Reading at the speed of speech: the rate of eye movements aligns with auditory language processing

Benjamin Gagl[1,2], Julius Golch[1], Stefan Hawelka[3], Jona Sassenhagen[1], David Poeppel[4,5] and Christian J. Fiebach[1,2,4]

[1] Department of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany
[2] Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany
[3] Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria.
[4] Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany.
[5] Department of Psychology, New York University, USA

Correspondence to gagl@psych.uni-frankfurt.de

Abstract

Across languages, the speech signal is characterized by a ~4-5 Hz rhythm of the amplitude modulation spectrum, reflecting the processing of linguistic information chunks approximately every 200 ms. Interestingly, ~200 ms is also the typical eye-fixation duration during reading. Prompted by this observation, we estimated the frequency at which readers sample text, and demonstrate that they read sentences at a rate of ~5 Hz. We then examined the generality of this finding in a meta analysis. While replicating the experimentally measured 5 Hz sampling rate in the language in which it was obtained, i.e., German, we observe that fixation-based sampling frequencies vary across languages between 3.1 and 5.2 Hz, with the majority of languages lying between 4 and 5 Hz. Remarkably, we identify a systematic rate reduction from easy to difficult writing systems. Reading in easy-to-process writing systems thus is aligned with the rhythm of speech, which may constitute an upper boundary for reading. We argue that reading is likely tuned to supply information to linguistic processes at an optimal rate, coincident with the typical rate of speech.

Significance Statement

Across languages, speech is produced and perceived at a rate of 4-5Hz. When listening to speech, our brain 'picks up' this temporal structure. Here we report that during reading, our eyes sample text at the same rate. We demonstrate this empirically in one language, and generalize this finding in a meta analysis of 124 empirical studies, covering 14 different languages. Reading rates vary between 3.1 and 5.2Hz – i.e., broadly in the range of the speech signal. We demonstrate that this variance is determined by the orthographical difficulty of different writing systems, and propose that the rate at which our brain processes spoken language acts as a driving force and upper limit for the voluntary control of eye movements during reading.

Keywords: reading, speech, eye-movements, linguistic information sampling, z-string scanning

Introduction

When we listen to speech, our brain entrains to the frequency at which linguistic information enters the auditory system. For example, when spoken words are presented at a consistent rate of 4 Hz, the frequency spectrum of simultaneously measured electrophysiological brain activity shows a prominent peak at around 4 Hz (1), suggesting that the temporal structure of the linguistic stimulus drives neural processes in the auditory and language systems. Natural speech – which is considerably more variable in the time domain – produces characteristic frequency correlates in the theta-band of human electrophysiological brain activity, i.e., between 4 Hz and 8 Hz (2). This in turn is consistent with the recent observation that, across languages, the speech signal amplitude changes at a frequency of about 4-5 Hz (3, 4). These findings suggest that roughly every 200 ms, a chunk of information is processed in the brain when listening to speech (5).

Interestingly – and we hypothesize not accidentally – the duration of a typical eye-fixation during reading is very similar, i.e. between around 200 ms for orthographically transparent languages like German or Finnish (6, 7) and 250 ms for character based writing systems like Chinese (7, 8). Abundant research has used eye movement recordings to study reading with high temporal resolution. This includes the influence of various characteristics of the perceived words (such as their length, frequency of occurrence, or predictability out of the sentence context; for a review see (8)), but also group comparisons, for example between dyslexic and typically developing readers (9, 10). Among various measures that can be derived from eye movement recordings, including saccade length or word skipping rate (8), timing measures like the *duration of fixations* are the most frequently examined variables and are generally considered precise markers of reading speed. However, unlike in other research fields (e.g., attention, (11)), eye-movements in natural reading have so far not been investigated from a frequency perspective. While the observed fixation durations in reading suggest that the temporal sampling during reading may take place in a very similar time domain as speech perception, the sampling rate of reading has to date not been explicitly examined. Consequently, important open questions emerge, including (i) whether written language is sampled at similar frequencies as spoken language, (ii) whether the sampling frequency is specific for eye movements in reading in contrast to non-linguistic tasks, and (iii) whether the sampling rate of reading differs between languages or writing systems.

Here we adopted a frequency perspective on reading behavior to answer these foundational questions. In a first step, the sampling frequencies of 50 native speakers of German were determined while reading sentences and during a non-linguistic control task, building on established empirical data (12). Remarkably, it has not yet been systematically and quantitatively investigated whether mean fixation durations during reading indeed translate into similar sampling frequencies as observed for speech perception. To determine the generality of our empirical

results and to investigate possible cross-linguistic differences in the sampling rate of reading, we next conducted a meta analysis. To this end, we implemented a frequency analysis for fixation durations extracted from published eye tracking studies of natural (sentence or paragraph) reading. The experimental and meta analytic results support the hypothesis of a common information uptake rate for reading and speech perception, which in reading, however, is modulated by the difficulty of the respective script.

Results: Empirical data

Fixation durations. Subject-specific sampling frequencies during reading were determined based on data from a previously published experiment (12). 50 healthy volunteers read sentences from the Potsdam Sentence Corpus (144 sentences; 1,138 words; (6); presented as a whole) while movements of their right eye were tracked at a resolution of 1,000 Hz. As a non-linguistic control task, participants scanned 'z-strings' that were constructed by replacing all letters of the sentence stimuli by the letter 'z'. For example, the sentence "*Ein berühmter Maler hat sich selbst ein Ohr abgeschnitten" (translation: A famous painter cut off his own ear.)* was transformed to "*Zzz zzzzzzzzz Zzzzz zzz zzzz zzzzzz zzz Zzz zzzzzzzzzzzzz*." (See Materials and Methods for details).

After data preprocessing (leading to a removal of 3.1% of the data), we estimated the mean fixation duration separately for each participant and experimental condition. Figure 1a shows that fixation durations (presented here as subject-specific means) are shorter for reading in contrast to scanning (197 ms vs. 249 ms, respectively; Cohen's d = 1.57; $t(49) = 11.1$; $p < .001$). This observation has been reported previously for this dataset (12) and is also consistent with previous results for reading in German from other groups (13).

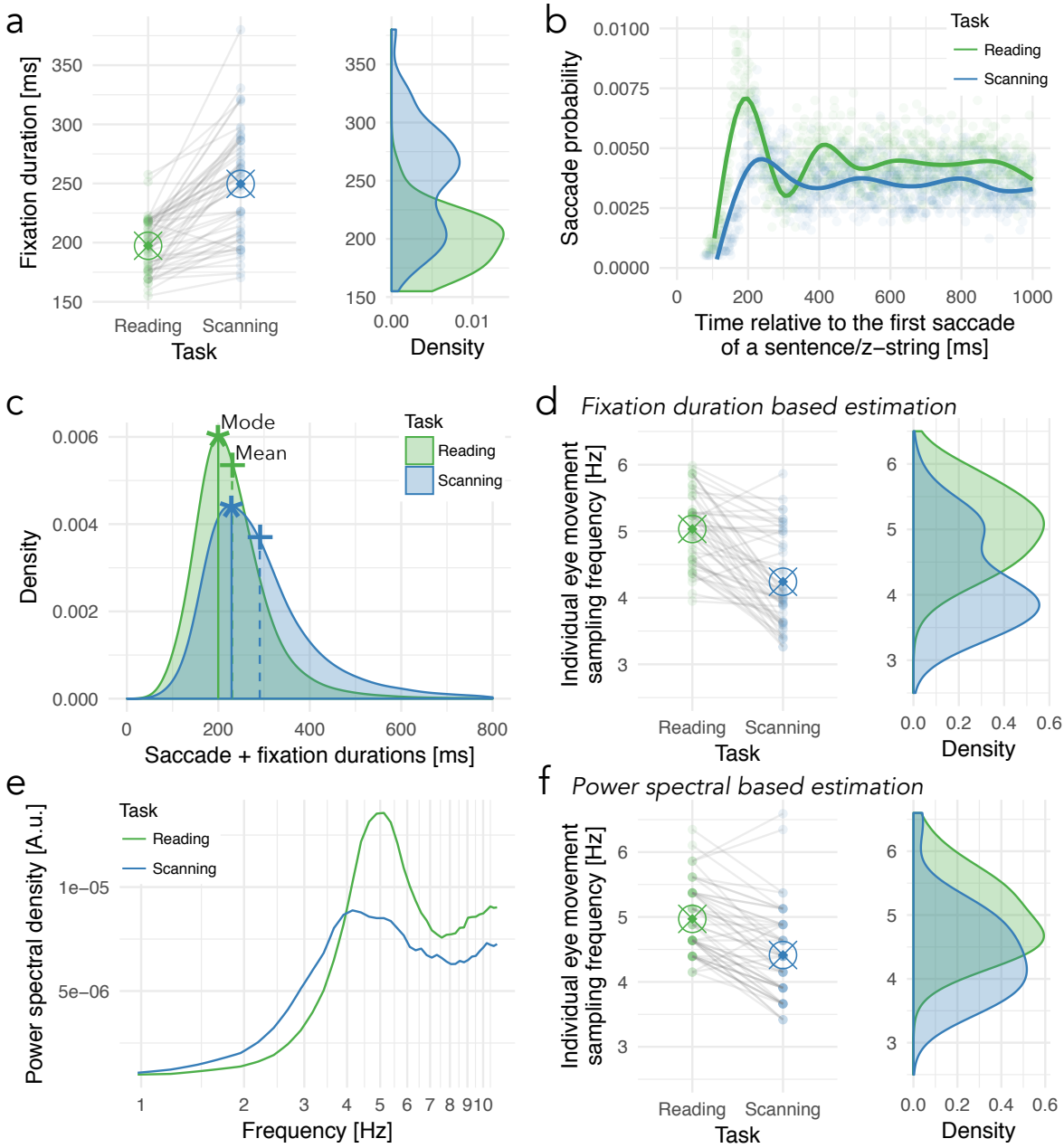

Figure 1. Empirical study of reading-related sampling rate. (a) Subject-specific mean fixation durations from 50 participants (dots) and the overall mean (cross) while reading sentences on the Potsdam sentence corpus (6) and scanning z-strings. Lines connect reading with z-string scanning data, per subject, to visualize effects at the single-subject level. Left panel shows the distribution of individual means separated by task (Blue: Scanning; Green: Reading) by using a kernel density estimation (similar in d and f). (b) Mean saccade probability (over all participants and stimuli, separated by task) relative to the first saccade of the sentence, with a non-linear regression line. (c) The sampling period t of one event is here defined as the duration of a fixation plus its preceding saccade. Displayed is the distribution of these sampling periods for sentence reading (green) and z-string scanning (blue), with estimated means (+ symbol and dashed lines) and modes (* symbol and solid lines). (d) Subject-specific mean sampling frequencies f (i.e., equals 1/t) and the overall mean (cross) based on the sampling periods shown in c. Lines connect reading with z-string scanning data, per subject, to visualize effects at the single-subject level. (e) Power-spectrum for reading and z-string

scanning, estimated across all participants, and (f) individual peak frequencies estimated from individual power spectra and their mean (cross). See Methods for further details.

Eye-movement sampling rate differs between reading and non-linguistic control task. As a first exploration of rhythmic eye movement patterns during reading, we plotted the saccade probability, relative to the first saccade of each stimulus (Figure 1b; one probability value per sample; sampling rate 1,000 Hz). This demonstrates prominent peaks visible at regular intervals, giving a first indication that eye-movements follow a rhythmic structure in reading and scanning. To quantitatively estimate the dominant sampling rate we adopted two procedures: on the one hand, sampling rates were estimated from the fixation durations. Secondly, we also applied a more classical frequency analysis. While the former approach is important because fixation durations are also the basis for the subsequent meta analysis, the second approach allows us to evaluate the validity of fixation-based frequency estimation.

To estimate sampling rates from fixation durations, we first added to each fixation duration (N = 112,547) the duration of the preceding saccade, to estimate the respective sampling period *t* (i.e., the time from the start of a saccade to the start of the next). Figure 1c shows the distribution of all sampling periods across all participants, separately for reading and z-string scanning. Note that the mode (solid line) represents the predominant sampling period better than the mean (dashed line). This is a typical characteristic of ex-Gaussian distributions, which are generally found for fixation durations (14) and reaction times (15). We then estimated an eye movement sampling frequency f for each participant and condition, by dividing 1 sec by the subject-specific mode of the sampling period. This analysis revealed a higher average sampling rate for reading relative to the control task, i.e., 5.03 Hz vs. 4.24 Hz, respectively (Figure 1d). This difference was significant (Cohen's d = -0.92; $t(49) = -6.5$; $p < .001$), and 45 out of 50 participants showed a numeric reduction of the sampling frequency from reading to scanning (grey lines in Fig. 1d).

We finally reproduced this result with a canonical frequency analysis which was implemented by estimating the power spectra of reading vs. z-string scanning. For each task, we created a time series starting with the first saccade of the first participant and ending with the last fixation of the final participant. This time series was set to 1 at the exact time of saccade onset, and 0 elsewhere (1000 Hz sampling rate). Subsequently, we estimated the power spectrum of this event time course via the Fourier Transform, separately for each task and participant (see Materials and Methods for details). The result indicates the periodic components in the signal. Corroborating the results of the first analysis approach, a prominent peak was found at 5 Hz for reading and a broader peak around 4 Hz for scanning (Figure 1e). To submit the frequency analysis result to a statistical test, we estimated separate power spectra for each participant. Individual peaks were retrieved, averaged (Figure 1f),

and submitted to a t-test. This analysis closely reproduces the sampling frequencies estimated from the mode of the fixation durations (4.97 Hz vs. 4.41 Hz for reading and scanning, respectively; Reading: r = .80; $t(48) = 9.3$; $p < .001$; Scanning: r = .62; $t(48) = 5.5$; $p < .001$), indicating a high validity of the eye movement based frequency estimations. In addition, the significantly higher frequency in reading could be found again (Cohen's d = 1.12; $t(49) = -7.9$; p < .001).

Interim summary. A quantitative frequency-domain characterization of eye-tracking data shows that the predominant sampling frequency during reading in German, across participants, is found at ~5 Hz. This frequency representation of the reading process converges with the predominant modulation frequency of the speech signal (3, 4) – which in turn has a clear reflection in the neuronal response to speech (2). We observed the ~5 Hz peak during reading using two different analysis strategies, i.e., when estimating sampling frequencies from saccade and fixation durations as well as when analyzing in the frequency domain the sequence of saccade events over time. Attentive scanning of z-strings, in contrast, showed a significantly lower sampling frequency at ~4 Hz, convergent with findings from attentional reorienting tasks (11). Note that the estimation of the sampling frequency in z-string scanning was less reliable when compared to reading, as indicated by a significantly lower across-person correlation between the two approaches to estimating the sampling rate (cp. r = .62 to r = .80, respectively; $z = 2.0659$, $p = 0.04$ (16)). The reduced reliability is caused by the larger variation of fixation durations (i.e. Reading: SD = 23 ms; Scanning: SD = 47 ms; Cohen's d = 4.2; $t(49) = 29.6$; p < .001), which indicates that the rhythmic structure of z-string scanning is less pronounced. Given that reading takes place at a higher rate despite the fact that it is cognitively more demanding than the control task (see Materials and Methods), and given the high degree of similarity to the sampling rate of processing spoken language, we hypothesize that 5 Hz may reflect the preferred rate at which neural language systems process linguistic information, independent of the modality of presentation. As such, the rate is likely to reflect a cortical computation principle.

To be sure, this hypothesis is based on only a single dataset. If our empirical finding of a ~5 Hz sampling rate indeed represents a more fundamental property of cortical computation that applies to reading, it should generalize across languages and writing systems. On the other hand, writing systems differ substantially between languages (7, 17), and even within writing systems, the mapping from orthography to meaning differs between languages (18). For example, the letter *a* in *cat* vs. in *ball* maps onto two different speech sounds in English, whereas it maps onto only one speech sound in the German translations of these words, i.e., *Katze* and *Ball*. This letter-to-sound correspondence strongly influences reading acquisition (19), so that opaque orthographies (i.e. writing systems with inconsistent letter-to-sound correspondence like English) are associated with lower reading accuracy during the first years of learning to read. Contrary to the claim for language-independent fixation durations, these differences between writing systems would be suggestive

of cross-linguistic differences in the efficiency of reading. Recent experimental evidence like, e.g., the observation of longer fixation durations for Chinese as compared to Finnish or English (7), seems to support this prediction. Given that German is a relatively transparent and thus easy-to-process orthography among alphabetic writing systems (e.g. 130 vs. 226 grapheme-to-phoneme rules for German and English, respectively; (18)), we conjectured that the empirically determined sampling rate of ~5 Hz when reading German may represent an upper bound for the sampling of text reading. To test these alternative predictions, we conducted a meta analysis of more than one hundred eye-tracking studies of reading covering 14 different languages.

Results: Meta analysis

In this meta analysis, we compared the sampling frequency of reading between 14 different languages, based on 1,420 fixation duration estimates extracted from 124 studies published between 2006 and 2016. In addition to this cross-linguistic comparison, we examined possible differences between character-based vs. alphabetic writing systems and, among the alphabetic writing systems, the effect of letter-to-sound correspondence.

All the studies selected for inclusion into the meta analysis reported mean fixation durations. However, as shown in Figure 1c, mean fixation durations are not a consistent and valid representation of the predominant sampling duration in the fixation data. We therefore developed a transformation function that allowed us to estimate the mode from the mean fixation durations reported in the original publications (see *Materials and Methods* and Supplement 1 for details). In brief, this involved (i) fitting ex-Gaussian distributions to 29 empirical fixation duration datasets, (ii) retrieving distributional parameters for each fitted distribution (specifically: μ, mean of the normally distributed component; σ, standard deviation of the normally distributed component; τ, parameter reflecting the rightward skew, i.e., representing the contribution of the exponential distribution), allowing us to (iii) implement a regression-based transformation function that is capable of converting a single mean into a mode fixation duration (see Materials and Methods and Supplementary Materials for more details). Figure 2a presents the final generalized mean-to-mode transformation function, applied to the full range of all possible fixation durations in the meta analysis. Figure 2b shows performance of this transformation, i.e., how well the modes of our 29 simulation datasets can be recovered by this function. The presented regression line nicely shows that, despite some unsystematic noise, the numeric transformation was nearly perfect (i.e., beta = 0.95; SE = 0.23; t = 4.1). To obtain the sampling period t (i.e., the interval from the onset of a saccade until the end of the following fixation; cf. the empirical study, above), a saccade duration estimate (28 ms; i.e., the mode of saccade durations from the reading dataset used in the first experiment) was added to each of the 1,420 mean fixation durations of the meta analysis dataset. This is feasible since

saccade durations do not differ much between participants in reading, which corresponds to previous findings (e.g. (20): range 20-39 ms, mean: 29 ms). Finally, the sampling period values (*t*) were transformed into frequency values ($f = 1 / t$; see Materials and Methods).

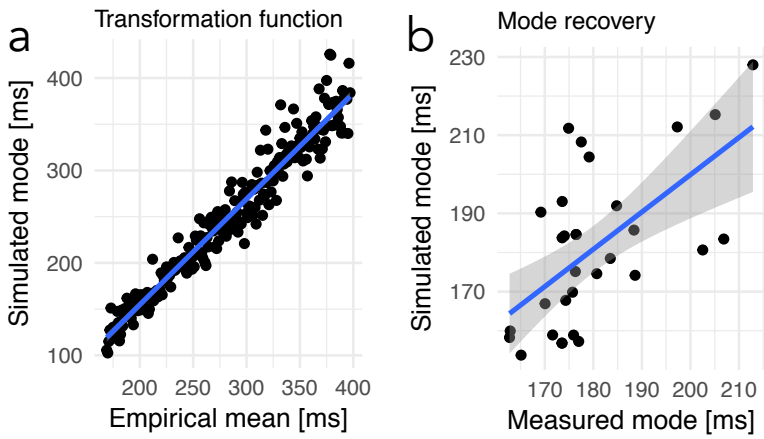

Figure 2. Transformation function for converting mean fixation durations into simulated mode values. The function was established using ex-Gaussian estimations on 29 available datasets containing fixation durations from 11 empirical studies (i.e., three German studies from our own lab, (e.g. (21)), and multiple openly available English studies. For details see Supplement 1. (a) Performance of the final mean-to-mode transformation function (blue line) demonstrated for 248 fixation durations, i.e., for each millisecond within the range of fixation durations included in the meta analysis (149 – 397 ms; x-axis: empirical means). (b) Performance of the final mean-to-mode transformation function, as demonstrated by the relationship between empirically measured (x axis) and simulated modes (y-axis) from the 29 datasets used for establishing the transformation function.

Fixation duration and sampling frequency: Descriptive statistics. Figure 3 shows that the majority of fixation durations derived from the reading studies were between 200 and 300 ms (upper panel), which transformed to sampling frequencies ranging from 3.1 to 5.2 Hz (lower panel). Relating these results back to our starting point, i.e., the language-independent peak in the speech modulation spectrum (3, 4), we observe (a) that the mean reading-related sampling rates of all included languages fall within one standard deviation from the peaks of the speech modulation spectrum (Figure 3, lower panel, dotted lines) and (b) that 10 of the 14 languages fall within the range of reported mean speech modulation rates (i.e., between the minimum and maximum reported, language-specific mean speech modulation rates; Figure 3, lower panel, dashed lines). Of the 1,420 individual sampling rate values derived from the included studies, only 3.0% were lower and only 0.3% higher than one standard deviation around the mean of the speech modulation spectrum reported by (3) (see Fig 3, lower panel, violin plot). Nevertheless, our observed mean sampling rate when averaging across all languages is at the lower bound of the mean range of the speech modulation spectrum (see Fig 3, lower panel, violin plot).

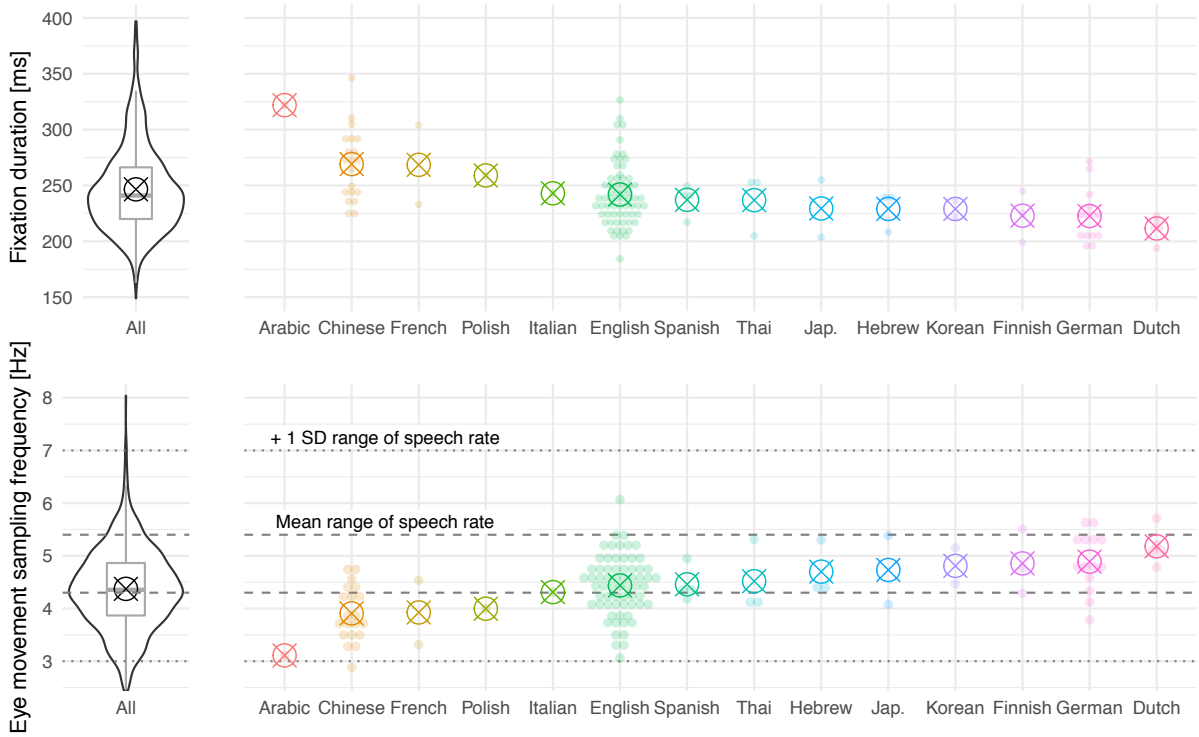

Figure 3. Meta analysis of reading-related sampling rates. Fixation durations (upper panel) and corresponding eye movement sampling frequencies (lower panel) for 14 different languages. Violin plots in the left part of the figure represent the distribution of all 1,420 duration / frequency values from all included studies, independent of language. Embedded boxplots show the median as well as one (box) and two (whiskers) standard deviations, crossed circles reflect the mean. For the right panel, dots reflect each study (mean number of fixation durations per study: 12.4); crossed circles reflect the mean across studies for each language. In the lower panel the dashed and dotted lines represent the mean range and standard deviations, respectively, for the amplitude modulation spectrum in speech (read out from Figure 3c in (3)).

Differences in sampling frequency between writing systems. The observed cross-linguistic differences, arguably, are related to differences in orthographic properties between languages. One plausible hypothesis is that the high perceptual complexity of character-based scripts (as opposed to alphabetic scripts; (17)) may modulate the rate at which written text is sampled. In line with this assumption, Figure 4a shows that the eye movement sampling frequency is significantly lower for Chinese (the only character-based language included; n = 256 sampling rates from 20 studies; mean: 3.9 Hz) in contrast to alphabetic languages (n = 1.215 sampling rates from 97 studies; mean: 4.5 Hz; effect size estimate (Est) of difference: -0.65 Hz; Standard error (SE): 0.14; t = 4.7; see Materials and Methods for details on linear mixed effects modeling).

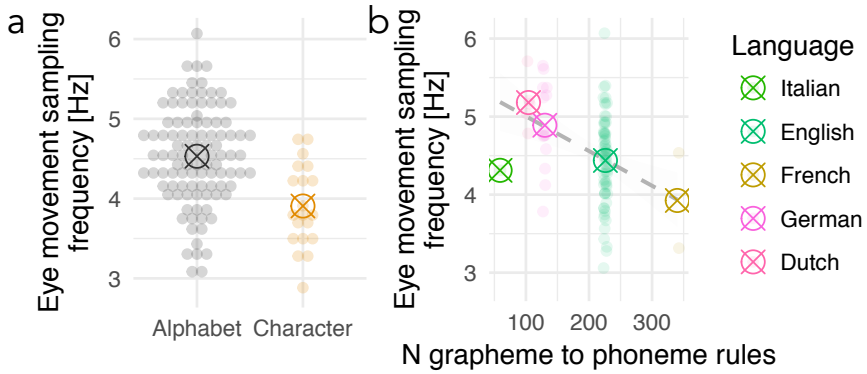

Figure 4. Comparison of writing systems. (a) Character vs. alphabetic contrast including 256 fixation durations from 20 studies of Chinese reading and 1,215 fixation durations from 97 studies of reading in alphabetic languages. (b) The effect of language transparency/opacity. Only studies from alphabetic languages for which the number of grapheme-to-phoneme rules could be quantitatively estimated (see Materials and Methods) were used (5 languages with a total of N = 1,026 fixation durations). Note that for Italian, the meta analysis contained only one study. Dots reflect each study and crossed circles reflect the mean across studies for each language. The dashed line in (b) represents the approximation of the language transparency/opacity effect based on a linear regression.

But also among alphabetic languages, the orthographic difficulty of a writing system is likely to influence the speed of sampling the visual input, as discussed above. To examine this, we quantified orthographic difficulty as a continuous predictor representing the number of grapheme-to-phoneme rules (graphemes: letters or letter combinations that map onto one or multiple phonemes, i.e. speech sounds; cf. the example introduced above: to map the letter *a* to phonemes in *cat* and *ball* two rules are needed, while for the German translations *Katze* und *Ball* only one rule is needed) as defined by computationally implemented dual-route models of visual word recognition (e.g. (22)). To date, such implementations are available for five out of the nine alphabetic languages included in this meta analysis, which restricts this test to Italian, English, French, German, and Dutch (n = 1, 965, 3, 48 and 45 sampling rates, respectively; see (18) for a detailed comparison of the Dual-Route models). Figure 4b demonstrates that an increase of the number of grapheme-to-phoneme rules significantly decreases the dominant eye movement sampling frequency during reading, indicating a reduced sampling rate for less transparent writing systems (Est: -0.14 Hz, SE = 0.04, t = 3.9; note that the interpretability of the apparent outlier result for Italian is limited since only one fixation duration was included). Interestingly, highly transparent orthographies like German or Dutch produce sampling rates around 5 Hz (Fig. 4b). For these easy-to-process scripts we assume that the analogy between speech and eye-movement sampling during reading can be investigated best since difficulties that arise from the orthographic code are minimal.

General discussion

In this (to our knowledge first) frequency-based investigation of eye movements during reading we show that reading operates in a comparable frequency domain as the auditory processing of natural speech. Our empirical data replicate in the frequency domain previous insights based on fixation duration measures (12, 13, 21, 22) that reading is implemented more efficiently even though linguistic processing is cognitively more demanding than non-linguistic scanning of letter strings (12) In addition, the sampling rate of reading is more variable than previously assumed. For example, Rayner (2009) claimed that even for very distinct orthographies like Chinese and English, "Chinese readers tend to have average fixations durations that are quite similar to readers of English." (p. 1461). Our results, in contrast, show sampling rates varying between around 3.1 and 5.2 Hz between languages. The modulation range of spoken speech varies more narrowly around 5 Hz (i.e., 4.3-5.4 Hz (3); 4.3-5.5 Hz (4)), excluding the lower range of frequencies observed for reading. Even though we have demonstrated that specifics (character vs. alphabetic or high vs. low grapheme-to-phoneme rules) of the writing system can slow down the sampling of written language within certain limits, it is remarkable that – even in the absence of any speech input – an internally controlled visual-perceptual process – reading – seems to be associated to the rate at which spoken language is produced and typically perceived (i.e. all language based eye-movement sampling frequencies within one standard deviation from speech rates; 3.1-5.2 Hz). This convergence of findings for speech and reading argues, perhaps, for the existence of a perceptual principle underlying the temporal structure of linguistic information processing irrespective of modality.

The frequency perspective is not needed for the analysis of eye movements during reading since the majority of saccades in reading can be detected by the naked eye. In essence, the frequency representation of reading-related eye tracking data is merely a transformation of fixation and saccade duration data. That being said, this novel approach opens up several interesting new research perspectives. For example, it becomes possible to more directly compare reading behavior to evidence from other domains, such as brain activation data, and to other linguistic (e.g., speech perception (3, 4)) or psychological processes (e.g., attention (11)).

The frequency spectrum of spoken language is broadly distributed, but mean peaks are surprisingly stable across languages at 4.3-5.5 Hz (3, 4). The auditory and linguistic systems of our brain entrain to this frequency range (more specifically: brain rhythms in the range of 4 to 8 Hz) when listening to naturalistic stimuli (e.g. (2)). The observation that the active sampling frequency of our eyes in reading is also located in that range indicates that information extraction from linguistic stimuli follows a similar temporal structure in time irrespective of modality. A plausible account for this striking overlap is to assume that the linguistic system has a preferred rate of information uptake and acts as an internal generator process

controlling the movement of the eyes from word to word during reading. Indirect support for this hypothesis comes, for example, from the observation that most readers experience an 'inner speech' during reading, and that manipulating the speed of inner speech during reading by associating text for example with fast or slow speakers has a causal effect on reading speed (23–25).

In conclusion, we show that the temporal domain of active sampling of visual input during reading resembles to a remarkable degree the well-established rate at which speech is produced and perceived. We propose that the rate at which linguistic systems of our brain process spoken language also acts as a driving force for active and voluntary control of eye movements during reading. We speculate that this enables the visual system to supply linguistic information at a rate that is optimal for the brain's language system. The novel frequency perspective on reading that we adopt here opens up new perspectives in reading research, for example for understanding slow or impaired reading or for second language learning.

Materials and Methods

Eye tracking experiment

Participants. Fifty (13 male; 18–47 years old; M = 24 years) native speakers of German participated after giving informed consent according to procedures approved by the local ethics committee. All participants had normal vision and were students at the University of Salzburg. See our original publication of this dataset (12) for more details. Note that, relative to the original study, one participant was added.

Procedure. Movements of the right eye were tracked with a sampling rate of 1,000 Hz (Eyelink 1000, tower mount system; SR-Research, Ontario, Canada). We used a forehead and chin rest to fixate the head of participants at a distance of 60 cm from a 21" CRT screen. In the reading task, we used the Potsdam Sentence Corpus (PSC; (6)) which consists of 144 sentences and a total of 1,138 words. Participants were instructed to read silently for comprehension, which was controlled by simple comprehension questions after 38 of the 144 sentences.

As a non-linguistic control task, participants performed a z-string scanning task using stimuli in which all letters of the sentence corpus were replaced by the letter z (preserving letter case, punctuation, and word boundaries; (see (26), and examples above). Participants were instructed to visually scan the meaningless z-strings as if they were reading, but for obvious reasons, no comprehension questions were administered in this condition. Z-string scanning has been used as control task in previous studies (12, 13, 26, 27). While it is difficult to find a reasonable control task for reading (for discussion, see, e.g., (13)), z-string scanning proved to be interesting because participants produce similar scan path

patterns (i.e. fixation position patterns) as when reading (12, 13, 26, 27). Interestingly, while z-string scanning produced longer mean fixation durations than reading, the pupil response indicated higher cognitive effort in reading, in the dataset used here (12). We take this dissociation between cognitive effort and reading time as evidence for the operation of reading-specific cognitive processes that go beyond mere attentional processes.

In each task, a 9-point standard calibration was performed before the 10 practice trials, before the experimental trials, and after a break halfway through the experiment. A calibration was considered accurate when the mean error was below 0.5° of visual angle. Visual stimuli were presented in black letters (mono-spaced, bold Courier New font; 14 pt., width ~0.3°) on white background with a 1,024 × 768 pixel resolution and a refresh rate of 120 Hz, using Experiment Builder software (SR Research, Ontario, Canada). In both tasks, a trial started when an eye-fixation was found at a dot presented 100 pixels from the left margin of the monitor, at the horizontal level of the fixation cross. For this fixation check, real-time analysis of eye-tracking data was used to present the sentence only when a fixation of at least 100 ms was identified on the position of the dot. If no fixation was registered on the dot for 10 seconds, a re-calibration procedure was initiated. Following the fixation check, the stimulus (i.e., sentence or z-string) appeared, with the center of the first word presented at the position of the fixation dot so that participants always landed first on the first word of the sentence. Stimulus presentation was terminated when participants fixated an X in the lower right corner of the screen after the sentence was read. As noted, in about 25% of sentences, the presentation was followed by a comprehension question to assure that participants processed sentences semantically. This procedure was practiced in ten trials prior to the main experiment.

Data analysis

Fixation durations. The first word of each sentence was excluded from analyses, since the first word is known to be contaminated by stimulus onset effects. A total of 994 words were analyzed per subject. For each participant, all fixation durations from all analyzed words were extracted. Words with fixation durations shorter than 60 ms and longer than 1,000 ms and saccade durations longer than 80 ms were removed from the analysis (3.1% of the data) since they likely reflect machine error. On the basis of the remaining fixation durations, the mean was calculated in order to estimate each participant's individual mean fixation duration, separately for the reading and scanning tasks. Note, to account for the ex-Gaussian distributions (see Figure 1c) for the statistical test we realized a log-transformation resulting in a normal distribution (Kolmogorov-Smirnov test not significant; $D < 0.14$; $p > .7$).

Estimation of the sampling frequency. To estimate the sampling frequency of eye movements in reading, first an event that has a repetitive nature (i.e., that takes place

more than once) has to be identified, in our case the repetitive pattern of saccades. Second, the time between the first and the subsequent occurrence is defined as the *sampling period*, which can be transformed into a frequency value. To this end, we added fixation and saccade duration, to calculate the respective sampling period. The distribution of the sampling period is ex-Gaussian, for both reading and z-string scanning (Figure 1c). Ex-Gaussian distributions are a convolution of a normal distribution and an exponential distribution reflecting the rightward skew. As Figure 1c shows, the central tendency is best represented by the mode, so that all subsequent fixation duration based frequency estimations are realized by a participant-specific mode (t). These subject-specific mode values are equivalent to the predominant sampling period of the respective participant, which in turn can be transformed to an individual eye movement sampling frequency ($f = 1 / t$).

Power spectrum. We performed a canonical frequency analysis by estimating a power spectrum for both tasks and all participants each. As mentioned before, for Figure 1e, we estimated the power spectrum based on a time series starting with the first saccade of the first participant and ending with the last fixation of the final participant for each task. For Figure 1f, the time series was cut into participant-specific time series, so that individual peaks could be recovered for each participant for each task. The time series was realized as a sparse sequence of zeros and ones (resolution: 1,000 entries per second), set to one at time points at which a saccade was initiated, and zero otherwise. Subsequently, a Fast Fourier Transform was used to estimate a power spectrum (power spectral density; *psd_welch* function from MNE-Python; (28) ; 0-100 Hz, length of the FFT used = 4096 samples) for each of event time courses separately. All data (also from the meta analysis) and analysis code will be made available when published.

Meta analysis

We included empirical studies that report eye-tracking results from natural reading tasks, published between 2006 and 2016. These studies were identified by the search term *eye movement in "natural reading" or "sentence reading" or "text reading"* in the *PubMed* (https://www.ncbi.nlm.nih.gov/pubmed) and *PsychInfo* (https://health.ebsco.com/products/psycinfo) databases. Additionally, 10 studies were manually identified (e.g., on the basis of reference lists in published papers). From the resulting sample of 124 articles we extracted 1,420 fixation durations, including mean fixation durations (all fixations on a word combined; 10% of the dataset), first fixation durations (duration of the first fixation on a word; 67%), and single fixation durations (fixation duration in case a word was fixated only once, which is the predominant case for normal readers; (e.g. (6); 23%). A full list of all included studies can be found in the Supplementary Materials 1. Note that the results of the above-reported experiment and its previous analysis (12) were not included. This meta analytic dataset encompassed 14 different languages, with a range from one (Arabic, Italian, and Polish) to 65 (English) retrieved papers.

Consistent with a general bias towards English in reading research (29), 68% of fixation durations in our dataset were from English.

Frequency estimation. In order to estimate the predominant sampling frequency, per published study, we have to take into account, once more, the ex-Gaussian distribution of fixation duration data. Following the general trend in the eye movement reading literature, most studies reported only mean fixation durations (see (14) for an exception since, in addition, the fitted ex-Gaussion paramaters were reported). However, as shown in Figure 1c, mean fixation durations are not a consistent and valid representation of the predominant duration in the fixation data. For the purposes of the present meta analysis, we developed a transformation function that allowed us to estimate the mode from the mean fixation durations reported in the original publications (for details see Supplement 1). This transformation function was then applied to transform mean fixation durations extracted from the published original studies into the mode. In the final transformation the sampling periods (mode fixation duration plus mode saccade duration) were converted in a frequency value.

Writing system comparisons. In order to explore whether or not the sampling frequency of reading is influenced by global characteristics of writing systems and languages, we implemented two tests. First, Chinese reading (256 fixation duration data points) was compared to all alphabetic writing systems (1,215 fixation durations). Note that the Korean (alphabetic syllabary orthography; (30)) and Japanese (using a mixture of Kana and Kanji) studies in the meta analysis could not be clearly assigned to the character or alphabet categories and therefore were not included into this contrast. Second, among the alphabetic scripts we examined the differences in transparency/opaqueness of the letter-to-sound relationship by a continuous predictor representing the number of grapheme-to-phoneme rules as defined by computationally implemented dual-route models (18). A low number of number of grapheme-to-phoneme rules reflects a high transparency, meaning that letters more consistently represent only one speech sound. For example Italian, Dutch, and German are considered transparent orthographies, with 59, 104, and 130 rules, respectively; see (18)). English and French, in contrast, are typically considered as intransparent with 226 and 340 rules, respectively, because letters map to multiple speech sounds on a regular basis.

Both effects were analyzed using linear mixed models (LMM; (31)). In addition to the parameters of interest, we accounted for the use of different eye-trackers in different labs, sampling rates of the eye-trackers, experimental settings (experiment vs. corpus based studies), and fixation measures presented (mean, single, or first fixation duration) by introducing these parameters into the LMM as fixed effects. In addition, we estimated the random effect on the intercept of study, to take into account unspecific differences between studies. *t*-values larger then 2 were interpreted as significant (cf. (32)).

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2013) under grant agreement n° 617891 awarded to CJF and from the European Community's Horizon 2020 Programme under grant agreement n° 707932 awarded to BG.

Author contributions

B.G., D.P. and C.J.F. designed research; B.G. and S.H. performed the eye-tracking study; B.G. and J.G. performed the meta analysis; B.G. and J.S. analyzed data; and B.G. and C.J.F. wrote the paper. All authors gave comments on the paper during the process.

References

1. Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical Tracking of Hierarchical Linguistic Structures in Connected Speech. *Nat Neurosci* 19(1):158–164.
2. Luo H, Poeppel D (2007) Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron* 54(6):1001–1010.
3. Ding N, et al. (2017) Temporal modulations in speech and music. *Neurosci Biobehav Rev*. doi:10.1016/j.neubiorev.2017.02.011.
4. Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C (2017) A cross-linguistic study of speech modulation spectra. *J Acoust Soc Am* 142(4):1976–1989.
5. Teng X, Tian X, Doelling K, Poeppel D Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. *Eur J Neurosci* 0(0). doi:10.1111/ejn.13742.
6. Kliegl R, Grabner E, Rolfs M, Engbert R (2004) Length, frequency, and predictability effects of words on eye movements in reading. *Eur J Cogn Psychol* 16(1–2):262–284.
7. Liversedge SP, et al. (2016) Universality in eye movements and reading: A trilingual investigation. *Cognition* 147:1–20.
8. Rayner K (2009) Eye movements and attention in reading, scene perception, and visual search. *Q J Exp Psychol* 62(8):1457–1506.
9. Gagl B, Hawelka S, Hutzler F (2014) A similar correction mechanism in slow and fluent readers after suboptimal landing positions. *Front Hum Neurosci* 8:355.
10. Hawelka S, Gagl B, Wimmer H (2010) A dual-route perspective on eye movements of dyslexic readers. *Cognition* 115(3):367–379.
11. Hogendoorn H (2016) Voluntary Saccadic Eye Movements Ride the Attentional Rhythm. *J Cogn Neurosci* 28(10):1625–1635.
12. Gagl B, Hawelka S, Hutzler F (2011) Systematic influence of gaze position on pupil size measurement: analysis and correction. *Behav Res Methods* 43(4):1171–1181.
13. Nuthmann A, Engbert R, Kliegl R (2007) The IOVP effect in mindless reading:

Experiment and modeling. *Vision Res* 47(7):990–1002.
14. Staub A, White SJ, Drieghe D, Hollway EC, Rayner K (2010) Distributional Effects of Word Frequency on Eye Fixation Durations. *J Exp Psychol Hum Percept Perform* 36(5):1280–1293.
15. Smith PL, Ratcliff R (2004) Psychology and neurobiology of simple decisions. *Trends Neurosci* 27(3):161–168.
16. Diedenhofen B, Musch J (2015) cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE* 10(4):e0121945.
17. Chang L-Y, Chen Y-C, Perfetti CA (2017) GraphCom: A multidimensional measure of graphic complexity applied to 131 written languages. *Behav Res Methods*:1–23.
18. Schmalz X, Marinus E, Coltheart M, Castles A (2015) Getting to the bottom of orthographic depth. *Psychon Bull Rev* 22(6):1614–1629.
19. Seymour PHK, Aro M, Erskine JM, collaboration with COST Action A8 network (2003) Foundation literacy acquisition in European orthographies. *Br J Psychol* 94(2):143–174.
20. Rayner K (1978) Eye movement latencies for parafoveally presented words. *Bull Psychon Soc* 11(1):13–16.
21. Gagl B (2016) Blue hypertext is a good design decision: no perceptual disadvantage in reading and successful highlighting of relevant information. *PeerJ* 4:e2467.
22. Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J (2001) DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol Rev* 108(1):204.
23. Alexander JD, Nygaard LC (2008) Reading voices and hearing text: talker-specific auditory imagery in reading. *J Exp Psychol Hum Percept Perform* 34(2):446–459.
24. Yao B, Scheepers C (2011) Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition* 121(3):447–453.
25. Stites MC, Luke SG, Christianson K (2013) The psychologist said quickly, "Dialogue descriptions modulate reading speed!" *Mem Cognit* 41(1):137–151.
26. Vitu F, O'Regan JK, Inhoff AW, Topolski R (1995) Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Percept Psychophys* 57(3):352–364.
27. Rayner K, Fischer MH (1996) Mindless reading revisited: Eye movements during reading and scanning are different. *Percept Psychophys* 58(5):734–747.
28. Gramfort A, et al. (2014) MNE software for processing MEG and EEG data. *NeuroImage* 86:446–460.
29. Share DL (2008) On the Anglocentricities of current reading research and practice: the perils of overreliance on an" outlier" orthography. *Psychol Bull* 134(4):584.
30. Taylor I (1980) The Korean writing system: An alphabet? A syllabary? a logography? *Processing of Visible Language*, Nato Conference Series. (Springer,

Boston, MA), pp 67–82.
31. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 67(1):1–48.
32. Kliegl R, Wei P, Dambacher M, Yan M, Zhou X (2011) Experimental effects and individual differences in linear mixed models: estimating the relationship between spatial, object, and attraction effects in visual attention. *Front Psychol* 1:238.

Supplement

Mean to mode transformation function

Eye fixation durations during reading are typically ex-Gaussian distributed (for an example see Figure 1c). This distribution can be best modeled by combining a normal distribution with an exponential distribution. Therefore, central tendency of this distribution can be best described by three parameters; μ: mean of the normally distributed component; σ: standard deviation of the normally distributed component; τ: parameter reflecting the rightward skew, i.e., representing the contribution of the exponential distribution. For the frequency analysis, which investigates if a reoccurring event (i.e. a saccade) has a temporal structure, the mode of the ex-Gaussian distribution of the sampling durations (fixation plus saccade duration) indicates the most common fixation duration. In the literature on eye tracking of reading, it is more typical to report mean fixation durations. A central problem of the current meta analysis, accordingly, was the transformation of the mean fixation durations extracted from the included papers into mode values. Therefore, we implemented a mean-to-mode transformation function based on empirical datasets in three steps: (i) Ex-Gaussian fitting to existing datasets, (ii) using fitted ex-Gaussian parameters to simulate new informed ex-Gaussian distributions and derive a transformation function, and (iii) optimizing the transformation function.

Ex-Gaussian fitting to existing datasets. In detail, the function was determined as follows: First, we used 29 empirical datasets containing fixation durations (from 11 published studies, i.e., three German studies from our own lab, (1–3), and multiple English studies (4–10) for which datasets were openly available) to fit an ex-Gaussian distribution to the fixation durations of each of the datasets (see Fig. S1a). Henceforth, these 29 datasets are jointly referred to as the 'simulation data'.

Using fitted ex-Gaussian parameters to simulate new informed ex-Gaussian distributions. After fitting, the three parameters describing the ex-Gaussian distribution were obtained for each simulation dataset. In a next step, we estimated one robust linear regression model (*rlm* function in R from the MASS package; (11)) for each of the three ex-Gaussian parameters (μ, σ, τ) where the respective parameter was predicted on the basis of the respective dataset's mean fixation

duration (Fig. S1b; μ: 0.40, SE = 0.09, t = 4.6; σ: 0.14, SE = 0.09, t = 1.5[1]; τ: 0.60, SE = 0.09, t = 6.7).

On the basis of the obtained linear regression coefficients (intercept, beta weight) for the fitted three ex-Gaussian parameter estimates, we are now able to simulate ex-Gaussian distributions (with 500 samples) for any mean value. For example, a mean fixation duration of 200 ms would translate to a μ value of 140, a σ value of 30 and a τ value of 40. With these parameters, one is able to randomly sample values form an ex-Gaussian distribution with exactly these parameters. The sampled values from the ex-Gaussian distributions then allow us to estimate the mode of the distribution, in our case around 150 ms, which allows us to directly relate this mode to the mean value of 200 ms.

To reduce estimation noise and increase robustness against outliers, we sampled ex-Gaussian distributions, not only for the 29 datasets available, but for the whole range of mean fixation durations (149 and 397 ms) that were entered into the meta analysis. From these 248 ex-Gaussian distributions, the mode values were derived connecting each mean to a mode value (e.g. as presented in Fig. S1d). Finally, these mean and mode values allowed us to generate a generalized mean-to-mode transformation function on the basis of only one linear regression (e.g. blue line in Fig. S1d). Note that for the present research question, we on purpose neglected the specific experimental manipulations of the different studies of our simulation data.

Optimizing the transformation function. For an initial quality control, we used the obtained linear regression coefficients (intercept, beta weight) from the transformation function to transform the mean fixation duration of each of the 29 simulation datasets into a mode. Since we were also able to measure the mode of these datasets were able to compare the resulting simulated modes to the measured modes of each dataset directly. The residual errors of the 29 estimated modes, relative to the measured modes in the simulation dataset, for this initial transformation step are presented in Fig. S1c (Level 1). The negative relationship between the measured mode and the residuals of the estimated relative to the empirically measured mode indicated a systematic overestimation for low modes and underestimation for high modes. To account for this systematic error, the simulated modes were corrected by a sequential procedure during which a linear model was used to describe this error in the residuals. The linear regression model

[1] The σ parameter, which represents the standard deviation of the normal distribution of the ex-Gauss distribution, shows the expected positive association with the mean. However, the increase was relatively small, resulting in a non-significant beta-value (i.e., not significantly different from zero). As σ, representing the width of the normal component, is nevertheless essential for the ex-Gauss simulations, for the purposes of the present transformation function, we nevertheless estimated sigma based on the presented regression model.

is then used to predict the over/underestimation of the mean in order to correct it. This was applied two times, and in Fig. S1c (Level 3) shows that in the residuals of the final transformation function no systematic error is present. Fig. S1d, accordingly, shows the final, i.e., corrected, transformation function, for all possible 248 mean fixation durations, and Figure S1e shows the final quality for the simulation dataset. Despite some unsystematic noise, this optimized transformation function showed a near-perfect beta of 0.95 (SE = 0.23; t = 4.1). Data and code for the reported steps in developing and optimizing the transformation function will be made available when published.

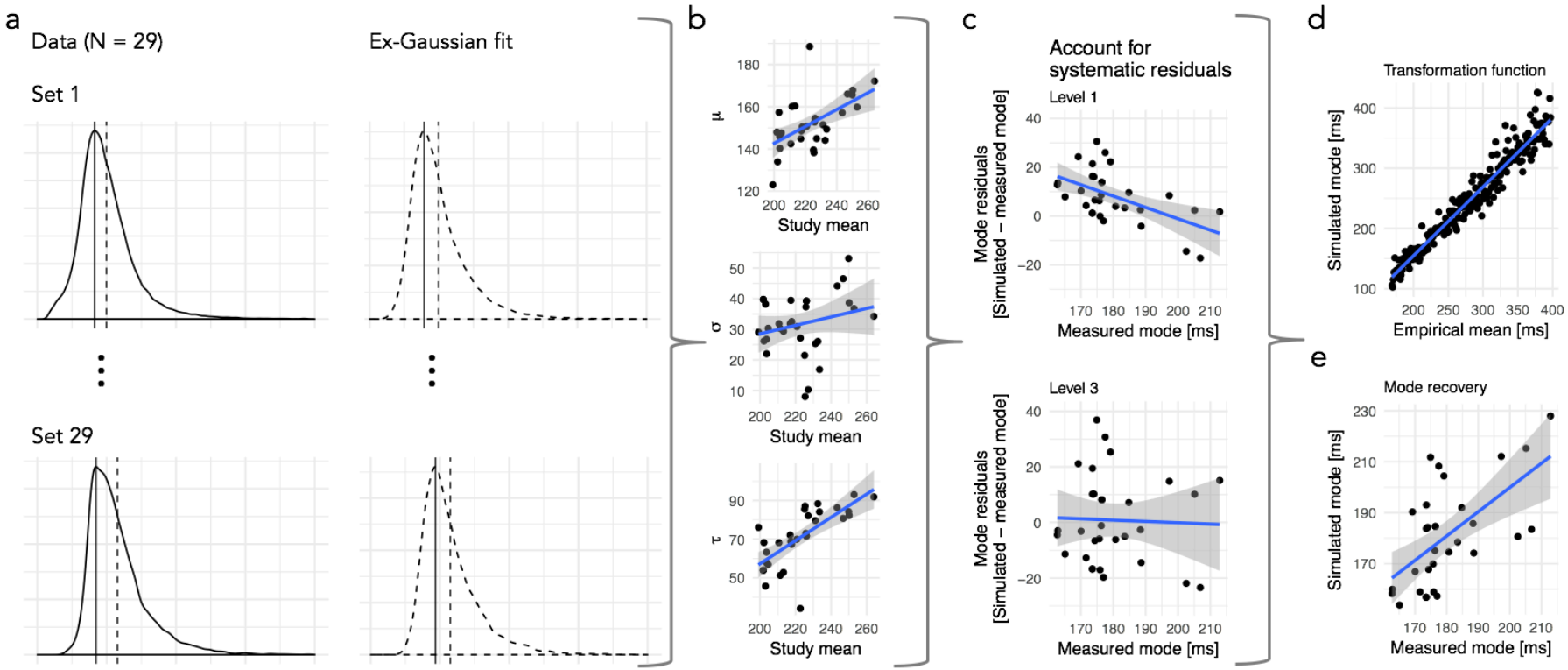
a
Data (N = 29)
Ex-Gaussian fit
Set 1
Set 29
b
μ
σ
τ
Study mean
c
Account for
systematic residuals
Level 1
Level 3
Mode residuals
[Simulated – measured mode]
Measured mode [ms]
d
Transformation function
Simulated mode [ms]
Empirical mean [ms]
e
Mode recovery
Simulated mode [ms]
Measured mode [ms]

Fig. S1. Development of a transformation function for converting mean fixation durations into simulated mode values. In order to estimate the predominant (i.e., mode) fixation duration from a single mean value extracted from a published empirical study, the following procedure was implemented: (a) First, an ex-Gaussian function was fitted separately to empirical fixation duration distributions of each of 29 independent datasets (left panel; see Materials and Methods), and the parameters μ, σ, τ describing the fitted ex-Gaussian distribution (right panel) were extracted. (b) Second, across these 29 datasets, the relationship between the ex-Gaussian parameters and the empirical study means were described by linear models, separately for each parameter. The intercepts and beta-weights resulting from these linear models, for each ex-Gauss parameter, were then used to simulate an ex-Gaussian for each of the 29 empirical mean fixation durations, so that empirical and simulated ex-Gaussian distributions could be compared. (c) Residuals for the mode estimation showed a systematic error, i.e., an overestimation for low modes and underestimation for high modes (upper panel / Level 1). This was sequentially and incrementally accounted for by two linear models describing this systematic error; see lower panel / Level 3 for residuals after accounting for this systematic estimation error. (d) Performance of the final mean-to-mode transformation function (blue line), after correcting for systematic bias, demonstrated for 248 fixation durations, i.e., for each millisecond within the range of the meta analysis (149 – 397 ms). (e) Performance of the final mean-to-mode transformation function, as demonstrated by the relationship between empirically measured and simulated modes from the 29 datasets used for establishing the transformation function.

References S1

1. Gagl B, Hawelka S, Hutzler F (2014) A similar correction mechanism in slow and fluent readers after suboptimal landing positions. *Front Hum Neurosci* 8:355.
2. Gagl B (2016) Blue hypertext is a good design decision: no perceptual disadvantage in reading and successful highlighting of relevant information. *PeerJ* 4:e2467.
3. Gagl B, Hawelka S, Richlan F, Schuster S, Hutzler F (2014) Parafoveal preprocessing in reading revisited: Evidence from a novel preview manipulation. *J Exp Psychol Learn Mem Cogn* 40(2):588–595.
4. Abbott MJ, Angele B, Ahn YD, Rayner K (2015) Skipping syntactically illegal the previews: the role of predictability. *J Exp Psychol Learn Mem Cogn* 41(6):1703–1714.
5. Angele B, Rayner K (2013) Processing the in the parafovea: are articles skipped automatically? *J Exp Psychol Learn Mem Cogn* 39(2):649–662.
6. Angele B, Tran R, Rayner K (2013) Parafoveal-foveal Overlap Can Facilitate Ongoing Word Identification During Reading: Evidence from Eye Movements. *J Exp Psychol Hum Percept Perform* 39(2):526–538.
7. Rayner K, Schotter ER (2014) Semantic preview benefit in reading English: The effect of initial letter capitalization. *J Exp Psychol Hum Percept Perform* 40(4):1617–1628.

8. Rayner K, Yang J, Schuett S, Slattery TJ (2014) The effect of foveal and parafoveal masks on the eye movements of older and younger readers. *Psychol Aging* 29(2):205–212.
9. Schotter ER, Bicknell K, Howard I, Levy R, Rayner K (2014) Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition* 131(1):1–27.
10. Schotter ER, Tran R, Rayner K (2014) Don't Believe What You Read (Only Once) Comprehension Is Supported by Regressions During Reading. *Psychol Sci*:0956797614531148.
11. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* (Springer) Available at: //www.springer.com/de/book/9780387954578 [Accessed December 6, 2017].