



HAL
open science

Audio-visual synchronization in reading while listening to texts: Effects on visual behavior and verbal learning

Emilie Gerbier, Gérard Bailly, Marie-Line Bosse

► To cite this version:

Emilie Gerbier, Gérard Bailly, Marie-Line Bosse. Audio-visual synchronization in reading while listening to texts: Effects on visual behavior and verbal learning. *Computer Speech and Language*, 2018, 47 (january), pp.79-92. 10.1016/j.csl.2017.07.003 . hal-01575227

HAL Id: hal-01575227

<https://hal.science/hal-01575227>

Submitted on 18 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SYNCHRONIZATION IN READING WHILE LISTENING

Audio-visual synchronization in reading while listening to texts: Effects on visual behavior and verbal learning.

Gerbier, Emilie^a, Bailly, Gérard^b, & Bosse, Marie Line^c

Running head: SYNCHRONIZATION IN READING WHILE LISTENING

May 31st, 2017

a – BCL Lab, UMR 7320, CNRS, Université Nice Sophia Antipolis, Department of Psychology, Nice, France. Campus Saint Jean d'Angély 3, 24 avenue des Diabls Bleus, 06357 NICE Cedex 4, France. emilie.gerbier@unice.fr

b- GIPSA-Lab, UMR 5216, CNRS, Université Grenoble-Alpes. 11, rue des Mathématiques, 38400 ST Martin d'Hères, France. gerard.bailly@gipsa-lab.fr

c- LPNC, UMR 5105, CNRS, Université Grenoble-Alpes. Bâtiment Sciences de l'Homme et Mathématiques BP47, 38040 Grenoble Cedex 9, France. marie-line.bosse@univ-grenoble-alpes.fr

Correspondence concerning this article should be addressed to Emilie Gerbier, B.C.L. emilie.gerbier@unice.fr. Phone number: +33(0)4 89 88 14 38.

SYNCHRONIZATION IN READING WHILE LISTENING

Abstract

Reading while listening to texts (RWL) is a promising way to improve the learning benefits provided by a reading experience. In an exploratory study, we investigated the effect of synchronizing the highlighting of words (visual) with their auditory (speech) counterpart during a RWL task. Forty French children from 3rd to 5th grade read short stories in their native language while hearing the story spoken by a narrator. In the non-synchronized (S-) condition the text was written in black on a white background, whereas in the synchronized (S+) RWL, the text was written in grey and the words were dynamically written in black when they were aurally displayed, in a karaoke-like fashion. The children were then unexpectedly tested on their memory for the orthographic form and semantic category of pseudowords that were included in the stories. The effect of synchronizing was null in the orthographic task and negative in the semantic task. Children's preference was mainly for the S- condition, except for the poorest readers who tended to prefer the S+ condition. In addition, the children's eye movements were recorded during reading. Gaze was affected by synchronization, with fewer but longer fixations on words, and fewer regressive saccades in the S+ condition compared to the S- condition. Thus, the S+ condition presumably captured the children's attention toward the currently heard word, which forced the children to be strictly aligned with the oral modality.

Keywords: “audio-assisted reading”, “supported e-text”, “assistive reading software”, “reading while listening”, “audio-visual synchrony”

Highlights

- Audiovisually-synchronized vs. unassisted Reading While Listening were compared.
- Children's orthographic and semantic learning did not benefit from synchronization.

SYNCHRONIZATION IN READING WHILE LISTENING

- Gaze data showed that synchronization triggered a capture onto the highlighted words.
- Preference for the synchronized condition was the highest for the poorest readers.

SYNCHRONIZATION IN READING WHILE LISTENING

1. Introduction

1.1. Reading while listening (RWL)

It is not unusual that adults read stories to children while the printed text is displayed in front of the child. The adult may even accompany the child through the text by pointing to the words being read. *Talking books* could be a means to automatize and generalize such an audio-visual reading experience. Indeed, talking books allow readers to autonomously perform a Reading While Listening (RWL) task. RWL consists of an experimental reading situation where one reads a text while one can hear it said aloud by a pre-recorded speaker or by a text-to-speech system. By providing oral rendition of the text, RWL can be conceived of as a reading assistance.

The first reports of the effect of RWL were very promising, especially for poor readers. In these non-experimental studies, RWL improved comprehension, word recognition, word meaning acquisition (Carbo, 1978), reading fluency (van der Leij, 1981), motivation to read (Carbo, 1978; Chomsky, 1976; Gamby, 1983), and the “*awareness of the end-product of learning to read*” (van der Leij, 1981).

Subsequently, a number of experimental studies compared RWL with reading only (RO) or listening only (LO) to the texts (see Table 1 for an overview). In Steele, Lewandowski, and Rustling (1996), comprehension and recall, but not word recognition, were boosted by RWL in 9-12 year-old poor readers. Also, RWL was preferred to the other conditions. In Reitsma (1988), however, reading fluency for the words included in the texts benefited *less* from RWL than from a situation where the 7-year old children actively read the text and could use assistance when needed. In contrast, in Shany and Biemiller (1995), RWL was equally efficient for subsequent fluency and comprehension as a situation where the 3rd and 4th grade children could request individual help from the teacher whenever they encountered some difficulty during reading.

SYNCHRONIZATION IN READING WHILE LISTENING

Article	Participants	Method and measures	Conditions	Main results
Torgesen, Dahlem, & Greenstein (1987)	Grade 9-12; learning disabled students.	Textbook passage readings. Comprehension.	RWL; RO; LO.	* Short-term comprehension on paragraph-length material: RWL & LO > RO. * Long-term comprehension on chapter-length material: no difference.
Reitsma (1988)	1st-grade.	Repeated reading of texts. Rate and accuracy of reading the 20 words included in the texts.	Guided reading; RWL; Independent reading with computer-generated speech feedback available; Control.	Guided reading & Independent reading with feedback > RWL & Control.
Shany & Biemiller (1995)	3rd and 4th grade, at risk-readers.	Text comprehension; Accuracy and speed of word reading.	Teacher-assisted reading (TAR); RWL; Control.	TAR & RWL > Control.
Montali & Lewandowski (1996)	Grade 8 and 9 (junior High School) (13-15 ans).	Textbook passages. Comprehension and word reading.	RWL ; RO ; LO. Of note, RWL was synchronized since words or groups of words were highlighted when they were heard	* Comprehension. Good readers: no difference; Poor readers: RWL > RO & LO. * Word reading: no difference. *Overall preference for RWL.
Steele, Lewandowski & Rustling (1996)	9-10 years old (N=8).	Reading text passages. Spontaneous recall, comprehension, short-term sight word recognition.	RWL ; RO ; LO.	* Spontaneous recall and comprehension: RWL > RO & LO. * Sight word recognition: no difference. *Preference for RWL.
Verlaan & Ortlieb (2012)	10-th grade High school students.	Comprehension of novels.	RWL ; RO.	RWL > RO.

SYNCHRONIZATION IN READING WHILE LISTENING

Table 1. Overview of the method and results of the experimental studies using texts in native language, comparing the effects of reading while listening (RWL) with other modes of interaction.

RO: reading only; LO: listening only.

Three studies investigated text comprehension in adolescents. Verlaan and Ortlieb (2012) showed a benefit of the RWL condition over RO in grade-10 students. In Torgesen, Dahlem, and Greenstein (1987), RWL was more beneficial than RO but not than LO for disabled readers in Grade 9 to 12. Finally, in Montali and Lewandowski (1996), RWL was superior to both RO and LO for poor readers in Grade 6 to 9; however, there was no such difference for good readers. Overall, both poor and normal readers preferred RWL. Of note, the latter study was the only one where RWL was synchronized since words or groups of words were highlighted when they were heard.

Studies investigating foreign language learning in adolescents and adults are consistent with the above studies. RWL proved more helpful than RO for vocabulary acquisition (e.g., Brown, Waring, & Donkaewbua, 2008; Webb & Chang, 2012), and for comprehension and reading rate (e.g., Chang & Millett, 2015). As in the previously reported studies in native language, students preferred RWL to the other reading conditions.

1.2. The effect of RWL may depends on the reading level

Some elements of the literature suggest that another variable, namely the child's reading level, might influence these effects. In particular, for good readers RWL might be less efficient than RO (Adesope & Nesbit, 2012; Holmes & Allison, 1985). More generally, there is evidence of a negative link between the reading level and the benefit obtained from RWL, that is, the lower the reading level, the higher the expected benefit (Gamby, 1983; Steele et al., 1996; Torgesen, Dahlem, & Greenstein, 1987; Verlaan & Ortlieb, 2012). This may be because RWL provides poor readers with the oral information that they cannot easily produce by reading.

SYNCHRONIZATION IN READING WHILE LISTENING

The oral rendition provided in RWL presumably allows the release of cognitive resources, typically devoted to decoding, towards higher-level processes like comprehension. Besides helping young and struggling readers by providing the oral version of the words they decode with difficulty, RWL may also help learning new words or help comprehension because it provides multimodal input of the same information. RWL provides an audio-visual sensory redundancy which allows for a reciprocal support from one modality to the other (Montali & Lewandowski, 1996). Research in fundamental cognitive psychology has shown that such an audio-visual redundancy favors further recall of words (e.g., Bird & Williams, 2002; Lewandowski & Kobus, 1993) compared to unimodal (i.e., either oral or visual) presentations. However, this redundancy may have no effect for good readers who are able to produce the correct phonological representations from the text. It may even interfere with the automatized reading process and the process of self-generating the phonological representations.

1.3. The effect of synchronized RWL

As indicated previously, only one study involved the audio-visual *synchronization* during RWL (Montali & Lewandowski, 1996), in which words or groups of words were highlighted when they were heard. However, synchronized RWL has never been compared with a non-guided, classical RWL condition that does not involve highlighting. Back to the introducing example, such synchrony would be the equivalent of an adult guiding the child by pointing to the text while he/she reads the book aloud.

A related line of research examined the use of assistive reading software (Elbro, Rasmussen, & Spelling, 1996; Elkind, Cohen, & Murray, 1993; Hecker, Burns, & Elkind, 2002; Medwell, 1998; for a review, see Anderson-Inman & Horney, 2007), that read aloud a piece of text through synthesized speech. In most cases, the current words are highlighted in contrasting color, which provides audio-visual synchrony during reading. The effects on

SYNCHRONIZATION IN READING WHILE LISTENING

literacy skills were overall positive but differed according to the measure of interest. For example, in Elkind et al. (1993), such technique was beneficial for reading comprehension in middle school dyslexic students, but not on vocabulary acquisition nor on the capacity to read printed texts. In general, the children's motivation to read was boosted by the technique, consistent with the previously cited studies. In young adults with attentional deficits, Hecker et al. (2002) showed that computer-assisted reading led to more improvements in attention, fatigue, and reading pace, but not in comprehension, compared to unassisted reading. However, in these studies, a whole set of functionalities were associated with synchronised RWL, such as the use of an electronic dictionary, or the possibility to click on a word to trigger its pronunciation or meaning (Anderson-Inman & Horney, 2007). Therefore, it is impossible to specifically determine the isolated effect of synchronisation in these studies.

To specifically address this issue, we developed a RWL application that allows words to be visually highlighted exactly when they are heard, in a synchronized, karaoke-like way (hereafter, *S+ condition*). From a cognitive perspective, the rationale is that the timely highlighting of words should trigger a visuo-attentional capture toward the word being highlighted (and simultaneously heard). With respect to literacy skills, this audio-visual synchrony may boost orthographic and semantic learning thanks to the temporal congruency between visual and audio inputs. Prior studies (Gerbier, Bailly, & Bosse, 2015) with middle-school students were encouraging in showing an advantage for the *S+* over the non-synchronized (hereafter, classical, *S- condition*) RWL on semantic learning, although not on orthography. The present study was designed to replicate that previous study with elementary-school children. This was motivated by the implications of the self-teaching hypothesis (e.g., Share, 1999). This theory assumes that learning the visual (orthographic) form of a word is favored by the joint activation of its oral and visual representations. In good readers, this joint activation is a consequence of a successful decoding. This hypothesis is thought to account

SYNCHRONIZATION IN READING WHILE LISTENING

for the implicit learning of lexical orthography during the process of a casual reading experience. However such learning is not possible for readers with lower levels of decoding abilities since they cannot activate the oral representation of words by themselves. The self-teaching hypothesis should then predict that younger readers in a condition of RWL would take advantage from the joint activation of the visual and oral modalities. The S+ condition might be even more profitable by providing guidance as to where to look during the RWL process.

In sum, previous studies have compared RWL with other modes of interaction with texts, and focused mostly on reading fluency and comprehension. In the present study we wished to extend the scope of previous research by studying only RWL and investigating what effect the word-based audio-visual synchronization would cause on short-term orthographic and semantic learning.

Because no previous study on RWL addressed the issue of the ocular behavior triggered by RWL in comparison to usual reading, we monitored the children's eye movements during RWL in the present study. We hypothesized that the concurrent oral version would trigger a fluent ocular trajectory, reflected for instance by few regressive saccades. Long fixations and regressive saccades would then be limited to new words or words that have a special interest in the given context. We also hypothesized that the attentional capture triggered by the word-by-word highlighting in the S+ condition would be reflected in the oculometric gaze pattern, which should differ from that of the unassisted (S-) RWL condition.

2. Experiment

2.1. Rationale and overview

In order to explore whether the addition of an audio-visual synchronization on words during RWL could facilitate the memorisation of the orthographic form and/or the semantic

SYNCHRONIZATION IN READING WHILE LISTENING

category of new words presented in the texts, we used *talking books* with French elementary-school children (grades 3 to 5) in a typical incidental self-teaching experiment (e.g., Share, 1999) in their native language. We specifically compared RWL with and without audio-visual synchronisation on the word level (S+ and S-, respectively). By doing so, we quasi replicated a prior study that used a very similar protocol in 6th graders (Gerbier et al., 2015). In addition, eye movements were monitored during reading. Subjective reports in previous research without oculometry indicated that the S+ condition was uncomfortable because it forced the readers to fixate the highlighted word. Such an effect was expected to be seen in our eye-tracking data.

We focused on two behavioural measures that we thought might be influenced by synchronization. First, we investigated incidental orthographic learning. Research on the self-teaching hypothesis (e.g., Share, 1999) assumes that learning the visual form of a word stems from the joint activation of both its oral and visual representations, which are generated by the decoding process, by good readers at least. In the synchronized (S+) version of the RWL, the strict synchrony between the attended auditory and visual inputs of any given word is forced, whereas it is not the case in the S- condition. Second, we investigated whether the acquisition of the meaning of new words (i.e., the pseudowords, PWs) was affected by synchronisation, and therefore we tested children's memory for their semantic category. The children's reading level child was also assessed, since previous work suggested that this variable could modulate the effect of RWL. Since we anticipated some degree of discomfort to be experienced by the children in the S+ condition, we also included a subjective assessment of the preference of children for either of the two versions of the RWL.

To this aim, children silently read short stories in French whose topics revolved around a thing or an animal, named by a specific pseudoword (PW, e.g., *landice*). The PW was presented four times in the text. The children read the texts on the screen and could hear a

SYNCHRONIZATION IN READING WHILE LISTENING

narrator tell the story aloud through a pre-recorded audio file. They were instructed to try their best to follow the text with their eyes according to the rhythm of the narrator, that is, look at a particular word when it was heard. Half of the texts were presented in S+ and the other half in S-. Eye movements were monitored during reading. After the readings, the children were unexpectedly tested on their memory for the orthography and for the semantic category of the PWs.

In addition, using oculometry, we investigated how the two reading conditions would affect gaze behavior. We hypothesized that, as in our prior study with 6-th graders, the S+ condition would dynamically capture the children's gaze toward the currently highlighted words and, therefore, that the gaze behaviour would be different from that observed in the S- condition. In particular, the number of between-word regressive saccades was expected to be lower in the S+ condition.

2.2. Method

2.2.1. Participants

40 children (20 girls and 20 boys), attending two classes in two elementary schools in the area of Grenoble, France, took part to the experiment. Their parents' written consents were obtained. They were all native French speakers. Nineteen children belonged to the first class, i.e., ten 3rd grade (5 female) and nine 4th grade (4 female). Twenty-one children belonged to the second class, i.e., six 4th grade (5 female) and fifteen 5th grade (6 female). Ages ranged from 8;5 (year; months) to 11;5 years, with an average of 10;1 years.

2.2.2. Material and apparatus

2.2.2.1. Texts and pseudowords

Twelve French stories, already used in previous unpublished experiments (Chaves, 2012), were used. They were 82 to 100 words in length and appeared on the screen all at once over 8 to 10 lines. Each story revolved around a fictional object or creature named with a

SYNCHRONIZATION IN READING WHILE LISTENING

pseudoword (e.g., *landice*), which appeared four times in the text, but neither in the title nor in the last four words. See Figure 1 for an example of two texts as they were presented on the screen. The translations of the text in English are presented in Appendix 1.

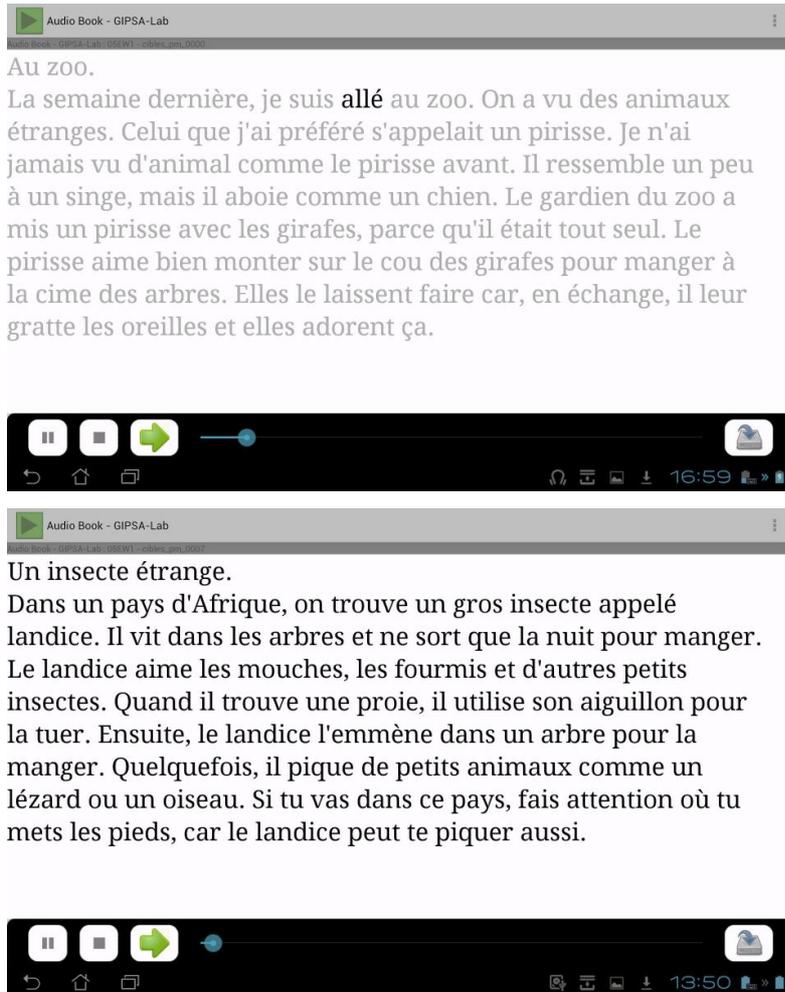


Figure 1. Screenshots of the screen for two RWL conditions used: Synchronous (S+; top) vs. non synchronous (S-; bottom). Pseudowords were respectively "pirisse" and "landice". In both cases, the child was instructed to follow the pre-recorded audio band with their eyes. The translations of the text in English are presented in Appendix 1.

Twelve two-syllable PW were used, that contained 5 to 9 letters and were controlled for trigram frequency (Peereman, Lété, & Sprenger-Charolles, 2007). Each syllable of each

SYNCHRONIZATION IN READING WHILE LISTENING

PW could be written using two homophonic graphemes, enabling us to create an alternative to the target orthography of the PW. For instance, the alternative orthography to the target PW *landice* was *lendisse*. The alternative PW was similar to the target PW with respect to trigram frequencies in French. The mean trigram frequency was 867 occurrences per million for the target PWs (min = 15; max = 3713) and 886 for the alternative PWs (min = 14; max = 3713). The PWs are presented in Table 2.

For counterbalancing purposes, the 12 PW were divided into two sets, A and B, of 6 PWs each. In order to control for grapheme frequency, target PWs from subsets A and B were paired to contain the same target phonemes but with both graphemic versions. For example, *mendint* belonged to Subset A, whereas *landice* belong to subset B (in French, “an” and “en” are alternative graphemes for /ã/). In addition, two different PW did not contain the same two target phonemes. For example, subset B included *landice*, and subset A contained *pirisse* and *mendint* (in French, “ce” and “sse” are graphemic alternatives for a final /s/).

Subset	Target PW	Alternative PW
A	claintond	cleintont
A	teinart	tainard
A	jaulu	jollu
A	fortie	phortit
A	pirisse	pyrice
A	mendint	mandin
B	phatin	fatint
B	lyonit	lionie
B	veingard	vaingart
B	naigont	neigond
B	solloi	sauloi
B	landice	lendisse

Table 2. List of the pseudowords (PW) used. The alternative PW was only used in the orthographic task, and differs from the target PW by two graphemes.

The texts were pre-recorded by a native male speaker in a narrative mode, at a slow speaking rate (4.81 syllables/s) and a phonation rate (duration of phonation / duration of

SYNCHRONIZATION IN READING WHILE LISTENING

utterances, including pauses) of 62%. The total duration of the texts were 36 to 40 seconds each.

2.2.2.2. Audio-visual synchronization

An Android® application was developed for the ASUS Transformer Pad, based on the web-based application for karaoke-style audiobook developed from Bailly and Barbour (2011). The application offers a complete control of the visual presentation and highlighting of the displayed text in synchrony with an audio file.

In order to synchronize the display of visual and audio components (i.e., having words highlighted exactly when they are heard), the system requires information about the nature and time course of the successive phonemes and pauses in the audio file. These data needed to be coded in the XML files in the Pad. To this aim, the audio signals were first semi-automatically aligned with the phonetic transcription of the read sentences (including pauses) using speaker-independent phone-sized HMM, together with a text-specific pronunciation dictionary and language model. These alignments were then hand-checked using Praat (Boersma, 2002). Such an alignment of audio signals with their phonetic transcriptions allows several levels of synchronization: phonemes, syllables, words (used in the present study) and breath groups. For more technical details, see Gerbier et al. (2015).

The same parameters as in the previous experiment with 6-th grade children were used. In the S- condition, the text was written in black over a white background (see Figure 1, bottom). In the S+ condition, the default color of the text was grey, whereas focus on the current word was triggered by switching the word color to black (i.e., highlighted; see Figure 1, upper). Thus, at any given time, the word currently heard by the subject was presented with the same contrast (i.e., black on white) in both conditions, the only difference being the other words in the text (black in S-, but grey in S+). In addition, the highlighting onset of the current word preceded its auditory onset by 300ms. This choice was based upon informal

SYNCHRONIZATION IN READING WHILE LISTENING

pilot experiments with fluent readers that suggested that a 300-ms advance of the visual over the audio input was more comfortable than a strict synchrony of 0ms. Such an asynchrony gives the reader enough time to prepare and execute the saccade towards the word in response to its highlighting, so that, at the time the word is being heard, it has been fixated and partly processed, which, we assume, allows a genuine audio-visual synchrony on a perceptual and cognitive level. (Of course, the optimal duration for this asynchrony might differ according to readers and circumstances.)

2.2.2.3. Oculometry

Using its HDMI port, the pad screen was mirrored on a monitor that embedded a SMI® RED250 eye tracker device. *SMI experiment Center* software was used to control for the display of visual stimuli.

The font size was set so that all texts fitted in one screen. This resulted in a maximum of 11 lines per screen and a large angular character size of 1.86° (i.e., with distance 60 cm and screen height 22") which allowed the children to read easily and naturally on the monitor (Legge & Bigelow, 2011) .

Information about the temporal relation between the audio-visual information provided by the pad, on one hand, and the gaze data recorded by the monitor-based eye tracker, on the other hand, was obtained by inserting audio triggers in one of the audio channel of the audio (stereo wave) file played by the Pad. A home-made electronic circuit converted these triggers into transistor-transistor logic (TTL) signals that were fed into an appropriate pin of the parallel port of the PC driving the eye-tracker device. This trigger signal was registered by the eye-tracker software, making the temporal alignment between both signals available for subsequent analyses.

SYNCHRONIZATION IN READING WHILE LISTENING

2.2.3. Experimental design

During the reading phase, two blocks of six texts were displayed, each either in the S+ or in the S- condition. In order to compensate for potential serial and/or item effects on subsequent memory performance, the condition (S+ vs. S-) and the pseudoword subset (A vs. B) associated with each block were counterbalanced across subjects, creating four different counterbalancing patterns: 1) Subset A as S+, then Subset B as S-. 2) Subset B as S+, then Subset A as S-. 3) Subset A as S-, then Subset B as S+. 4) Subset B as S-, then Subset A as S+.

Each subject was assigned a counterbalancing pattern. In each counterbalancing pattern, the presentation order of the six pseudowords of a block was randomly determined.

2.2.4. Procedure

The experiment included six phases: 1) The general presentation. 2) The calibration phase. 3) The reading phase. 4) The test phase. 5) The subjective assessment. 6) The reading level assessment. They are described below.

Children were tested individually in a quiet, isolated room, in their school. The child was seated in front of the monitor, within a distance appropriate to record oculometry, namely about 60 cm for the 22-inch monitor. After the general presentation phase, the child was asked to put the headset on and to put their chin on a chinstrap, in order to ensure immobility during the recording. The experimenter (i.e., the first author) was seated on the right to the children, in front of the computer. She manipulated the eye-tracker software and the Pad.

First, the calibration phase was conducted, in which the child followed a moving point on the screen for a few seconds. Then, the specific instruction followed: the child was informed that 12 texts corresponding to children short stories would appear on the monitor, one after the other, and that they would be said aloud by a narrator in the headset at the same time. Children were then instructed to try to move their eyes over the text according to the

SYNCHRONIZATION IN READING WHILE LISTENING

pace that the narrator would adopt while also paying attention to the stories depicted in the texts. In other words, they had to look at the words whenever they would be heard. Before each block of six texts, the experimenter described the settings of the forthcoming block, that is, the colors of the words, depending on the S+ or S- condition. A very short, basic text that did not contain any PW was presented as an example of those settings. The child was not informed that a test would occur after the reading period.

During the reading phase, the experimenter could monitor the child's eye movements on-line through the computer screen, and ensure that the instruction to follow the narrator pace was followed. The experimenter manually triggered the transition from one text to the next on the Pad.

Immediately after the reading phase, the child was tested on an orthographic test. A sheet of paper was used that read the PWs and their alternative homophones (e.g., *pirisse* and *pyrice*; Table 2). The sheet was covered with a second one containing an empty window that let only the two alternatives of any given PW visible at the same time. The children had to circle the correct pseudoword (i.e., "*how it was written in the stories*"). The words were not pronounced by the experimenter, and there was no feedback provided during the test. Four different random orders were used alternatively for consecutive children.

Next, in the semantic task, each PW was pronounced by the experimenter, and the child had to choose the right category among a choice of two (e.g., "*landice, is it a cake or an insect?*"). Any given category only appeared in two questions in total. The question order was randomized for all children.

The subjective assessment task consisted for the children to express their preference for either reading condition, with five options: "*Generally, during reading, did you prefer: no preference, S+ (a lot), S+ (a little), S- (a lot), or S- (a little)?*".

SYNCHRONIZATION IN READING WHILE LISTENING

Finally, the *Test de l'Alouette* standardized reading fluency test (Lefavrais, 2006) was administered. It consists of reading aloud a meaningless text for three minutes, as accurately and quickly as possible. The number of words erroneously read is subtracted from the total number of words read, so that the child's reading age is then determined.

Of note, the method was very similar to that used in our prior study with 6-th graders (Gerbier et al., 2015), except that 1) the orthographic and semantic tests were made easier by reducing the number of options to two instead of four or twelve, and 2) the reading pace was much slower, to adjust to the average reading level of elementary-school children.

2.3. Results

2.3.1. Behavioral measures

2.3.1.1. Learning tasks

In the orthographic task, there was no effect of the condition ($M = 3.95$ out of 6, $MSE = 0.23$; $M = 4.00$, $MSE = 0.22$; respectively for S+ and S-; Wilcoxon $U = 225.5$, $p = .89$). In the semantic task, there was a significant *disadvantage* associated with the S+ condition compared to the S- condition ($M = 4.75$, $MSE = 0.20$; $M = 5.25$, $MSE = 0.14$; Wilcoxon $U = 117.5$, $p = .025$; Figure 2). Thus, our hypotheses were not confirmed. Although the absence of effect for the orthographic task is consistent with our previous study with middle-school students (Gerbier et al., 2015), the disadvantage for S+ in the semantic task is contradictory with it. The latter result might be accounted for by the visual capture triggered by the highlighting in the S+ condition that could have prevented the children from semantically integrating the text and therefore memorising the pseudowords' semantic category. This is also consistent with their overall preference for the S- condition (see below). The reason why the results were different in both studies remains unknown, but may be related to the narration pace, which was faster for the middle-school students.

SYNCHRONIZATION IN READING WHILE LISTENING

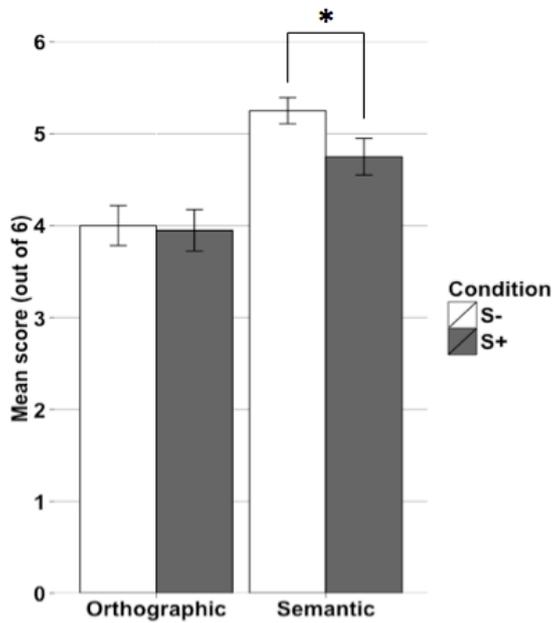


Figure 2. Mean scores for the orthographic and the semantic tasks for each of the 2 conditions (S-: not synchronized; S+: synchronized). Errors bars report standard errors of the mean. The asterisk indicates a significant difference ($p < .05$)

2.3.1.2. Preference

11 children (28%) expressed no preference for any synchronisation condition. Among the 29 who expressed a preference, 20 (69%) preferred the S- and 9 (31%) preferred the S+ (Figure 3). Thus, there was no consensus on the preference. Of note, the modal response was a strong preference for the S- condition. This suggests that, for some children, the S+ was associated with discomfort, which is consistent with previous pilot studies.

SYNCHRONIZATION IN READING WHILE LISTENING

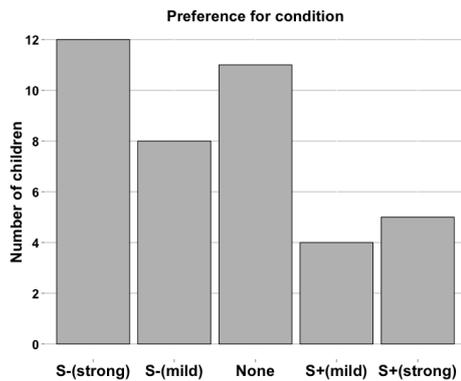


Figure 3: Number of children expressing different preferences for the RWL conditions (S-: not synchronized ; S+: synchronized).

2.3.1.2. Link between reading age and performance

The reading age, as assessed by the *Test de l'Alouette*, ranged from 7;0 to 13;10 years, with a mean of 9;11. Reading age positively correlated with the total score in the orthographic task (Bravais-Pearson's $r = +0.32$, $p < .05$) but not in the semantic task ($r = 0.05$). Reading age was not significantly correlated with the *benefit* provided by the audio-visual synchrony (i.e., score in S+ minus score in S-), neither in the orthographic (Bravais-Pearson's $r = 0.06$) nor in the semantic task ($r = 0.13$).

Interestingly, the correlation between the reading age and the preference for S+ was significant (Spearman's $r = -.52$; $p < .001$), indicating that the lower the reading age, the more likely the children were to prefer the S+ condition (see Figure 4). This suggests that synchronisation might be useful for poor readers to follow the text, which was what they were asked to do. This is also consistent with the observation, yielded by the on-line monitoring of eye movements by the experimenter, that some of the poorest readers sometimes lost track of the text in the S- condition, but less frequently and for a shorter duration in the S+ condition.

SYNCHRONIZATION IN READING WHILE LISTENING

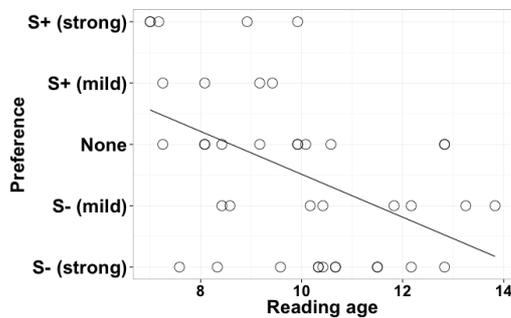


Figure 4. Correlation between children's reading age and preference for either reading condition. (S-: not synchronized ; S+: synchronized).

2.3.2. Oculometric measures

2.3.2.1. Methods

The experimental settings involved the projection of the audio-visual information generated in the Pad (i.e., text and sound) onto the oculometric-associated monitor (on which the calibration was performed and the measures collected). This prevented us from using classic oculometric analysis algorithms provided by the software. We needed a means to match information from those two sources in order to assign the ocular fixations to their localisation on the projected Pad screen, and to specify the time at which they occurred. To this aim, we used the Dynamic Time Warping (DTW) algorithm (Rabiner, Rosenberg, & Levinson, 1978). In addition, DTW also allows for compensating potential calibration issues. Calibration issues sometimes produce apparent fixation pathways that drift from the actual fixation pathway. For instance, the measured pathway for horizontal lines may appear to artefactually drift below the lines.

The DTW algorithm aims at determining, among all the possible gaze trajectories, which one is the most likely given a set of constraints. First, DTW uses spatial information. "Bounding boxes", that is, rectangular coordinates, were specified for each word. Second, DTW uses timing information, that is, the time when each word was aurally presented. Based

SYNCHRONIZATION IN READING WHILE LISTENING

on both information, the strategy is to minimize the sum of two types of costs. The “local cost” is the cost associated with the distance (in pixels) between a fixation and its associated bounding box after their matching by the algorithm. The “transition cost” refers to the speed of transition from one fixation to the next one. This speed, namely the ratio between the temporal difference between the two fixations (inter-fixation duration) and the physical difference between those two fixations (inter-fixation distance), should be reasonable given the current context. For each of these two costs and for each fixation, a z-score is computed relative to all the fixations in the entire study, and then the two z-scores are summed. Finally, for a given text for a given child, the DTW method determines the trajectory that minimizes the sum of these two costs. [Figure 5a and 5b display a result of this method in cases that do not \(upper part\) and do \(bottom part\) present a drift. No compensation for drifts was performed before further analyses. Appendix 2 presents the source code for the algorithm.](#)

Mis en forme : Anglais (États Unis)

Mis en forme : Anglais (États Unis)

SYNCHRONIZATION IN READING WHILE LISTENING

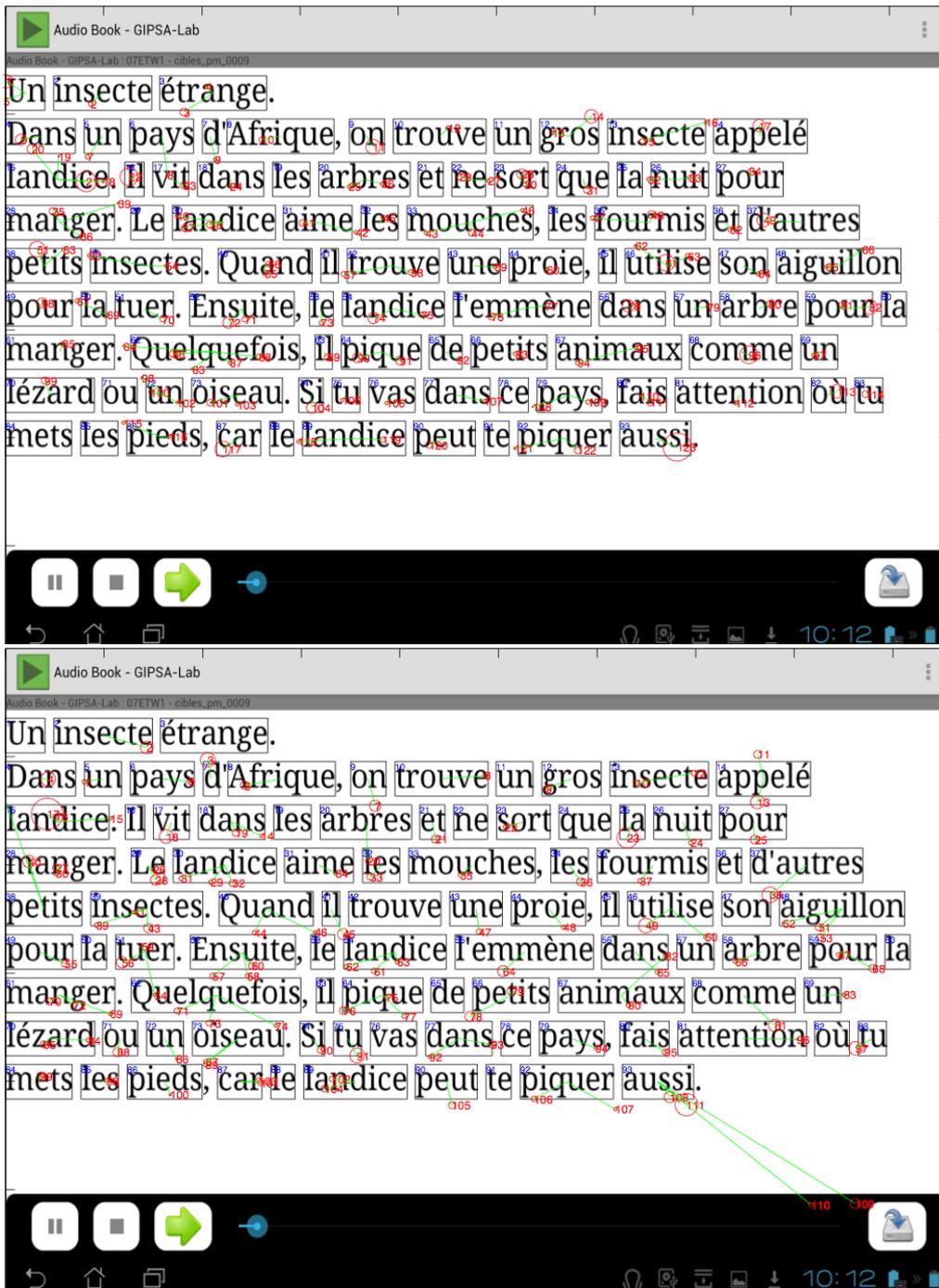


Figure 5. Screenshot for the oculometric results for two texts in two children. The green lines indicate the associations between the fixations (numbered red circles) and the center of the words' bounding boxes (rectangles with a blue number) Upper part : no drift is present ; bottom part : a drift is present.

SYNCHRONIZATION IN READING WHILE LISTENING

2.3.2.1. Pre-analyses

Because there was no fixation point to prime the child's gaze before the display of the texts, we excluded the first three eye fixations, as well as the fixations on the first three bounding boxes (none of which consisted of PW), in each text. We also excluded the fixations for which the distance between its location and its associated bounding box (as determined by the DTW method) was larger than 50 pixels. Finally we excluded fixations that lasted more than 1000ms.

Some fixations patterns were still problematic however, given that some children either displayed calibration issues or did not manage to follow the text as instructed. To check for such problematic data, we computed, for each text of each child, a "local accuracy ratio", the ratio between the number of fixations located exactly inside their associated bounding box (that is, a "correct" fixation) and the total number of fixations. We considered that a ratio inferior to .40 indicated that the associations between the fixations and their bounding boxes were forced by the DTW method, and that such data were not reliable. For some of the children (N=13) all of their texts indicated such a small ratio, and therefore we excluded the data from these children. Not surprisingly, they were the same children with calibration issues noticed during the experiment.

For eight other children, a maximum of three texts presented this ratio. We excluded those problematic texts from the analyses (between 1 and 3 texts per child) but kept their other texts. Of note, among the 16 discarded texts, 11 (69%) were in the S- condition.

In addition, we examined the temporal difference between the actual fixation on any given word and the time when it was aurally displayed. This indicated whether the reading trajectory was synchronized with its oral counterpart, and thus indicate whether the children did follow the instruction to follow the text as it was heard. The examination of this variable revealed that one child (that had otherwise normal parameters as described above) presented a

SYNCHRONIZATION IN READING WHILE LISTENING

pattern that contrasted with the regular pattern. This was the case in all the texts of the S- condition, but in none of the S+ condition. Although the data from this child was further discarded from the analyses, this analysis illustrates the fact that the instruction was more easily followed, at least for some children, in the S+ condition. In particular, this child was 8;10 years old and presented a reading age of 7;0 (the lowest in the sample):

For two other children, a maximum of two texts presented this ratio. We excluded those problematic texts from the analyses (1 text per child) but kept their other texts. Of note, these 2 texts were in the S- condition, and these children already had discarded texts based on the .40 local accuracy ratio above.

Finally, in the group of 14 children of which the data were discarded, ten children were from the first class: five 4th grade (2 girls) and five 5th grade (1 girl), and four children were from the second class: one 5th grade (1 girl) and three 6th grade (2 girls). The average reading age of the discarded group (Mean = 9;0, Median = 8;6, Min = 7;0, Max = 13;3) was inferior to that of the 26 remaining children (Mean = 10;4, Median = 10;4 Min = 7;12, Max = 14;0), suggesting that younger readers tended to have more problematic collected oculometric data. This makes sense given that any issue in the child's ocular movements would have consequences both on reading age, oculometric calibration, and gaze behavior during reading.

2.3.2.2. Results

Firstly, we want to stress that only very general results are reported here because the experiment was conducted in conditions that were not optimal to test hypotheses on eye movements very precisely. For example, the texts extended over several lines and there was no specific control over the visual properties of the text. Thus, the study was only meant to be exploratory with respect to eye movements.

To compare the effect of the S+ and S- conditions on gaze behaviour, several dependant variables were examined (e.g., duration of fixation) in the final sample of 26

SYNCHRONIZATION IN READING WHILE LISTENING

children. For each variable, the mean was first computed by child and by text, and then across the 6 (or 5, 4, or 3) texts of each condition for each child. Figure 6 presents the mean (M) and standard error (SE) of six variables for each child in both conditions. For each variable, a bilateral paired Student t test was performed to compare the children's means in the S+ and S- conditions. The reported means and standard errors reported below were computed across all the 26 children separately in the two conditions.

The rate of words that were totally skipped (i.e., not fixated at all) was not significantly different in the S+ (22.22% of words, SE = 0.009) and the S- (20.98%, SE = 0.010) conditions in S- ($t(25) = 1.76$, $p = 0.09$). The rate of words that were initially skipped and subsequently received a regressive saccade from another word was lower in the S+ (M = 9.15% of words, SE = 0.0067) than in the S- (M = 11.61%, SE = 0.0084; $t(25) = 4.58$, $p < .001$; Figure 6A). Regressive saccades on previous words, when considering all the fixations (i.e., even when the target word had already been fixated), were less frequent in the S+ (M = 15.11% of words, SE = 0.013) than in the S- condition (M = 19.86%, SE = 0.015), $t(25) = 5.88$, $p < .001$ (Figure 6B). Taken together, the two previous results indicate that the S+ condition prevents the usual between-word regressive saccades, especially on words that were initially skipped.

The average number of fixations per word (for the words that were fixated) was smaller in the S+ condition (M = 1.54, SE = 0.022) than in the S- condition (M = 1.62, SE = 0.020; $t(25) = 6.05$, $p < .001$ Figure 6C). (Note that the analyses run for all words, i.e., even those not fixated, led to the same pattern).

The average duration of the first fixation on words was higher in the S+ (M = 293ms, SE = 7.69) than in the S- (M = 270ms, SE = 6.55) condition, $t(25) = 5.15$, $p < .001$; Figure 6D). (Note that the analyses run after excluding the words that were initially skipped and then received a regressive saccade from another word led to the same pattern). The average total

SYNCHRONIZATION IN READING WHILE LISTENING

fixation duration on any given fixated word was higher in the S+ ($M = 423$ ms, $SE = 6.47$) than in the S- condition ($M = 410$ ms, $SE = 6.34$), $t(25) = 2.67$, $p = .013$; Figure 6E).

Since the total duration of texts was constrained by their oral rendition, the previous results are overall consistent with each other. In the S+ condition, each word was, on average, fixated less often and, therefore, fewer within-words saccades were performed. Thus, the total duration of fixation was longer.

Figure 6F presents the average fixation delay of the first fixations on words (excluding the words that were initially skipped and then received a regressive saccade from another word). Fixation delay is the temporal delay (ms) between a fixation and the auditive onset of the corresponding word. A fixation delay of 0 ms refers to a fixation that occurred exactly when the word started to be heard. A fixation delay of -300 ms refers to a fixation that occurred exactly when the word started to be highlighted (remind that a 300-ms audiovisual delay was set). A fixation delay of -400 ms refers to a fixation that was anticipated with 100ms relative to the highlighting of the word. A fixation delay between -300 and 0 ms refers to a fixation that occurred between the highlighting of the word and its auditive onset, suggesting a visual attentional capture. It can be seen in the figure that the average delay varies substantially between the children, with a minimum delay of -8 ms and a maximum of -746 ms. In addition, the delays were larger in the S+ ($M = -412$ ms, $SE = 31.6$) than in the S- ($M = -328$, $SE = 38.6$), $t(25) = 4.21$, $p < .001$, indicating that the children anticipated more in the synchronized condition. Visual examination of Figure 6 suggests that: a) roughly one third of children ($N=10$ on the left part) fixated the words in S+ *before* they were highlighted (delay of approximately -500ms, i.e., larger than -300ms), with a similar anticipation in S-; b) one third ($N=9$ in the middle) fixated the words in S+ with an anticipation of approximately 100ms before the highlighting (-400ms) while in S- they fixated the words 300ms before their auditive onset (-300ms); c) one third ($N=7$, far right of the figure) fixated words

SYNCHRONIZATION IN READING WHILE LISTENING

approximately 100ms *after* they were highlighted in the S+ (-200ms), and approximately 100ms before their auditory onset in the S- condition (-100ms). Thus, the highlighting helped these children to fixate the words with a delay similar to that of the intermediate group. Interestingly, the average reading level was different in those three subgroups, respectively 11;1, 10;4, and 9;4 years. These results are consistent with the hypothesis that the very good readers (group on the left) anticipated the reading to a great extent because of possible discomfort or constraints (in both conditions and perhaps even more in S+) and that the poorest readers (group on the right) benefitted from the attentional capture provided by the highlighting of words.

Overall, these oculometric results are consistent with the data generally observed in children's reading (Rayner, Ardoin, & Binder, 2013), to the exception of regressive saccade rates which are inferior to what is observed in other studies (around 25%). This may be due to the provided oral version of the text that somewhat forces a forward oculomotor trajectory both in S- and in S+. The results are also similar to those obtained with middle-school students (Gerbier et al., 2015).

As hypothesized, the synchronization seemed to have triggered the visual capture of visual attention to a certain extent, and forced the children to follow the imposed visual pace. This was reflected in fewer fixations per word, longer initial fixations on words, a longer total fixation per word, and fewer regressive saccades in the S+ than in the S- condition. These indications of attentional capture in the S+ condition are consistent with the preference for the S- condition expressed by the best readers in the sample. The assumed attentional capture in the S+ condition may be experienced by the best readers as a source of discomfort. In contrast, this attentional capture may be a precious help for poor readers that have not yet automatized the oculomotor strategies associated with reading.

SYNCHRONIZATION IN READING WHILE LISTENING

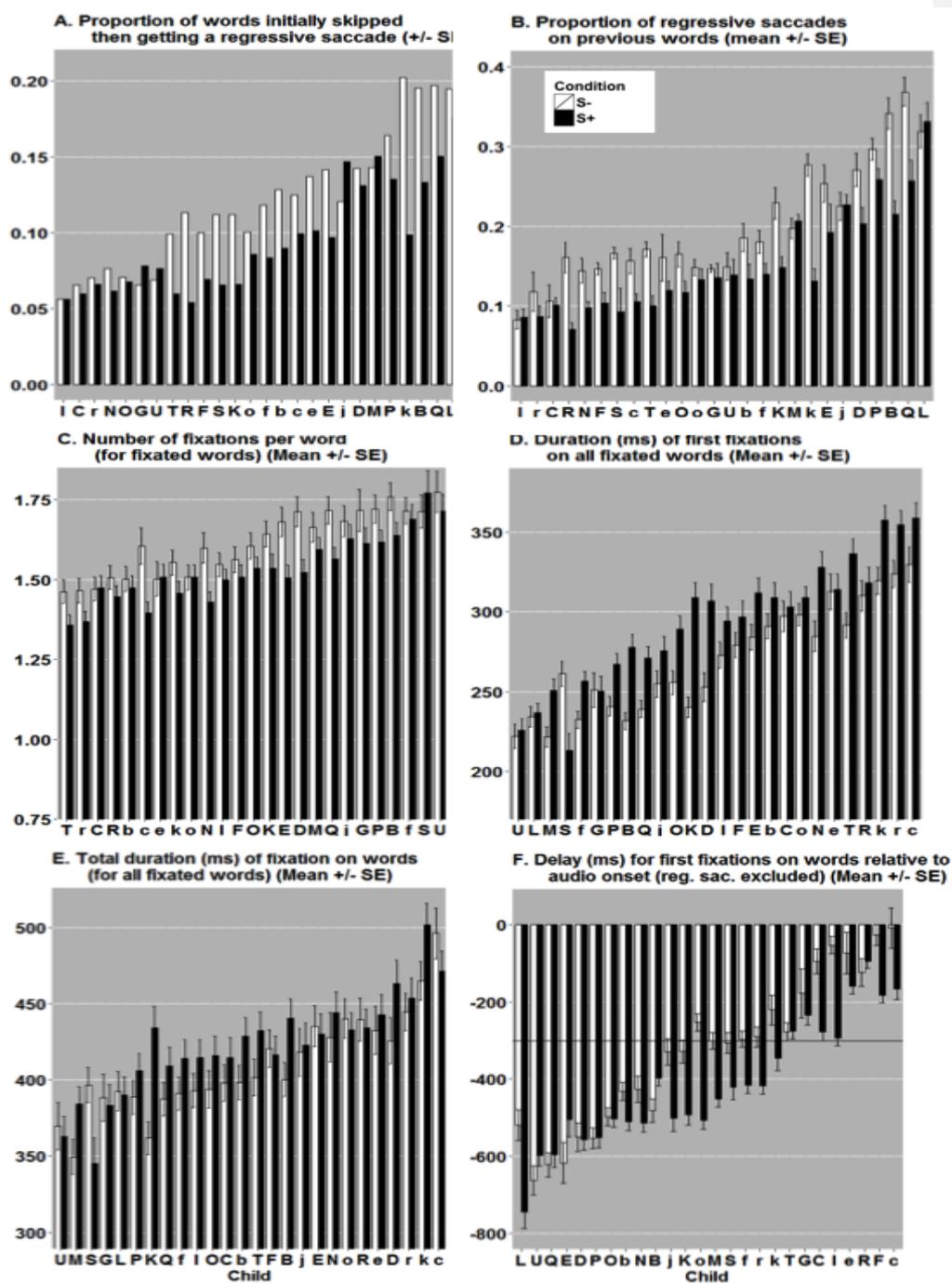


Figure 6. Representation of the gaze data for six oculometric variables in the 26 children. For the sake of readability, the display of the children on the x-axis was ordered with respect to their average level on this variable and, as a consequence, the order is different on every plot. As can be seen, the effects are generally consistent across the children. Error bars represent standard errors of the means.

SYNCHRONIZATION IN READING WHILE LISTENING

On a final note, although the experimental setting was not suited for analyses focused on specific words (specifically, the position of the PWs on the different lines was had not been controlled), we present here some exploratory results on the pseudowords. First, the same S+/S- difference as those for the words were found, except for the number of regressive saccades and for the total duration of fixation, for which the differences between S+ and S- were no longer significant. Second, we compared the first occurrence of the PW in each text with the three subsequent occurrences grouped together. The first occurrence has a specific status, as it is both a novel string of letters and a novel entity, which suggest that they would be treated in a specific manner (e.g., fixated longer). The number of fixations per PW was higher for the first occurrence ($M = 2.96$, $SE = 0.118$) than for the next three occurrences ($M = 2.03$, $SE = 0.050$; $t(25) = 9.52$, $p < .001$). The duration of the first fixation on the first occurrence ($M = 390$, $SE = 24.3$) was higher than for the next three occurrences ($M = 314$, $SE = 9.54$; $t(25) = 3.31$, $p < .01$). The total duration of fixation on each PW was higher for the first occurrence ($M = 1074$, $SE = 37.3$) than for the other occurrences ($M = 586$, $SE = 12.9$). The anticipation of the first fixation on PW relative to the auditive onset of the PW was higher for the first occurrence ($M = -560$, $SE = 53.6$) than for the last three occurrences ($M = -332$, $SE = 37.4$; $t(25) = 6.04$, $p < .001$). Overall, these data indicate that the first occurrence of a PW in a text is processed more intensely than its subsequent occurrences, confirming the idea that it has a special status.

3. Discussion

The present results are mixed regarding the influence of audio-visual synchronization during RWL in children aged between 8;5 and 11;5. On one hand, we found a detrimental influence of audio-visual synchronization on the short-term memorisation of the semantic category of pseudowords read in the text. This result contradicts our hypothesis and prior results observed with a similar paradigm with middle-school students (Gerbier et al., 2015).

SYNCHRONIZATION IN READING WHILE LISTENING

The reasons for this discrepancy are far from clear but could be related to the reading pace that was much slower in the present study, with 4.81 vs. 5.59 syllables/sec, and 62% vs. 55% phonation rates respectively for the present and the previous study, which indicates a slower pace with shorter pauses in the present study. In addition, the average reading age was close in both samples (10.47 vs. 9.89, respectively). It may be that higher comprehension is attained when the oral pace is adapted to the reading level of elementary students. It can be hypothesised that the S+ condition is damageable when the pace is slow, but beneficial when the pace is faster, relative to the child's reading level, but more exploration is desirable. In addition, there was no effect of synchronisation on the memorization of the orthographic form, consistently with previous findings.

On the other hand, we found that the lower their reading level, the more the children expressed a preference for the synchronized condition. This finding is consistent with our previous study with middle-school students (Spearman's $r = -.52$ and $-.24$, respectively). Thus, synchronized RWL is preferred by the poorest readers above non-synchronized RWL. This is an extension of prior studies showing that RWL is preferred to LO or RO especially by the poorest readers (Holmes & Allison, 1985; Montali & Lewandowski, 1996). Subjective reports from children indicated that they preferred S+ because it helped them follow the text on the screen. This is also consistent with a study from Hecker et al. (2002) where children with attentional disorders reported that computer-assisted learning (including RWL with synchrony) helped them to concentrate efficiently. This suggests that synchronized RWL may be helpful to improve the poor readers' motivation to read. In the case of good readers, in contrast, the synchronisation may have impeded their natural and automatized way of reading, including their tendency to read ahead of the oral version (Gamby, 1983). As a consequence, the constant word-based time-locked synchronisation used here caused discomfort for them. Such discomfort was reported by some children and by adults in pilot studies. This issue

SYNCHRONIZATION IN READING WHILE LISTENING

raises the idea that one needs to dissociate the effect of the genuine audio-visual synchrony from the effect of providing a guide helping readers to follow the text. In the present study, both effects are entangled.

In support to the latter findings, we found that synchronisation did influence the children's gaze pattern. In the S+ condition, words were fixated less often but for a longer period of time. That is, words tended to receive a unique and long fixation, instead of multiple and short fixations in the S- conditions. There were also fewer regressive saccades on previous words in the S+ condition. This is consistent with the idea that synchronized RWL did compel the readers to follow the pace imposed by the oral rendition.

More research is needed to extend these exploratory results. First, further research will need to focus more specifically on struggling or beginning readers. Indeed, good readers seem to neither benefit from synchronisation nor to enjoy it, in contrast to poor readers. The best modalities of synchronized RWL should be investigated in order to know how it could be used as an effective and efficient way to encourage poor readers to read. Second, other measures of learning need to be investigated with respect to synchronisation. For example, would synchronized RWL enhance reading fluency, prosody, high-level comprehension, or vocabulary development? Third, the effect of the oral reading pace relative to the child's reading level should be explored more. We experimentally set the reading pace based on our intuition of which pace would be the most appropriate for children's age in our samples. Shany and Biemiller (1995) stated that the oral pace should not be more rapid than the child's reading speed. In addition, in practical settings, the oral pace should be a parameter that the user can modify (Bergman, 1999). Fourth, the delay between oral rendition and visual highlighting could also be modulated. In the present study, we set a constant audio-visual asynchrony of 300ms (based on previous pilot studies with expert readers) but it might be different for each reader, and even more specifically, modulated for each word and/or

SYNCHRONIZATION IN READING WHILE LISTENING

sentence. This modulation could be obtained by collecting specific audio-visual synchronization profiles – between acoustic onsets and fixation onsets - of reading experts for these texts, and to investigate whether learning and preference patterns are improved when words in experimental texts are highlighted according to these fixation profiles.

Another modulation that could be of importance is the spatial span of synchronisation. The synchronisation in the present study occurred on the word level because we were primarily interested in the acquisition of lexical orthography. However, highlighting individual words creates a rapid succession of attentional captures, which might be the source of discomfort and the reason for the overall preference for the S- condition. Larger units of synchronisation might be more appropriate, like groups of words (e.g., “il y a”, i.e., “there is” in French), phrases, or even the whole sentences. In contrast, smaller units (e.g., graphemes) might be more appropriate for beginning readers (Wise et al., 1989; Wise, 1992), although they may be associated with a “blinking light” effect that is to be avoided. Related, the effects of visual parameters such as the colours and the contrasts between colours deserve more investigation. All these modulations are currently possible in our application.

A few other points also deserve discussion. For instance, the learning timescale under consideration is important. The present study investigated immediate retention of verbal information after a unique episode of reading. It would be interesting to study the effect of synchronized versus unassisted RWL in long-term studies involving training and testing over several days, as was done in most of the research presented in the introduction. In addition, the instructions that the children had to follow during reading were dictated by the collection of eye-tracking data. Specifically, we invited the children to visually follow the oral pace in both conditions, so that we could reasonably compare both conditions in the eye-tracker analyses. It will be useful to dissociate the studies about verbal acquisition, in which the instructions will prompt a more natural reading behaviour, from the studies about gaze

SYNCHRONIZATION IN READING WHILE LISTENING

movements, in which the reading behaviour must be constrained. Finally, it would be interesting to extend our findings by applying more experimental control over various aspects of text content and presentation on eye movements, such as presenting isolated sentences with words of controlled length, and controlling their position on the screen.

4. Conclusions

One of the interests of computer-assisted reading is that it does not require the full-time presence of a teacher or an adult to accompany the child during reading. Overall, the present and previous research showed that RWL is economical since it can be used in the classroom with several children using the system in parallel. In addition, some results are encouraging regarding its potential benefits for beginning or struggling readers in particular. If the benefits of *synchronized* RWL on literacy skills are to be further demonstrated in the future for beginning or struggling readers, it could be used in combination with other existing tools to improve children's learning to read (Shany & Biemiller, 1995) and to help reading-disabled children (Wise et al., 1989). Our motivation for using computer tablets was to enable autonomous reading sessions at school or at home at the children's pace. We expect to conduct evaluations of the efficiency of synchronized RWL using embedded eye tracking technology on mobile devices.

5. Acknowledgements

This work was supported by the ANR-12-BSH2-0013 ORTHOLEARN project. We are thankful to the schools, teachers, and children who welcomed us to perform the experiments. We warmly thank Julien Minet for the design of the electronic circuit that triggers the eye tracker.

SYNCHRONIZATION IN READING WHILE LISTENING

6. References

- Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology, 104*(1), 250–263. doi: 10.1037/a0026147
- Anderson-Inman, L., & Horney, M. A. (2007). Supported eText: Assistive technology through text transformations. *Reading Research Quarterly, 42*(1), 153–160. doi: 10.1598/RRQ.42.1.8
- Bailly, G., & Barbour, W. (2011). Synchronous reading: learning French orthography by audiovisual training. In *12th Annual Conference of the International Speech Communication Association* (pp. 1153–1156).
- Bergman, O. (1999). *Wait for me! Reader control of narration rate in talking books*. Newark, DE: International Reading Association.
- Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509–33.
- Boersma, P. P. G. (2002). Praat, a system for doing phonetics by computer. *Glott international, 5*.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136–163.
- Carbo, M. (1978). Teaching reading with talking books. *Reading Teacher, 32*(3), 267–73.
- Chang, A. C.-S., & 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

SYNCHRONIZATION IN READING WHILE LISTENING

Chaves, N. (2012). *Rôle du traitement visuel simultané dans l'acquisition des connaissances orthographiques lexicales*. Université Toulouse II Le Mirail, Toulouse, France.

Chomsky, C. (1976). After decoding: What? *Language Arts*, 53(3), 288–314.

Elbro, C., Rasmussen, I., & Spelling, B. (1996). Teaching reading to disabled readers with language disorders: A controlled evaluation of synthetic speech feedback. *Scandinavian Journal of Psychology*, 37(2), 140–155. doi: 10.1111/j.1467-9450.1996.tb00647.x

Elkind, J., Cohen, K., & Murray, C. (1993). Using computer-based readers to improve reading comprehension of students with dyslexia. *Annals of Dyslexia*, 43, 238–259. doi: 10.1007/BF02928184

Gamby, G. (1983). Talking books and taped books: Materials for instruction. *Reading Teacher*, 36(4), 366–69.

Gerbier, E., Bailly, G., & Bosse, M.-L. (2015). Using Karaoke to enhance reading while listening: impact on word memorization and eye movements (pp. 59–64). Presented at the Speech and Language Technology for Education (SLaTE).

Hecker, L., Burns, L., & Elkind, J. (2002). Benefits of assistive reading software for students with attention disorders. *Annals of Dyslexia*, 52, 243–272.

Holmes, B. C., & Allison, R. W. (1985). The effect of four modes of reading on children's comprehension. *Reading Research and Instruction*, 25(1), 9–20.

Lefavrais, P. (2006). (2006). Test de leximétrie de l'Alouette. Paris: ECPA.

Legge, G. E., & Bigelow, C. A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision*, 11(5), 8-8.

Lewandowski, L. J., & Kobus, D. A. (1993). The effects of redundancy in bimodal word processing. *Human Performance*, 6(3), 229–239. doi: 10.1207/s15327043hup0603_3

Medwell, J. (1998). The Talking Books Project: Some Further Insights into the Use of Talking Books to Develop Reading. *Reading*, 32(1), 3–8.

SYNCHRONIZATION IN READING WHILE LISTENING

Montali, J., & Lewandowski, L. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities, 29*(3), 271–279. doi: 10.1177/002221949602900305

Peereman, R., Lété, B., & Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme—phoneme mappings, and infralexical and lexical units in child-directed written material. *Behavior Research Methods, 39*(3), 579–589. doi: 10.3758/BF03193029

Rabiner, L., Rosenberg, A. & Levinson, S. (1978). Considerations in dynamic time warping algorithms for discrete word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 26*(6), 575-582.

Rayner, K., Ardoin, S. P., & Binder, K. S. (2013). Children's eye movements in reading: A commentary. *School Psychology Review, 42*(2), 223–233.

Reitsma, P. (1988). Reading practice for beginners: Effects of guided reading, reading-while-listening, and independent reading with computer-based speech feedback. *Reading Research Quarterly, 23*(2), 219–235. doi: 10.2307/747803

Shany, M. T., & Biemiller, A. (1995). Assisted reading practice: effects on performance for poor readers in grades 3 and 4. *Reading Research Quarterly, 30*(3), 382–395.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72*(2), 95–129. doi: 10.1006/jecp.1998.2481

Steele, E., Lewandowski, L., & Rustling, E. (1996). *The effectiveness of bimodal text presentation for poor readers*. Presented at the Annual convention of the National Association of School Psychologists, Atlanta, GA.

Torgesen, J. K., Dahlem, W. E., & Greenstein, J. (1987). Using verbatim text recordings to enhance reading comprehension in learning disabled adolescents. *Learning*

SYNCHRONIZATION IN READING WHILE LISTENING

Disabilities Focus, 3(1), 30–38.

Torgesen, J. K., Dahlem, W. E., & Greenstein, J. (1987). Using verbatim text recordings to enhance reading comprehension in learning disabled adolescents. *Learning Disabilities Focus*, 3(1), 30–38.

Van der Leij, A. (1981). Remediation of reading-disabled children by presenting text simultaneously to eye and ear. *Bulletin of the Orton Society*, 31(1), 229–243. doi: 10.1007/BF02658613

Verlaan, W., & Ortlieb, E. (2012). Reading while listening: Improving struggling adolescent readers' comprehension through the use of digital audio recordings. In *What's hot in literacy: Trends and issues* (Dubuque, pp. 30–36). IA: Kendall Hunt.

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

Wise, B., Olson, R., Anstett, M., Andrews, L., Terjak, M., Schneider, V., ... Kriho, L. (1989). Implementing a long-term computerized remedial reading program with synthetic speech feedback: Hardware, software, and real-world issues. *Behavior Research Methods, Instruments & Computers*, 21(2), 173–180. doi: 10.3758/BF03205578

Wise, B. W. (1992). Whole words and decoding for short-term learning: Comparisons on a “talking-computer” system. *Journal of Experimental Child Psychology*, 54(2), 147–167. doi: 10.1016/0022-0965(92)90033-3

7. Appendixes :

Appendix 1 : Translations of two experimental texts

Text in Figure 5 - Upper part. At the zoo. Last week, I went to the zoo. We saw weird animals. The one I preferred was called a pirisse. I had never seen an animal like the pirisse before. It looks a little like a monkey, but it barks like a dog. The zookeeper placed the pirisse with the giraffes, because it was lonely. The pirisse likes to climb up the giraffes' neck to eat on top of the trees. They let it, because, in exchange, it scratches their ears and they love it.

Text in Figure 5 - Bottom part. A weird insect. In a country in Africa, one can find a big insect named landice. It lives in the trees and goes outside only at night to eat. The landice likes flies, ants, and other small insects. When it finds a prey, it uses its sting to kill it. Then, the landice brings it up to a tree to eat it. Sometimes, it stings small animals like a lizard or a bird. If you ever go to this country, be careful where you put your feet because the landice can sting you too.

Appendix 2: Source code for the DTW algorithm

```

algorithm align-fixations-with-words is
input:
  fix(1:nb_fix,2) // columns are locations of fixations % to the screen size
  tfix(nb_fix) // fixation timestamps (s)
  box(1:nb_box,4) // columns are locations of the corners of the boundingbox of the written
word % to the screen size
  tbox(nb_box) // box enlightening timestamps (s)
  wt = 0.5 // close-shadowing weight
  wf = 0.1 // neighbourhood weight
output:
  algn(1:lg_algn,1) // index of the word aligned with each fixation
use:
  d = dist_fixation2boundingbox(i_fix,i_box) // distance between the fixation i_fix and
boundingbox i_box; =0 if the fixation is inside the boundingbox
  [v,rg] = min(t) // value and index of the minimum in t

for i_fix=1:nb_fix do
  for i_box=1:nb_box do
    DL(i_fix,i_box) = dist_fixation2boundingbox(i_fix,i_box) + wt * abs(tfix(i_fix)-
tbox(i_box)) // taking into account close-shadowing
  end
end

```

SYNCHRONIZATION IN READING WHILE LISTENING

```
Dcp = DL(1,:);
bpt(1,:) = 0; // backtracks the alignment
for i_fix=2:nb_fix // forward pass
    for i_box=1:nb_box
        [Dc(i_box),bpt(i_fix,i_box)] = min(Dcp + wf * abs((1:lg_box)-i_box+1)) +
        DL(i_fix,i_box)) // favors previous fixations from neighbouring boxes
    end
    Dcp=Dc
end

[v,i_box] = min(Dc) // initiate backtracking
algn=[];
while i_fix > 0
    algn(i_fix)=i_box;
    i_box=bpt(i_fix,i_box);
    i_fix = i_fix - 1;
end
```

8. Vitae

Emilie Gerbier is assistant professor at Université de Nice Sophia Antipolis, at the Bases, Corpus, Language lab. She has a PhD in Cognitive Science and has expertise in the cognitive psychology of learning and memory, especially on the effect of repetition and sleep on learning. She is also interested in reading acquisition, in particular through the use of computer technology. She teaches statistics and cognitive psychology in the department of Psychology.



SYNCHRONIZATION IN READING WHILE LISTENING



Gérard Bailly is a senior CNRS Research Director at GIPSA-Lab, Grenoble-France. He was deputy director of the lab in 2007-2012.

He has been working in the field of speech communication for 30 years, supervised 27 PhD Thesis, authored 40 journal papers, 24 book chapters and more than 170 papers in major international

conferences. He coedited *Talking Machines: Theories, Models and Designs* (Elsevier, 1992), *Improvements in Speech Synthesis* (Wiley, 2002) and *Audiovisual speech processing* (CUP, 2012). His current interests include the conception and evaluation of interactive systems, in particular social robots and virtual conversational agents.

Marie-Line Bosse is professor of cognitive psychology at the University Grenoble-Alpes, Grenoble, France. She is the team leader of the Language team at the LPNC (Laboratoire de Psychologie et Neurocognition, UMR 5105 CNRS). She works on the cognitive factors of child reading and orthographic acquisition.

SYNCHRONIZATION IN READING WHILE LISTENING

