

THIS IS AN EARLIER VERSION OF THE MANUSCRIPT. FOR THE FINAL VERSION, PLEASE CHECK THE JOURNAL WEBSITE: The published version of this paper should be considered authoritative, and any citations or page references should be taken from it:
<https://journals.sagepub.com/home/slr>

What eye-tracking tells us about reading-only and reading-while-listening in a first and second language

Kathy Conklin^{1,2}, Sara Alotaibi¹, Ana Pellicer-Sánchez³, and Laura Vilkaitė-Lozdienė⁴

¹ University of Nottingham, ² Max Planck Institute for Psycholinguistics, ³University College London, ⁴Vilnius University

Abstract

Reading-while-listening has been shown to be advantageous in second language learning.

However, research to date has not addressed how the addition of auditory input changes *reading* itself. Identifying how reading differs in reading-while-listening and reading-only might help explain the advantages associated with the former. The aim of the present study was to provide a detailed description of reading patterns with and without audio. To address this, we asked first (L1) and second (L2) language speakers to read two passages (one in a reading-only mode and another in a reading-while-listening mode) while their eye movements were monitored. In reading-only, L2 readers had more and longer fixations (i.e. slower reading) than L1 readers. In reading-while-listening, eye-movement patterns were very similar in the L1 and L2. In general, neither group of participants fixated the word that they were hearing, although the L2 readers' eye movements were more aligned to the auditory input. When reading and listening were not aligned, both groups' eye movements generally preceded the audio. However, L2 readers had more cases where their fixations lagged behind the audio. We consider why reading slightly ahead of the audio could explain some of the benefits attributed to reading-while-listening contexts.

Keywords

eye-tracking; reading; reading-while-listening; first language; second language

Introduction

While reading is a fairly recent development in human history, existing for only a few thousand years (Immordino-Yang and Deacon, 2007), it has become an essential life skill in modern society. For second language learners, reading is a gateway to learning new vocabulary, more colloquial language, and new grammatical constructions (Wilkinson, 2012). Because of its importance, teachers ask students to read in, and outside of, the classroom. Researchers and teachers have explored ways of supporting the reading process and maximizing its learning potential, and have suggested that combining reading and auditory input aids comprehension and leads to larger learning gains. In this context, second language (L2) learners may encounter audio books, in which they are asked to listen to a story while following along with the text. Indeed, researchers have demonstrated that reading-while-listening is beneficial for comprehension and fluency (Chang 2009; Chang and Millet, 2014, 2015; Lightbown, 1992; c.f. Woodall, 2010, who found that it only aided comprehension and perceived fluency, but not actual fluency) and incidental vocabulary learning (Brown, Waring and Donkaewbua, 2008; Webb and Chang, 2012; 2014), as well as providing a more positive experience for learners (e.g. Brown et al., 2008; Chang, 2009; Lightbown, 1992; Tragant, Muñoz and Spada, 2016; Tragant and Vallbona, 2018). While it appears that reading-while-listening is beneficial for L2 learners, applied linguists know very little about how *reading* itself is impacted by the simultaneous presentation of audio. Having a better understanding of how reading patterns change with the presence of auditory input might help researchers and L2-theorists understand the advantages conferred by audio in such reading contexts. With this in mind, the aim of the current study is threefold: (1) to compare reading behavior in reading-only and reading-while-listening conditions; (2) to investigate differences in reading in a first language (L1) and L2 in the two conditions; and (3) to examine how eye movements align with the written input in a reading-while-listening condition. In order to contextualize the current research, we first present an overview of L1 and L2 reading and eye movements. We then look at some research on the benefits of reading-while-listening.

Eye-tracking reading

Reading involves a series of eye movements (*saccades*) that are sometimes back to a previously read part of a text (*regressions*), and brief pauses (*fixations*). Eye-tracking technology tells us where people's eyes fixate when they read, how many times they land in that position or region (*fixation/regression count*), and how long each fixation lasts (*fixation duration*), as well as measuring saccade duration and length. In reading, not all words are directly fixated, which is referred to as *skipping*. It appears that readers only directly fixate about 70% of the words in a text and skip the other 30% (Schotter, Angele and Rayner, 2012; and see Rayner, 2009 for a comprehensive overview of the factors that lead to skipping).

In reading, eye location provides an index of attention (Rayner, 2009), meaning that people's eyes can indicate what they are paying attention to. The number and length of fixations provide an indication of how much effort is being expended to process the input at the fixation point. Thus, eye-movement patterns provide an index of the difficulty and complexity of what people are looking at while reading (Castelhano and Rayner, 2008). Importantly, for the current study, factors other than difficulty can influence fixations, regressions and saccades (Rayner and Pollatsek, 1989). For example, readers' goals—reading a text for understanding versus skimming a text—affect the pattern of eye-movement behavior. Similarly, there may be key differences between reading a text with and without simultaneous audio, which is a situation that arises when L2 learners encounter a text that has an accompanying audio version. In fact, Rayner (2009) wrote that it is somewhat hazardous to generalize across tasks in terms of eye-movement behavior.

Fixations tend to be longer and saccades shorter when reading aloud, as compared to silent reading (oral reading mean fixation duration: 275-325 milliseconds (ms) and mean saccade

length 6-7 letter spaces, versus silent reading: 225-250 ms and 7-9 letter spaces; Castelhana and Rayner, 2008). This is because skilled readers can read words silently more quickly than they can say them aloud, and to prevent their eyes from getting too far ahead of what they are saying, people fixate longer and make shorter saccades when reading aloud (Laubrock and Kliegl, 2015). A question that arises is whether the same is true in reading-while-listening. Are fixations longer and saccades shorter so that readers' eyes align with the speed of another person's word production? If eye movements are synchronized with speech production, it would mean that learners are getting two potentially valuable sources of information simultaneously, visual and auditory. In fact, it has been claimed that audio-visual synchrony can have a positive effect on the development of literacy skills, by boosting orthographic and semantic learning (Gerbier, Bailly and Posse, 2018). If, however, fixations do not align with the audio, it could indicate that the comprehender is not exploiting both sources of information, potentially because the redundant information may add to the cognitive load of the task (Diao and Sweller, 2007).

Importantly, eye-tracking has provided considerable insight into silent L1 reading. In general, more and longer fixations and regressions, less skipping, and/or shorter saccades indicate that more processing effort is needed, while fewer and shorter fixations and regressions, more skipping, and/or longer saccades indicate that less processing effort is being expended. Based on this, one would expect that when reading the same text in a L1 and L2, there would be more and longer fixations, more regressions, less skipping and shorter saccades in the L2 because it is more 'difficult' than the L1. While there is some evidence for this, studies in which authors directly compared eye movements when reading authentic texts in the L1 and L2 are limited. A notable exception is the work by Cop, Drieghe and Duyck (2015). They explored word reading of all of the words in an authentic text by monolingual English speakers (in their L1) and non-natives of English (in their L1 Dutch and L2 English). Participants read an Agatha Christie novel in four sessions while their eye movements were recorded. The monolinguals read the book

entirely in English, while the non-native participants read chapters one to seven in one language, and eight to thirteen in the other. The non-native speakers had more fixations, shorter saccades and less word skipping in their L2 than in their L1, while their regression rates were the same in the two languages. There were no important differences in reading patterns for Dutch speakers in their L1 and monolingual English speakers in English. The evidence from Cop and her colleagues demonstrated that in silent reading, eye-movement patterns are different in a L2 and a L1, and this could be taken as an indication of more effortful reading in a L2.

A number of eye-tracking measures have been used in eye-tracking research to explore reading performance. In the current study, we selected a range of well-established reading measures to gain a good overview of reading in a L1 and L2, as well as of reading-only versus reading-while-listening. Here we briefly introduce these measures, but a more detailed discussion of eye-tracking measures can be found in Conklin, Pellicer-Sánchez and Carrol (2018) and Godfroid (2020). We looked at *first fixation duration*, which refers to the length of the first fixation made on a word or region of interest (ROI). *First pass reading time* considers the duration of all fixations made on a word or ROI before the gaze exits (to the left or right). These measures are thought to provide an index of lexical access, or how easily words are recognized and retrieved from the mental lexicon. We also considered skipping behavior: the chance that a word will not receive a fixation during first pass reading. The summary data provided in Table 2 is averaged performance: the probability (or percentage) of skipping, calculated as the total number of trials where a word is skipped during the first pass, divided by the total number of trials (which can be multiplied by 100 to give a percentage). The models reported in the paper consider skipping as a bimodal variable (1=skipped, 0= not skipped) and predict the probability of a word being skipped, which we refer to as the *probability of skipping*. *Regression count* is the number of visits to a previously viewed ROI. *Total reading time* is the sum of all fixations made on a word or ROI. It includes both first fixation/gaze durations and any subsequent re-reading. In addition

to assessing the duration of all of the fixations, we also looked at *total fixation count*, which is the total number of fixations to an ROI. These last two measures index the initial retrieval of a word from the lexicon as well as its subsequent integration into the larger context.

Reading-only versus reading-while-listening

Books with accompanying audio are commonplace in the L2 classroom, and nowadays L2 learners have access to numerous language learning apps outside of the classroom that combine written and auditory input. Thus, learners in a variety of contexts will be familiar with bimodal (written + audio) input. Crucially, reading-while-listening is thought to aid learning in a number of ways, to have advantages over unimodal input, and to be engaging for learners (see Tragant and Vallbona, 2018).

There is considerable research demonstrating that combining written and auditory input is beneficial for L2 learners (for a notable exception see Diao and Sweller, 2007). Lightbown's (1992) research on an extensive reading-while-listening program by young L2 learners demonstrated initial benefits in terms of comprehension, receptive vocabulary knowledge and some measures of oral production (see also Trofimovich, Lightbown, Halter and Song, 2012). Similarly, Chang (2009) demonstrated better comprehension by L2 participants after reading-while-listening than reading-only. Research by Chang and Millet (2014, 2015) showed that L2 learners had better listening fluency and comprehension for texts presented in a reading-while-listening mode compared to a reading-only mode. A study by Brown et al. (2008) demonstrated more learning of new words from context in reading-while-listening than reading-only and in turn from reading-only than from listening-only at an immediate post-test. Other work has demonstrated that reading-while-listening is more advantageous than reading-only for single word learning (Webb and Chang, 2014). Webb and Chang (2012) suggested that the use of both

auditory and visual input helps learners link the written and spoken forms of words, which could contribute to greater vocabulary learning because of the increased associative links between form (phonological and orthographic) and meaning.

A recent study using eye-tracking showed that the addition of auditory input changes the way learners read a text. Pellicer-Sánchez et al. (2018) presented children (aged 11 -12 years old) and adult L1 and L2 readers with an illustrated text in reading-only and reading-while-listening modes. When auditory input was present there was less reading of the text (fewer and shorter fixations) and more time was spent looking at the images relative to the reading-only mode. The different input modes did not lead to differences in comprehension for either the child or adult learners. Similar patterns were found by Serrano and Pellicer-Sánchez (2019) when children (aged 11 -12 years old) read an authentic, illustrated graded reader in their L2. However, the authors of the two studies did not look in depth at the difference in eye-movement patterns to the text in the two input modes, which is the focus of the current study.

We have discussed some of the advantages conferred by reading-while-listening, as well as possible explanations for them. Prior research has shown that reading-while-listening is advantageous because the spoken words provide auditory support for the visual input. This helps learners segment individual words from the continuous speech stream (i.e. visual spaces between written words delineate word boundaries). It also helps them match spoken and written forms and develop letter-sound correspondences. In the case of collocations and potentially other types of multiword units, the audio may help learners segment the text into larger meaningful chunks. Chang and Millet (2015) suggested that appropriately adjusting the rate of the auditory presentation can help develop reading fluency. Finally, Tragant and colleagues (2016) pointed out that learners inevitably vary in their ability to process spoken and written text; presenting a

text in both auditory and visual modalities allows them to approach the tasks according to their own strengths.

A number of the studies on reading-while-listening have solicited participants' opinions about presenting audio and visual text simultaneously: learners appear to find reading-while-listening to be more engaging, and they generally have more positive attitudes about it than other input modes (e.g. Brown et al., 2008; Chang, 2009; Lightbown, 1992; Tragant et al., 2016; Tragant and Vallbona, 2018). Thus, it appears that reading-while-listening has important learning benefits, as well as being positively perceived by learners. Gaining a better understanding of how reading differs in reading-while-listening and reading-only should help us to explain some of the reasons for these observed benefits and increased learners' enjoyment. In addition, an examination of alignment of readers' eye movements to the auditory input in reading-while-listening would provide better understanding of how readers make use of the two sources of input when they are provided simultaneously. In the current study we explore how reading patterns change in the presence of auditory input and address the following questions:

1. How does the presence of auditory input change reading?
2. Are there differences between L1 and L2 readers' reading in the reading-only and reading-while listening modes?
3. Is the processing of the audio and written text aligned in reading-while-listening in the L1 and L2?

Study

Methods

Participants. We collected data from 32 L2 speakers of English and 31 L1 speakers of English. Due to high levels of track loss, the recordings from 4 L2 and 3 L1 participants were discarded before any analyses were carried out, leaving a final sample of 28 L2 participants (3 males; 25 females; mean age = 25.6) and 28 L1 participants (7 males; 21 females; mean age = 20.1). All participants were students at the University of Nottingham who received either course credit or were paid £6 for taking part. The L2 participants were advanced learners of English who had met the university's entry requirement for English proficiency (6.0 or higher on the International English Language Testing System (IELTS) or equivalent). They had various L1s (Arabic $n = 14$; Chinese $n = 8$; Turkish $n = 1$; Hindi $n=1$; Dutch $n=2$; and German $n=2$). All participants had normal or corrected-to-normal vision and reported normal hearing.

Both the L1 and L2 speakers completed an online vocabulary size test with 200 items with a maximum score of 10,000 (Meara and Miralpeix, 2016). The L1 speakers achieved a level that corresponds to a “very high level of proficiency” (Meara and Miralpeix, 2016, p. 118) ($M = 8596.00$, $SD = 624.91$) and the L2 speakers were at a level indicating a “good level of competence” (Meara and Miralpeix, 2016, p. 118) ($M = 6182.32$, $SD = 1744.65$). The L1 group had significantly higher vocabulary test scores than the L2 group, $t(53) = 6.87$, $p < .001$, $d = 1.83$. The L2 speakers also completed a language-background questionnaire (see Table 1) and self-rated their speaking, listening, reading and writing proficiency on a 7-point scale (1 = very low; 7 = native-like). They gauged what percentage of the time (0% to 100%) they used English in a range of contexts (speaking with family, friends etc.; writing to family, friends, etc., reading for academic purposes, work, etc.; watching / hearing / listening TV, radio, etc.).

Table 1. Summary of the L2 participants (means with standard deviations in parentheses) in terms of their age, years studying English, self-ratings of proficiency on a 7-point scale (1 = very

low; 7 = native-like), percentage of time using English in various contexts and performance on a vocabulary test (maximum possible score of 10,000).

	Age	Years studying English	Speaking	Listening	Reading	Writing	Use of English	Vocab test
Mean	25.57	12.79	5.61	6.21	5.96	5.54	66.35%	6182.32
SD	5.38	3.81	1.03	0.79	0.88	1.10	22.26	1744.65
Max.	41	25	7	7	7	7	100.00	8474
Min.	19	8	3	3	4	4	14.71	3356

Materials. There were two story passages that were each approximately 1,500 words long. To ensure that the passages would not be too difficult for the L2 participants, all of the words that made up the texts were high frequency words, belonging to the first 3000 most frequent word families in English. For the reading-while-listening task, the passages were recorded by a native speaker of British English at a normal rate of speech of 3.5 words per second for both stories (with a typical speech rate being 3.7 words second for native speakers of English; Goldman-Eisler, 1961). The recordings were approximately six minutes long. The written stories were presented across eight or nine screens with a maximum of 220 words per screen. The texts were in 18-point Courier New font with 2.5 line-spacing.

An ASIO sound card provided accurate audio timing, allowing the presentation of the text and audio to be time-locked. Following De Luca et al. (2013), the temporal onsets and offsets of words were extracted using Audacity software, and these values were input into Experiment Builder, which sent time-stamped messages to the eye-tracker indicating precisely when a participant heard a specific word during the audio presentation.

Procedure. Participants sat in front of a computer monitor with their head stabilized via a desk-mounted chinrest. Eye movements were recorded monocularly, tracking each participant's left eye at a sample rate of 1000Hz with an Eyelink 1000+ system from SR Research. Participants sat approximately 60 cm from a wide-screen computer monitor having a resolution of 1280 x 1960 and a refresh rate of 144 Hz. Before the experiment, accuracy was verified using a nine-point calibration and validation grid and again before the second passage. During the experiment, between each screen a fixation point appeared to allow for trial-by-trial drift checking and recalibration was carried out if required. Participants read the texts as normally as possible for comprehension. In the reading-only mode, they pressed the spacebar when they finished a page, and in the reading-while-listening mode, they pressed the spacebar only when they were done reading and the audio had finished playing. Following each story, a series of five yes/no comprehension questions appeared to ensure that participants had attended to the text. The passages were presented in the reading-only and reading-while-listening modes in a counterbalanced design. The order of presentation was fixed on each list: List 1 reading-while-listening preceded reading-only and List 2 reading-only preceded reading-while-listening. After the eye-tracking session, all participants completed the online vocabulary size test, and the L2 participants filled out the language-background questionnaire.

Results

Both the L1 and L2 speakers showed good understanding of the two stories on the comprehension questions (L1 = 96.43% correct, $SD = .06$; L2 = 95.71%, $SD = .07$) and they did not differ from each other ($t(54) = 0.45$, $p = .69$, $d = .02$). Performance was also good following both modes (reading-only = 97.50%, $SD = .06$), reading-while-listening = 94.64%, $SD = .07$) and did not differ ($t(54) = 1.59$, $p = .12$, $d = .03$).

Analysis. Following Cop et al. (2015) we excluded from our analyses any words preceded or followed by punctuation, as well as the first and last word of every line (because the return sweep to the beginning of the line might not land where expected, and refixation(s) may be required, and because fixations to the final word encompass the programming of a longer saccade; see Conklin et al. 2018 for a discussion). This left us with 2,253 words for our analyses (Story 1 = 1,145 words; Story 2 = 1,108 words), which we analyzed using *R* software, Version 3.4.4 (R Core Team, 2013). Linear mixed effects models were fitted using the *lme4* package (version 1.1-17, Bates, Maechler, Bolker and Walker, 2014), *p*-values were estimated using the *lmerTest* package (Kuznetsova, Brockhoff and Bojesen Christensen, 2015) and interactions were inspected using the *phia* package (Rosario-Martinez, 2015). All of the continuous variables (first fixation duration, first-pass reading time, and total reading time) were log-transformed before the analysis. The data were trimmed by deleting data points that fell above or below 3 standard deviations for each experimental condition (reading or reading-while-listening) in each language group (L1 or L2) separately. This led to a loss of 1.23% of the data. The models were also fitted with the full dataset. The same predictors were significant and the same pattern of results appeared. However, the model fits were poorer with the full dataset. The models on the trimmed data are reported here and have the best Akaike information criterion (AIC), an estimator of the quality of each model relative to other models. In terms of the predictors, word frequencies were taken from the SUBTLEX-UK corpus, transformed into Zipf scale values (van Heuven, Mandera, Keuleers and Brysbaert, 2014). Word length was calculated in letters for the analysis of Research Questions 1 and 2, and in number of syllables for Research Question 3, where the focus was on spoken words (i.e. alignment of the audio and fixations). Part of speech (POS) information was based on the SUBTLEX-UK corpus, and participants' vocabulary scores were log-transformed to help ensure that predictors were on the same scale.

We carried out two types of analyses. First, we compared reading behavior in the L1 and L2 in the reading-only and reading-while-listening modes. We carried out this analysis on all of the words in the texts after the exclusions outlined above. Second, we looked at eye-movement patterns in the reading-while-listening mode to explore how L1 and L2 speakers' reading was aligned with the audio recording. Here we looked at a subset of the words in the story. We randomly selected 20 of the content words on each of the experimental screens, including verbs, nouns, adjectives and verbs (Total = 377 words, 199 from Story 1 and 178 from Story 2).

Comparing L1 and L2 reading behavior in reading-only vs. reading-while-listening

conditions. We analyzed six eye-tracking measures (first fixation duration, first-pass reading time, total reading time, total fixation count, regression count, and skipping probability) to look at the reading behavior in reading-only and reading-while-listening conditions. Table 2 provides a summary of the data for each of these eye-tracking measures.

Table 2. Mean with standard deviations presented in the parentheses for first fixation duration, first-pass reading time, total reading time, total fixation count, regression count, and skipping probability in reading-while-listening and reading-only for L1 and L2 participants.

	Reading-while-listening		Reading-only	
	L1	L2	L1	L2
First fixation duration	232.22 (100.35)	243.18 (97.59)	203.04 (72.51)	240.48 (93.80)
First-pass reading time	251.47 (123.82)	272.18 (132.09)	212.55 (84.57)	276.83 (139.75)
Total reading time	322.77 (191.13)	325.04 (186.52)	242.90 (118.54)	336.10 (196.30)
Total fixation count	0.92 (0.81)	0.92 (0.78)	0.66 (0.64)	0.98 (0.80)
Skipping probability	0.32 (0.47)	0.31 (0.46)	0.43 (0.49)	0.28 (0.45)
Regression count	0.25 (0.43)	0.19 (0.39)	0.19 (0.38)	0.16 (0.36)

To investigate reading behavior, separate linear mixed-effects models were fit for continuous outcome variables (total reading time, first fixation duration and first-pass reading time) and generalized linear mixed-effects model for count (total fixation count, regression count) or binary (skipping probability) outcome variables. The core model included experimental condition (reading-only or reading-while-listening), language group (L1 or L2) and an interaction between these two variables. As potential covariates, the models also included word frequency, word length (in letters), POS (baseline condition was adjective), participants' vocabulary scores, the version of the experiment the participants were assigned (to account for counterbalancing of the two input modes and the presentation order of the two modes), the story the word was used in (1st or 2nd), and the interactions between the language group and all of the other predictors. In order to select the best-fitting model, we started from the full model and then removed non-significant covariates one by one. We compared each new model to the previous one using likelihood ratio tests and estimating AIC values. We also checked for any potential indications of multicollinearity in the models, but the VIF scores indicated no problems in any of the reported models. Random effect structures of the models included random intercepts for target words as well as for participants, random by-word slopes for experimental condition and for language group, and random by-participant slopes for experimental condition (as all the participants were presented with one story in the reading-only condition, and one in the reading-while-listening condition). We report the full structure of both fixed and random effects for each model. The summary of the models for each of the eye-tracking measures is in Table 3. The vocabulary score data was missing for one of the L1 participants, so we excluded his data from the models that had vocabulary score as a significant predictor. We tried fitting models with his vocabulary score replaced by the mean vocabulary score of the L1 group; this did not change the significant predictors in the models, which we believe justified the exclusion.

Table 3. Mixed effects model estimates for each of the eye-tracking measures, with indications

of significance *** $p < .001$, ** $p < .01$, * $p < .05$.

Fixed effects		First fixation duration				First-pass reading time				Total reading time			
		<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept		7.31	0.55	13.20	.000***	7.74	0.55	13.98	.000***	7.07	0.42	16.96	.000***
Language (L2)		0.17	0.04	4.06	.000***	0.09	0.06	1.65	.100	0.09	0.05	1.94	.057
Mode (reading+listening)		0.12	0.02	8.04	.000***	0.15	0.02	8.05	.000***	0.25	0.03	7.77	.000***
Length						0.00	0.00	2.00	.046*	0.02	0.00	5.65	.000***
Part of Speech	(adverb)	0.03	0.01	2.32	.021 *	0.04	0.01	2.71	.007**	0.03	0.02	1.51	.131
	(conjunction)					0.04	0.03	1.59	.113	-	0.03	-0.88	.379
						0.01	0.02	0.56	.576	0.03			
	(determiner)	0.01	0.02	0.90	.369	0.04	0.02	2.02	.045*	0.02	0.02	0.86	.390
	(marker)					0.02	0.05	0.38	.709	-	0.06	-0.27	.787
						0.00	0.04	0.08	.940	0.02			
	(noun)					-0.02	0.01	-1.90	.058	-	0.01	-4.27	.000***
	(number)					-0.01	0.01	-0.70	.482	0.06			
						0.04	0.04	0.84	.404	-	0.05	-0.22	.828
						0.02	0.04	0.64	.523	0.01			
Frequency	(preposition)	0.02	0.02	1.40	.164	0.03	0.02	1.46	.147	0.00	0.02	0.01	.991
	(pronoun)	0.03	0.02	1.97	.050	0.05	0.01	2.73	.007**	0.03	0.02	1.14	.256
	(verb)	0.02	0.01	1.78	.076	0.02	0.01	1.94	.053	0.01	0.01	0.93	.351
					.000***	-0.03	0.01	-5.23	.000***	-		-6.00	.000***
						-0.02	0.00	-3.71		0.04	0.01		
	Vocabulary					-0.25	0.06	-4.16	.000***	-	0.05	-3.55	.001**
						-0.22	0.06	-3.56	.001**	0.16			
	Story (2)					-0.14	0.03	-5.35	.000***	-	0.02	-2.47	.017*
										0.05			
	Language * Mode				.000***	0.02	0.00	9.24	.000***	-	0.04	-5.40	.000***
					-0.10	0.02	-4.64		0.24				
Language * Length					-0.01	0.00	-2.52	.012*	0.03	0.00	12.20	.000***	
Language * Frequency				.000***	7.74	0.55	13.98	.000***					

Random effects		Variance	<i>SD</i>	Variance	<i>SD</i>	Variance	<i>SD</i>
Word	(intercept)	0.002	0.040	0.003	0.052	0.004	0.066
	Mode	0.002	0.040	0.002	0.041	0.011	0.103
	Language	0.001	0.028	0.001	0.031	0.002	0.040
	Mode * Language	0.002	0.039				
Participant	(intercept)	0.011	0.104	0.015	0.122	0.022	0.149

Residual	Mode	0.006				0.076				0.009				0.094				0.026				0.162			
		0.131				0.361				0.167				0.409				0.216				0.465			
		Total Fixation count				Regression count				Skipping probability															
Fixed effects		<i>b</i>	<i>SE</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>z</i>	<i>p</i>								
Intercept		-			.000**	-			.000**	-				-0.15	0.26	-0.58	.562								
Language (L2)		0.33	0.09	-3.53	*	8.02	2.08	3.86	*	-0.15	0.26	-0.58	.562												
Mode (reading+listening)		0.37	0.07	5.25	*	0.15	0.14	1.07	.283	-1.09	0.26	-4.11	.000***												
Length		0.34	0.05	6.59	*	0.40	0.07	5.73	*	-0.56	0.12	-4.46	.000***												
Part of Speech	(adverb)	0.09	0.00	2	*	0.03	0.01	2.99	.003**	-0.30	0.01	22.66	.000***												
	(conjunction)	0.01	0.03	0.46	.650	0.05	0.07	0.79	.432	-0.04	0.08	-0.54	.591												
	(determiner)	-				0.11	0.06	-2.01	.045*	0.14	0.13	1.06	.290	0.32	0.15	2.15	.032*								
	(marker)	-				0.05	0.04	-1.17	.241	0.21	0.09	2.30	.021*	0.03	0.11	0.29	.770								
	(noun)	0.20	0.10	-2.12	.034*	0.21	0.23	0.91	.361	0.35	0.30	1.17	.243												
	(number)	-				0.07	0.02	-3.06	.002**	0.13	0.06	2.15	.031*												
	(preposition)	-				0.00	0.09	-0.03	.980	0.02	0.21	0.09	.932												
	(pronoun)	0.09	0.04	-2.35	.019*	0.21	0.08	2.52	.012*	0.16	0.09	1.67	.094												
	(verb)	0.12	0.04	-3.27	.001**	0.23	0.08	2.74	.006**	0.17	0.10	1.75	.081												
Frequency		0.02	0.02	0.86	.392	0.12	0.05	2.22	.027*	-0.04	0.06	-0.77	.443												
Vocabulary		-			.000**	-				-															
Story		0.06	0.01	-5.04	*	0.68	0.23	2.97	.003**	0.14	0.03	4.35	.000***												
Language * Mode		0.08	0.03	-2.80	.005**	-				-															
Language * Length		-			.000**	-				-															
Language * Frequency		0.37	0.07	-5.12	*	0.22	0.10	2.19	.028*	0.80	0.17	4.54	.000***												
Random effects		Variance		SD		Variance		SD		Variance		SD													
Word	(intercept)	0.011	0.106	0.078	0.279	0.088	0.297																		
	Mode	0.001	0.037	0.004	0.063	0.009	0.097																		

Participant	Language	0.000	0.021	0.004	0.067	0.045	0.211
	(intercept)	0.065	0.255	0.181	0.425	0.474	0.689
	Mode	0.067	0.259	0.120	0.346	0.418	0.647

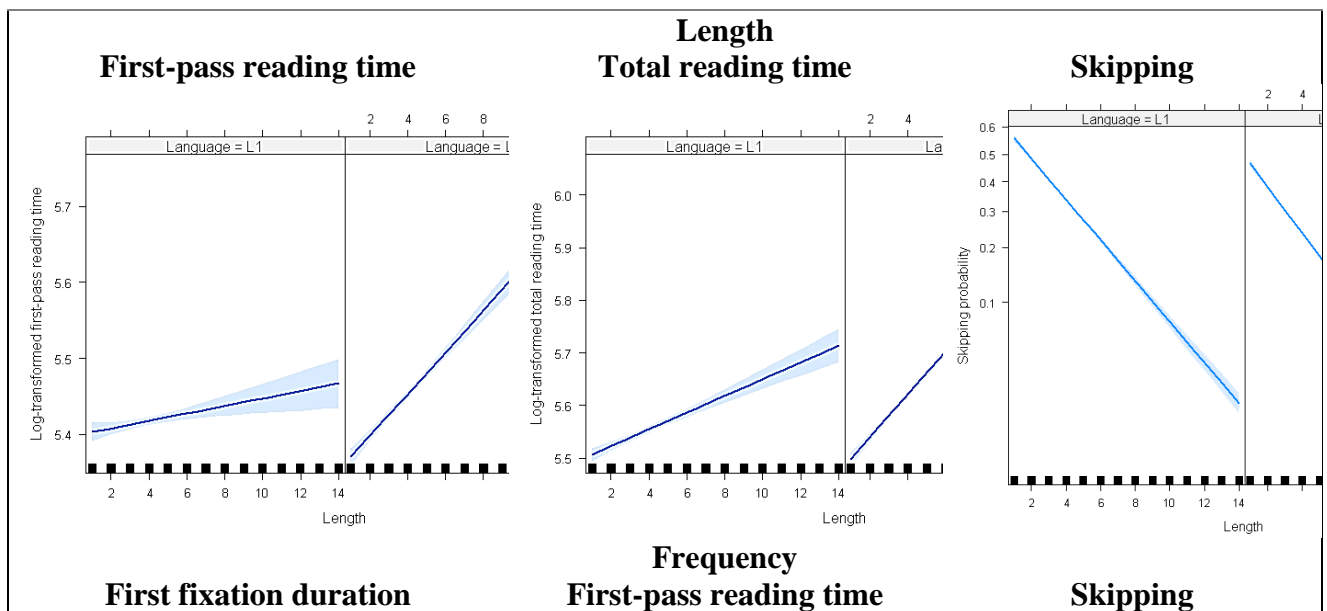
Looking at the results in Table 3, we see that all of the models had a significant interaction between the mode (reading-while-listening and reading-only) and language group (L1 and L2). We explored these interactions using the *phia* package (Rosario-Martinez, 2015) by looking at the differences between the language groups in each mode and between the modes in each language group, which are summarized in Table 4. For L1 speakers, the reading-only condition elicited consistently fewer and shorter fixations, as well as fewer regressions and more skipping compared to the reading-while-listening condition. For L2 speakers, the two modes did not differ in terms of number or duration of fixations. However, there were fewer regressions in reading-only. The L1 and L2 speakers had very similar performance in the reading-while-listening mode but varied in a number of measures in reading-only, with L1 speakers generally having fewer and shorter fixations.

Table 4. Interactions between mode (reading-only (RO) and reading-while-listening (RWL)) and language group (L1 and L2).

	Language group		Mode	
	L1	L2	RO	RWL
First fixation duration	RO < RWL ($X^2 = 64.55$, $p < .001$)	RO = RWL ($X^2 = 2.61$, $p = .11$)	L1 = L2 ($X^2 = 3.22$, $p = .14$)	L1 = L2 ($X^2 = 0.87$, $p = .35$)
First-pass reading time	RO < RWL ($X^2 = 64.801$, $p < .001$)	RO = RWL ($X^2 = 0.40$, $p = .52$)	L1 < L2 ($X^2 = 8.01$, $p = .009$)	L1 = L2 ($X^2 = 0.57$, $p = .45$)
Total reading time	RO < RWL ($X^2 = 64.44$, $p < .001$)	RO = RWL ($X^2 = 0.14$, $p = .71$)	L1 < L2 ($X^2 = 19.91$, $p < .001$)	L1 = L2 ($X^2 = 2.44$, $p = .12$)
Total fixation count	RO < RWL ($X^2 = 43.45$, $p < .001$)	RO = RWL ($X^2 = 0.30$, $p = .58$)	L1 < L2 ($X^2 = 27.55$, $p < .001$)	L1 = L2 ($X^2 = 0.00$, $p = .99$)
Regression count	RO < RWL ($X^2 = 32.78$, $p < .001$)	RO < RWL ($X^2 = 6.80$, $p = .01$)	L1 = L2 ($X^2 = 1.25$, $p = .56$)	L1 = L2 ($X^2 = 0.22$, $p = .64$)

Skipping probability	RO > RWL ($X^2 = 19.87$, $p < .001$)	RO = RWL ($X^2 = 3.64$, $p = .06$)	L1 > L2 ($X^2 = 23.01$, $p < .001$)	L1 = L2 ($X^2 = 1.75$, $p = .19$)
----------------------	---	---	--	--

Notably, some of the models reported in Table 3 had a significant interaction between language group (L1 or L2) and target word length and/or language group and target word frequency. These interactions are plotted in Figure 1. As can be seen in the figure, the direction of the effect was always the same in both language groups: longer words yielded longer reading times and less skipping, while more frequent words led to shorter reading times and more skipping. The strength of the effect was larger in L2 group (as indicated by steeper lines in the plot).



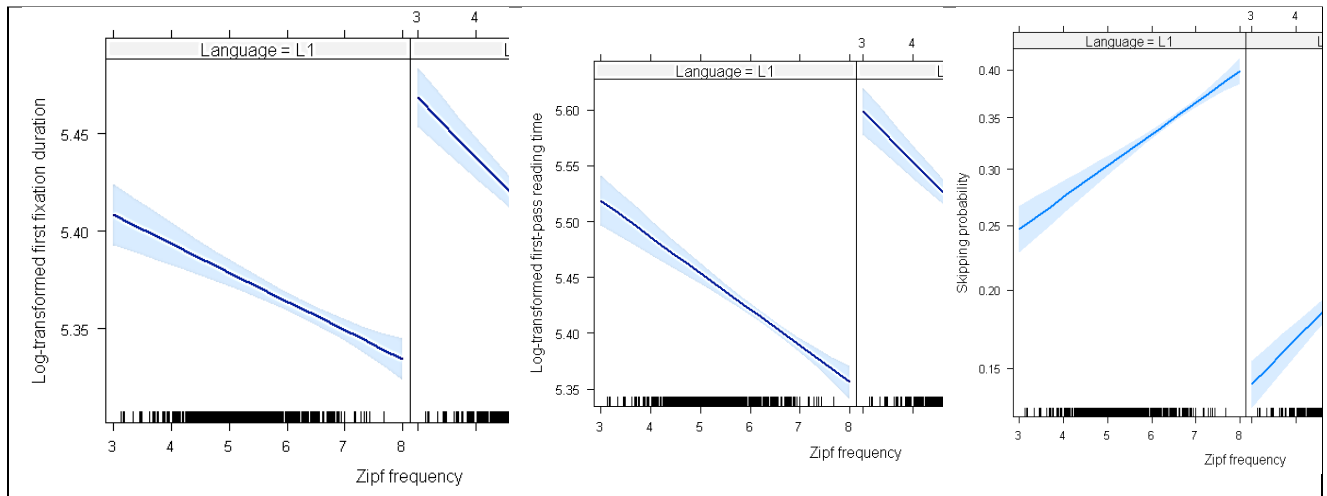


Figure 1. Interactions between the language group (L1 or L2) and target word length and frequency

Alignment of reading and listening behavior. We also looked at whether fixations were aligned with the audio in the L1 and L2. In other words, were readers fixating a word when they heard it? If they were *not* fixating the word they were hearing, were their eyes ahead or behind the audio? For this analysis, only the reading-while-listening part of the dataset was relevant. For each of the words in this analysis, we identified where the participants fixated precisely when the word was present in the audio. More specifically we looked at the onset and offset of the word in the audio file and determined whether a fixation occurred to the word during that time period, which is illustrated in Figure 2. If the onset and/or offset of the fixation occurred in the window of the audio word it was coded as ‘aligned’. Also, if the onset of the fixation was before the start of the audio and the offset was after the audio, the word was coded as ‘aligned’. If the onset and offset occurred before the audio it was coded as ‘behind’ and if they occurred after the audio it was coded as ‘ahead’.

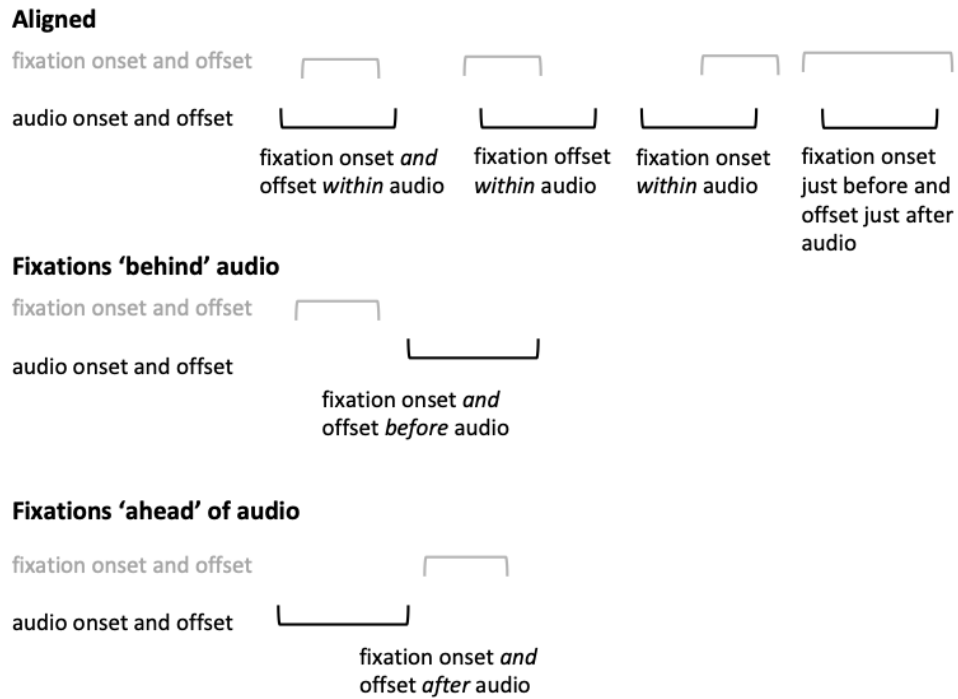


Figure 2. Assessing the alignment of eye fixations to a word and its audio presentation with fixation onset and offset depicted by a gray line and audio by a black line.

We fit generalized linear mixed effects models to the data as the outcome variable was binomial: the audio and the reading behavior was either aligned or not. The predictors were the same as in the first part of the analysis: language group, word length (in syllables), POS, word frequency, story (whether participants encountered story 1 or 2 in reading-while-listening), and vocabulary score. We started by looking at the odds of participants fixating a target word at the time of hearing it. In general, neither group was fixating the word that they were hearing. L1 speakers' reading aligned with the audio about 17% of the time, and L2 speakers' reading did about 33% of the time. We used this binary outcome variable (aligned = 1 or not = 0) to fit a generalized linear mixed effects model in order to see which variables affected it (see top half of Table 5). The model showed that longer words had larger odds of being fixated at the time they were

heard. Also, words in story 2 seemed to be fixated more often when they were heard. Vocabulary score was a significant predictor, with participants that had larger vocabularies having less alignment between the visual and auditory words. Language group was not a significant predictor in this case, but vocabulary score was correlated with language group ($R=0.68$, $p=.000$), and L1 speakers tended to have higher vocabulary scores than the L2 speakers.

Table 5. Odds of fixating the word being heard and odds of fixating ahead of the audio, with indications of significance *** $p < .001$, ** $p < .01$, * $p < .05$.

Odds of fixating the word heard (with not fixating a word as reference category)					
Fixed effects	<i>b</i>	<i>exp(b)</i>	<i>SE</i>	<i>z</i>	<i>p</i>
Intercept	10.96	57526.44	3.06	3.59	.000***
Vocabulary score	-1.44	0.24	0.34	-4.20	.000***
Length	0.27	1.31	0.05	5.98	.000***
Story 2	0.41	1.51	0.15	2.71	0.007**
Random effects	Variance		SD		
Word (intercept)	0.101		0.329		
Participant (intercept)	0.482		0.699		
Odds of fixating ahead of listening (fixating behind as reference category)					
Fixed effects	<i>b</i>	<i>exp(b)</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	-14.91	0.00	3.74	-3.99	.000***
Vocabulary score	1.89	6.62	0.42	4.49	.000***
Random effects	Variance		SD		
Word (intercept)	0.380		0.616		
Participant (intercept)	0.688		0.830		

Following on this analysis, we looked at the subset of the data where the reading and listening were aligned to determine when the participants started and ended fixating a word in relation to its onset in the audio. When log-transforming the data, we added a minimum value to eliminate negative values (negative values indicated that a participant had started fixating the word before

hearing it). Then we fitted the model, which is summarized in Table 6. Vocabulary score was a significant predictor both for fixation start and end times. Looking at Figure 2, this means that participants with larger vocabularies started and stopped fixating aligned words earlier than those with smaller vocabularies. Part of speech also seemed to play a role in fixation start time for verbs, such that verbs were fixated earlier. There were no significant differences between the L1 and L2 groups.

Table 6. Mixed effects model estimates for fixation start and end times, with indications of significance *** $p < .001$, ** $p < .01$, * $p < .05$

Fixed effects		Fixation start				Fixation end time			
		<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept		7.36	0.43	17.05	.000***	7.24	0.35	20.77	.000***
Vocabulary score		-0.12	0.05	-2.47	.017*	-0.14	0.04	-3.42	.001**
POS	Adverb	-0.12	0.06	-1.89	.060				
	Determiner	-0.07	0.17	-0.43	.666				
	Noun	-0.03	0.05	-0.64	.525				
	preposition	-0.09	0.14	-0.72	.473				
	Pronoun	-0.09	0.14	-0.70	.487				
	Verb	-0.09	0.05	-2.06	.041*				
Random effects		Variance	SD	Variance		SD			
Word	(intercept)	0.019	0.138	0.020	0.143				
Participant	(intercept)	0.007	0.086	0.004	0.067				
Residual		0.127	0.356	0.100	0.316				

Finally, we looked at the cases where a word was not fixated during its audio occurrence to determine whether reading was ahead or behind the audio. When reading was not aligned with the audio, both groups were generally reading ahead, but this was more so for the L1 speakers (ahead about 89% of the time) than the L2 speakers (ahead about 79% of the time). The analysis of this data showed that the only significant predictor of reading ahead or behind was vocabulary score (see bottom half of Table 5). Participants with larger vocabularies had higher odds of reading ahead.

Discussion

We were interested in how reading differs when a corresponding audio text is present or absent, as well as how reading behavior differs in the L1 and L2. In the L1, there were differences in reading patterns across the early and late eye-tracking measures in the reading-only and reading-while-listening conditions. L1 readers had fewer and shorter fixations, as well as more word skipping and fewer regressions, in the reading-only versus the reading-while-listening mode. For L2 readers, the only difference in the two modes was in terms of the number of regressions, with there being fewer in the reading-only mode. Notably, in reading-while-listening, where the presence of audio may moderate reading speed, performance was the same for the L1 and L2 speakers across the early and late reading measures. However, in the reading-only mode, performance was different, with L1 speakers generally having fewer and shorter fixations and more skipping. This suggests that reading is faster and more fluent in the reading-only condition in the L1, as one would expect. The presence of the audio slows L1 reading, especially for already efficient readers. In the case of the L2, reading is slower in the reading-only mode and there is no additional slow down in reading-while-listening.

Importantly, reading in the L2 was modulated by two factors that are well established in the L1 literature. Both L1 and L2 readers had shorter and fewer fixations to high frequency and shorter words. Reading in the L2 was slower for long and low frequency words than reading in the L1, while reading high frequency words was equivalent in the two groups. Vocabulary knowledge influenced eye movements, such that with greater vocabulary knowledge there were generally fewer and shorter fixations.

Similar to what the literature has demonstrated about oral reading, reading-while-listening in the L1 elicited longer and more fixations and regressions and less skipping than reading-only (silent

reading). Notably, our fixation means for the L1 readers line up with figures in the field for silent reading and reading aloud. Castelhana and Rayner (2008) said that mean fixation durations during silent reading range from 225-250 ms; our mean fell within this at 243 ms. However, our L2 readers were well outside of this, with a mean fixation duration of 332 ms. Castelhana and Rayner reported a range for oral reading (reading aloud) of 275-325 ms, which is very similar to our mean for reading-while-listening of 323 ms. Crucially, the L2 speakers also fell within this, with a mean of 325 ms, and they were very similar to the L1 readers in our study. Castelhana and Rayner hypothesized that, in reading aloud, readers fixate more and longer to keep their eyes from getting too far ahead of what they are saying. This appears to be the same when listening: Readers do not get too far ahead of what they are hearing someone else say. It is important to point out that a pattern of slowed reading in the reading-while-listening mode is not evident in L2 speakers, likely because reading-only itself is relatively slow in the L2.

While reading was slower in the reading-while-listening condition, readers' gaze was *not* generally aligned with the audio text. However, the L2 readers' fixations were aligned more than those of the L1 readers (33% vs. 17%). This might indicate that L2 comprehenders are making more use of the visual text to aid listening than L1 comprehenders and/or that the speed of the audio was better matched to the L2 readers' reading rate. Also, readers with smaller vocabularies (some of the L2 readers) had greater alignment of eye fixations with the audio. This could indicate that those with a smaller vocabulary size used the audio to help segment, decode, parse and/or make the form-meaning link for words in the text. This supports Gerbier et al.'s (2018) view that reading-while-listening is more beneficial to less skilled readers. Notably, there were differences in the two groups in terms of where the eyes fixated when reading was *not* aligned. For both groups the eyes generally were ahead of the audio, but this was less so for the L2 speakers: When reading was unaligned, L2 speakers were ahead about 79% of the time versus L1

speakers 89% of the time. L2 comprehenders fixating ahead of the audio is similar to what Wisniewska and Mora (2018) found when investigating eye movements to captioned videos. Although they only looked at audio-visual alignment for ten words in the captions, they found that L2 comprehenders' fixations were ahead of the audio about 70% of the time. (They did not investigate L1 speakers.) In the current study, L2 speakers lagged behind the audio more than the L1 speakers. This could have been because the audio was too fast for them. However, given that they were also generally ahead of the audio, this is unlikely to be the case.

An important question is how reading somewhat ahead might account for some of the advantages of reading-while-listening that are found in the literature (e.g. Brown et al., 2008; Chang 2009; Chang and Millet, 2014, 2015; Lightbown, 1992; Tragant, Muñoz and Spada, 2016; Tragant and Vallbona, 2018; Webb and Chang, 2012; 2014; Woodall, 2010)? In English, a written text provides listeners with a visual cue for the boundaries of *upcoming* auditory words. This strong word segmentation cue may help speed word identification. It may also help learners to develop letter-sound correspondences. Since they have seen the visual form of the word they are about to hear, it may help them link the two forms. This may very well, as Webb and Chang (2012) suggested, strengthen the form (phonological *and* orthographic) and meaning connections that contribute to vocabulary development. Although entirely speculative, having a good visual cue to auditory word segmentation—something that is challenging for language learners in listening tasks—might contribute to learners' enjoyment and engagement with a text.

While the results indicate how audio *might* benefit reading, key questions remain about how exactly a benefit arises from audio-visual (near) synchrony, as well as about how the difficulty of the text (i.e. both in terms of the difficulty of the language and rate of presentation) and the proficiency of the comprehender impact reading in a reading-while-listening mode. A recent

study investigating looking patterns to captioned video suggests that in order to be beneficial to L2 learners, the videos should be matched with language learners' proficiency levels (Gass, Winke, Isbell and Ahn, 2019). More research is needed to determine how factors such as the rate of audio presentation, difficulty of the text, proficiency, etc. influence reading patterns in reading-while-listening tasks. It seems plausible that having visual cues about segmenting upcoming auditory information might be helpful to listeners. However, the misalignment of the incoming visual and auditory information could conceivably hinder processing. The fact that comprehenders see one word while hearing another means they are processing two (competing) words simultaneously, which might disrupt comprehension. Any explanation for the benefit of reading-while-listening would need to propose how the processing system deals with this misalignment of the visual and auditory words.

In sum, in the reading-only mode, we found L2 readers were generally slower (having more and longer fixations and less skipping) than L1 readers. In reading-while-listening, performance was largely similar across the L1 and L2 groups. For the most part, neither group of participants fixated the word that they were hearing, although the L2 readers did so more than the L1 readers. In general, both groups' eye movements preceded the audio. However, L2 readers had more cases where their fixations lagged behind the audio. The L2 readers' pattern of reading with the audio may be able to account for some of the benefits that have been seen in the literature for reading-while-listening, but also raises important questions about how the processing system deals with two competing sources of information.

References

- Bates D, Maechler M, Bolker B and Walker, S (2014) lme4: Linear mixed-effects models using Eigen and S4_. R package (Version 1.1-7) Retrieved from <http://CRAN.R-project.org/package=lme4>>
- Brown R, Waring R and Donkaewbua S (2008) Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a foreign language* 20(2): 136-163.
- Castelhano MS and Rayner, K (2008) Eye movements during reading, visual search, and scene perception: An overview. In: Rayner K, Shen D, Bai X and Yan G (eds) *Cognitive and cultural influences on eye movements*. Tianjin: Tianjin People's Publishing House, pp. 3-33.
- Chang A (2009) Gains to L2 listeners from reading while listening only in comprehending short stories. *System* 37: 652-663.
- Chang A and Millett S (2014) The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal* 68(1): 31-40. doi: 10.1016/j.system.2009.09.009
- Chang A and Millett S (2015) Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System* 52: 91-102. doi: 10.1016/j.system.2015.05.003

- Cop U, Drieghe D and Duyck W (2015) Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS One* 10(8): 1-38. doi: 10.1371/journal.pone.0134008
- Conklin K, Pellicer-Sánchez A and Carrol G (2018) *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.
- De Luca M, Pontillo M, Primativo S, Spinelli D and Zoccolotti P (2013) The eye-voice lead during oral reading in developmental dyslexia. *Frontiers Human Neuroscience* 7(696): 1-17. doi: 10.3389/fnhum.2013.00696
- Diao Y and Sweller J (2007) Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and Instruction* 17(1): 78-88. doi: 10.1016/j.learninstruc.2006.11.007
- Gass S, Winke P, Isbell DR and Ahn J (2019) How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning & Technology* 23(2): 84-104/
- Gerbiera E, Bailly G and Bosse ML (2018) Audio-visual synchronization in reading while listening to texts: Effects on visual behaviour and verbal learning. *Computer Speech and Language* 47(1): 74-92. doi: doi.org/10.1016/j.csl.2017.07.003
- Godfroid A (2020) *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*. Routledge.

- Goldman-Eisler F (1961) The significance of changes in the rate of articulation. *Language and Speech* 4(3): 171-174. doi: 10.1177/002383096100400305
- Immordino-Yang M and Deacon T (2007) An evolutionary perspective on reading and reading disorders. In: Fischer KW, Bernstein JH and Immordino-Yang MH (eds) *Mind, brain, and education in reading disorders*. Cambridge University Press, pp. 16-29.
- Kuznetsova A, Brockhoff PB and Bojesen Christensen B (2015) lmerTest: Tests in Linear mixed effects models (Version 2.0-29) Retrieved from <http://CRAN.R-project.org/package=lme4>>
- Laubrock J and Kliegl R (2015) The eye-voice span during reading aloud. *Frontiers in Psychology* 6: 1432. doi: 10.3389/fpsyg.2015.01432
- Lightbown PM (1992) Can they do it themselves? A comprehension-based ESL course for young children. In Courchène R, St John J, Therrien C and Glidden J (eds) *Comprehension-based second language teaching: Current trends*. Ottawa: University of Ottawa Press, pp. 353-370.
- Meara P and Miralpeix I (2016) *Tools for Researching Vocabulary*. Bristol: Multilingual Matters.
- Pellicer-Sánchez A, Tragant E, Conklin K, Rodgers M, Llanes A and Serrano R (2018) L2 reading and reading-while-listening in multimodal learning conditions: An eye-tracking study. *ELT Research Papers*. British Council, pp. 1-28.

R Core Team (2013) R: A language and environment for statistical computing. Vienna, Austria:

R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>

Rayner K (2009) Eye movements and attention in reading, scene perception, and visual search.

Quarterly Journal of Experimental Psychology 62(8): 1457-1506. doi:

10.1080/17470210902816461

Rayner K and Pollastek A (1989) *The Psychology of Reading*. Englewood Cliffs, NJ: Prentice

Hall.

Rosario-Martinez H (2015) phia: Post-hoc interaction analysis. R package (Version 0.2-0)

Retrieved from <http://CRAN.R-project.org/package=phia>

Schotter ER, Angele B and Rayner K (2012) Parafoveal processing in reading. *Attention,*

Perception and Psychophysics 74(1): 5-35. doi: 10.3758/s13414-011-0219-2

Serrano R and Pellicer-Sánchez A (2019) Reading vs. Reading while listening in young learners:

An eye-tracking investigation. *Applied Linguistics Review*. Available first view:

<https://doi.org/10.1515/applirev-2018-0102>

Tragant E, Muñoz C and Spada N (2016) Maximizing Young Learners' Input: An Intervention

Program. *The Canadian Modern Language Review / La revue canadienne des langues*

vivantes 72(2): 234-257. doi: 10.3138/cmlr.2942

Tragant E and Vallbona A (2018) Reading while listening to learn: young EFL learners' perceptions. *ELT Journal* 72(4): 395-404. doi: 10.1093/elt/ccy009

Trofimovich P, Lightbown PM, Halter R and Song, H (2009) Comprehension based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition* 31: 609-639. doi: 10.1017/S0272263109990040

van Heuven WJB, Mandera P, Keuleers E and Brysbaert, M (2014) SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67(6): 1176-1190. doi: 10.1080/17470218.2013.850521

Webb S and Chang A (2012) Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review* 68(3): 267-290. doi: 10.3138/cmlr.1204.1

Webb S and Chang A (2014) Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research* 19: 667-686. doi: 10.1177/1362168814559800

Wilkinson D (2012) A data-driven approach to increasing student motivation in the reading classroom. *Language Education in Asia*, 3(2): 252-262. doi: 10.5746/LEiA/12/V3/I2/A13/Wilkinson

Wisniewska N and Mora, J C (2018) Pronunciation learning through captioned videos. In: Levis J (ed) *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference*. Ames, IA: Iowa State University, pp. 204-215. ISSN 2380-9566

Woodall B (2010) Simultaneous listening and reading in ESL: Helping second language learners read (and enjoy reading) more efficiently. *TESOL journal* 1(2): 186-205. doi: 10.5054/tj.2010.220151