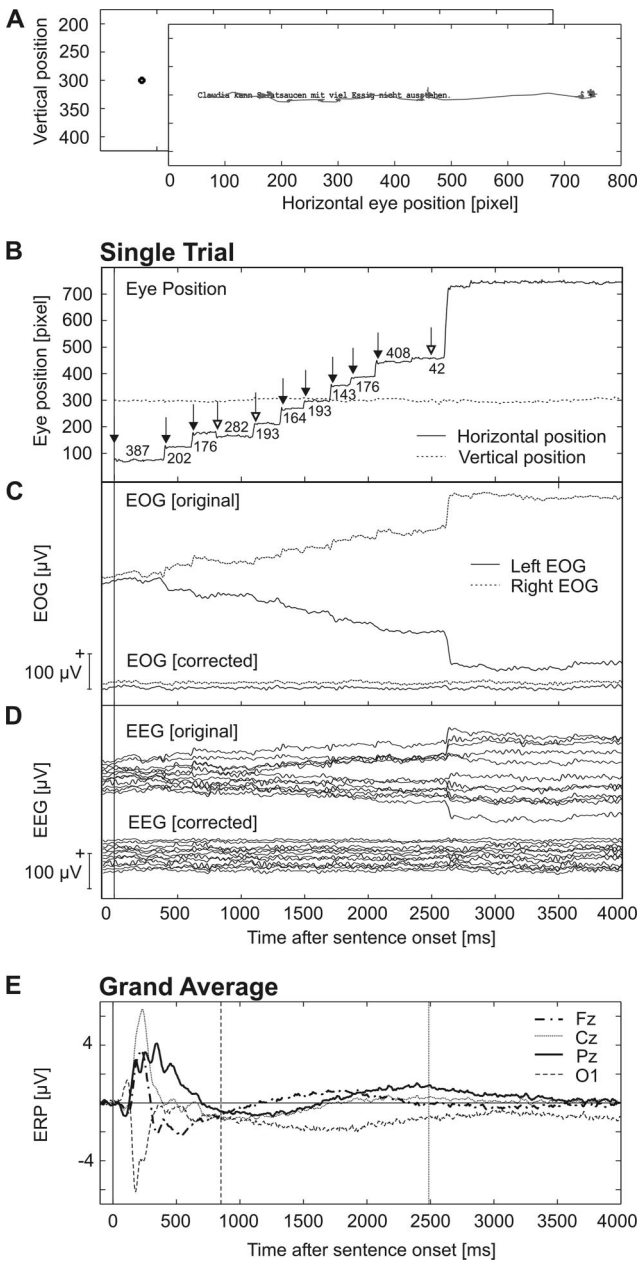

Figure 1 (opposite). Panel A. Trial scheme and data for a typical sentence: Each trial began with a fixation point on the left. Once the eye tracker detected a precise fixation, a single sentence was displayed as one line of text. Subjects read the sentence at their individual pace, moving the eyes freely over the text. Eye movements are plotted for one subject who is reading the sentence "Claudia kann Salatsaucen mit viel Essig nicht ausstehen [Claudia cannot stand salad dressings with lots of vinegar]." After reading the sentence, subjects looked at a point on the right. Gaze-controlled presentation ensured a continuous reading flow, including leftwards return saccades to read a new sentence. Panel B. Horizontal and vertical gaze positions. The sentence appeared at Time 0. Solid arrows indicate the onsets of first fixations; open arrows mark refixations. Fixation durations are given in milliseconds. Panel C. Signal at left and right horizontal electro-oculographic (EOG) electrode before and after corneo-retinal artifact correction. Panel D. Synchronized electroencephalographic (EEG) record for a subset of channels before and after correction. Panel E. Grand average artifact-corrected event-related potentials, time-locked to sentence onset. The dotted line indicates mean sentence-reading duration. To avoid overlap between potentials evoked by sentence onset and those evoked by individual reading fixations, we considered only fixations that occurred more than 700 ms after sentence onset (dashed line). Electrodes: Fz, Cz, Pz, and O1.

Fixation-Locked EEG

Around each fixation, we cut a 1,600-ms segment of EEG (from 600 ms before to 1,000 ms after fixation onset) and baseline-corrected it by subtracting the mean voltage in the 100-ms interval prior to fixation onset. To reject muscle or drift artifacts, we discarded segments with absolute voltages in any channel greater than 100 μV or with a peak-to-peak voltage difference greater than 150 μV. We performed joint EM and EEG analyses in MATLAB (Mathworks, Natick, MA) using selected functions of the EEGLAB toolbox (Delorme & Makeig, 2004).

Evaluation of Ocular Correction

The synchronized eye-tracking data served as a new criterion to assess ocular correction quality. For this purpose, a corresponding set of fixation-locked segments was also cut from the original, uncorrected EEG. For each corrected and uncorrected segment, the correlation between each EEG channel (downsampled to 240 Hz) and the horizontal component of the eye track was computed in an interval from −100 to 1,000 ms around fixation onset. Correlation coefficients for individual segments were Fisher's z transformed, averaged within each subject and then across subjects. Correlations before and after ocular correction were tested against zero for each channel with a paired t test.

Analysis of Predictability Effects

Artifact-corrected segments were averaged according to the predictability of the fixated word. To test for the presence of an N400 effect, we submitted mean amplitude in the traditional N400 window (300–500 ms after fixation onset) to a repeated-measures ANOVA on the factors of predictability and electrode. To estimate a discrete time point for the onset and peak of the N400 effect, we used the difference wave between the two extreme predictability conditions (low predictability − high predictability) at electrode Pz, low-pass-filtered at 10 Hz (zero-phase). N400 onset was determined with consecutive, sample-by-sample t tests on this difference wave between −300 and 600 ms around fixation onset. The t-max permutation test of Blair and Karniski (1993) was used to control for multiple testing.[3]

N400 peak latency was defined as the time of the maximum absolute voltage in the grand average difference wave between 0 and 800 ms. To test for a lateralization of N400 effects, we aggregated effect amplitude (low-predictability words – high-predictability words, 300–500 ms) over all 15 left- and all 15 right-hemisphere electrodes and compared them with a t test, leaving out the three midline sites.

Analysis of EM–EEG Relationship

To explore the relative timing of EM and EEG measures, we analyzed on which word subjects were fixating at the onset and peak of the N400 effect. For the same purpose, we also computed an additional average, aligned to the saccade that terminated the first fixation on the target. In analogy to response-locked averages in traditional ERPs, this SRP reveals whether, on average, N400 effects arise prior to the initiation of the next saccade, that is, during the initial fixation on the word. For this analysis, the baseline remained identical; SRP segments were baseline corrected, with the baseline still placed before the onset of the preceding target fixation. An analogous t-max statistic was computed also for this average.

To test for a between-subject linear relation between predictability effects in EM and N400 amplitude, we correlated the size of the predictability effect on behavior (FFD and GD) with that on the FRP across subjects. For this analysis, target words were categorized as having low or high predictability via a split at the median cloze probability of 0.024.

Of special concern was the relation between fixation duration and N400 amplitude at the level of individual fixations. We specified linear mixed models, with the N400 amplitude following each individual fixation as dependent variable, and word and sentence characteristics (predictability, frequency, length, constraint, word position)—as well as either the log of FFD or the log of GD—as linear covariates (fixed effects). Predictability values were logit-transformed (logit cloze $p = .5 * \ln[\text{cloze } p / (1 \text{ cloze } p)]$; see Kliegl et al., 2006) before they entered the model. Subjects and words were included as crossed random factors. For parameter estimation, we used the lmer program of the lme4 package (Bates & Maechler, 2009) in the R system for statistical computing (R Development Core Team, 2009). In these regression analyses, the variance of the N400 differences between subjects and between words were modeled on the assumption that they were normally distributed.

Results

Results are organized in six sections. First, we present standard EM effects that establish the ecological validity of the data. Second, we report measures for the quality of ocular artifact correction. Passing these checks was a precondition for the validity of FRP results. Third, we describe the FRP over the course of reading an entire sentence. In particular, sentence onset evoked a potential that spilled over to different degrees in the FRPs, forcing the exclusion of some fixations from the following analyses. Fourth, we introduce the artifact-corrected FRP as an EEG measure. Since FRPs have not been described in detail for natural viewing situations, we describe how presaccadic motor potentials, postsaccadic visual potentials, and overlapping potentials influence the results. Fifth, we present word predictability effects on FRPs. Finally, we compare EM and EEG effects and map them onto a common timeline.

Eye Movements

Eye movements for a typical trial are shown in Figure 1. On average, subjects read the sentence for 2,490 ms (SE = 130 ms) before they initiated the final saccade toward the right fixation

[3] During 10,000 data permutations, the sign of each single-subject difference wave was randomly assigned, t values were again computed for every sampling point, and the t value with the maximum absolute value was stored. This resulted in a distribution of 10,000 maximum t values expected under the null hypothesis (i.e., with randomly shuffled conditions). N400 onset was defined as the first sample of the recorded waveform where the t value was below (more negative than) the 5th percentile of the t-max distribution ($t = -3.33$ for the FRP). This tested the directed hypothesis of more negative voltages for low predictable words.

point. Mean FFD on target words was 224 ms (SE = 5 ms). Targets received at least one refixation in 24% of the cases, resulting in a mean gaze duration of 278 ms (SE = 8 ms).

Word predictability clearly influenced EM behavior.[4] On average, low-predictability words were fixated 22 ms longer than high-predictability words upon first fixation: low: 235 ms; medium: 219 ms; high: 213 ms; $F(2, 58) = 39.2$, $p < .001$, $\varepsilon = .90$, $\eta_p^2 = .58$. Gaze duration was 57 ms longer for words with low predictability than for words with high predictability low: 304, medium: 268, high: 247; $F(2, 58)=57.5$, $p < .001$, $\varepsilon = .87$, $\eta_p^2 = .67$. Due to these effects, total sentence reading duration also differed between predictability levels, $F(2, 58) = 29.1$, $p < .001$, $\varepsilon = .96$, $\eta_p^2 = .50$ (Table 1).

Of special methodological importance were differences in the amplitude of the incoming and outgoing saccade. This is because saccade amplitude per se influences the morphology of saccade-related visuomotor brain potentials, independent of the distortion by EM artifacts (see the Methodological Discussion). Outgoing saccade amplitude did not differ as a function of word predictability. Incoming saccade amplitude was slightly smaller for fixations on words with medium predictability: low: 2.0°, medium: 1.9°, high: 2.0°; $F(2, 58) = 3.5$, $p < .05$, $\varepsilon = .95$, $\eta^2 = .27$; see Table 1), but later we will show that this difference of less than 0.1° was unlikely to have a relevant impact on the FRP.

Quality of Artifact Correction

EEG correction quality was assessed with three criteria: (a) visual impression of the continuous EEG, (b) voltage differences between left- and right-hemisphere electrodes in the averaged FRP, and (c) residual correlations between gaze position and EEG after MSEC correction. Figure 1 provides an example of the continuous EOG (Fig. 1C) and EEG (Fig. 1D) before and after correction. After correction, it was generally not possible for us to visually identify residual artifacts in the continuous data. Figure 2A shows the FRP, superimposed for all channels, before and after correction. Before correction, large distortions from the predominantly rightward-going saccades were evident, with positive distortions at right-hemisphere electrodes and negative distortions at left-hemisphere electrodes. As the incoming saccade was usually followed by more right-going saccades, artifacts from multiple saccades summed up toward the end of the segment. At 1,000 ms, temporal electrodes T7 and T8 differed in voltage by about 100 μV.

After correction, signals at all channels were in the typical ERP amplitude range (Figure 2A). Although artifacts were drastically reduced, some frontolateral channels still showed indications of reversed polarities on opposite sides of the head toward the very end of the segment. This suggests that correction across several saccades was not perfect for these electrodes. As a new quantitative measure, we computed linear correlations between each EEG channel and the horizontal component of the eye track. Before correction, 31 of 33 channels correlated significantly, $|t(29)| > 2.05$, $p < .05$, with gaze position, with maximum correlations at electrodes near the eyes (see Table 2). The horizontal EOG electrodes each showed a near-perfect correlation with gaze position ($r = \pm .97$), which increased to .99 in a bipolar EOG montage (right − left). After MSEC, correlations at all channels were close to zero (maximum $|r| = .07$ and maximum $R^2 = .005$), suggesting that the residual variance accounted for by horizontal saccades was small. Nevertheless, correlations remained significantly different from zero for about half (18 of 33) of the channels. Electrodes on the posterior sagittal midline were least affected by horizontal saccades, but Cz and Pz were the only two electrodes not significantly correlated with gaze, even in the uncorrected data. Because Pz contained little or no corneoretinal artifact in the first place, effects of correction were minimal at this electrode.

While MSEC removed most of the corneoretinal artifact, it only partially removed the brief muscle spike potential (Keren, Yuval-Greenberg, & Deouell, 2010) at saccade onset. Nevertheless, MSEC attenuated the spike potential because its topography overlaps with that of corneoretinal artifacts (see also the Methodological Discussion).

Sentence-Onset ERP

Figure 1E shows the artifact-corrected ERP time-locked to the screen onset of the sentence. Sentence onset elicited a large ERP, whose dominant component was a temporally extended P300 with a characteristic centroparietal-positive scalp distribution. To avoid a carryover of this stimulus-evoked ERP into the fixation-locked segments, we analyzed only fixations that occurred later than 700 ms after sentence onset, when this ERP had returned to baseline. Still, on average, the remaining low-predictability words occurred at earlier sentence positions and were fixated sooner after sentence onset than high-predictability words (see Table 1). To ensure that the EEG background activity was not different at the time the target was fixated, we calculated the mean amplitude in the 100-ms *prefixation* baseline relative to a 100-ms *prestimulus* baseline before sentence onset. An analysis of variance of these "absolute" prefixation baseline amplitudes as dependent variable and of predictability level and electrode as factors yielded a nonsignificant result, $F(64, 1856) < 1$. This suggests that FRPs occurred against a similar baseline in all predictability conditions.

Figure 1E shows that between 700 ms and the mean reading duration of 2,490 ms, the sentence-locked ERP was characterized by a slow positive shift at frontal and central electrodes and a relative negativity at occipital sites. This sentence-level ERP during normal reading has not been described before and may be theoretically interesting as a phenomenon on its own. However, the present data did not allow us to determine whether these late ERP fluctuations reflect sentence-level processing demands (e.g., working memory load) or merely the superposition of many individual FRPs over the course of reading.

Fixation-Related Potential in Natural Vision

Figure 2 summarizes the features of the FRP after artifact correction. The postfixation waveshape was dominated by the

[4] To ensure that predictability accounted for variance in EM measures under concurrent control of other word and sentence properties, we specified two control models with either FFD or GD as the dependent variable and the six variables from Model 2 (pred, freq, Pred × Freq, length, constraint, and word position; see Table 3) as predictors. Predictability was a significant predictor of both FFD ($t = -4.5$) and GD ($t = -6.0$). Other significant predictors of FFD were constraint (−3.7), and the Pred × Freq (2.8) interaction. Other significant predictors for GD were word length (13.6), word position (3.2), and Pred × Freq (4.7).

visually evoked lambda response (Kazai & Yagi, 2003), which peaked after 104 ms (SE = 1.3 ms). It was largest at lateral-occipital electrodes over right visual cortex (amplitude at PO10: M = 4.4 μV, SE = 0.3 μV), but also influenced the waveshape at frontal electrodes with a reversed polarity (cf. also Figure 3A). The lambda response is considered primarily a visual response (Riemslag, van der Heijde, & van Dongen, 1987; Thickbroom, Knezevic, Carroll, & Mastaglia, 1991) that is most likely generated in striate or early extrastriate cortex (Dimigen, Valsecchi, Sommer, & Kliegl, 2009; Kazai & Yagi, 2003). Because the average interval between any two reading fixations was only 233 ms, the waveshape of the FRP was characterized by overlapping lambda responses from preceding and subsequent fixations. As shown in Figure 2B, a second occipital peak after 280 ms reflected the summation of a potential evoked by the current fixation n and the lambda response elicited by fixation $n + 1$. Due to normal variance in fixation duration, overlapping contributions from adjacent fixations are time jittered. The FRP in natural vision therefore resembles a damped oscillation with an increasingly jittered occipital peak about every 250 ms.

The prefixation waveshape was influenced by correlates of oculomotor preparation and execution, in particular the presaccadic spike potential. The spike potential is a sharp, biphasic spike at saccade onset that is believed to reflect summated electric activity of the oculomotor nerves or muscles. It is best seen in SRPs (cf. Figure 2C and Figure 4A), and smeared in FRPs due to variance in saccade duration. The spike potential showed the typical topography, which is reversed relative to the corneoretinal artifact: a frontal negativity, shifted ipsilateral to saccade direction, and a parietal positivity, shifted contralateral to saccade direction. For large reading saccades (>3°), there was also some indication of an earlier posterior positivity that culminated into the spike potential (Figure 2C). This may be the *presaccadic positivity* (also called antecedent potential) a slow, ramp-like potential found prior to voluntary saccades and believed to reflect saccade preparation in cortical structures (Everling et al., 1997; Richards, 2003).

Figure 2C shows comparisons of FRPs and SRPs for different saccade sizes. Both the spike potential and the visual lambda

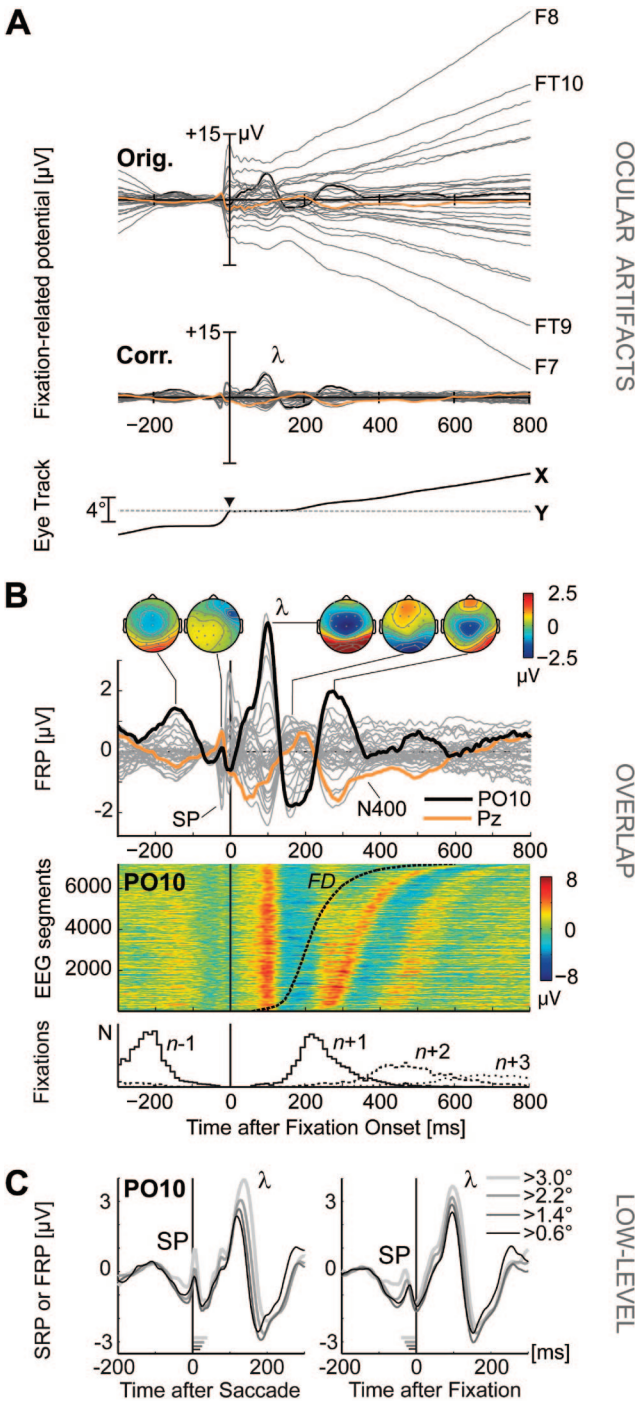

Figure 2 (opposite). The fixation-related potentials (FRP) in multisaccadic vision. Panel A. Grand average FRP for 7,113 target fixations before and after corneoretinal artifact correction. Time 0 marks fixation onset. Signals at all electroencephalographic (EEG) electrodes are shown superimposed. Prior to multiple-source eye correction (MSEC; figure labeled "Orig."), frontal and temporal channels were strongly distorted. After MSEC (figure labeled "Corr."), artifacts were largely eliminated, and genuine brain activity was visible. Midline electrode Pz (orange line) was unaffected by corneoretinal artifacts, even before correction. Panel B–Top image: Properties of the grand average FRP. Maps depict scalp topographies at five component peaks: (a) the visually evoked lambda response evoked by previous fixation $n - 1$, (b) the myogenic spike potential (SP) at saccade onset, (c) the lambda response evoked by current fixation n, which peaked after 104 ms at parieto-occipital electrodes PO9 and PO10, (d) the equivalent of the N170 component, and finally (e) the N400 component, overlapped by the lambda response from fixation $n + 1$. Panel B–Middle image: Each horizontal line in the erpimage represents one of the 7,113 EEG segments that entered the average shown above. Amplitude is coded as color. Segments were sorted by fixation duration (FD) and then smoothed vertically with a moving average across 50 adjacent segments. Sorting shows how the second positive peak at around 280 ms is partially explained by the lambda response from fixation $n + 1$. Because contributions from successive fixations are increasingly time jittered, the overall waveshape of the FRP resembles a damped oscillation. Panel B–Bottom image: Histogram of onset latencies of preceding and subsequent fixations. Panel C. Effect of saccade size on the morphology of saccade-related potentials (SRPs) and FRPs. The amplitude of the spike potential and the lambda response increased with saccade size.

Table 2
Correlations Between Electrodes and Horizontal Eye Track

Hemisphere/electrode	Original	Corrected
Midline		
Fz	**.07**	**.05**
Cz	−.02	.02
Pz	.01	.03
Left		
hEOG	**−.97**	**−.04**
T7	**−.63**	**−.05**
C3	**−.32**	.00
P3	**−.17**	.00
O1	**−.06**	.01
Right		
hEOG	**.97**	−.02
T8	**.57**	.00
C4	**.24**	.01
P4	**.14**	**.03**
O2	**.07**	**.04**

Note. Shown are mean correlations for selected electrodes before and after multiple-source eye correction. Correlations that differ significantly from zero are printed in bold. hEOG = horizontal electro-oculogram electrode.

response increased with increasing saccade size (see also the Methodological Discussion).

Predictability Effects in FRPs

Figure 3A shows that predictability clearly modulated the FRP evoked after the initial fixation on the word: Words with little contextual support elicited more negative voltages at centroparietal scalp sites. The presence of an N400 effect was confirmed by a Predictability × Electrode interaction, $F(64, 1856) = 4.84$, $p < .001$, $\varepsilon = .13$, $\eta_p^2 = .14$, in the 300- to 500-ms time window (effects across the whole scalp are only meaningful in interaction with electrode, because the average reference sets the mean of all electrodes to zero). Pairwise comparisons between predictability levels showed that the interaction with electrode was significant for the contrast low versus high, $F(32, 928)=4.82$, $p < .001$, marginally significant for low versus medium ($p = .07$), and not significant ($p = .19$) for medium versus high-predictability words. All three comparisons were significant (at $p < .01$) when the main effect of predictability was tested only at electrode Pz.

The centroparietal distribution of the N400 with a maximum over Pz (Figure 3C) resembled that observed in many SVP experiments (Kutas et al., 2006). However, while effects in SVP studies are often shifted slightly toward the right hemisphere, no evidence for lateralization was found in normal reading; effect amplitude did not differ between the left and the right hemispheres ($p = .49$).

For comparison with fixation durations, a discrete time point for the onset and peak of the N400 effect was determined in the difference wave at Pz (see Figure 4B). At Pz, a sustained N400 effect began 248 ms after fixation onset and peaked 384 ms after fixation onset with an effect amplitude of 1.53 μV. It is interesting that much weaker N400-like central negativities, qualitatively resembling the topography between 300 and 500 ms, could also be seen in earlier intervals (in particular between about 120 and 160 ms; see Fig. 3C) but did not survive correction for multiple comparisons. Nevertheless, they indicate that the semantic processing underlying the N400 may begin earlier—possibly only on a subset of fixations—than suggested by our strict onset criterion. No predictability effects were observed prior to the onset of the first fixation (i.e., there was no parafovea-on-fovea effect). This was also the case when the baseline interval was moved further away from fixation onset.

Finally, N400 onset was also determined relative to the offset of the first fixation by time-locking backward to the following saccade (SRP). In this analysis, a sustained N400 effect began 20 ms after the end of the first fixation. Additionally, an only temporary significant effect was observed in an early interval from −228 ms to −192 ms before fixation offset.

EM–EEG Relationship

A unique feature of the data set was that it offered us the possibility of comparing FRP effects to corresponding modulations in EM behavior. We explored this relationship in four analyses: First, we investigated at what time FRP effects arose relative to those on behavior. Figure 4 shows the predictability effect at Pz, relative to the beginning and end of the first fixation. When the N400 effect peaked in FRPs (384 ms), readers had already terminated the initial fixation on the target word (fixation n) in 96% of the cases. Instead, as Figure 4D shows, readers were typically already engaged in fixation $n + 1$ (75%) or $n + 2$ (19%; saccade intervals were assigned to the following fixation in this analysis). On the level of words, we found that in only 25% of the cases, readers were still looking at the target (word w) at the latency of the N400 peak. These were mostly cases where the word was refixated. Instead, readers had typically moved on to the next words $w + 1$ (45%) or $w + 2$ (20%). A somewhat different picture emerged when N400 onset latency was considered: In 30% of the cases, the statistical onset latency of the N400 effect (248 ms) fell into the first fixation on the word. In many more cases (67%), it fell only into the following fixation $n + 1$. However, because fixation $n + 1$ was sometimes a refixation (in 24% of the cases), readers were still looking at the target word in about half (47%) of the cases at the statistical N400 onset latency.

Second, we compared EM and FRP effects over five levels of predictability. As Figure 5A shows, both N400 amplitude and GD (as well as refixation probability, which is one aspect of GD) were monotonic functions of predictability. Only for FFD was there a discontinuity in the higher predictability range. While N400 amplitude was an approximately linear function of logit-scaled predictability, Figure 5A indicates that behavioral measures, in particular FFD, differentiated better in the low- than in the high-predictability range.

Third, we investigated whether subjects with large predictability effects in FRPs also exhibit large behavioral effects. Figure 5B shows that of 30 subjects, 25 showed a predictability effect in the expected direction in both measures, supporting the reliability of the coregistration data. However, when these difference scores were correlated across subjects; N400 amplitude correlated neither with FFD ($r = -.07$, $p = .70$) nor GD ($r = .15$, $p = .42$).

Fourth, in separate linear mixed models, we regressed FRP amplitude after each individual fixation in the N400 time window on the two EM measures, FFD and GD. Gaze duration was found to be a strong and significant (i.e., $|t| > 2$) predictor of N400 amplitude ($b = -0.62$, $SE = 0.17$, $t = -3.7$) whereas FFD showed only a numerical

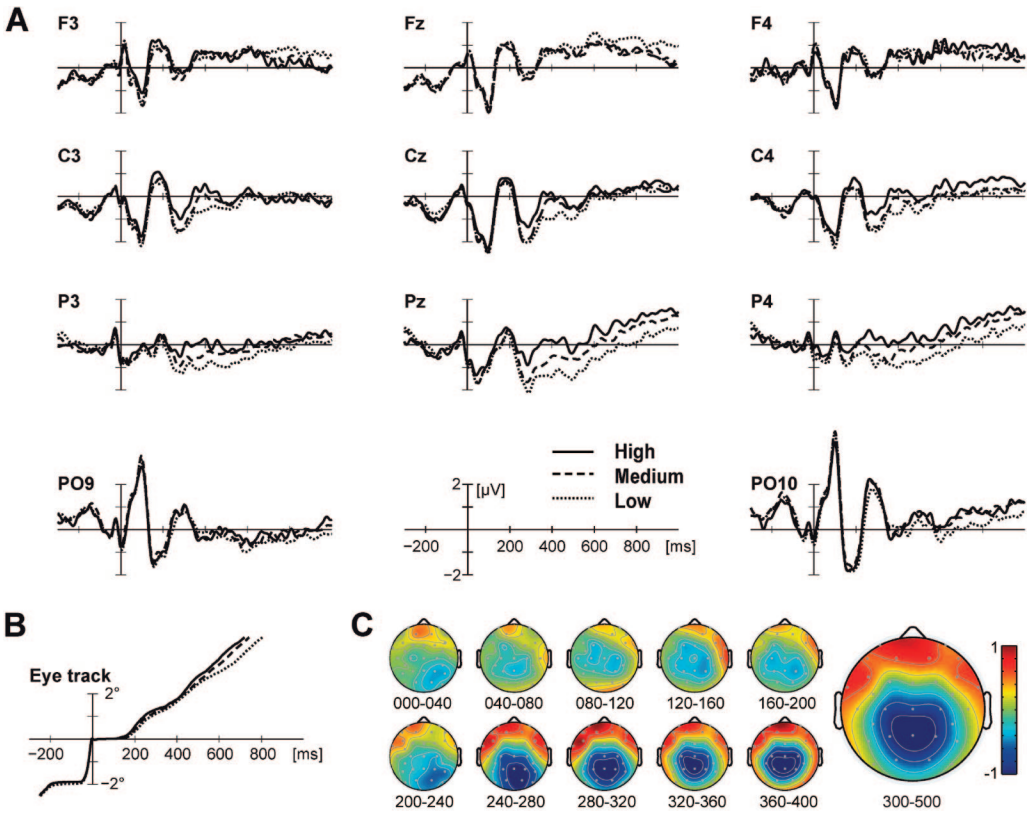

Figure 3. Predictability effect in fixation-related potential (FRP). Panel A. The grand-average FRP, time-locked to first fixation on the target word (Time 0) shows a graded effect of word predictability that is largest at centroparietal electrode Pz. Panel B. Mean horizontal component of the eye track. Panel C. Scalp distributions of the predictability effect (low predictability − high predictability) are shown for successive 40-ms windows after fixation onset and for the traditional N400 window (300–500 ms).

trend in the expected direction ($b = -0.37$, $SE = 0.21$, $t = -1.73$; see Model 1 in Table 3). In a second step, we included lexical and sentential properties into the model. Both N400 amplitude and the EM measures are known to relate to the (logit of) word predictability and the (logarithm of) word frequency of the fixated words, as well as to the interaction between both variables (Dambacher et al., 2006). Predictability, frequency, and their multiplicative interaction were therefore included as predictors in the model. Additionally, we included the covariates word length, contextual constraint, and word position.

Results are shown in Table 3 (Model 2). Of the newly included predictors, predictability, frequency, word position, and the Predictability × Frequency interaction were highly significant, while constraint and length had no effect on N400. Predictability was therefore a significant predictor of N400 amplitude under statistical control of other variables correlated with predictability in the sentence material. GD remained a significant predictor in Model 2. Thus, there is shared variance between GD and N400 amplitude that is not covered by the word or sentence properties included in our model.[5]

Discussion

Psycholinguistic Discussion

In the present experiment, subjects read sentences from left to right while eye movements and EEG were recorded. By time-locking the EEG to fixations on words that were expected to various degrees in the sentence context, we could replicate robust effects of word predictability on behavior and concurrent brain activity. The demonstration of predictability effects in an ordinary reading situation with heterogeneous sentence materials and in- and out-going saccades suggests that EEG recordings in natural vision are feasible in principal. In this first part of the Discussion, we comment on the psycholinguistic aspects of our results. In a second, methodologically oriented part of the Discussion, we review the technical challenges that emerged in the present experiment.

N400 effects. Decreasing word predictability increased a parietal negative-going component in the FRP that reached a maximum at 384 ms. It should be noted that this effect was observed despite the limited range of cloze probabilities in the Potsdam Corpus, which contains normal sentences and mostly

[5] The same inferences resulted from likelihood-ratio tests of these models. Adding sentence and word properties as predictors significantly improved the model fit both with FFD, $\chi^2(6) = 58.0$, $p < .0001$, and with GD as first predictor in the model, $\chi^2(6) = 101.1$, $p < .0001$. Conversely, dropping FFD from the model did not significantly decrease the fit, $\chi^2(1) = 2.1$, $p = .143$, whereas dropping GD did decrease $\chi^2(1) = 24.6$, $p < .0001$.

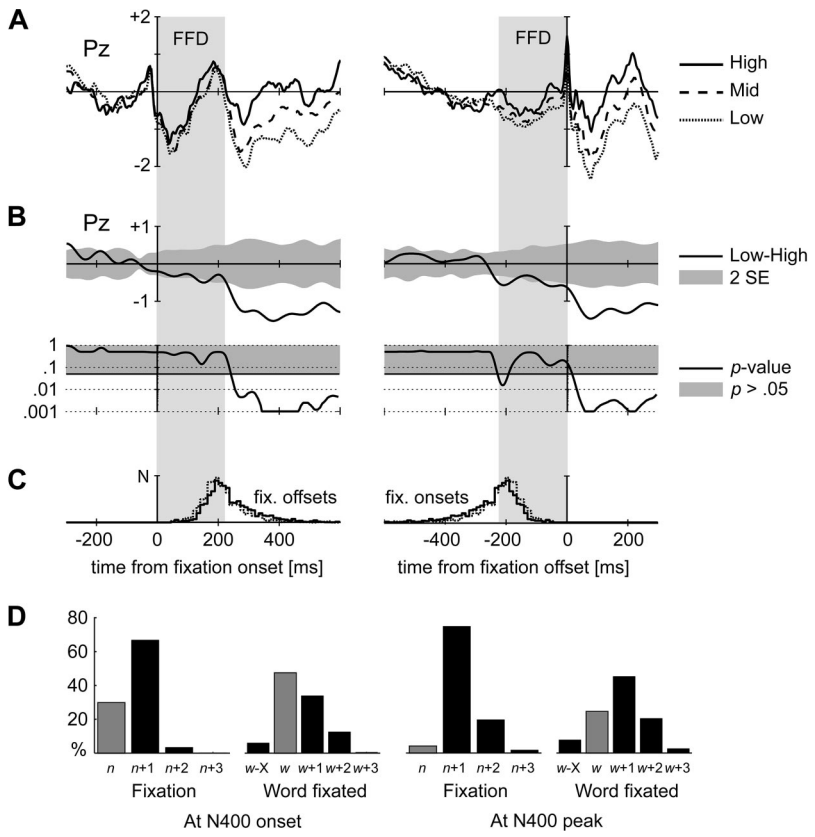

Figure 4. Relative timing of predictability effects. Panel A. Grand-mean fixation-related potentials (FRP) at Pz aligned to the onset (FRP, left side) and offset (saccade-related potentials [SRP], right side) of the first fixation on the word. Gray boxes indicate mean first-fixation duration (FFD). Panel B–Upper panel: Difference wave between low- and high-predictability words, which shows the effect of word predictability devoid of common overlapping activities. Shading indicates the 95% confidence interval without correction for multiple testing. Panel B–Lower panel: Corresponding p values from the permutation test. Significant effects ($p < .05$) are indicated by points outside the gray shaded area. In the FRP, an effect of word predictability was observed starting 248 ms after fixation onset, which peaked at 384 ms in the grand average. Panel C: Distribution of fixation offset and onset latencies relative to the time-locking point. Panel D: Gaze position at onset (248 ms) and peak (384 ms) of the N400 predictability effect. The notation $w - X$ indicates that the reader regressed to an earlier word in the sentence.

words of low and moderate cloze probability. Because of its time course, polarity, scalp distribution, and sensitivity to word predictability, we take this component to reflect the N400. The fact that the topography of the effect was very similar to the N400 effect commonly observed during word-by-word presentation is reassuring evidence for the ecological validity of ERP data collected in traditional SVP paradigms.

Although in the present study, we were mainly concerned with demonstrating the feasibility of this approach and did not primarily aim at covering new psycholinguistic ground—clearly a topic for future research—it also provided one indication that coregistration may yield somewhat different results than SVP. This concerns the N400 time course: The present N400 appeared to begin earlier than what is commonly reported from SVP. In visual word presentation, N400 effects typically arise at around 200–250 ms (Kutas et al., 2006). While our conservative onset criterion yielded a latency in this range (248 ms), we also observed much weaker N400-like effect topographies in earlier intervals after fixation onset, which did not survive correction for multiple comparisons. As we did not specify mixed models for these early intervals, we cannot exclude the possibility that this pattern was due to other variables correlated with predictability (e.g. constraint). However, such early deviations were not observed by Dambacher et al. (2006) who presented the same sentences word-by-word and tested for early predictability effects in the P200 time window. In contrast, in an unpublished follow-up experiment in which an experimental manipulation of predictability (Dimigen, Sommer, Dambacher, & Kliegl, 2008) was used, we could replicate the observation of a comparatively early N400 onset in natural reading. An early onset of N400 effects under natural reading conditions was also reported in a recent study by Kretzschmar, Bornkessel-Schlesewsky, & Schlesewsky (2009). These authors co-registered eye movements and EEG from centroparietal electrodes while subjects read sentences that contained an entirely unpredictable target word. In a condition where the unpredict-

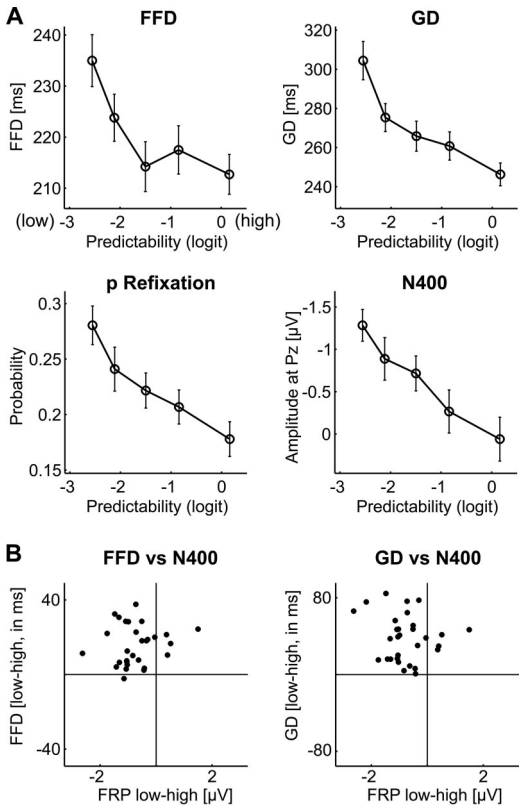

Figure 5. Comparison of eye-movement (EM) and electroencephalographic (EEG) effects. Panel A. EM behavior (first-fixation duration [FFD], gaze duration [GD], and refixations probability) and N400 amplitude (at Pz between 300 and 500 ms) is plotted across five levels of word predictability (logit scaled). Mean cloze probabilities in the five bins were .00, .01, .05, .16, and .58. Note that negative voltages are plotted upwards in this panel only. Compared with N400 amplitude, FFDs showed stronger modulations in the low- than in the high-predictability range. Panel B. Size of the predictability effect (low- predictability words minus high-predictability words) in fixation times and FRPs. Each point indicates the data of one subject. Of 30 subjects, 25 showed effects in the expected direction in both measures: FRPs were more negative, and fixation times were prolonged for low- predictability words. The size of EM and EEG effects did not correlate across subjects.

able target word was also semantically unrelated to the most expected word, an N400 effect arose soon after the first fixation on the target word (i.e., between 250 and 400 ms after the onset of the last pretarget fixation, which lasted 186 ms on average).

Although caution is necessary in the absence of a within-subject comparison to SVP, these observations indicate that the time line of word recognition in normal reading can differ from that commonly found in SVP experiments. An earlier N400 onset in normal reading is very plausible because of the parafoveal preview obtained during the previous fixation, a benefit absent in SVP. This could also explain the early N400-like deviations in the present study. But there are also other reasons why processing speed could differ in normal reading. For example, the fact that saccades are self-initiated should reduce temporal uncertainty about the arrival of new visual input. In the absence of parafoveal preview, Marton et al. (1985) still observed faster word-categorization after a 24° saccade, compared with foveal presentation. Similarly, Dimigen, Schild, Hohlfeld, Berg, & Sommer (2011) compared manual reaction times to small symbols presented either at fixation or at 10° eccentricity. Although parafoveal preview was unavailable in the 10° condition, postsaccadic reaction time (RT; measured from saccade offset to reaction) was 30–70 ms shorter than RT to the same stimulus presented at fixation. Both results indicate that the time to prepare and execute a saccade can act as a foreperiod, which allows subjects to optimize temporal preparation (e.g., Niemi & Näätänen, 1981) and thus enhances postsaccadic processing.

EM–EEG relationship. To our knowledge, the present study is the first to offer a detailed comparison of oculomotor and electrophysiological effects of a fixated item in free vision. We carried out several exploratory analyses to investigate the EM–FRP relationship. As expected, both EM and FRP measures were sensitive to word predictability, suggesting that they are driven by common underlying processes. While GD and especially refixation probability aligned well with N400 amplitude across five levels of logit predictability (Figure 5A), a different function was observed for FFD. In an earlier study, Dambacher and Kliegl (2007) compared fixation times and ERPs for the same words but measured them in different groups of subjects; EMs originated from natural

Table 3
Regressions of N400 Amplitude on Eye Movement Behavior and Word or Sentence Properties

Variable	*b*	*SE*	*t*
	First-fixation duration as predictor		
Model 1			
Intercept	1.12	1.15	0.98
log (FFD)	−0.37	0.21	−1.73
Model 2			
Intercept	3.03	1.30	2.34
log (FFD)	−0.31	0.21	−1.47
pred	1.00	0.30	3.34
freq	−0.35	0.13	−2.77
length	0.02	0.03	0.69
constraint	−0.01	0.01	−1.63
pos	0.29	0.05	6.09
pred × freq	−0.15	0.06	−2.48
	Gaze duration as predictor		
Model 1			
Intercept	2.63	0.94	2.81
log (GD)	−0.62	0.17	−3.71
Model 2			
Intercept	4.95	1.20	4.11
log (GD)	−0.67	0.18	−3.71
pred	0.91	0.33	2.73
freq	−0.34	0.14	−2.42
length	0.05	0.04	1.38
constraint	−0.01	0.01	−1.59
pos	0.30	0.05	5.72
pred × freq	−0.13	0.07	−2.05

Note. No. of observations: 7,113; no. of subjects: 30, no. of unique words: 499. Dependent variable was always N400 amplitude (mean amplitude 300–500 ms at Pz). Predictors: first-fixation duration (FFD), gaze duration (GD), logit of predictability (pred), log of frequency (freq), word length (length), number of words predicted (constraint), word position (pos), and interaction of pred and freq (pred × freq).

reading whereas ERPs were collected in SVP. Dambacher and Kliegl reported remarkably similar functions for the duration of single fixations and N400 amplitude across five levels of log frequency and four levels of logit predictability (their Figure 2). The profiles included even a disordinal trend with slightly longer single-fixation durations and slightly larger N400 amplitude for words of medium log frequency. The study of Dambacher and Kliegl differs in several details from the present one (e.g., N400 from SVP rather than simultaneous recordings; aggregation over words, not over identical fixations) so it is difficult to speculate about the reasons for this difference in the exact relationship between fixation durations and N400 amplitude.

A model of N400 amplitude at the level of individual fixations (Table 3) provided no evidence for common variance between FFD and N400 amplitude that was not explained by the properties of the word or preceding sentence fragment. However, such covariance was observed between GD and N400 amplitude. Of course, this relationship could be mediated by other lexical variables not included in our model. The alternative explanation is that N400 amplitude and GD are directly related to each other, for example due to moment-to-moment fluctuations in the efficiency of word processing, which would affect both measures.

Since both EM and FRP measures are sensitive to word predictability, one might expect readers who show strong N400 effects to also show strong behavioral effects, and vice versa. For example, good readers should make better use of sentence context, and this may show up as larger predictability effects in EMs and FRPs. This was not the case: We found no evidence indicating that the size of the behavioral effects correlated with N400 effect amplitude across subjects (Figure 5B). This result was surprising because Dambacher and Kliegl (2007) established such a correlation *across words* by using data from separate experiments. The lack of a correlation *across subjects* is most likely caused by the notoriously unreliable difference scores entering the correlations, which may represent too weak a signal to overcome individual differences in brain anatomy (e.g., cortical folding). Such anatomical differences among subjects may influence the strength at which activity of an additional neural generator propagates to the scalp and could be a stronger source of N400 amplitude variation than differences in underlying brain activation, possibly concealing any existing relationship.

A final set of comparisons concerned the relative timing of predictability effects. Researchers have argued on the basis of SVP data that there is a discrepancy between the latency of the N400—the primary and so far the only robust index of semantic processing in psycholinguistic ERP research—and the fixation durations measured in eye-tracking studies. The N400 typically peaks at around 400 ms in ERP studies, a time when the eyes have already left the critical word in natural reading (Sereno & Rayner, 2003). As Rayner and Clifton (2009) have pointed out, this conundrum is difficult to explain: How can the eyes react faster than the brain? We were able to address this time lag question within the same data set. Despite a relatively early N400 peak (386 ms) in normal reading, the pattern was still the same as in SVP: While predictability clearly influenced the duration of the first fixation on the target word, this fixation had almost always ended when the predictability effect peaked in the FRP (Figure 4D).

One common view on the N400 assumes that it reflects a late, postlexical process of semantic context integration. Likewise, it is a commonly held view in EM research that these processes are reflected in GD, which is seen as a measure of late processing. It is therefore interesting to note that the N400 peak did not fall into the mean GD (278 ms) either. Of course, reading a low predictable word increases not only FFD and GD, but can also prolong later fixations on the following words (*spillover* or *lag effects*; Rayner & Duffy, 1986; Kliegl et al., 2006). Nevertheless, the present data make it hard to conceive the measurable neural effects of predictability as being causal in some way for the behavioral effects, because the bulk of the predictability effects in ERPs only followed those in behavior. This raises questions about the functional interpretation of the N400 peak, whose latency does not seem to correspond to the maximum processing difficulty as reflected in the EM record.

Alternatively, one could consider the onset of the N400, rather than its peak, as the critical event. At the statistical N400 onset latency, readers were still looking at the target in more than half of the cases and were still in the first fixation in 38% of the cases. However, in order for lexicosemantic processing to influence FFD, it must do so before saccadic motor programming enters the nonlabile stage, that is, at an estimated 80 ms before the end of the fixation (Becker, 1991; Findlay & Harris, 1984). The onset of the N400 effect therefore still seems to occur surprisingly late in comparison to the FFD effect. However, in contrast to the FRP analyses, the analyses of the SRP aligned to fixation offset provided some evidence that N400 effects may begin within the first fixation. The temporal contingencies observed here therefore do not rule out completely the possibility that the processes reflected in the N400 onset are also responsible for—or "driving"—the early effects in FFD. An answer to this question requires the design of dedicated experiments with strongly expected or unexpected words. Such experiments will allow very precise measurements of onset latency and possibly indicate that N400 onsets can occur early enough to influence behavior (see also Dimigen et al, 2008; Kretzschmar et al., 2009). They will also clarify whether N400 effects can arise before the direct fixation of a target word due to parafoveal preview, a hypothesis that was not supported by the current results.

Possible applications. Apart from validating traditional ERP findings for more natural reading situations, coregistration can be used to investigate aspects of the reading process that are difficult or impossible to study with SVP. As described in the introduction, one such aspect is the availability of parafoveal information in natural reading. The timing and extent to which upcoming words are preprocessed is still controversial and can be studied in greater detail with FRPs. A promising approach in this context is the combination of simultaneous recordings with gaze-contingent changes of the computer display, as they are often used in eye-tracking studies to manipulate preview (e.g., the boundary technique; Rayner, 1975). The question of whether word meaning can be extracted from parafoveal words is one issue that could be further investigated with this technique. Coregistration is also the only viable approach to study EEG correlates of complex reading behavior. Interesting questions concern the EEG signature before a word is skipped rather than fixated, the FRPs that precede and follow the decision to trigger a regressive saccade (and their relationship to established syntactic ERP components), or the functional localization of individual differences in reading ability

and reading speed. The final section is a review of the relevant technical aspects for conducting such studies.

Methodological Discussion

Researchers who wish to record the EEG during reading or other free-viewing tasks are faced with several technical and data-analytical problems, which are the likely reason why such recordings have rarely been attempted. The four main challenges we identified are (a) the need for precise coregistration of gaze position (b) the correction of corneoretinal and myogenic eye movement artifacts, (c) varying degrees of overlap between successive FRPs as well as between FRPs and background ERPs, and (d) variation of saccade-related cortical potentials according to low-level visuomotor factors. In the following, we will discuss each problem and possible solutions in some detail.

Coregistration of gaze position. A basic requirement for fixation-based averaging is accurate information about the latency and location of each fixation. Traditionally, ERP researchers have used EOG (Oster & Stern, 1980) electrodes near the eyes to control for a steady fixation. Basis of the EOG is an electrical gradient of 0.4–1 mV (Young & Sheena, 1988) between cornea and retina, which can be modeled by an equivalent electric dipole near the eyeball (Berg & Scherg, 1991). Because changes in the orientation of the eyeball change the potential at periocular electrodes, the EOG is well suited as a tool to determine the onset latency of single, large saccades. However, with a spatial accuracy of ± 1.5°–2.0° (Malmivuo & Plonsey, 1995; Young & Sheena, 1988; see Joyce, Gorodnitsky, King, & Kutas, 2002, for a method to optimize EOG accuracy), it does not provide absolute gaze position with the single-letter accuracy required for reading analysis.

Current video-based eye trackers afford spatiotemporal resolutions up to 0.01°/2 kHz, and both table-mounted (Baccino & Manunta, 2005; Bodis-Wollner et al., 2002; Dimigen et al., 2009; Hutzler et al., 2007; Kennett, Van Velzen, Eimer, & Driver, 2007; Kretzschmar et al, 2009; Valsecchi, Dimigen, Sommer, Kliegl, & Turatto, 2009; for MEG, see Herdman & Ryan, 2007) and head-mounted (Graupner et al., 2007; Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008) systems have been used for coregistration. Technical concerns about concurrent eye tracking are (a) pressure artifacts from contact between electrodes and eye tracker, (b) muscle artifacts resulting from head stabilization or unnatural sitting positions, (c) proper synchronization of the data records, and (d) electromagnetic artifacts from an electric device operating close to the EEG sensors. In the present study, these problems were minimized by (a) foam-cushioning forehead electrodes, (b) careful adaptation of the subject's sitting position, and (c) synchronization of EM and EEG records with a shared TTL pulse every few seconds. To double-check proper synchronization, we also use an analog-to-digital converter in the eye tracker that feeds an analog copy of the gaze position as an additional channel into the EEG record. To test for (d) electromagnetic artifacts, we compared the EEG spectrum during steady fixation while the eye tracker was either recording or disconnected from power. We found that eye tracker operation introduced a weak 50-Hz line noise artifact at frontal electrodes near the eye tracker. However, this high-frequency artifact was irrelevant for the present FRP analyses and can be minimized by a notch filter (Yuval-Greenberg et al., 2008) or by a remote eye tracker outside the shielded cabin.

In summary, advances in video-based eye tracking allow the routine recording of high-resolution EMs without obstructing EEG recordings. Moreover, two recent studies suggest that eye tracking can improve EEG data quality even in experiments that require steady fixation, as it allows to identify myogenic (Yuval-Greenberg et al., 2008) and visuocortical (Dimigen et al., 2009) potentials from involuntary microsaccades.

EM artifacts. EM artifacts in the EEG are generated by three mechanisms: rotation of the eyeball's corneoretinal dipole (Berg & Scherg, 1991), relative movements of the eyelid during blinks and upward saccades (Picton et al., 2000), and electrical eye muscle activity at saccade onset, which propagates to the EEG as a spike potential (Thickbroom & Mastaglia, 1986). In normal vision, the strongest artifacts are corneoretinal. The changes in the corneoretinal potential, which provide the basis for the EOG, also propagate to the EEG electrodes, although they attenuate with increasing distance to the eyes (Picton et al., 2000). The horizontal saccades that are dominant in reading produce the largest distortions at lateral-frontal channels and smallest distortions at posterior electrodes along the sagittal midline. While excluding of contaminated trials is not an option during natural vision, there are simple ways to minimize EEG contamination despite saccades. In early studies, saccades were recorded only from occipital midline sites (e.g., Gaarder, Krauskopf, Graf, Kropfl, & Armington, 1964), or equal numbers of left- and right-going saccades were used, based on the assumption that both artifacts cancel out during averaging (e.g., Kurtzberg & Vaughan, 1982; Marton et al., 1985). Others have avoided the problem by limiting data analyses to the short interval of EEG before the first saccade (Baccino & Manunta, 2005; Simola, Holmqvist, & Lindgren, 2009) or after the terminal saccade in a sequence of saccades (e.g., Hutzler et al., 2007; Marton & Szirtes, 1988a, 1988b). Obviously, these approaches place severe limitations on the time segment, electrode site, and study design.

A large variety of algorithms have been proposed to correct mathematically for ocular artifacts (for reviews, see Brunia et al., 1989; Croft & Barry, 2000; Delorme, Sejnowski, & Makeig, 2007; Gratton, 1998; Ille et al., 2002). It is interesting that the application of these algorithms has been largely restricted to correcting blink artifacts and accidental saccades in experiments that require fixation. Here, we applied surrogate MSEC (Berg & Scherg, 1994; Ille et al., 2002) to correct the heavily contaminated data. Although the collection of clean calibration EMs from each subject is time-consuming, the method was chosen for four reasons. First, MSEC can reduce the elimination of genuine brain activity compared with traditional regression-based methods (Berg & Scherg, 1994), because brain activity is modeled. Second, MSEC can be applied to continuous rather than averaged data, which supports a flexible resegmentation of the corrected EEG to different time-locking points (e.g., the onset of fixation $n - 1$). Third, the method is not based on assumptions on the spatial or temporal orthogonality or independence of artifact and brain activities (the application of PCA during preprocessing is optional). Fourth, it does not require subjective choices from the experimenter apart from the one-time selection of a surrogate brain model.

Inspection of the continuous EEG, analysis of the averaged FRP, and analysis of residual correlations with the eye tracker converged to suggest that it was feasible to compensate for most

of the artifact. While artifacts from the incoming saccade were completely abolished, residual artifact remained toward the very end of the fixation-locked segment, when several saccade artifacts had summated. This became apparent as small but significant correlations between about half of the corrected EEG channels and the eye track. Also, MSEC did not fully remove the spike potential (cf. Figure 4A, right panel), because its topography was not predefined as an artifact source. In summary, residual artifacts were small, and correction quality was sufficient for the analyses that were being conducted. The use of more electrodes and a more realistic surrogate model may further improve MSEC correction.

Utility of eye tracking to improve ocular correction. Although correction worked well for the present purposes, other algorithms may have performed equally well or better. However, a fundamental problem with ocular correction methods is a lack of objective external criteria (Brunia et al., 1989) to compare and evaluate their performance on experimental data (for simulated data, see Delorme et al., 2007; Klemm, Haueisen, & Ivanova, 2009; Wallstrom, Kass, Miller, Cohn, & Fox, 2004). Choice of an appropriate correction method is particularly important for natural vision recordings. We therefore propose that concurrent eye tracking is useful to evaluate, compare, and improve correction methods:

First, unlike the EOG, the eye track provides a measure of eye position that is electrically independent of the EEG. Correlations between eye track and EEG after correction are therefore likely to result from residual corneoretinal or myogenic artifacts (or, less likely, from saccade-related brain activity occurring in synchrony with the saccade). The degree to which the EEG depends on eye position after correction (exemplified here by a linear correlation) can help the researcher to evaluate correction quality across studies and algorithms.

Second, eye tracking may provide information about whether an algorithm overcorrected the data and distorted genuine brain activity. Provided that an experiment contains at least some intervals with steady fixation, high-resolution eye tracking allows the researcher to select EEG intervals objectively free of any ocular artifact (eye blinks, saccades, and microsaccades). These intervals should not be altered by ocular correction and therefore provide a test case to quantify the distortion of brain signals introduced by the method.

Third, eye tracking may directly improve correction. In correction methods based on PCA or independent component analysis (ICA), the EEG is decomposed into multiple uncorrelated (e.g., Lagerlund, Sharbrough, & Busacker, 1997) or statistically independent (Delorme et al., 2007) signal components, and then correction is then performed via removal of components classified as artifact. This classification is typically performed manually and is based on criteria such as the component's scalp distribution and spectrum (Li, Ma, Lu, & Li, 2006; Okada, Jung, & Kobayashi, 2007; Rong & Contreras-Vidal, 2006), or correlation with the EOG (Joyce, Gorodnitsky, & Kutas, 2004; Wallstrom et al., 2004). However, classification can be ambiguous, especially when many components are produced, and many studies do not report selection criteria (Fatourechi, Bashashati, Ward, & Birch, 2007). Considering the relation between the components' time series and gaze position (i.e., eye tracker-informed ICA) should greatly improve the reliability of component selection.

Finally, we propose that ocular correction is entirely unnecessary for certain research questions. Reading studies with SVP have provided some evidence of early ERP correlates of lexicosemantic processing within 200 ms after stimulus onset (Dambacher et al., 2009; Hauk & Pulvermüller, 2004; Penolazzi, Hauk, & Pulvermüller, 2007; Sereno, Brewer, & O'Donnell, 2003; for review, see Pulvermüller, Shtyrov, & Hauk, 2009). An interesting question will be whether similar effects, within the duration of the current fixation, can be reliably established in FRPs and whether they predict the current fixation duration or the upcoming saccade target. Each fixation, however, is by definition a short interval of EEG that is free of EMs and that can be analyzed without prior correction. This approach requires a sufficient number of fixations with a minimum duration (e.g., 200 ms) after shorter fixations have been excluded. Second, because the artifact of the incoming saccade precedes fixation onset, a prefixation baseline—as in the present study—is not feasible. Instead, the baseline could be placed in the first few milliseconds of the fixation itself (provided that effects are not already present in the baseline due to parafoveal preview). Finally, direct current amplifiers should be used to prevent a spillover of the prefixation artifact into the postfixation segment due to the time ~~constraint~~ of the amplifier (Yagi, Kazai, & Takeda, 2000).

In summary, results suggest that corneoretinal artifacts are not a principal—and not the most serious—obstacle for EEG recordings during natural vision: Correction was sufficient for the present analyses and can likely be improved further. Eye tracking provides new criteria to validate and optimize correction while certain analyses do not require artifact correction at all.

Differential overlap. Serial presentation allows full control over the stimulus that is presented to the visual system at any time. In normal vision, on the other hand, the experimenter has little control over the spatiotemporal pattern of fixations, in particular over the latency and duration at which a subject chooses to fixate a target item. This leads to two problems of EEG overlap:

The first problem is the varying degree of temporal overlap between the potentials evoked by a target fixation and existing EEG *background activity* that is related, for example, to the stimulus onset. In the present experiment, sentence onset evoked a P300, which returned to baseline level only after about 700 ms. If target items in two conditions are fixated at systematically different latencies after stimulus onset—for example, salient versus nonsalient parts of a picture during scene perception—fixation-related potentials occur against a different background. This will distort the FRP waveshape and topography between conditions.[6] It is therefore important to ensure that target fixations do not differ in terms of overlapping background activity. In the current case, we accomplished this by excluding early fixations and by ensuring

[6] Consider this example: In normal sentences, low-predictability words tend to occur at earlier word positions and will be fixated sooner after sentence onset. As a consequence, the prefixation baseline interval for these fixations will overlap more often with the sentence-evoked P300. Via the process of baseline correction, which involves the subtraction of the baseline voltage from each channel, the sentence-onset P300 (a centroparietal positivity) will carry over into the fixation-locked segment with a reversed polarity (as a centroparietal negativity), thereby creating a bogus, N400-like effect for low-predictability words.

that the prefixation baseline activity did not differ between predictability levels.

The second problem is temporal overlap between the potentials elicited by successive reading fixations. With interfixation intervals of around 250 ms, reading proceeds considerably faster than typical SVP paradigms, leading to massive overlap between the potentials evoked by subsequent fixations. This means that late, endogenous components from the previous fixation $n - 1$ overlap early, exogenous components from fixation n. Likewise, late components from n will overlap early components from $n + 1$ (cf. Figure 2B). Temporal overlap and, hence, summation of successive fixation-related responses over the duration of the trial was also the likely reason for the highly significant effect of word position on EEG amplitude in the N400 window (see Table 3); an effect that did not reach significance when the same sentences were presented in SVP (Dambacher et al, 2006).

Due to variance in fixation duration, the overlapping potentials will be latency jittered relative to the current fixation and therefore low-pass filtered (Woldorff, 1993). While overlap is also encountered in fast SVP experiments (in which words have been presented at up to 10 Hz; Kutas, 1987), overlap is problematic if it differs systematically between conditions. This is clearly the case in reading and many other viewing tasks, where readers modulate fixation time according to processing difficulty. Because any fixation duration effect translates into a change in the phase of the overlapping potentials, EEG effects that reflect stimulus processing can be easily confused with changes that merely result from different degrees of overlap. Because differential overlap occurs with any fast and self-paced stimulation, the problem also applies to self-paced SVP (Ditman et al., 2007).

The extent to which overlap is a problem depends on the exact paradigm and the size of the behavioral effect. For example, the last fixation in a sequence is only influenced by overlap from previous fixations but not from subsequent fixations. In addition, in the absence of behavioral effects on fixation $n - 1$ (e.g., a parafovea-on-fovea effect), different overlap will influence the FRP only after the outgoing saccade is executed (this is not entirely true, because correlates of saccade preparation like the presaccadic positivity may precede the outgoing saccade).

We are not aware of a simple solution to the problem of differential overlap in FRPs, and this problem has been ignored in previous SRP/FRP studies. Several deconvolution methods have been proposed to separate overlapping potentials in ERP experiments with fast stimulation and variable inter-stimulus intervals (Delgado & Özdamar, 2004; Hansen, 1983; Jewett et al., 2004; Wang, Özdamar, Bohorquez, Shen, & Cheour, 2006; Woldorff, 1993). For example, in the time-domain technique ADJAR (level 2 implementation; Woldorff, 1993), overlapping waveforms are iteratively deconvolved on the basis of information about the temporal distribution of the ERP-eliciting events. In natural vision, this information is provided by the eye tracker. However, deconvolution methods typically rely on high signal-to-noise ratios (Talsma & Woldorff, 2004), require knowledge about the nonoverlapped template waveform, or are based on the assumption that each successive event evokes an identical response. As discussed in the next section, the last condition, in particular, is unlikely to be met during free vision and it needs to be tested whether deconvolution can be successfully applied to FRPs.

However, the influence of differential overlap can at least be approximated by convolving an estimate of the nonoverlapped FRP with the latency distribution of fixation onsets in each condition (here: low-, medium-, and high-predictability words).[7] As an alternative, the large pool of fixations that is easily obtained during natural vision allows the selection of fixation subsamples from each condition that are matched in terms of fixation duration. Because experimental effects on fixation duration are often small in reading, only a small percentage of fixations must be excluded to equalize the distribution of fixation durations post hoc. Obviously, the resulting fixation samples, matched for fixation duration, present a biased selection, which may exclude the fixations that are theoretically the most interesting. However, this procedure provides a simple test whether FRP effects persist once overlap is controlled.

Low-level influences on saccade-related brain potentials. Saccade- and fixation-related brain potentials are not only modulated by higher cognitive processing demands, but their waveshape is also influenced by visuomotor low-level factors that cannot be controlled during natural scanning behavior. These influences must be carefully delineated from effects that reflect higher level cognitive processing (e.g., semantic processing) of the fixated item. Among these low level influences are (a) the exact visual input at saccade offset and (b) the kinematics of the incoming saccade.

In most ERP studies of higher cognitive processing, great care is invested to match the visual stimuli in terms of low-level properties such as contrast, luminance, and spatial frequency. In contrast, retinal inputs vary considerably across different fixation locations (e.g., the center vs. the edge of a bright screen). Both amplitude and latency of the visually driven lambda response around 100 ms after fixation onset (cf. Figure 2) vary as a function of the luminance and contrast (Gaarder et al., 1964; Kazai & Yagi, 1999, 2005; Marton & Szirtes, 1982) and spatial frequency (Armington & Bloom, 1974; Kazai & Yagi, 2005) of the background. The lambda response therefore resembles the P1 component in ERPs, which is generated in overlapping areas of visual cortex (Kazai & Yagi, 2003) and modulated by the luminance, size, contrast, and frequency content of the field of presentation (Tobimatsu & Celesia, 2006). Visual influences on the lambda response have been observed with stimuli that covered large parts of the visual field (Riemslag et al., 1987). To our knowledge, it has not been investigated whether differences in the foveal input across different fixation locations (e.g., local changes in luminance and contrast when viewing different parts of a scene) cause relevant modulations of the fixation-related EEG. For the sentence stimuli presented in the current study, we assumed that visual field properties would be comparable for different fixation locations on the screen. In addition, we attempted to reduce purely visually driven effects by covering the peripheral visual field with a mask.

A second important low-level factor that influences SRPs and FRPs is saccade size. Saccade amplitude modulates not only the size of the presaccadic muscle spike (Boylan & Ross Doig, 1989)

[7] Such a simulation (not shown here) that we conducted for the present data set with the grand-average FRP (averaged across conditions) as the waveform estimate suggested that effects of differential overlap were small compared with the much larger effect of word predictability.

but also the waveshape of the postsaccadic EEG. The visual lambda response, in particular, comprises both saccade onset- and saccade offset-related responses (Kurtzberg & Vaughan, 1977; Thickbroom et al., 1991). For long saccades (>10°), it dissociates into two subcomponents, a first subcomponent, presumably related to visual changes at saccade onset, and a larger second one, presumably evoked by the inflow of new visual input at saccade offset (Thickbroom et al., 1991). However, as Figure 2C shows, saccade amplitude modulates the lambda response even for the limited range of saccade amplitudes found in reading. Therefore, one needs to be cautious when conditions are compared that differ in terms of incoming saccade amplitude. This is not necessarily a fundamental limitation because reading saccades have fairly constant amplitudes of around 7–9 characters (Rayner, 1998), and eye tracking allows the post hoc exclusion of very short or long saccades. In the present case, condition differences in incoming or outgoing saccade amplitude were too small (<0.1°) to cause relevant changes in FRP waveshape. However, when we specified a mixed model (not shown here) in which we added incoming saccade amplitude as an additional predictor, this predictor explained EEG variance at electrode Pz in all time windows between 40 and 280 ms after fixation onset. We therefore propose to include incoming saccade amplitude as a covariate in FRP analyses. Finally, we are not aware of any study that has tested the influence of saccade direction on FRPs, a question that seems important for applications of the technique to visual search or scene perception.

Summary: feasibility of coregistration. Fixation-related potentials are influenced by corneoretinal artifacts, differential overlap, and visual- and motor-related brain potentials, all of which vary with oculomotor behavior. These indirect influences of EM behavior on the EEG can be easily mistaken for genuine condition differences in the brain's processing of the fixated item. Special care must be taken during the analysis and interpretation of multisaccadic EEG experiments, and researchers should account for how these problems were addressed. In particular, details about oculomotor behavior (fixation durations and saccade amplitudes) on the fixations preceding and following a target fixation should be provided. The aforementioned problems are not exclusive to the analysis of FRPs in the time domain but also apply to EEG analyses in the frequency domain, that is, to saccade- or fixation-related oscillations (SROs or FROs). However, EEG recordings in natural vision also have the major advantage in that many fixations can be collected in a short time and with little strain on the subject. In future studies, these large fixation pools could be used to model and disentangle the influences of artifacts, low-level processing, and higher level processing on the fixation-related EEG.

Conclusions

Visual perception outside the ERP laboratory is fundamentally transsaccadic and involves an active sampling of the environment several times per second. What types of information are obtained on each fixation, and how this information is integrated with the data from previous and subsequent fixations? Despite their advantages, EEG recordings have been largely precluded from natural viewing conditions. We demonstrated here that EEG indices of semantic processing can be obtained in natural reading and compared with EM behavior. With the appropriate consideration of technical and data-analytic issues, concurrent recordings may contribute new answers to long-standing questions.

References

Abe, T., Ogawa, K., Nittono, H., & Hori, T. (2004). Lack of presaccadic positivity before rapid eye movements in human REM sleep. *NeuroReport, 15,* 735–738. doi:10.1097/00001756-200403220-00033

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24,* 477–492. doi:10.3758/BF03200936

Armington, J. C., & Bloom, M. B. (1974). Relations between the amplitudes of spontaneous saccades and visual responses. *Journal of the Optical Society of America, 64,* 1263–1271. doi:10.1364/JOSA.64.001263

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baccino, T., & Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology, 19,* 204–215. doi:10.1027/0269-8803.19.3.204

Bach, M. (1996). The Freiburg Visual Acuity Test: Automatic measurement of visual acuity. *Optometry and Vision Science, 73,* 49–53. doi:10.1097/00006324-199601000-00008

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17,* 364–390. doi:10.1016/0010-0285(85)90013-1

Barber, H. A., Donamayor, N., & Kutas, M. (2010). Parafoveal N400 effect during sentence reading. *Neuroscience Letters, 479,* 152–156. doi:10.1016/j.neulet.2010.05.053

Barlow, J. S. (1971). Brain information processing during reading: Electrophysiological correlates. *Diseases of the Nervous System, 32,* 668–672.

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effect models using S4 classes. (R package version 0.999375–31). [Computer software]. Retrieved from http://cran.r-project.org/

Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Vision and visual Dysfunction* (pp. 95–137). Boca Raton, FL: CRC Press.

Berg, P. (2002). Eye artifact treatment. In P. Berg (Ed.): *Konstanz raw data format handbook: EEG and psychophysiology data analysis programs.* PDF document version (pp. 7–1 to 7–3). Retrieved from ftp://klipsy.psychologie.uni-konstanz.de/KNFORMAT/KNbook-04-12-02.pdf

Berg, P., & Scherg, M. (1991). Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology, 79,* 36–44. doi:10.1016/0013-4694(91)90154-V

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology, 90,* 229–241. doi:10.1016/0013-4694(94)90094-9

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology, 30,* 518–524. doi:10.1111/j.1469-8986.1993.tb02075.x

Bodis-Wollner, I., Von Gizycki, H., Avitable, M., Hussain, Z., Javeid, A., Habib, A., . . . Sabet, M. (2002). Perisaccadic occipital EEG changes quantified with wavelet analysis. *Annals of the New York Academy of Sciences, 956,* 464–467. doi:10.1111/j.1749-6632.2002.tb02856.x

Boylan, C., & Ross Doig, H. (1989). Effect of saccade size on presaccadic spike potential amplitude. *Investigative Ophthalmology and Visual Science, 30,* 2521–2527.

Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5,* 34–44. doi:10.1162/jocn.1993.5.1.34

Brunia, C. H. M., Möcks, J., Van den Berg-Lenssen, M. M. C., Coelho, M.,

Coles, M. G. H., Elbert, T., . . . Gratton, G. (1989). Correcting ocular artifacts in the EEG: A comparison of several methods. *Journal of Psychophysiology, 3,* 1–50.

Burdette, L. J., Walrath, J., Gross, B. J., & Stern, J. A. (1986). A comparison of saccade evoked potentials recorded during reading and tracking tasks. *Physiology & Behavior, 37,* 527–532. doi:10.1016/0031-9384(86)90279-9

Croft, R. J., & Barry, R. J. (2000). Removal of ocular artifact from the EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology, 30,* 5–19. doi:10.1016/S0987-7053(00)00055-1

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research, 1155,* 147–162. doi:10.1016/j.brainres.2007.04.027

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084,* 89–103. doi:10.1016/j.brainres.2006.02.010

Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., & Jacobs, A. M. (2009). Event-related potentials reveal rapid verification of predicted visual input. *PLoS ONE, 4,* 35047. doi: 10.1371/journal.pone.0005047

Debruille, J. B. (2007). The N400 potential could index a semantic inhibition. *Brain Research Reviews, 56,* 472–477. doi:10.1016/j.brainresrev.2007.10.001

Delgado, R. E., & Özdamar, O. (2004). Deconvolution of evoked responses obtained at high stimulus rates. *Journal of the Acoustical Society of America, 115,* 1242–1251. doi:10.1121/1.1639327

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21. doi:10.1016/j.jneumeth.2003.10.009

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage, 34,* 1443–1449. doi:10.1016/j.neuroimage.2006.11.004

Dimigen, O., Schild, U., Hohlfeld, A., Berg, P., & Sommer, W. (2011) *Auditory language comprehension during saccadic eye movements: An ERP investigation*. Manuscript submitted for publication.

Dimigen, O., Sommer, W., Dambacher, M., & Kliegl, R. (2008). Simultaneous recording of eye movements and ERPs indicates an early access to word meaning in natural, left-to-right reading [Abstract]. *International Journal of Psychology, 43*(3/4), 47. Available at http://www.psypress.com/9781841698601

Dimigen, O., Valsecchi, M., Sommer, W., & Kliegl, R. (2009). Human microsaccade-related visual brain responses. *Journal of Neuroscience, 29,* 12321–12331. doi:10.1523/JNEUROSCI.0911-09.2009

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology, 44,* 927–935. doi:10.1111/j.1469-8986.2007.00593.x

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 954–959. doi:10.1037/0096-1523.31.5.954

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 7192–7197. doi:10.1073/pnas.0509557103

Everling, S., Krappmann, P., & Flohr, H. (1997). Cortical potentials preceding pro- and antisaccades in man. *Electroencephalography and clinical Neurophysiology, 102,* 356–362. doi:10.1016/S0013-4694(96)96569-4

Fatourechi, M., Bashashati, A., Ward, R. K., & Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology, 118,* 480–494. doi:10.1016/j.clinph.2006.10.019

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44,* 491–505. doi:10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146,* 75–84. doi:10.1016/j.brainres.2006.06.101

Findlay, J. M., & Harris, L. R. (1984). Small saccades to double-stepped targets moving in two dimensions. In A. G. Gale & F. Johnson (Eds.) *Theoretical and applied aspects of eye movement researc*h (pp. 71–78). Amsterdam, the Netherlands: Elsevier.

Friedman, D., Simson, R., Ritter, W., & Rapin, I. (1975). The late positive component (P300) and information processing in sentences. *Electroencephalography and Clinical Neurophysiology, 38,* 255–262. doi:10.1016/0013-4694(75)90246-1

Gaarder, K., Krauskopf, J., Graf, V., Kropfl, W., & Armington, J. C. (1964, December 11). Averaged brain activity following saccadic eye movement. *Science, 146,* 1481–1483. doi:10.1126/science.146.3650.1481

Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments & Computers, 30,* 44–53. doi:10.3758/BF03209415

Graupner, S. T., Velichkovsky, B. M., Pannasch, S., & Marx, J. (2007). Surprise, surprise: Two distinct components in the visually evoked distractor effect. *Psychophysiology, 44,* 251–261. doi:10.1111/j.1469-8986.2007.00504.x

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15,* 883–899. doi:10.1162/089892903322370807

Hansen, J. C. (1983). Separation of overlapping waveforms having known temporal distributions. *Journal of Neuroscience Methods, 9,* 127–139. doi:10.1016/0165-0270(83)90126-7

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology, 115,* 1090–1103. doi:10.1016/j.clinph.2003.12.020

Herdman, A. T., & Ryan, J. D. (2007). Spatio-temporal brain dynamics underlying saccade execution, suppression, and error-related feedback. *Journal of Cognitive Neuroscience, 19,* 420–432. doi:10.1162/jocn.2007.19.3.420

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30,* 47–61. doi:10.1111/j.1469-8986.1993.tb03204.x

Hutzler, F., Braun, M., Vö, M. L. H., Engl, V., Hofmann, M., Dambacher, M., & Jacobs, A. M. (2007). Welcome to the real world: Validating fixation-related brain potentials for ecologically valid settings. *Brain Research, 1172,* 124–129. doi:10.1016/j.brainres.2007.07.025

Huynh, H., & Feldt, L. S. (1976). Estimation for the box correction for degrees of freedom from sample data in the randomized block and split plot design. *Journal of Educational Statistics, 1,* 69–82. doi:10.2307/1164736

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology, 19,* 113–124. doi:10.1097/00004691-200203000-00002

Jewett, D. L., Caplovitz, G., Baird, B., Trumpis, M., Olson, M. P., & Larson-Prior, L. J. (2004). The use of QSD (q-sequence deconvolution) to recover superimposed, transient evoked-responses. *Clinical Neurophysiology, 115,* 2754–2775. doi:10.1016/j.clinph.2004.06.014

Joyce, C. A., Gorodnitsky, I. F., King, J. W., & Kutas, M. (2002). Tracking eye fixations with electroocular and electroencephalographic recordings. *Psychophysiology, 39,* 607–618. doi:10.1111/1469-8986.3950607

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology, 41,* 313–325. doi:10.1111/j.1469-8986.2003.00141.x

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye

fixations to comprehension. *Psychological Review, 87,* 329–354. doi:10.1037/0033-295X.87.4.329

Kazai, K., & Yagi, A. (1999). Integrated effects of stimulation at fixation points on EFRP (eye-fixation related brain potentials). *International Journal of Psychophysiology, 32,* 193–203. doi:10.1016/S0167-8760(99)00010-0

Kazai, K., & Yagi, A. (2003). Comparison between the lambda response of eye-fixation–related potentials and the P100 component of pattern-reversal visual evoked potentials. *Cognitive, Affective &Behavioral Neuroscience, 3,* 46–56. doi:10.3758/CABN.3.1.46

Kazai, K., & Yagi, A. (2005). Contrast dependence of lambda response. *International Congress Series, 1278,* 61–64. doi:10.1016/j.ics.2004.11.053

Kennedy, A., Pynte, J., & Ducrot, S. (2002). Parafoveal-on-foveal interactions in word recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 55,* 1307–1337.

Kennett, S., Van Velzen, J., Eimer, M., & Driver, J. (2007). Disentangling gaze shifts from preparatory ERP effects during spatial attention. *Psychophysiology, 44,* 69–78. doi:10.1111/j.1469-8986.2006.00470.x

Keren, A. S., Yuval-Greenberg, S., & Deouell, L. Y. (2010). Saccadic spike potentials in gamma-band EEG: Characterization, detection and suppression. *NeuroImage, 49,* 2248–2263.

Klemm, M., Haueisen, J., & Ivanova, G. (2009). Independent component analysis: Comparison of algorithms for the investigation of surface electrical brain activity. *Medical and Biological Engineering and Computing, 47,* 413–423.

Kliegl, R. (2007). Toward a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery, and Reichle (2007). *Journal of Experimental Psychology: General, 136,* 530–537. doi:10.1037/0096-3445.136.3.530

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. doi:10.1080/09541440340000213

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135,* 12–35. doi:10.1037/0096-3445.135.1.12

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *NeuroReport, 20,* 1613–1618. doi:10.1097/WNR.0b013e328332c4f4

Kurtzberg, D., & Vaughan, H. G. (1977). Electrophysiological observations on the visuomotor system and neurosensorium. In J. E. Desmendt (Ed.), *Visual evoked potentials in man: New developments* (pp. 314–331). Oxford, England: Claredon Press.

Kurtzberg, D., & Vaughan, H. G., Jr. (1979). Maturation and task specificity of cortical potentials associated with visual scanning. In D. Lehmann & E. Callaway (Eds.), *Human evoked potentials: Applications and problems* (pp. 185–199). New York, NY: Plenum.

Kurtzberg, D., & Vaughan, H. G., Jr. (1982). Topographic analysis of human cortical potentials preceding self-initiated and visually triggered saccades. *Brain Research, 243,* 1–9. doi:10.1016/0006-8993(82)91115-5

Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. *Electroencephalography and Clinical Neurophysiology: Supplement, 40,* 406–411.

Kutas, M., & Hillyard, S. A. (1980, January 11). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205. doi:10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984, January 12). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163. doi:10.1038/307161a0

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. A. &. T. M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York, NY: Elsevier.

Lagerlund, T. D., Sharbrough, F. W., & Busacker, N. E. (1997). Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *Journal of Clinical Neurophysiology, 14,* 73–82. doi:10.1097/00004691-199701000-00007

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De) constructing the N400. *Nature Reviews Neuroscience, 9,* 920–933. doi:10.1038/nrn2532

Li, Y., Ma, Z., Lu, W., & Li, Y. (2006). Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. *Physiological Measurement, 27,* 425–436. doi:10.1088/0967-3334/27/4/008

Malmivuo, J., & Plonsey, R. (1995). Electro-oculogram. In J. Malmivuo & R. Plonsey (Eds.), *Bioelectromagnetism: Principles and applications of bioelectric and biomagnetic fields.* New York, NY: Oxford University Press.

Marton, M. (1991). Saccade-related brain potentials during reading. In G. Lázár (Ed.), *Advances in biological research in Hungary (1986–1990)* (pp. 66-69). Budapest, Hungary: Akaprint.

Marton, M., & Szirtes, J. (1982). Averaged lambda potential and visual information processing. *Studia Psychologica, 24,* 165–170.

Marton, M., & Szirtes, J. (1988a). Context effects on saccade-related brain potentials to words during reading. *Neuropsychologia, 26,* 453–463. doi:10.1016/0028-3932(88)90098-X

Marton, M., & Szirtes, J. (1988b). Saccade-related brain potentials during reading correct and incorrect versions of proverbs. *International Journal of Psychophysiology, 6,* 273–280. doi:10.1016/0167-8760(88)90014-1

Marton, M., Szirtes, J., & Breuer, P. (1985). Electrocortical signs of word categorization in saccade-related brain potentials and visual evoked potentials. *International Journal of Psychophysiology, 3,* 131–144. doi:10.1016/0167-8760(85)90033-9

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81,* 899–917. doi:10.1037/h0037368

Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin, 89,* 133–162.

Nuthmann, A., Engbert, R., & Kliegl, R. (2005). Mislocated fixations during reading and the inverted optimal viewing position effect. *Vision Research, 45,* 2201–2217. doi:10.1016/j.visres.2005.02.014

Okada, Y., Jung, J., & Kobayashi, T. (2007). An automatic identification and removal method for eye-blink artifacts in event-related magnetoencephalographic measurements. *Physiological Measurement, 28,* 1523–1532. doi:10.1088/0967-3334/28/12/006

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9,* 97–113. doi:10.1016/0028-3932(71)90067-4

O'Regan, J. K., & Lévy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. In M. Coltheart (Ed.), *Attention and performance XII*: *The psychology of reading* (pp. 363–383). Hillsdale, NJ: Erlbaum.

Oster, P. J., & Stern, J. A. (1980). Measurement of eye movement: Electrooculography. In I. Martin & P. H. Venables (Eds.), *Techniques in psychophysiology* (pp. 275–309). Chichester, England: Wiley.

Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology, 74,* 374–388. doi:10.1016/j.biopsycho.2006.09.008

Picton, T. W., Van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology, 111,* 53–65. doi:10.1016/S1388-2457(99)00227-8

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language, 110,* 81–94. doi:10.1016/j.bandl.2008.12.001

Raney, G. E., & Rayner, K. (1993). Event-related brain potentials, eye movements, and reading. *Psychological Science, 4,* 283–286. doi:10.1111/j.1467-9280.1993.tb00565.x

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65–81. doi:10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the EZ Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 720–732. doi:10.1037/0096-1523.30.4.720

Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research, 41,* 943–954. doi:10.1016/S0042-6989(00)00310-2

Rayner, K., & Clifton, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology, 80,* 4–9. doi:10.1016/j.biopsycho.2008.05.002

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191–201. doi:10.3758/BF03197692

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 53,* 1061–1080. doi:10.1080/02724980050156290

Rayner, K., Pollatsek, A., Drieghe, D., Slattery, T. J., & Reichle, E. D. (2007). Tracking the mind during reading via eye movements: Comments on Kliegl, Nuthmann, and Engbert (2006). *Journal of Experimental Psychology: General, 136,* 520–529. doi:10.1037/0096-3445.136.3.520

R Development Core Team. (2009). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Richards, J. E. (2003). Cortical sources of event-related potentials in the prosaccade and antisaccade task. *Psychophysiology, 40,* 878–894. doi:10.1111/1469-8986.00106

Riemslag, F. C., van der Heijde, G. L., & van Dongen, M. M. (1987). Are eye movement evoked potentials different from pattern reversal evoked potentials? *Documenta Ophthalmologica, 66,* 279–289. doi:10.1007/BF00213656

Rong, F., & Contreras-Vidal, J. L. (2006). Magnetoencephalographic artifact identification and automatic removal based on independent component analysis and categorization approaches. *Journal of Neuroscience Methods, 157,* 337–354. doi:10.1016/j.jneumeth.2006.04.024

Ross, J., Morrone, C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24,* 113–121. doi:10.1016/S0166-2236(00)01685-4

Scherg, M. (2003). Artifacts: Using calibration data to generate artifact coefficients. In *Manual for BESA: Brain-Electrical Source Analysis software* (Version 5.0) [Computer software]. Munich, Germany: Megis Software.

Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science, 14,* 328–333.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences, 7,* 489–493. doi:10.1016/j.tics.2003.09.010

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *NeuroReport, 9,* 2195–2200. doi:10.1097/00001756-199807130-00009

Simola, J., Holmqvist, K., & Lindgren, M. (2009). Hemispheric differences in parafoveal processing: Evidence from eye-fixation related potentials. *Brain and Language, 111,* 101–113. doi:10.1016/j.bandl.2009.08.004

Takeda, Y., Sugai, M., & Yagi, A. (2001). Eye fixation related potentials in a proof reading task. *International Journal of Psychophysiology, 40,* 181–186. doi:10.1016/S0167-8760(00)00185-9

Talsma, D., & Woldorff, M. G. (2004). Methods for the estimation and removal of artifacts and overlap in ERP data. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 115–148). Cambridge, MA: MIT Press.

Thickbroom, G. W., Knezevic, W., Carroll, W. M., & Mastaglia, F. L. (1991). Saccade onset and offset lambda waves: Relation to pattern movement visually evoked potentials. *Brain Research, 551,* 150–156. doi:10.1016/0006-8993(91)90927-N

Thickbroom, G. W., & Mastaglia, F. L. (1986). Presaccadic spike potential: Relation to eye movement direction. *Electroencephalography and Clinical Neurophysiology, 64,* 211–214. doi:10.1016/0013-4694(86)90167-7

Tobimatsu, S., & Celesia, G. G. (2006). Studies of human visual pathophysiology with visual evoked potentials. *Clinical Neurophysiology, 117,* 1414–1433. doi:10.1016/j.clinph.2006.01.004

Valsecchi, M., Dimigen, O., Sommer, W., Kliegl, R., & Turatto, M. (2009). Microsaccadic inhibition and P300 enhancement in a visual oddball task. *Psychophysiology, 46,* 635–644. doi:10.1111/j.1469-8986.2009.00791.x

Vitu, F. (1991). The influence of parafoveal preprocessing and linguistic context on the optimal landing position effect. *Perception & Psychophysics, 50,* 58–75. doi:10.3758/BF03212205

Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., & Fox, N. A. (2004). Automatic correction of ocular artifacts in the EEG: A comparison of regression-based and component-based methods. *International Journal of Psychophysiology, 53,* 105–119. doi:10.1016/j.ijpsycho.2004.03.007

Wang, T., Özdamar, Ö., Bohórquez, J., Shen, Q., & Cheour, M. (2006). Wiener filter deconvolution of overlapping evoked potentials. *Journal of Neuroscience Methods, 158,* 260–270. doi:10.1016/j.jneumeth.2006.05.023

Woldorff, M. G. (1993). Distortion of EEG averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology, 30,* 98–119. doi:10.1111/j.1469-8986.1993.tb03209.x

Yagi, A., Kazai, K., & Takeda, Y. (2000). Spatial and temporal variations in eye-fixation-related potentials. *Japanese Psychological Research, 42,* 69–75. doi:10.1111/1468-5884.00132

Young, L. R., & Sheena, D. (1988). Eye movement measurement techniques. In J. G. Webster (Ed.), *Encyclopedia of medical devices and instrumentation* (pp. 1259–1269). New York, NY: Wiley.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron, 58,* 429–441. doi:10.1016/j.neuron.2008.03.027

Received November 4, 2009
Revision received March 28, 2011
Accepted March 28, 2011 ■