Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Using Simulations to Understand the Reading of Rapidly Displayed Subtitles

Permalink

https://escholarship.org/uc/item/77f4h3qn

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN

1069-7977

Authors

Reichle, Erik D
Yu, Lili
Liao, Sixin
et al.

Publication Date

2021

Peer reviewed

Using Simulations to Understand the Reading of Rapidly Displayed Subtitles

Erik D. Reichle (erik.reichle@mq.edu.au)
Department of Psychology, Macquarie University
Sydney, NSW 2109 Australia

Lili Yu (lili.yu@mq.edu.au)
Department of Psychology, Macquarie University
Sydney, NSW 2109 Australia

Sixin Liao (sixin.liao@students.mq.edu.au)
Department of Linguistics, Macquarie University
Sydney, NSW 2109 Australia

Jan-Louis Kruger (janlouis.kruger@mq.edu.au)
Department of Linguistics, Macquarie University
Sydney, NSW 2109 Australia

Abstract

Liao et al. (2020) reported an eye-movement experiment in which subtitles were displayed at three different rates, with a key finding being that, with increasing speeds, participants made fewer, shorter fixations and longer saccades. To understand why these eye-movement behaviors might be adaptive, we completed simulations using the E-Z Reader model (Reichle et al., 2012) to examine how subtitle speed might affect word identification and sentence comprehension, as well as the efficacy of six possible compensatory reading strategies. These simulations suggest that the imposition of a lexical-processing deadline and/or strategy of skipping short words may support reading comprehension in impoverished conditions.

Keywords: eye-movement control; E-Z Reader; reading; strategies; subtitles

Introduction

Given the increasing importance of subtitles in film and educational videos, Liao, Yu, Reichle, and Kruger (2020) recently reported an eye-movement experiment to directly examine how subtitle speed affects comprehension, as well as various on-line indicators of language processing. Some of the key findings of this experiment will be reviewed below, but before doing that, it is important to note that the main goal of this article is to report simulation results, using the *E-Z Reader* model (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Pollatsek, & Rayner, 2012), to more carefully examine how subtitle speed might affect lexical processing and higher-level sentence comprehension. We also report the results of simulations that explore the feasibility of compensatory strategies that readers might adopt to aid understanding of rapid subtitles. Finally, we briefly discuss the broader implications of our results. Before doing all of this, however, we will first review what has been learned from empirical studies on subtitle reading, and how this then relates to specific guidelines that have been provided by government and industry in relation to the rates at which subtitles are displayed. We then provide a brief description of the E-Z Reader model—one that is sufficient to understand the method and implications of our simulation results.

Reading Subtitles in Film

An important aspect of subtitled film is that the text (i.e., the subtitles) is only displayed on screen for a limited period of time. This means that the pace at which people read the subtitles is not entirely under their control, but is constrained by the speed at which subtitles are delivered. *Subtitle speed* is thus an important parameter in subtitling guidelines, and has most often been defined using *words per minute* (*wpm*) or *characters per second* (*cps*). Because the latter measure has been found to be more language independent (Martí Ferriol, 2013), cps is more often used in the subtitling industry.

Although subtitle speed plays an important role in both video comprehension and enjoyment (Mikul, 2014), there is little empirical evidence for the establishment of an optimal subtitle speed. For example, although the *six-second rule*, which suggests that a two-line subtitle with 37 characters per line should remain on screen for six seconds (≈ 12 cps), has been widely adopted as the 'gold standard' (Diaz-Cintas & Remael, 2007), this convention is largely based on some (intuitive) notion of what a typical viewer would find to be a comfortable reading speed, and that would permit the reading of the full subtitle while also allowing time to inspect the image. For that reason, industry standards vary substantially between countries and broadcasting companies (cf., Spain = 15 cps vs. Canada = 20 cps; AENOR, 2012; CAB, 2008). This is problematic because subtitles are increasingly being used in educational videos. For example, in a national survey that investigated the use and perceptions of subtitles among 2,839 students from 15 institutions across the United States, 34.9% of students reported that they "always" or "often" watch subtitles when they are available (Linder, 2016).

The lack of solid evidence about subtitle speed conventions motivated Liao et al. (2020) to conduct an eye-movement experiment in which participants read subtitles displayed at three different rates (12 cps vs. 20 cps vs. 28 cps). In this article, we will focus on three key findings from that study: As the overall subtitle speed increased from 12 cps to 28 cps, the participants made longer saccades, as well as fewer fixations that were shorter in duration. Although these seemingly simple behaviors might not appear to warrant further investigation, as we will demonstrate below, these behaviors are unlikely to reflect simple adjustments to the 'parameters' that control saccadic program or execution (e.g., fewer automatic refixations) but instead likely reflect cognitive 'strategies' (e.g., skipping short words) that are adopted to maximize comprehension under impoverished reading conditions[1]. This conjecture is based on the simulations reported below.

The E-Z Reader Model

The simulations reported below were completed using the E-Z Reader model of eye-movement control during reading (Reichle et al., 2012). This model explains how vision, attention, language processing, and eye movements are coordinated to support skilled reading. The following description is not detailed but is sufficient to understand the simulations reported below. It is also important to note that these reported simulations should be viewed as being exploratory because of both the number and nature of our simplifying assumptions, and our emphasis on reproducing and explaining the qualitative patterns of several key eye-movement behaviors reported by Liao et al. (2020).

Figure 1 is a schematic diagram of the model. As shown, information from the text that is being read (e.g., subtitles) is propagated from a pre-attentive stage of visual processing to the cognitive systems responsible for word identification, sentence processing, and programming saccades. This propagation of information requires some amount of time, $t(V) = 60$ ms, as estimated from the eye-mind lag (Reichle & Reingold, 2013). Some portion of the visual information (e.g., letter features) is then selected by attention for further lexical processing, while other information (e.g., blank spaces between words) is used for selecting saccade targets. By assumption, attention is allocated to only one word at a time, in a strictly serial manner. The word that is attended is then identified across two successive stages, L_1 and L_2, as described by Equations 1-3. L_1 corresponds to a global sense of *familiarity* which indicates that lexical access is imminent, and that the oculomotor system can start programming a saccade to move the eyes to the next word. L_2 then corresponds to the retrieval of a word's phonological, semantic, and syntactic codes, or *lexical access*.

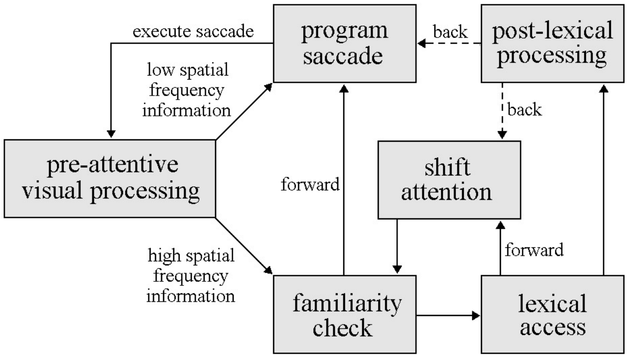

Figure 1: Schematic diagram of E-Z Reader.

As Equation 1 shows, the time (in ms) to complete L_1, $t(L_1)$, for word N is a function of that word's frequency of occurrence in printed text (i.e., $freq_N$, as tabulated in corpora), its predictability within a given sentence context (i.e., $pred_N$, as measured using cloze-predictability), and three free parameters, α_1 (= 115), α_2 (= 2.2), and α_3 (= 13). The model thus predicts that common or predictable words will be processed more rapidly than uncommon or unpredictable words.

$$(1) \qquad t(L_1) = \alpha_1 - \alpha_2 \ln(freq_N) - \alpha_3\, pred_N$$

The time (in ms) required to complete L_2, $t(L_2)$, is then given by Equation 2. As shown, $t(L_2)$ is a fixed proportion, Δ (= 0.22), of $t(L_1)$.

$$(2) \qquad t(L_2) = \Delta \cdot t(L_1)$$

L_1 (but not L_2) is also modulated by visual acuity, as specified by Equation 3. There, the value of $t(L_1)$ given by Equation 1 is updated by multiplying its value (given by Equation 1) by a parameter, ε (= 1.15), that has as its exponent the mean distance (in character spaces) between each of the M letters in the word being processed and the current fixation location (i.e., center of vision). This allows the model to predict that short words or words close to the center of vision will require less time to identify than long words or words far from the center of vision.

$$(3) \qquad t(L_1) \leftarrow t(L_1) \cdot \varepsilon^{\sum_{i=1}^{M} |(fixation - letter_i)/M|}$$

Equations 2 and 3 provide mean values of $t(L_1)$ and $t(L_2)$, respectively, for a word of a given frequency, predictability, and fixation location; the actual values used during any given Monte Carlo simulation are then sampled from gamma distributions having those means and standard

---

[1] Our use of scare quotes acknowledges that we are agnostic about whether readers are consciously aware of these adaptive behaviors (i.e., strategies), or whether they can, e.g., be explicitly taught.

deviations equal to 0.22 of the means. (The latter is true of all gamma distributions used in the model.)

As previously indicated, the completion of L_1 for word N causes the oculomotor system to initiate saccadic programming to move the eyes to word N+1. The subsequent completion of L_2 then causes attention to shift to word N+1, with post-lexical processing of word N then continuing. The remainder of the model's assumptions are thus related to saccade programming and execution, the shifting of attention, and post-lexical processing. These will now be briefly discussed in turn.

E-Z Reader assumes that saccades are programmed in two stages: an initial *labile stage*, M_1, that can be cancelled by the initiation of a subsequent saccade, followed by a *non-labile stage*, M_2, that cannot be cancelled. The mean times to complete M_1 and M_2 are respectively given by $t(M_1) = 125$ ms and $t(M_2) = 25$ ms, with the actual times during any given Monte Carlo simulation being sampled from gamma distributions. The time to physically move the eyes from one viewing location to the next, S, is for the sake of simplicity constant, $t(S) = 25$ ms.

The remaining assumptions about saccades are related to their execution. By assumption, saccades are directed towards the center of the upcoming word, word N+1, because this *optimal viewing position* (*OVP*) affords efficient lexical processing (O'Regan, 1992). But the saccade that is executed will be a function of the *intended saccade length* (*ISL*), or the distance (in character spaces) between the launch-site fixation and the OVP of the targeted word, as well as both *systematic* and *random* motor error, as shown by Equation 4.

$$saccade = ISL + systematic + random \tag{4}$$

The systemic error is given by Equation 5, where Ψ (= 7 characters) is the saccade length that, in English, neither over nor undershoots the intended saccade. For each character space greater/less than Ψ, the executed saccade will under/overshoot its target by an amount given by the righthand term of Equation 5; this quantity is a function of two scaling parameters, Ω_1 (= 6) and Ω_2 (= 3), and the fixation duration on the launch site word. The bias is thus on average about half a character space, but with the precise amount being reduced following longer fixations.

$$systematic = (\Psi - ISL)\left[\frac{\Omega_1 - \ln(duration)}{\Omega_2}\right] \tag{5}$$

The random error component in Equation 4 is sampled from a Gaussian distribution with $\mu = 0$ and a standard deviation that increases with the ISL, as described by Equation 6, where η_1 (= 0.5) and η_2 (= 0.1) are free parameters. Longer saccades are thus more prone to error.

$$\sigma = \eta_1 + \eta_2 ISL \tag{6}$$

Upon fixating a word, an efference copy of the intended saccade is compared to the actual saccade and any discrepancy has a probability p of initiating the programming of a corrective saccade to move the eyes to a better viewing location, the OVP. This tendency is described by Equation 7. As shown, the propensity to refixate increases with the absolute distance (in character spaces) between the initial fixation and OVP, scaled by the parameter λ (= 0.25). This assumption allows E-Z Reader to explain why refixations are more likely after initial fixations near the beginning or ending of a word, as well as the inverted-OVP effect (Vitu, McKonkie, Kerr, & O'Regan, 2001).

$$p = \max(\lambda|fixation - OVP|, 1) \tag{7}$$

The time required to shift attention from one word to the next, $t(A)$, is sampled from a gamma distribution with $\mu = 25$ ms. Similarly, the time required to integrate a word's meaning into the overall sentence representation, $t(I)$, is also sampled from a gamma distribution with $\mu = 25$ ms. Although the model does not specify how this post-lexical integration process actually happens, it can fail in two ways, both causing the eyes and attention to be directed back to the location (word) of integration failure. The first type of integration failure is assumed to reflect factors related to sentence processing (e.g., syntactic mis-parsing; Frazier & Rayner, 1982) and occur with some small probability, $p_F = 0.01$. The second type of failure occurs whenever word_{N+1} has been identified (i.e., L_2 has completed for word_{N+1}) before word_N has been integrated. By assumption, this situation is problematic because word_{N+1} cannot itself be integrated without the prior integration of word_N. Although this might suggest that this second type of integration failure is modulated by gross differences in the rate of lexical processing (and thus reading skill; Reichle et al., 2013), integration failure only occurs if: (1) $t(I) - t(A) > 0$ (to allow preview of word_{N+1}); and (2) $t(L_1) + t(L_2)$ for $\text{word}_{N+1} < t(I) - t(A)$. These constraints mean that this second type of integration failure is only modestly affected by reading skill. Finally, the use of sentence context to aid the identification of word_N (i.e., its cloze predictability; Equation 1) is conditional upon word_{N-1} having been successfully integrated; if word_{N-1} has not been integrated prior to starting lexical processing of word_N, then the value of $pred_N$ in Equation 1 is set equal to 0.

Simulation Method & Results

E-Z Reader was used to complete the three simulations reported below. Unless otherwise indicated, the simulations were completed using 1,000 statistical subjects per condition and the model's default parameter values. The simulations were also completed using the 48 sentences from the Schilling, Rayner, and Chumbley (1998) corpus. The rationale for this was twofold. First, participants in this study were moderately skilled readers (i.e., university undergraduate students) who silently read the sentences at their own pace, thereby allowing good comprehension of the sentences (mean accuracy > 0.95). Second, the model's default parameter values have been previously selected to maximize the goodness-of-fit to the corpus (see Reichle et al., 2012, Appendix B). The corpus thus provides an ideal baseline of 'normal' reading, against which the effects of sentence (subtitles) presentation rates and/or reading skill can be evaluated. Finally, the simulations excluded trials in which an inter-word regression occurred due to a failure of post-lexical integration; this precaution means that the simulations thus provide conservative estimates of how well the sentences were understood as a function of subtitle speed because they correspond to situations wherein integration was completed accurately and rapidly enough to *not* interrupt the forward progression of the eyes.

Simulation 1

The model was first used to examine the possible consequences of Liao et al.'s (2020) subtitle speed manipulation on eye movements, word identification, and sentence comprehension. This was done by running the model on the Schilling et al. (1998) sentences, but halting each simulation after a time interval corresponding to how long each sentence would have been displayed had it been a subtitle presented at 12 cps, 20 cps, or 28 cps. For example, the sentence *Erik took his sick parakeet to the veterinarian on Tuesday.* is 59 characters in length, including the spaces between words and the period at the end of the sentence. At presentation rates of 12 cps, 20 cps, and 28 cps, the sentence would thus be displayed for 4,917 ms, 2,950 ms, and 2,107 ms, respectively. The model was allowed to read the sentence for each of these intervals, and each time two dependent variables were recorded: (1) the proportion of words in the sentence that were identified (i.e., processed to the completion of L_2); and (2) the proportion of words that were successfully integrated (i.e., processed to the completion of I). The mean values of these two measures were calculated across sentences and statistical subjects to determine how the three subtitle speeds might be expected to affect word identification and comprehension. Note that the second of these two measures likely overestimates sentence comprehension because each word only receives minimal post-lexical processing and not all of the linguistic processing required to fully understand a text (for discussion of these processes, see Reichle, 2021). This overestimate may be attenuated, however, by our assumption that lexical and post-lexical processing abruptly halts with the removal of the subtitles. This assumption is a simplification because evidence indicates that reading can proceed even if each fixated word only remains visible for 60 ms (Liversedge et al., 2004), the time required for visual word information to be converted into some type of more stable representation.

Finally, previous simulations have shown that the patterns of eye movements exhibited by less skilled (e.g., beginning) readers can be simulated within the model by increasing the values of a single parameter, α_1, to slow lexical processing (Reichle et al., 2013). Because the model's default parameter values were selected to simulate the reading of moderately skilled readers (i.e., university students in the Schilling et al., 1998 study), it was important to examine how our manipulation of subtitle speed might interact with reading skill. This was done by repeating the simulation as described using three different values of α_1: (1) the default value, α_1 = 115; (2) a value (α_1 = 138) that would slow lexical processing by 20% to simulate less skilled readers; and (3) a value (α_1 = 92) that would speed lexical processing by 20% to simulate highly skilled readers.

Table 1 shows the results of Simulation 1. As shown, subtitle speed adversely affected word identification and comprehension, but these effects were largely limited to the fastest subtitle speed, particularly with the less skilled readers. At 28 cps, the values of α_1 that respectively correspond to the highly, moderately, and less skilled readers resulted in 0.98, 0.91, and 0.81 of the words being identified. Similarly, using the same respective values of α_1, the model predicted that 0.96, 0.82, and 0.67 of the words would be successfully integrated. That the model was less successful at integrating than identifying words was expected due to the fact that word integration lags behind and is dependent upon word identification. Note, however, that the relative disparity between these two measures increases as reading skill declines, being negligible for highly skilled readers (= 0.02) but pronounced for less skilled readers (= 0.14). It is important to note that, because the model was 'calibrated' on moderately skilled readers and the simulations excluded trials containing inter-word regressions, the simulations likely underestimate how subtitle speed affects comprehension. Finally, because the Schilling et al. (1998) sentences averaged 11.17 words in length (with 8-14 words per sentence), the simulated less-skilled readers in the 28-cps condition on average failed to identify approximately 2.12 words and failed to integrate 3.69 words.

Table 1: Mean (standard deviation in parentheses) proportion of lexical and post-lexical processing completed as a function of reading skill (i.e., values of α_1) and subtitle speed (in *characters per second*, or *cps*).

α_1	Processing Completed	Subtitle Speed (cps)		
		12	20	28
92	Lexical	1.00 (0)	1.00 (0)	0.98 (0.06)
	Post-Lexical	1.00 (0)	1.00 (0.01)	0.96 (0.11)
115	Lexical	1.00 (0)	1.00 (0.01)	0.91 (0.11)
	Post-Lexical	1.00 (0)	1.00 (0.02)	0.82 (0.19)
138	Lexical	1.00 (0)	0.99 (0.03)	0.81 (0.14)
	Post-Lexical	1.00 (0)	0.98 (0.07)	0.67 (0.19)

Simulation 2

Liao et al.'s (2020) experiment indicates that readers do not simply read each subtitle from its beginning until it disappears, but instead seem to adopt 'strategies' or visual routines (Ullman, 1984) that presumably allow them to compensate for the dynamic nature of how the text is displayed. More specifically, as the subtitle presentation rate increased from 12 cps to 28 cps, three key findings were that: (1) the fixation durations decreased; (2) the number of fixations decreased; and (3) the saccade length increased.

Given that these eye-movement behaviors may have been compensatory, allowing the participants to maintain some minimal level of comprehension under otherwise difficult conditions, it is important to understand *how* these behaviors might support comprehension.

The second simulation examined this issue by first using E-Z Reader to simulate the reading of the Schilling et al. (1998) sentences using the model's default parameter values under normal conditions (i.e., unlimited sentence-viewing times). This 'baseline' of performance (see Table 2) was used to generate predictions about three dependent measures: (1) mean fixation duration; (2) mean number of fixations per sentence; and (3) mean saccade length. We then tested specific hypotheses about how these measures might be modulated by subtitle speed by systematically manipulating the model's parameters to determine if the model would reproduce the qualitative pattern reported by Liao et al. (2020) (i.e., fewer, shorter fixations and longer saccades). (This basic approach has previously been used to test possible accounts of, e.g., skill-related differences in readers' eye movements; Reichle et al., 2013.)

Table 2: Three eye-movement measures of reading performance as a function of six possible eye-movement strategies, with asterisks denoting values that are qualitatively correct (relative to the default 'strategy').

Strat-egy	Para-meter Values	Summary Explana-tions	Mean Eye-Movement Measures		
			# of Fixa-tions	Fixation Duration (ms)	Saccade Length (char-acters)
N/A	default	baseline	8.37	230	6.38
1	λ = 0.125	fewer refixations	7.63*	249	6.99*
2	η_1 = 0.25, η_2 = 0.05	more accurate saccades	7.89*	241	6.77*
3	Ψ = 10.5	longer preferred saccades	8.72	221*	6.13
4	θ = 92	lexical-processing deadline	8.09*	218*	6.61*
5	default	skip short words	8.97	217*	5.96
6	default	skip short words & $t(L_1) = 0$ ms	6.66*	227*	7.88*

The first three strategies are oculomotor in nature because they entail the adjustment of parameters that control saccadic programming. As Table 2 shows, the first strategy corresponded to the hypothesis that, as subtitle speed increases, readers make fewer refixations. This strategy was implemented by reducing the value of the parameter that controls the propensity to making automatic refixations, λ (= 0.25; Equation 7) to 50% of its default value (i.e., $\lambda = 0.125$). As shown, this change reduced the number of fixations and increased the saccade lengths (both consistent with what Liao et al., 2020 reported with faster subtitles), but also increased fixation durations (inconsistent with Liao et al.) due to an increased need to identify words from suboptimal viewing locations.

The next strategy corresponded to the hypothesis that, as subtitle speed increases, readers make more accurate saccades. This was done by reducing the values of two parameters that modulate the random saccadic error ($\eta_1 = 0.5$ and $\eta_2 = 0.1$; Equation 6) to 50% of their default values (i.e., $\eta_1 = 0.25$ and $\eta_2 = 0.05$). As Table 2 shows, this second strategy also resulted in fewer fixations, longer saccades, but longer fixation duration. This likely reflects a trade-off between the benefit that comes from processing words near their OVPs and the cost that comes from the fact that these viewing locations will (on average) tend to be farther from the OVPs of subsequent words.

The final oculomotor strategy corresponded to the hypothesis that readers might also reduce saccadic error by increasing the preferred saccade length by making (more accurate) longer saccades. This was done by increasing the value of the Ψ (= 7; Equation 5) by 50% of its default value (i.e., $\Psi = 10.5$). As Table 2 shows, this resulted in shorter fixations, but also increased the number of fixations and reduced saccade length. This result again likely reflects complex trade-offs; although longer saccades are more accurate (e.g., exhibit less systematic error), the distribution of word lengths in English likely means that the preferred saccade length of seven characters is already optimal for the task of reading English efficiently.

The next three strategies listed in Table 2 are perhaps more 'cognitive' in that they entail parameters that control lexical processing and/or more complex visual routines (Ullman, 1984). For example, strategy 4 tests the assumption that, as subtitle speed increases, readers adopt a threshold or deadline (represented by a new parameter, θ) for the maximal time spent engaged in the lexical processing of a word. Thus, if $t(L_1)$ exceeds θ ms, then the first stage of lexical processing terminates, allowing the second stage to continue using whatever information was made available (which may be erroneous) and the initiation of a saccadic program to move the eyes forward. A value of $\theta = 92$ ms was used because this corresponds to the mean maximum $t(L_1)$ duration of a skilled reader in Simulation 1 (i.e., using $\alpha_1 = 92$ for Equation 1). As shown, this strategy reproduced the pattern of eye movements observed by Liao et al. (2020), with fewer, shorter fixations and longer saccades.

Although a lexical-processing deadline has been previously suggested (e.g., Henderson & Ferreira, 1990), its plausibility here might be questioned on the grounds that it is only being posited to explain Liao et al.'s (2020) findings. Furthermore, the implications of its auxiliary assumption—that the second stage of lexical processing can continue using whatever information was obtained from an abbreviated first stage—are also unclear, although it is reasonable to assume that words processed in this manner would likely be prone to misidentification and thus reduce overall comprehension. The final simulations therefore explore the efficacy of two other possible strategies related to saccadic targeting during reading.

These two strategies entail using parafoveal word-length information to skip short (1-3 letter) words, but with strategy 6 also assuming that readers do not identify the words per se but instead generate predictions about their likely identities using semantic and syntactic constraints (i.e., $t(L_1) = 0$ ms for these words). Although one might question the plausibility of this final strategy, short words tend to be function words (and thus highly predictable), and there is evidence that such words are skipped even under conditions wherein a word is only visible if it is directly fixated (e.g., using a 1-word moving-window paradigm; Rayner, Well, Pollatsek, & Bertera, 1982).

As Table 2 shows, the final strategy is promising; by adopting the assumption that readers skip short words and infer their likely meaning from context, the model generates the pattern of eye movements reported by Liao et al. (2020): fewer, shorter fixations and longer saccades. In our final simulation, we will examine how this 'skimming' strategy might affect comprehension.

Simulation 3

We examined how the skimming strategy might support word identification and integration by running a final simulation in which the model was used to simulate a less skilled reader (i.e., $\alpha_1 = 138$) in the most rapid subtitle condition (i.e., 28 cps). As Table 3 shows, this simulation was completed twice: with and without using the skimming strategy.

Table 3: Performance of a simulated less-skilled reader in the most rapid subtitle-speed condition (28 cps) with and without the 'skimming' strategy (#6) from Simulation 2.

Mean Performance Measures	'Skimming' Strategy?	
	No	Yes
Number of Fixations	8.86	7.10
Fixation Duration (ms)	247	240
Saccade Length (character spaces)	6.09	7.52
Prop. of Words Identified	0.81	0.96
Prop. of Words Integrated	0.67	0.91

As shown, the strategy allowed the model to generate the pattern of eye movements reported by Liao et al. (2020), with fewer, shorter fixations and longer saccades. Moreover, this strategy increased the mean proportion of words identified (by 0.15) and integrated (by 0.24). However, it is important to acknowledge that these estimates of how much the skimming strategy enhanced performance are inflated due to our assumption that short words can simply be 'guessed' from context, with no negative repercussions for sentence- or discourse-level comprehension. But even if this assumption is only approximately correct, it suggests a simple strategy that *might* be explicitly taught to readers to support their comprehension of subtitles that might otherwise be displayed too rapidly to support normal reading.

General Discussion

The simulations reported above were completed using E-Z Reader, a model that has been used to simulate and explain a variety of phenomena related to eye-movement control during the reading of static text (Reichle et al., 2012; for a review, see Reichle, 2011). Although we contend that the model's application to dynamic text in the present context has been informative, we also acknowledge that the simulations only address half of the film-viewing experience—the reading of subtitles. A complete model of multimodal reading obviously requires additional assumptions about, for example, how characters and/or objects in a film are identified and tracked over time, how information from the auditory modality (e.g., spoken dialog and other sounds) is integrated with the subtitles and visual elements of the film, and so on. We are currently working to develop such a theoretical framework.

One example to illustrate this point is the simple fact that the contents of the videos, by virtue of being at least partially redundant with the contents of the subtitles, should enhance the predictability of many of the words in the subtitles[2]. The resulting increase in word predictability would be particularly beneficial for identifying the more difficult, low-frequency content words. For example, in the context of watching a documentary about polar bears, the images of polar bears would make the low-frequency and hence relatively unlikely words "polar" and "bears" more predictable, thereby facilitating their identification. To the extent that this facilitation reduces the times that would otherwise be required to identify the more difficult words in the subtitles, one might posit that the enhancement of word predictability would result in the same behaviors that were generated by our simulated skimming strategy.

To test this hypothesis, we completed one final simulation in which the cloze predictability of all completely unpredictable words (i.e., words with $pred_N = 0$; see Equation 1) were set equal to 0.2. This change was sufficient to produce the three key eye-movement behaviors reported by Liao et al. (2020): relatively to normal reading (i.e., see the baseline condition in Table 2), the mean number of fixations decreased (= 7.76 fixations), the mean fixation duration decreased (= 226 ms), and the mean saccade length increased (= 6.82 character spaces). This final simulation thus suggests that the eye-movement behaviors observed by Liao et al. with

---

[2] We wish to acknowledge one of our anonymous reviewers for making this suggestion.

rapid subtitles may have reflected an increased tendency to rely upon the contents of the videos to guess the identities of (at least some of) the more difficult-to-identify words.

Given this final simulation result, one might ask how an increased reliance upon video contents to predict words (as just described) differs from our skimming strategy (as discussed earlier)? The key difference is that the former mainly entails using the video to aid the identification of difficult content words, whereas the latter mainly entails ignoring short function words so as to dedicate more time towards the processing of content words. Of course, these two ways of compensating for rapid subtitles are not mutually exclusive, and both might be used by some readers. Future work will be required to evaluate this hypothesis.

Finally, despite the many acknowledged limitations of our work, we contend that our simulations, in combination with Liao et al.'s (2020) empirical findings, collectively demonstrate that subtitle speed is important and cannot simply be ignored, particularly if the intended audience consists of less skilled readers (e.g., students watching educational video in a second language). The practical significance of this conclusion cannot be overstated, particularly when education has become so reliant upon the use of subtitled videos.

Our simulations also suggest that readers may use simple strategies to enhance their comprehension of rapidly presented subtitles. On some level, this almost has to be true, although it is equally true that relatively little is known about these strategies. Many basic questions (e.g., Are these strategies consistent across readers? How do the strategies actually support comprehension? Can the strategies be taught?) are certainly worthy of future investigation. We hope that the work reported here sparks interest in this theoretically interesting and pedagogically important topic.

References

AENOR. (2012). *Subtitulado para personas sordas y personas con discapacidad auditiva* [*Subtitling for the deaf and hard of hearing*], Madrid: AENOR.

CAB (Canadian Association of Broadcasters). (2008). *Closed Captioning Standards and Protocol for Canadian English Language Television Programming Services*.

Díaz Cintas, J. & Remael, A. (2007). *Audiovisual translation: Subtitling*. Manchester: St Jerome.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Henderson, J. M. & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 417-429.

Liao, S., Yu, L., Reichle, E. D., & Kruger, J.-L. (2020). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading*. Advanced online publication, DOI: 10.1080/10888438.2020.1823986.

Linder, K. (2016). *Student uses and perceptions of closed captions and transcripts: Results from a national study*. Corvallis, OR: Oregon State University Ecampus Research Unit.

Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. W. (2004). Eye movements when reading disappearing text: Is there a gap effect in reading? *Vision Research*, *44*, 1013-1024.

Martí Ferriol, J. L. (2013). Subtitle reading speeds in different languages: the case of Lethal Weapon. *Quaderns*, *20*, 201-210.

Mikul, C. (2014). *Caption quality: International approaches to standards and measurement*. Media Access Australia, Sydney.

O'Regan, J. K. (1992). Optimal view position in words and the strategy-tactics model of eye movements in reading. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading*. New York, NY, USA: Springer-Verlag.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effects of word frequency, word predictability, and font difficulty on the eye movements of young and elderly readers. *Psychology and Aging*, *21*, 448-465.

Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation during reading. *Perception & Psychophysics*, *31*, 537-550.

Reichle, E. D. (2011). Serial attention models of reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook of eye movements* (pp. 767-796). Oxford, UK: Oxford University Press.

Reichle, E. D. (2021). *Computational models of reading: A handbook*. Oxford, UK: Oxford University Press.

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*, 110-149.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z Reader to simulate eye movements in non-reading tasks: A unified framework for understanding the eye-mind link. *Psychological Review*, *119*, 155-185.

Reichle, E. D. & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, *7:361*.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*, 1270-1281.

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97-159.

Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research*, *41*, 3513-3533.