Research Methods in Applied Linguistics 2 (2023) 100086

Contents lists available at ScienceDirect

Research Methods in Applied Linguistics

journal homepage: www.elsevier.com/locate/rmal

Eye-tracking reading-while-listening: Challenges and methodological considerations in vocabulary research

Kathy Conklin[a,*], Sara Alotaibi[b]

[a] University of Nottingham, School of English, Nottingham NG7 2RD, United Kingdom
[b] University of Jeddah, English Language Institute, United Kingdom

A R T I C L E I N F O

Keywords:
Eye-tracking
Reading
Listening
Reading while listening
Simultaneous written and auditory input
Vocabulary

A B S T R A C T

Learners need to know a considerable number of words to function in a second or foreign language. To help increase their word knowledge, learners are encouraged to engage in activities that provide a rich source of vocabulary like listening to music and audio books, and watching films, television, and video. In many of these types of activities, learners can listen to and read 'matched' content (i.e., text is both written and aural). For example, viewing television programs and films is often accompanied by subtitles that closely adhere to the auditory input. While reading and listening to matched content may be a fairly common experience, we have little understanding of how comprehenders process the two sources of information, nor how the addition of audio changes word reading or might impact word learning. Eye-tracking provides a means of measuring the effort associated with processing words, yet very few studies have explicitly investigated written-word processing while listening and even fewer have examined this in the context of word learning. The technology allows researchers to synchronize eye-movements in reading to an auditory text, but requires technical know-how. The goal of this research methods paper is to provide methodological and technical guidance on the use of eye-tracking in reading-while-listening with an emphasis on investigating vocabulary learning and processing.

Introduction

In today's world second language learners have a multitude of opportunities to encounter the target language in- and out-side of the classroom. This can be via more traditional activities like reading, but also includes contexts like: listening to audiobooks and podcasts, watching films, television programs and YouTube videos, and engaging with learning apps. In many of these situations, learners encounter audio input along with its written representation or a translation of it. For example, when listening to an audiobook, a learner can follow along with the written text. When watching video content, learners can turn on subtitles that match the language of the audio or provide its translation. Crucially, these kinds of activities expose learners to new vocabulary and already known vocabulary in new contexts, which benefits their vocabulary knowledge.

A lot is known about processing when comprehenders read in more traditional contexts (e.g., reading books or other written texts), with eye-tracking technology having provided much of our insight on silent reading. Decades of eye-tracking research has afforded researchers and educators alike with a good understanding of reading patterns in children and adults in their first (L1) and L2 (second)

* Corresponding author.
E-mail address: kathy.conklin@nottingham.ac.uk (K. Conklin).

https://doi.org/10.1016/j.rmal.2023.100086
Received 3 April 2023; Received in revised form 13 October 2023; Accepted 13 October 2023
Available online 18 October 2023

languages (for an overview see Conklin et al., 2018; Godfroid, 2020; Liversedge et al., 1998). While eye-tracking has shed light on reading in traditional contexts, we still have little understanding of how the simultaneous presentation of audio influences reading nor how reading behavior in reading-while-listening might relate to vocabulary learning. Eye-tracking can provide us with this insight, however, there are methodological and technical considerations when monitoring eye movements to written text with simultaneous audio. After providing an overview of eye-tracking, the focus of the discussion will be on key methodological and technical concerns associated with eye-tracking reading-while-listening when investigating vocabulary learning.

Eye-tracking

Eye-tracking technology permits processing to be monitored while engaging in familiar tasks from a computer screen (e.g., reading, online tests, looking at webpages, watching videos, etc.). The technology shows what is being attended to, and how much cognitive effort is being expended to process the input by providing a moment-to-moment record of eye fixations and movements when engaged in looking behavior. More specifically, eye-tracking technology allows researchers to determine where people's eyes land (*fixate*), how many times they land in that position or region (*fixation/regression count*), and how long each fixation lasts (*fixation/regression duration*), as well as indicating which words are skipped (*skipping*) and measuring the duration and length of the eye movements (*saccades*) (for a fuller description see Conklin et al., 2018). In eye-tracking, greater processing effort for particular words or regions of text is indexed by longer and/or more fixations, longer and/or more regressions, less skipping, and shorter saccades (for a discussion, see Castelhano & Rayner, 2008). Conversely, fewer and/or shorter fixations and regressions, and more skipping, as well as longer saccades, indicate that less processing effort is expended. In vocabulary research, the processing effort associated with target items can be investigated in terms of learning. For example, does greater attention and processing of an item lead to increased learning gains? Similarly, when new items are encountered multiple times, do changes in attention across repetitions provide an index of learning? Eye-tracking, used in concert with measures of learning gains, provides researchers with the means to address such questions.

Eye-tracking data is generally described in terms of fixations to a region of interest (ROI, sometimes referred to as an area of interest, AOI). Data is generally reported in terms of number of fixations to an ROI (i.e., fixation count), the duration of fixations to an ROI (i.e., fixation duration), which calculates the amount of time spent fixating it, and the likelihood or probability of fixating an ROI. There are a variety of eye-tracking measures, which are thought to index different aspects of processing and are often referred to as 'early' or 'late' (e.g., Altarriba et al., 1996; Staub & Rayner, 2007). Early measures are said to tap into automatic processes and the initial stages of processing (e.g., lower-level processes like word recognition in reading). Late measures reflect strategic processing and include revisits and reanalysis that result from processing difficulty. Late measures signal more effortful and/or conscious processing (e.g., lexical integration in reading).

When examining eye-tracking data, we see that fixations tend to be longer and saccades shorter during reading aloud than reading silently (Castelhano & Rayner, 2008) because skilled readers can read words silently more quickly than they can say them. Thus, 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 & Kliegl, 2015). In a seminal study, Conklin et al. (2020) found that L1 readers had different processing patterns for reading-only than reading-while-listening, with fewer and shorter fixations, as well as more word skipping and fewer regressions in reading-only. In contrast, L2 reading patterns were largely the same in the two modes, with the only difference appearing in regressions (i.e., fewer regressions in reading-only). In reading-only, processing is slower for L2 than L1 readers. However, because L2 processing is already (relatively) slow in reading-only, there is no additional slowdown for L2 readers in reading-while-listening. Notably, in their study, Conklin and colleagues examined the alignment of the eye movements to words in the written text and the audio occurrence of them. In general, neither the L1 nor L2 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.

Very little other eye-tracking research has been done on reading-while-listening. A notable exception is a set of studies investigating reading-only and reading-while-listening, which compared the amount of time spent looking at written text versus images in illustrated stories with L1 and L2 adults (Pellicer-Sánchez et al., 2021) and L2 children (Pellicer-Sánchez et al., 2020). For the adult participants (Pellicer-Sánchez et al., 2021), simultaneous presentation of the audio and written story allowed both L1 and L2 comprehenders to spend more time looking at pictures, supporting better integration of text and pictures. Notably, greater text processing time was related to better comprehension in L2 adult readers, while it was associated with lower comprehension for the L1 readers. Processing time on images was positively related to comprehension for L1 readers, but not L2 readers. For the L2 children, Pellicer-Sánchez et al. (2020) demonstrated that the presence of audio in the reading-while-listening condition allowed the children to look at the images more often. Further, increased processing time for the text was related to lower levels of comprehension, whereas processing time on images was positively related to comprehension. While these two studies provide some indication of reading behavior in contexts with simultaneous presentation of written and audio content, they do not provide a detailed examination of eye-movement patterns to individual words nor an exploration of how reading patterns might relate to vocabulary learning and processing.

Considerably more research needs to be done on reading-while-listening in varied contexts (e.g., while watching video context, gaming, using apps and reading audiobooks) to understand it more fully, as well as to establish how it might contribute to vocabulary learning – both of single words and multiword sequences (e.g., *fish and chips*). An important question left open by Conklin et al. (2020), is whether reading ahead of the audio occurs in all contexts and in varied populations of readers and whether this might confer some advantage to readers. For example, the written text may provide listeners with a visual cue for the boundaries of upcoming words, which may speed word identification as well as learning. It may help learners to develop letter-sound correspondences, and as Webb and Chang (2012) suggest, strengthen the form (phonological *and* orthographic) and meaning connections that contribute to vocabulary development. For multiword sequences (e.g., idioms, collocations and binomials), the prosodic cues and intonation contours in the accompanying auditory input could help learners recognize them as meaningful chunks, overcoming the visual cue (spaces) indicating that the words are separate items.

It is plausible that visual cues about segmenting upcoming auditory information could be beneficial to learning words, as could be the prosodic information indicating when visibly separate information (i.e., individual words) should be treated as chunks. However, the misalignment of the incoming visual and auditory information caused by reading ahead could conceivably hinder processing. More specifically, there could be a processing cost associated with the auditory system hearing one word while the visual system attends a different word. The fact that comprehenders see one word while hearing another means they are processing two (competing) words simultaneously, which might disrupt comprehension. These open questions about reading-while-listening have important implications for theories of learning such as: a) Cognitive Load Theory (Sweller, 1988), which postulates that because working memory is limited, instructional methods should avoid overloading it to maximize learning, and b) Dual Coding Theory (Paivio, 1990), which explains the relationship between verbal and nonverbal systems and suggests that the simultaneous activation of verbal and nonverbal systems supports learning (for an overview discussion of these two theories and their relevance to reading-while-listening see Wang & Tragant, 2022). In sum, it is clear that in a world where reading-while-listening is increasingly commonplace, more research needs to be done to fully understand it and its theoretical implications.

Eye-tracking reading-while-listening: challenges and methodological considerations

In this section we turn to some methodologial challenges and considerations. First, we touch on some general concerns before turning to key considerations when using written and audio materials. Finally, we talk about considerations when syncing looking behavior to audio material.

One of the advantages of eye-tracking is that it permits processing to be monitored from a computer screen in familiar contexts using authentic materials such as films with subtitles, which is an increasingly investigated form of reading-while-listening. While authentic materials have high ecological validity, making them attractive to study, it is challenging to draw conclusions directly from the eye-movement record because factors known to influence eye movements are not controlled for (e.g., sentence length, syntactic complexity, words matched for word class, length, frequency, familiarity, age of acquisition, number of meanings/polysemy, plausibility, contextual constraint, predictability; Clifton et al., 2007). In eye-tracking studies where participants engage with authentic materials, researchers generally need to use sophisticated analyses that allow for a variety of uncontrolled factors to be accounted for (Conklin et al., 2018 and see Godfroid, 2020 for an introduction and worked example). Unfortunately, our current analytic tools cannot fully account for the wide range of variability that exists in authentic language and visual input, posing a challenge for the use of authentic materials (Conklin et al., 2018; Conklin & Guy, 2020).

There are additional concerns when designing eye-tracking studies that investigate written material (for a fuller discussion see Conklin et al., 2018); these are important to consider when investigating reading-while-listening. In studies of vocabulary learning, researchers will be interested in single words or multiword sequences. Carrol and Conklin (2015) provide a good overview of additional considerations when eye-tracking multiword sequences. For both single words and muliword sequences, it important that an ROI does *not* occur at the end of a clause, sentence, or paragraph or at the beginning or end of a line on the screen. It is best to exclude fixations to words or ROIs in these positions. Precision and accuracy – the ability of the eye-tracker to reliably reproduce a gaze point – are always a concern. Eye-trackers are generally less accurate toward the edges of the computer screen, so ROIs close to the screen margin should be avoided or excluded. This may be of particular concern for video content, as subtiles often appear near the edge of the screen. Because eye-trackers are less accurate at measuring vertical than horizontal movement, double or triple line-spacing is recommended to more reliably attribute fixations to a particular line. Fonts like Courier New are often used so that letters take up the same amount of horizontal space, generally with a font-size of 14–18 pt. When using authentic materials (e.g., film subtitles), it may not be possible to follow this guidance unless the research produces their own version of the material (i.e., create their own version of the subtitles) with appropriate formating and placement of target items. Crucially, researchers need to carefully consider their ROIs and whether eye movements can be accurately and precisely attributed to them.

Reading-while-listening is often compared to other contexts like reading-only and listening-only. In listening-only, the audio generally proceeds automatically as the auditory text unfolds. In reading-while-listening, the input may advance from one page/screen to the next seamlessly without the need to do anything. In other words, the page/screen of an audiobook advances once all the text has been heard and, similarly, video frames advance automatically. Thus, the audio input provides a clear indication of when the page/screen will advance (i.e., once the audio is finished), indicating to comprehenders how quickly they need to read to reach the end of the text. Alternatively, participants can be asked to press a button to advance from one page/screen to the next. Again, accompanying audio provides an indication of when it is appropriate to advance (i.e., once the audio has finished). In reading-only, participants can indicate via a button press when they have finished reading and are ready to move forward. Without the accompanying audio, there is less constraint on reading rate, making it more variable. This makes it challenging to compare it to auditory contexts that have less variability in reading rate and (likely) more uniform eye-movement patterns. To address concerns about variability, and to ensure that reading-only and reading-while-listening conditions are well-matched in terms of the amount of time comprehenders spend on them, researchers may have pages/screens advance automatically at the same timepoint in both conditions. However, for the reader in a reading-only mode, there is no clear cue about when pages/screens will advance, which may encourage them to read more quickly to ensure they reach the end before the text advances, or they may read too slowly and not reach the end. Thus, a key methodological consideration is how and when text should advance, in particular when comparing reading-while-listening to reading-only.

In investigations of reading-while-listening, some of the research questions may involve the synchronization of the audio input with eye movements to the screen. For example, when studying vocabulary learning, researchers may be interested in exploring looking behavior when a particular vocabulary item occurs in the audio track. This can then be related to other behavior. For example, researchers could investigate how pre-teaching interventions relate to looking patterns to taught and untaught vocabulary items in a subsequent reading-while-listening activity. They can explore whether pre-teaching influences subsequent alignment of looks to items when they are heard in the audio. They can assess whether alignment of listening and looking behavior leads to better learning of the items or, alternatively, whether reading slightly ahead leads to better learning (as measured by subsequent performance on vocabulary tests). To address these types of research questions, researchers will need to align the occurrence of each word of interest in the audio input to fixations on the screen. For video content, researchers may want to do the same, as well as considering looks to objects and entities in the visual scene. For example, in an audiobook, when hearing the word 'car' is the participant reading the word 'car' or another word? In a video, when hearing the word 'car', are they looking at the written word 'car' in the subtitle, another word in the subtitle, or the object car on the screen and does this depend on whether the car is in a fixed position or moving across the screen? To carry out such an analysis, researchers must know precisely when a word occurs in the audio record (i.e., its onset and offset time), and this needs to be timelocked to looking behavior.

To timelock audio files with visual stimulus presentation, there are some important methodological considerations. First, researchers need a computer with good audio timing. For Windows machines, the default Direct X Windows drivers are not adequate; researchers need a computer with Windows 7, 10 (or a Mac). Second, an audio stream input/output (ASIO) sound card driver that is supported and recognized by the eye-tracking software needs to be installed. The ASIO sound card driver allows the eye-tracking software to connect directly to the computer's sound card hardware to provide predictable and low latency audio playback and recording, which is required for experiments that need high audio playing precision (e.g., those needing audiovisua synchronization). The ASIO sound card driver gives the experimental software precise audio timings, which allows for the start of the audio to be timelocked with the presentation of the visual stimulus. Step-by-step instructions for sound card installation may be available with an eye-tracker or software's user manual.

For studies investigating vocabulary learning and processing, researchers will likely want to identify what participants fixate when they hear a particular word – that word itself, another word close by or a word more distant to it in the text. To do this, researchers will need to use audio editing software to extract the onset and offset times (in milliseconds) in the audio stream for any words or phrases of interest. In the current example, we will use Audacity as it is a widely used audio software that is free, open-source and works across platforms, as well as having been used in previous reading-while-listening research (Conklin et al., 2020). In the Audacity audio editor, each audio file needs to be opened and the word or phrase of interest identified in it. Looking at the example in Fig. 1, we see that once the word of interest has been identified (top panel), the 'zoom in button' can be used. The zoom function (bottom panel) allows researchers to see the precise starting point of a word. As illustrated in Fig. 1, the onset for the word of interest is 687 ms after the start of

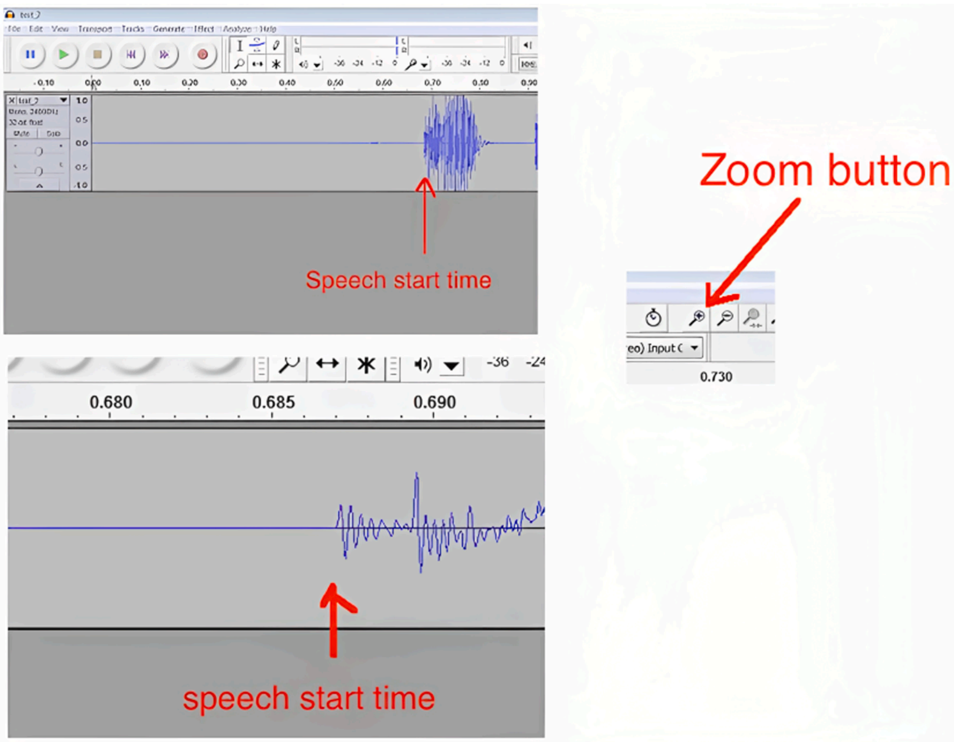

Fig. 1. Extracting audio timings using the software Audacity with the top panel depicting word identification and the bottom panel showing use of the 'zoom' function to ascertain the precise start (or finish) timing of a word.

the audio file. The zoom function can also be used to establish the word's offset. Extracting the onset and offset time in the audio stream needs to be done for every word of interest.

The extracted start and finish timings for words or phases of interest are input into the eye-tracking system's software, which sends time-stamped messages to the eye-tracker indicating precisely when a participant hears a specific word. More specifically, the extracted timings are used to define 'interest periods' – predefined time windows in which researchers are looking for effects (Ito & Knoeferle, 2022). The interest periods isolate the part of a trial that corresponds to certain words or phrases in the audio file. For example, researchers can create a period of interest for a word like 'car' based on its start and end time in an audio file. Later, a fixation report can be outputted that identifies where participants were fixating during that particular interest period (as well as during any other defined interest periods). In other words, it tells us what participants were fixating when the word 'car' occurred in the audio input.

There are some things to consider when timelocking audio input with the eye-movement record. Because extracting the precise timings of words or phrases is done manually, it is a time-consuming process. If researchers want to look at eye-audio synchronization for an entire text, they will need to manually extract the onset and offset times for hundreds, or even thousands, of words. Therefore, researchers should make sure that their materials are 'perfect' and that the experiment runs well before extracting the timings, as any later changes to the audio file would necessitate extracting them again. Researchers should carefully check their materials, including the quality of the audio input. It is important to ensure that the recordings are clear and free of hums, buzzes, hisses, echoes, or other unwanted sounds and that there are no fluctuations in the sound file. It is also important to consider whether the speech is at an appropriate rate for the participants that will be tested, as speech rate is often considered to be the most important variable in listening comprehension (for a discussion of the implication of different speech rates in L2 listening see Chang, 2018). Thus, it is recommended that all aspects of the experiment, and particularly the audio input, are thoroughly checked and tested before extracting the timings. Once the onset and offset timings have been extracted, they need to be entered in the experimental software to create an audio interest period for each target item. This will be used to generate an output report for every word or phrase that has a defined interest period.

It is important to keep in mind that a command sent by the experimental software timelocks the appearance of the visual material with the start of the audio file. This is problematic if the onset of the audio file contains silence. For example, if the first 3 ms of the audio file have silence, participants will have a 3 ms head start on reading before they hear the linguistic input. This can influence eye-movement patterns. Once the speech starts, participants may reread from the beginning to synchronize the audio and visual information. Alternatively, they may read ahead because of the offset in visual and audio information, or they may ignore the audio because they have already read what they are hearing. Thus, it is important to ensure that the occurrence of the linguistic input in the audio is synchronized with the visual input. The easiest way to achieve this is by using audio editing software to check that the onset of speech coincides with the beginning of the sound file and remove any silence before speech onset as required.

Another key consideration is how the instructions given to participants, as well as their perception of the aims of the study, might influence eye-movement patterns. For example, if participants believe that the focus of the study is only on listening, will they attend to the written text as they normally would in a reading-while-listening context? Similarly, might they ignore the audio if they deem it is unimportant to their performance? Particularly when researchers are interested in examining natural processing behavior, it is essential that any explanation of the study, as well as the instructions, make it clear that participants should engage with the task as they normally would without placing more importance on either the audio or written aspect of the task. For example, participants should simply be asked to engage with an audiobook or a film with subtitles as they usually would. To help ensure that behavior is as close to normal as possible, it is helpful to include a practice session to familiarize participants with the nature of the experiment. At this stage, it is also essential to ensure that the audio is working well.

Thus far we have discussed some of the main concerns when designing and carrying out a reading-while-listening eye-tracking study. There are also important considerations associated with processing the data. As we have seen, researchers need to create audio interest periods for each target word or phrase to obtain eye-movement data during the time window when the word was uttered. More specifically, we obtain fixation reports for predetermined audio interest periods. With some additional steps, the fixation reports can tell us about eye-audio synchronization. Taking a concrete example from a fixation report (see Table 1), we know that a word of interest (in this case 'brother') was uttered at 31,303 ms after audio onset and ended at 31,412 ms, while fixations to it occurred at 31,359 ms and ended at 31,594 ms. As can be seen in Table 1, this information can be used to determine whether fixations to 'brother' overlapped with its audio occurrence. We see that looks to the word started 56 ms after its onset in the audio and ended 182 ms after its offset. Thus, fixations to the visual word 'brother' coincided, or aligned, with its occurrence in speech.

The process outlined in Table 1 allows researchers to determine whether fixations are 'aligned' with the audio. Notably, different eye-movement patterns may be classified as aligned. In Fig. 2, we see that if a word is fixated at any time during its audio occurrence, it is considered 'aligned'. However, more stringent criteria could be imposed, such that not all of the scenarios depicted in Fig. 2 would be considered aligned. For example, only items for which the fixation started during the audio could be counted as aligned (i.e., the first and third aligned scenarios in Fig. 2). Conklin et. al (2020) considered all four scenarios depicted in Fig. 2 to be aligned. However, there

Table 1
Data about audio occurrence and visual fixation times, which are used to determine whether a word is fixated during an audio period of interest (e.g., "brother").

Current fixated item	Interest period start time	Interest period end time	Word fix start	Word fix end	Fix start rel to audio start	Fix end rel to audio end
brother	31,303	31,412	31,359	31,594	+56	+182

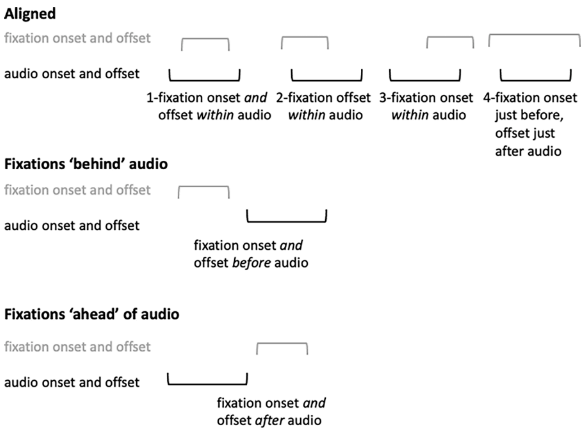

Fig. 2. Assessing the alignment of eye fixations to a word and its audio presentation with fixation onset and offset depicted by a gray line and audio by a black line (from Conklin et al., 2020).

is no standard practice for determining alignment, as there is very little eye-tracking research on reading-while-listening. Crucially, researchers should report what constitutes alignment in their study. When the eyes and audio are not aligned, it is important to consider whether fixations are ahead of or behind the audio. As can be seen if Fig. 2, if the onset and offset of fixations to a word occurs before the audio (pre-auditory fixation), reading is ahead of the audio. In contrast, if the fixation occurs after the audio the audio period (post-auditory fixation), reading is behind. Again, it is important to report what constitutes reading ahead or behind.

When investigating reading-while-listening it is important to be aware of the key challenges associated with examining eye movements to simultaneous audio presentation. This section has briefly described some important considerations researchers should keep in mind.

Conclusions

Eye-tracking provides precise information about the position and duration of a reader's eye gaze while processing various types of verbal and non-verbal input and has been used extensively to investigate reading-only over the past decades. However, relatively little work has been done on reading-while-listening. Given the pervasiveness of contexts in which learners encounter the simultaneous presentation of auditory and visual linguistic input, it is important to gain a greater understanding of how reading-while-listening impacts learning and processing. For researchers interested in vocabulary, eye-tracking offers a means of investigating how concurrent auditory and visual input affects word processing and potentially supports vocabulary learning. To help researchers who would like to investigate reading-while-listening with eye-tracking, this paper has discussed key considerations associated with the methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Altarriba, J., Kroll, J., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24*(4), 477–492.

Carrol, G., & Conklin, K. (2015). Eye-tracking multi-word units: Some methodological challenges. *Journal of Eye Movement Research, 7*(5), 1–11.

Castelhano, M. S., Rayner, K., Rayner, K., Shen, D., Bai, X., & Yan, G. (2008). Eye movements during reading, visual search, and scene perception: An overview. Eds.. *Cognitive and cultural influences on eye movements* (pp. 3–33) Tianjin: Tianjin People's Publishing House

Chang, A. C.-S. (2018). Speech rate in second language listening. In J. I. Liontas, T. International Association, & M. DelliCarpini (Eds.), *The TESOL Encyclopedia of English Language Teaching*.

Clifton, C., Staub, A., Rayner, K., & van Gompel, R. (2007). Eye movements in reading words and sentences. Ed.. *Eye movements: A window on mind and brain* (pp. 341–372) Amsterdam, the Netherlands: Elsevier

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research, 36*(3), 257–276. https://doi.org/10.1177/0267658320921496

Conklin, K., Guy, J., Adolphs, S., & Knight, D. (2020). English language and literature: Investigating literariness using a new psycholinguistic methodology. Eds. *Routledge handbook of english language and digital humanities* (pp. 494–510). Routledge.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge, UK: Cambridge University Press.

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. New York: Routledge.

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, 1–33. Online first.

Laubrock, J., & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology, 6*, 1432.

Liversedge, S., Paterson, K., Pickering, M., & Underwood, G. (1998). Eye movements and measures of reading time. Ed.. *Eye guidance in reading and scene perception* (pp. 55–75) Oxford, UK: Elsevier

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Pellicer-Sánchez A., Conklin K., Rodgers M., Parente F. (2021). The effect of auditory input on multimodal reading comprehension: An examination of adult readers' eye movements. Modern Language Journal, 105(4), 936–956. [DOI: 0.1111/modl.12743].

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, A. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition, 42*(3), 665. -665.

Staub, A., Rayner, K., & Gaskell, M. G. (2007). Eye movements and online comprehension processes. Ed.. *The oxford handbook of psycholinguistics* (pp. 327–342) Oxford, UK: Oxford University Press

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Wang, X., & Tragant, E. (2022). The effect of written text on comprehension of spoken english as a foreign language: A replication study. *International Review of Applied Linguistics in Language Teaching, 60*(3), 625–645.

Webb, S., & Chang, A. (2012). Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review, 68*(3), 267–290.