The role of audiovisual speech in the early stages of lexical processing as revealed by the ERP word repetition effect

Anahita Basirat[a], Angèle Brunellière[a], and Robert Hartsuiker[b]

aUniv. Lille, CNRS, CHU Lille, UMR 9193 – SCALab – Sciences Cognitives et Sciences Affectives, Lille, France and bDepartment of Experimental Psychology, Ghent University, Ghent, Belgium

Final author copy

Language Learning 68:s1, June 2018, pp. 80–101 80

DOI: 10.1111/lang.12265

Abstract

Numerous studies suggest that audiovisual speech influences lexical processing. However, it is not clear which stages of lexical processing are modulated by audiovisual speech. In this study, we examined the time-course of the access to word representations in long-term memory when they were presented in auditory-only and audiovisual modalities. We exploited the effect of the prior access to a word on the subsequent access to that word known as the word repetition effect. Using event-related potentials, we identified an early time window at about 200 ms and a late time window starting at about 400 ms related to the word repetition effect. These time windows might respectively reflect the early stages of contact with the lexicon and the late stages of access to lexical and semantic representations. Our results showed that the word repetition effect over the early time window was modulated by the speech modality while this influence of speech modality was not found over the late time window. Visual cues thus play a role in the early stages of lexical processing.

Keywords: Audiovisual speech, lexical processing, repetition effect, ERP

1. Introduction

Speech perception is driven by both acoustic information from the auditory signal and visual cues from speakers' articulatory gestures. It is well known that speech visual cues affect the perceptual processing of speech (for a review, see Campbell, 2008). When auditory information is degraded, e.g. in noisy situations, seeing the speaker's articulatory gestures provides listeners with complementary information about what is said, thus enhancing speech identification (e.g. Sumby & Pollack, 1954; Benoit, Mohamadi, & Kandel, 1994). The beneficial effect of visual speech is not limited to situations where auditory speech is degraded. For example, the comprehension of a clearly audible story improves when the speaker's face can be seen (Arnold and Hill, 2001). There is a growing body of evidence suggesting that these cues contribute to word recognition and lexical access (for a review, see Peelle & Sommers, 2015). However, the interaction between the processes underlying audiovisual speech integration and those related to lexical processing remains elusive. For example, a recent event-related potential (ERP) study (Baart & Samuel, 2015) showed that audiovisual speech integration and lexical processing affected speech processing but did not interact together. The goal of our study was to further investigate the impact of visual speech on lexical processing.

Most of the behavioral and electrophysiological studies showing the contribution of visual speech cues to lexical processing looked for a contrast between words and pseudo/non-words or low- and high-frequency words. For instance, using a lexical decision task in a priming paradigm, Kim, Davis and Krins (2004) observed that seeing the speaker's articulatory gestures of a word (prime) sped up the lexical decision made for the next presentation of that word (target) presented in auditory-only or in written modalities (e.g. when "back" was first presented in visual-only as prime and then in auditory-only or in written modalities as target). This priming effect did not exist for non-word targets. The same results were observed in an oral production task (i.e. participants were

asked to name the targets). The fact that a priming effect was found only for word targets supports the idea that visual speech influences the activation of lexical representations. This interpretation is consistent with other studies showing that the influence of visual cues on speech processing is modulated by lexicality. For example, in audiovisual modality when the auditory signal is degraded, it is easier to recognize a phoneme embedded in a word than in a pseudo-word (Fort, Spinelli, Savariaux, & Kandel, 2010). Seeing the onset of low-frequency words (and not that of high-frequency words) primed the auditory recognition of that word (Fort et al., 2013). In addition, the McGurk effect (e.g. auditory /ba/ dubbed onto visual /ga/ may result in the perception of a totally new syllable such as /da/. McGurk & MacDonald, 1976) was found to be more frequent when the visual stimulus formed a word and the auditory stimulus formed a non-word (Brancazio, 2004). However, some contradictory results about the influence of lexicality on the McGurk effect have been found (e.g. Dekle, Fowler, & Funnell, 1992; Sams, Manninen, Surakka, Helin, & Kättö, 1998; Windmann, 2004; Barutchu, Crewther, Kiely, Murphy, & Crewther, 2008). The aim of our study was to provide insights into whether or not speech visual cues play a role in lexical processing.

According to psycholinguistic models of spoken-word recognition (e.g. Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994), access to lexical representations can be understood in two stages: (1) activation of a set of candidates that are phonologically similar to the input and (2) selection of the most relevant candidate among the set of activated candidates. To understand whether or not visual speech influences lexical processing, it is important to disentangle these stages of lexical processing during audiovisual speech processing: does audiovisual speech modulate either of these stages? To answer this question, we exploited the word repetition effect, which reflects the facilitatory effect of prior processing of a word on its subsequent processing. For example, in a lexical decision task, words were recognized more rapidly as a function of the number of repeated presentations (Forbach, Stanners, Hochhaus, 1974). The authors reported that this

facilitatory effect did not exist for non-words. This effect has been generally linked to the access to word representations in long-term memory. According to the Logogen model suggested by Morton (1969), the word repetition effect can be explained by a decrease in the threshold of lexical representations. In other words, less excitation is needed in a subsequent presentation of a word for it to become activated.

In the current study, we looked at the word repetition effect on ERPs. The repetition effect may be revealed by a two-stage brain response: an early response at about 200 ms and a later response starting at about 400 ms after the onset of a repeated word (Rugg, 1987). The N400 component in response to the second presentation of a word is more positive than the N400 wave in response to the first presentation of that word, especially on posterior electrodes (Rugg, 1985; Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991; Rugg, Doyle, & Melan, 1993). Interestingly, the N400 is a well-known ERP component related to lexico-semantic processing of words (for a review, see Kutas & Federmeier, 2011). In line with behavioral measurements, the repetition effect on the N400 indicates that a prior access to a word representation in long-term memory can modulate the subsequent access to that representation. Although a word repetition effect starting around 400 ms post-stimulus is typically observed, an earlier effect around 200 ms has also been reported in the literature (Rugg, 1987; Snijders, Kooijman, Cutler & Hagoort, 2007; Cheng, Schafer, & Riddell, 2014). Like the late repetition effect, this early response in the P200 time window seems to index mechanisms underlying lexical processing. Indeed, Almeida & Poeppel (2013) observed a repetition effect on this time window whose amplitude in response to words differed from pseudo-words. Therefore, the ERP word repetition effect appears to be a good tool to investigate lexical processing during two different time windows. The early and late time windows would reflect respectively early stages of lexical processing (roughly, activation of candidates) and access to the word representation (roughly, final selection of the best candidate). Moreover,

studying the word repetition effect on ERPs offers the opportunity to study lexical processing in an implicit manner since it is not needed to draw participants' attention to the lexicality by using words, pseudo(non)-word stimuli and/or any explicit task.

We examined two-word repetition conditions in this study: (1) word repetition in isolation, i.e. a silence between two successive words, a procedure commonly used in word repetition paradigms; and (2) word repetition in sentence contexts, i.e. the critical word is embedded in a sentence and its subsequent presentations are embedded in novel sentences. Testing the word repetition effect in sentence context is interesting as this condition involves segmentation of continuous speech. Since audiovisual speech is known to be helpful in finding word boundaries in continuous speech (Mitchel & Weiss, 2014), the repetition effect on ERPs for words repeated in isolation and those repeated in sentence context could be modulated differently by auditory-only and audiovisual modalities. Using auditory-only stimuli, Snijders et al. (2007) studied the word repetition effect both in isolation and sentence contexts. Native and foreign listeners of Dutch listened to a repetition of a familiarized word following several sentences. Some of these sentences contained the familiarized word and some of them contained a novel word. In both groups, the authors observed a repetition effect on ERPs (i.e. a more positive N400 in the second compared to the first presentation of words) for the familiarized word repeated in isolation and in sentence context. For novel words that were only repeated in sentence context, the repetition effect was observed in native listeners but not in foreign ones. The word repetition effect is thus modulated by lexical knowledge and presentation context.

If seeing the speaker's articulatory gesture contributes to lexical processing of speech, an interaction between the word repetition effect (first presentation/second presentation) and the speech modality (auditory-only/audiovisual) should be observed. An interaction during the early time window would reflect the role of audiovisual speech in early contact with the lexicon, while an interaction during

the late time window would reflect the role of audiovisual speech in the late stages of lexical processing, i.e. access to lexical and semantic representation of that word.

2. Materials and methods

2.1 Participants

Sixteen (13 females) right-handed native French speakers participated in this study. Their mean age was 22 years (SD=3). They all had normal hearing by self-report, normal or corrected-to-normal vision, and no known history of neurological disease. The experiment was approved by the local ethical committee. Before testing, all participants were informed about the experiment by a written document and signed a consent statement.

2.2 Stimuli

A list of 40 CVCV (C: consonant, V: vowel) French words was selected from the Lexique 3.8 database (New, Pallier, Brysbaert & Ferrand, 2004). All words were singular masculine nouns. Among the selected words, 20 semi-arbitrary pairs were formed such that the two members of each pair had different onsets (e.g. "*Pompon (*=pompom*)* - *Furet (*=ferret*)"*). The mean lemma frequency of selected nouns, based on film subtitles, was 3.3 occurrences per million of words (SD =3.4). Another 20 pairs of words with the same properties as the first pairs were also selected (mean lemma frequency per million = 3.7, SD = 3.9). They were matched to the first pairs for their onset (e.g. "*Ponton (*=pontoon*) – Fusain (*=charcoal*)"* was matched to *"Pompon-Furet"*) (see Appendix for a complete list of pairs used in the experiment). This matching was done because one set of pairs was used in the auditory-only modality and the other was used in the audiovisual modality (see Experimental design and procedure). For each word, four short sentences containing that word were constructed. The position of the critical word and the word that preceded the critical word were matched within pairs (see table 1 for an example). Sentences provided no constraining

semantic information prior to the critical word. Stimuli (words in isolation and sentences) were produced by a female native French speaker and recorded audiovisually by a camera in front of her at a rate of 50 frames/second. The head, neck and shoulders of the speaker were visible in the videos. The videos were edited using Adobe Premier Pro. The size of the face on the screen was 6 x 8 cm. The mean duration of the words, calculated with Praat (Boersma, 2002), was 583 ms (SD = 83) in isolation and 479 ms (SD = 111) in sentence context. The sound level was 61 dB (SD = 2).

2.3 Experimental design and procedure

The experimental design was adapted from the study of Snijders et al. (2007) which is described in the introduction. During the experiment, participants received 20 blocks in auditory-only modality and 20 blocks in audiovisual modality. Table 1 shows an example of two matched blocks. Each block began with the presentation of an isolated word that was repeated 7 more times (= familiarized word). There was a 2-second silence between each word repetition. After the eighth presentation of the familiarized word, four sentences that contained the familiarized word and four others containing a novel word (= unfamiliarized word) were delivered to the participants in a randomized order (see Table 1 for an example). There was a 3-second silence between each sentence.

The experiment (i.e. 40 blocks) was divided into 20 parts, each comprised of two blocks presented in the same modality: an auditory-only part followed by an audiovisual part and vice-versa. The order of parts was counterbalanced across participants. After the delivery of each part, a picture was presented on the screen to maintain the attention of the participants. They were asked to judge whether the picture represented France (e.g. Eiffel Tower) or another country (e.g. Colosseum) by pressing F or J keys. No further explicit task was requested of them.

Four versions of the experiment were constructed so that each word appeared in both familiarized and unfamiliarized modality and in auditory-only and audiovisual modality across the participants. Thus, the familiarized status of a word of each pair and the modality of pairs were counterbalanced across participants. During the experiment, participants were asked to listen carefully to the sentences through headphones and to watch the computer screen carefully. The sound was presented at a comfortable level and the screen was placed at about 50 cm from the participants. They were asked to avoid eye-blinking and body movements during the presentation of words and sentences. Each participant received 20 auditory-only and 20 audiovisual blocks. To familiarize them with the procedure, all participants received one block in auditory-only and one in audiovisual modality before the experimental blocks. Stimuli were presented using Psychtoolbox in Matlab.

[Insert table 1 about here]

2.4 EEG recording and analyses

Scalp voltages were acquired using a 128-channel Biosemi system (Biosemi, Amsterdam, the Netherlands). They were amplified and sampled at 1024 Hz. Horizontal and vertical electro-oculograms were recorded simultaneously to monitor eye movements. Two other electrodes were placed on the left and right mastoids. Mastoid electrode activity was averaged to re-reference the EEG signal off-line. EEG data was filtered by a band-pass filter between 1 and 30 Hz and a notch filter at 50 Hz. Recordings were segmented 200 ms before (baseline) and 1200 ms after the onset of critical words. Epochs contaminated by eye or motion artifacts were automatically rejected (±70 μV). ERPs were averaged per participant and per condition. For words in isolation, ERPs to the first and second presentations of words in auditory-only and audiovisual modalities were obtained for each participant (= 2x2 conditions). For words in sentence context, ERPs to the first and second presentations of novel (unfamiliarized) words in auditory-only and audiovisual modalities were

calculated (= 2x2 conditions). We also calculated ERPs to the first presentation of a familiarized word in sentence context which were compared to the first presentation of a novel word as a control (see EEG analysis). The average number of trials kept after artifact detection ranged from 13 to 20 (mean: 18) for each condition. The pre-processing analysis was performed using the Cartool software (brainmapping.unige.ch/cartool).

Our goal was to test whether there was any interaction between the word repetition effect and the speech modality. ANOVAs were performed using nine representative channels: D4 (left anterior), Fz (middle anterior), C4 (right anterior), D19 (left central), Cz (middle central), B22 (right central), A7 (left posterior), Pz (middle posterior) and P4 (right posterior). As discussed in the introduction, previous studies on the word repetition effect reported an early brain response at around 200 ms and a late response starting at around 400 ms. Based on this literature and visual inspection of our data (Figure 1), we pre-selected two time windows: an early time window from 170 ms to 280 ms and a late time window from 400 ms to 700 ms post-stimulus. The aim was thus to test a possible interaction between the repetition effect and the speech modality over each time window. For words presented in isolation, we performed a repeated measures ANOVA with Modality (auditory-only/audiovisual), Order (first/second presentation of familiarized word), Position (anterior/central/posterior) and Lateralization (left/middle/right) as factors over the early and the late time window. When significant interactions were shown, paired Student t-tests were used for post-hoc comparisons. To ensure that there was no Modality x Order interaction over earlier time windows, 0-80 ms and 80-150 ms time windows were also analyzed using ANOVAs, as described above.

In sentence context, critical words were embedded in continuous speech and did not follow a period of silence, so the early ERP responses to these words were not clearly visible. This is in line with previous studies showing a reduction/disappearance of early ERP peaks due to the lack of silence

between successive auditory stimuli (e.g., Connolly et al., 1992; Hagoort & Brown, 2000; Näätänen & Picton, 1987). Thus, we performed our analysis only over the 400-700 ms time window. Two comparisons were made to examine any influence of speech modality on repetition effect. First, an ANOVA was performed with Modality (auditory-only/audiovisual), Order (first/second presentation of unfamiliarized word), Position (anterior/central/posterior) and Lateralization (left/middle/right) as within-subject factors. Second, we compared the first presentation of the familiarized words in sentence context (ninth overall presentation) and the first presentation of unfamiliarized words using an ANOVA with Modality (auditory-only/audiovisual), Order (first presentation of familiarized word in sentence context/first presentation of unfamiliarized word), Position (anterior/central/posterior) and Lateralization (left/middle/right) as within-subject factors. When significant interactions were shown, paired Student t-tests were used for post-hoc comparisons.

3. Results

3.1 Words in isolation

Figure 1 shows the time course of ERPs to critical words in four conditions of isolated presentation (auditory-only/audiovisual modalities x first/second presentations). For the sake of brevity, only significant results of ANOVAs are reported below (see also Table 2). Analyses over the 0-80 and 80-150 ms time windows showed that despite a significant difference between the auditory and audiovisual speech modality over 0-80 ms time window ($F(1,15)$=9.65, p<0.01), the Modality x Order interaction was not significant over either the 0-80 ms or the 80-150 ms time window (other significant results in these time windows are presented in Table 2).

[Insert Table 2 about here]

During the early time window of lexical processing (170-280 ms), the interaction between Lateralization and Modality was significant ($F(2,30)=4.05$, $p<0.05$): the ERPs in the AV modality were less positive over the middle electrodes than over the right electrodes ($t(15)=-2.92$, $p<0.05$). The ANOVA yielded a significant interaction between Lateralization and Order ($F(2,30)=4.09$, $p<0.05$): the ERPs to the second repetition of a critical word were less positive over the left electrodes than over the middle and right electrodes (left versus middle electrodes: $t(15)=-2.65$, $p<0.05$; left versus right electrodes: $t(15)=-2.57$, $p<0.05$). There was also a significant interaction between Order and Position ($F(2,30)=6.78$, $p<0.01$): the ERPs to the second repetition of a critical word were less positive over the anterior electrodes (anterior versus central electrodes: $t(15)=-2.62$, $p<0.05$; anterior versus posterior electrodes : $t(15)=-2.22$, $p<0.05$).

Critically, the ANOVA yielded a significant interaction between Modality and Order ($F(1,15)=7.75$, $p<0.05$). In the auditory-only modality, ERPs to the second presentation of a critical word were less positive than those to the first presentation of that word ($t(15)=2.44$, $p<0.05$). Analysis performed separately on the first and second presentation of words (i.e. auditory-only versus audiovisual for each Order) showed that ERPs to the first presentation of a critical word were more positive in the auditory-only modality than in the audiovisual one ($t(15)=3.30$, $p<0.01$). Note that ERPs to the second presentation of a critical word in the audiovisual modality did not differ from those to the first presentation of that word ($t(15)=-1.15$, $p=0.27$). Moreover, no difference in modality was observed for the second presentation of the critical words ($t(15)=-0.45$, $p=0.66$).

During the late time window, the ANOVA yielded a main effect of Modality ($F(1,15)=5.04$, $p<0.05$): ERPs in the auditory-only modality were less negative than those in the audiovisual modality. There was a significant Lateralization x Position interaction ($F(4,60)=3.05$, $p<0.05$): ERPs over Cz were more negative than over its left and right homologous electrodes (left vs. middle: $t(1,15)=2.78$, $p<0.05$; right vs. middle: $t(1,15)=2.73$, $p<0.05$). The interaction between

Order and Position was also significant ($F(3,30)=17.91$, $p<0.001$). Further analysis showed that the repetition effect (first vs. second presentation) was significant only over the posterior electrodes ($t(15)=-2.84$, $p<0.05$).

[Insert figure 1 about here]

3.2 Words in sentence context

Figure 2 shows the time-courses of ERPs to the first presentation of a novel (unfamiliarized) word in sentence context and its second presentation in another sentence context. The ANOVA over the 400-700 ms time window showed a main effect of Order ($F(1,15)=10.71$, $p<0.01$). ERPs to the second presentation of an unfamiliarized word were less negative than those to the first presentation of that word. Moreover, the interaction between Order and Position was significant ($F(2,30)=5.54$, $p<0.001$). The repetition effect (first versus second presentation) was significant over the central and posterior electrodes (respectively, $t(15)=-3.18$, $p<0.01$ and $t(15)=-4.23$, $p<0.001$). We also analyzed the repetition effect by comparing the first presentation of an unfamiliarized word and the first presentation of a familiarized word in sentence context (ninth overall presentation). The only significant effect was the main effect of Order ($F(1,15) = 11.83$, $p<0.001$). ERPs to the first presentation of the familiarized word were less negative than those to the first presentation of the unfamiliarized word.

[Insert figure 2 about here]

4. Discussion

In investigating whether speech visual cues contribute to lexical processing, most of the behavioral and electrophysiological studies to date have sought a contrast between words and pseudo/non-words or low- and high-frequency words. Our approach in this study was different as we used the

word repetition effect. As discussed in the introduction, the ERP word repetition effect reflects access to the word representation in long-term memory and makes it possible to distinguish early (i.e. activation of a set of candidates) and late (i.e. selection of the best candidate/access to its lexico-semantic representation) stages of word processing. Our goal was to investigate whether there is an interaction between the word repetition effect and the speech modality in either of these stages. Such an interaction would suggest that audiovisual speech may influence access to lexical representations during on-line speech processing. Our main findings are discussed below.

4.1 Speech modality and word repetition effect

As expected, a positive shift over the N400 was observed after the second repetition of words in isolation and in sentence context. This effect did not differ between auditory-only and audiovisual modalities. The N400 is commonly related to lexico-semantic processing (For a review, see Kutas & Federmeier, 2011). Importantly, it seems to be independent of the input modality since it is elicited by auditory words, written words and signed language (Kutas, Neville & Holcomb, 1987). The repetition effect on the N400 may thus reflect late stages of access to lexical and semantic representations stored in long-term memory. These stages of processing would not have been affected by the speech modality in our participants as the audiovisual word repetition effect did not differ from the auditory-only word repetition effect. This was also true for novel words presented and repeated in sentence context. In fact, audiovisual speech is known to be helpful in finding word boundaries in continuous speech (Mitchel & Weiss, 2014). However, contrary to our hypothesis, we did not observe any difference between auditory-only and audiovisual modalities in sentence context.

In a cross-modal repetition paradigm, Kaganovich, Schumaker and Rowland (2016) observed that the N400 is sensitive to the match between auditory and visual speech cues. In their study, participants received auditory-only words followed by visual-only words that either matched the

initial auditory-only words or not. Seeing unmatched articulatory gestures elicited a larger N400. Their cross-modal repetition effect on the N400 therefore reflects the mismatch between expected and unexpected articulatory gestures at the lexico-semantic level of processing. Contrary to their finding, we did not observe any effect of visual modality on the N400, perhaps owing to differences in the experimental procedure. Kaganovich et al. (2016) asked their participants to determine whether the visual-only word matched the word they had heard just before. This is different from the implicit processing of on-going speech without any mismatch as performed by our participants. Altogether, these data suggest that the access to lexico-semantic representations involved in the word repetition effect are performed independently of speech modality.

We observed an early repetition effect only with the auditory-only modality. This repetition effect over the P200 time window has been linked to early stages of lexical processing, as words and not pseudo-words elicit this effect (Almeida & Poeppel, 2013). However, ERP responses over both the P200 and the N400 time windows are frequently observed during word/sentence processing, these components reflecting different underlying processes. Dambacher, Kliegl, Hofmann and Jacobs (2006) observed that the amplitude of the P200 during sentence processing was affected by the lexical frequency of words but not by contextual predictability. On the contrary, the amplitude of the N400 was affected by contextual predictability, especially for low-frequency words. In line with the results of Dambacher et al. (2006), we believe that the early repetition effect reflects exclusively the first stages of lexical processing, i.e. the contact with the lexicon. Snijders et al. (2007) using auditory-only stimuli observed that a repetition effect started at 240 ms both for foreign and native listeners of Dutch. This is consistent with our interpretation that the early ERP repetition effect does not reflect late stages of lexical and semantic access, as the foreign listeners in their study had very little knowledge of Dutch. Altogether, these findings suggest that the contact with the lexicon is modulated, at least to some extent, by the speech modality.

The difference between auditory-only and audiovisual word repetition effects over the early time window could be explained by two hypotheses. First, only general memory retrieval processes might have been influenced by audiovisual speech. For example, the retrieval of a word might be enhanced by using visual cues in addition to auditory-only information during repeated presentation of that word. Second, in addition to general memory retrieval processes, the activation level of a lexical representation during the initial presentation of an audiovisual word could differ from that of an auditory-only word. If a lexical representation had been more strongly activated during its prior presentation in the audiovisual modality thanks to visual cues, it would have inhibited other lexical candidates more strongly (McClelland & Elman, 1986). Thus, during subsequent presentations, it would be temporarily easier to re-activate that lexical representation as its activation level would be high and other candidates would be inhibited. The beneficial effect of visual primes reported in behavioral studies using prime-target pairs (e.g. Kim and al., 2004; Fort et al., 2013) could be understood in this framework, i.e. seeing visual gestures of a word (prime) might activate the lexical representation of that word. This would facilitate the activation of that representation when the word is presented as target.

We believe that our results are consistent with the second hypothesis. In fact, our post-hoc analyses showed that the ERPs to the first presentation of audiovisual words were smaller than those to the first presentation of auditory-only words, while the ERPs to their second presentations were similar. Thus, audiovisual words differed from auditory-only words during their first presentations. In future, a cross-modal repetition paradigm (i.e. presentation of an auditory-only word following an audiovisual word) would be helpful to ensure that the interaction between speech modality and repetition effect is due to the modality of first presentations.

4.2 How does the speech modality modulate contact with the lexicon?

As described in the introduction, there is a growing body of evidence suggesting that visual speech plays a role in lexical processing (e.g. Brancazio, 2014; Fort et al. 2010, Fort et al., 2013). However, the influence of visual speech cues on lexical processing remains debated (e.g. Dekle et al., 1992; Sams et al., 1998; Windmann, 2004; Barutchu et al., 2008). Importantly, in an ERP study, Baart and Samuel (2015) did not observe any interaction between the effect of visual speech and lexicality. In that experiment, participants received 3-syllabic words and pseudo-words in auditory-only, visual-only and audiovisual modalities. The critical syllable was the last syllable which indicated whether a sequence was a word or pseudo-word. Although both the modality and the lexical status of the stimuli affected the ERPs in the same time-window, the authors did not observe any interaction between these two effects. They suggested that lip-read and lexical contexts both influence auditory speech processing but act differently on speech processing such that the former could play a role in the perceptual analysis of speech and the latter in the linguistic encoding of speech.

The findings of Ostrand, Blumstein, Ferreira, & Morgan (2016) may shed light on the debate regarding the role of speech modality in lexical processing. In their study, participants were presented with McGurk stimuli as prime (e.g. auditory "beef" dubbed onto visual "deef" which led to the perception of "deef*"*) and performed a lexical decision task on auditory-only targets. Primes and targets were semantically related or unrelated. The results showed a semantic priming effect of the words presented in the auditory track of the McGurk stimuli although the participants perceived a sequence, i.e. McGrurk percept, which was different from the auditory word. The access to word representations was thus driven only by the auditory information and not by the audiovisual integrated percept, i.e. McGurk percept. This suggests that mechanisms underlying the access to word representations and audiovisual speech integration are independent, since lexical access to words occurred earlier than audiovisual integration. However, a different trend was observed when the auditory track was not a real word (e.g. auditory non-word "bamp" dubbed onto visual real word

"damp"). In this condition, the authors observed a priming effect of McGurk percepts. This showed that when the auditory-only track was not a real word, integrated audiovisual percepts drove lexical access. The effect of speech visual modality on lexical access is thus complex and depends on the lexical status of auditory stimuli.

What is the role of speech modality during lexical processing? As shown by post-hoc analyses, the effect of the speech modality on lexical processing is related to the fact that the ERPs were already reduced in the first presentation of audiovisual words compared to the first presentation of auditory-only words. This could be due to perceptual analysis which in turn influences lexical processing. Numerous electrophysiological studies have shown that brain responses to auditory speech are modulated by speech visual cues (e.g., Colin et al., 2002; Saint-Amour, Sanctis, Molholm, Ritter & Foxe, 2007). For example, van Wassenhove, Grant and Poeppel (2005) showed that visual cues reduced the amplitude and sped up the latency of brain responses to auditory syllables. Similar results were observed in sentence context (Authors, xxxx). The effect of visual cues on ERPs was frequently observed on a negative peak elicited at about 100 ms (N100) and a positive peak at about 200 ms (P200) post-stimulus over the centro-parietal electrodes (for a review, see Baart, 2016). This property of brain response to audiovisual speech (i.e. reduced ERPs to audiovisual speech) might explain why the ERPs to the second presentation of a critical word in auditory-only modality was similar to the first presentation of the critical word in the audiovisual modality.

Considering the findings described above and our current results, it seems probable that visual speech affects lexical processing through perceptual processing. In this view, visual cues are integrated with auditory-only information early on during speech perception and later mechanisms underlying lexical processing are driven by integrated audiovisual percepts. Our results may thus reflect an enhancement of phoneme identification by visual cues, which in turn facilitates lexical access in a bottom-up manner without any direct effect of visual cues in constraining lexical access.

This issue about the impact of the time-course of auditory and visual integration on the repetition effect is beyond the scope of this study and requires future investigation. Interestingly, the time-course of speech auditory and visual integration seems to depend on the lexical properties of the stimuli (Ostrand et al., 2006). In future studies, it would be helpful to investigate whether the interaction between the speech modality and the early repetition effect is modulated by properties such as phonological neighborhood density. If neighborhood density influences the modality-specific repetition effect, it would support the idea that this effect is not purely perceptual.

In this study, our goal was to investigate the time-course of lexical processing in the audiovisual modality without drawing the participants' attention to the lexical status of the stimuli. For this reason, we did not ask them to perform any explicit task during word and sentence presentations. Further studies on lexical processing could use behavioral measurements to examine whether the interaction between the early word repetition effect and the speech modality reflects a beneficial role of audiovisual speech in speech processing (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017). It is also important to note that, before making a comparison between auditory-only and audiovisual modalities, the ERPs in the visual-only modality are usually subtracted from those in the audiovisual modality. In our study, participants did not receive any visual-only blocks and we could not compute AV-V ERPs. However, AV-V subtraction does not seem to be necessary to observe the known effects of speech modality on N100/P200 peaks (Ganesh, Berthommier, Vilain, Sato, & Schwartz, 2014; see also Baart, 2016). While it is interesting to examine the repetition effect in the visual-only condition and use AV-V subtraction in statistical analyses, we believe that the lack of visual-only condition does not influence the interpretations of our current results.

5. Conclusion

In summary, we used the word repetition effect to study the influence of seeing speakers' articulatory gestures on auditory word processing. As expected, we identified a two-stage brain response related to word processing and access to word representations in long-term memory. Our findings suggest that the late stages of access to lexical and semantic representations are accomplished independently of speech modality. Crucially, audiovisual speech cues influence the first stages of contact with the lexicon.

Acknowledgment

Hidden for review purposes

References

Authors, xxxx.

Almeida, D., & Poeppel, D. (2013). Word-specific repetition effects revealed by MEG and the implications for lexical access. *Brain and language*, *127*(3), 497-509.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339-355.

Baart, M. (2016). Quantifying lip□read□induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology*, *53*(9), 1295-1306.

Baart, M., & Samuel, A. G. (2015). Turning a blind eye to the lexicon: ERPs show no cross-talk between lip-read and lexical context during speech sound processing. *Journal of Memory and Language*, *85*, 42-59.

Barutchu, A., Crewther, S. G., Kiely, P., Murphy, M. J., & Crewther, D. P. (2008). When /b/ill with /g/ill becomes /d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Psychology, 20*, 1–11.

Benoit, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech, Language, and Hearing Research*, *37*(5), 1195-1203.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5(9/10): 341-345.*

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 445.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1493), 1001-1010.

Cheng, X., Schafer, G., & Riddell, P. M. (2014). Immediate auditory repetition of words and nonwords: an ERP study of lexical and sublexical processing. *PloS one*, *9*(3), e91988.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*(4), 495-506.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and language*, *43*(1), 1-18.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain research*, *1084*(1), 89-103.

Davis, C., & Kim, J. (2001). Repeating and remembering foreign language words: Implications for language teaching systems. *Artificial Intelligence Review*, *16*(1), 37-47.

Dekle, D. J., Fowler, C. A., & Funnell, M. G. (1992). Audiovisual integration in perception of real words. *Attention, Perception, & Psychophysics*, *51*(4), 355-362.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, *2*(2), 337-339.

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*(8), 1207-1223.

Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2010). The word superiority effect in audiovisual speech perception. *Speech Communication*, *52*(6), 525-532.

Ganesh, A. C., Berthommier, F., Vilain, C., Sato, M., & Schwartz, J.-L. (2014). A possible neurophysiological correlate of audiovisual binding and unbinding in speech perception. Frontiers in Psychology, 5, 1340.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, *38*(11), 1518-1530.

Kaganovich, N., Schumaker, J., & Rowland, C. (2016). Matching heard and seen speech: an ERP study of audiovisual word recognition. *Brain and language*, *157*, 14-24.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480-490.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review in Psychology, 62, 621–647.

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalography and Clinical Neurophysiology Supplement*, *39*, 325-330.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology*, *10*(1), 29-63.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.

Mitchel, A. D., & Weiss, D. J. (2014). Visual speech segmentation: using facial cues to locate word boundaries in continuous speech. *Language, Cognition and Neuroscience*, *29*(7), 771-780.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375-425.

Morton, J. (1969). Interaction of information in word recognition. *Psychological review*, *76*(2), 165-178.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234.

New, B., Pallier, C., Brysbaert, M., Ferrand, L. (2004) Lexique 2: A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, 36 (3), 516-524.

Ostrand, R., Blumstein, S. E., Ferreira, V. S., & Morgan, J. L. (2016). What you see isn't always what you get: Auditory word signals trump consciously perceived words in lexical access. *Cognition*, *151*, 96-107.

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. Cortex, 68, 169-181.

Rugg, M. D. (1985). The Effects of Semantic Priming and Word Repetition on Event□Related Potentials. *Psychophysiology*, *22*(6), 642-647.

Rugg, M. D. (1987). Dissociation of semantic priming, word and non-word repetition effects by event-related potentials. *The quarterly journal of Experimental Psychology*, *39*(1), 123-148.

Rugg, M. D., Doyle, M. C., & Melan, C. (1993). An event-related potential study of the effects of within-and across-modality word repetition. *Language and Cognitive Processes*, *8*(4), 357-377.

Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö R. (1998). McGurk effect in Finnish

syllables, isolated words, and words in sentences: effects of word meaning and sentence context. *Speech Communication, 26*, 75-87.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*(3), 587-597.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, *12*(11), 411-417.

Snijders, T. M., Kooijman, V., Cutler, A., & Hagoort, P. (2007). Neurophysiological evidence of delayed segmentation in a foreign language. *Brain research*, *1178*, 106-113.

Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212–215.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 394–417.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of cognitive neuroscience*, *3*(2), 131-150.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *PNAS, 102*(4), 1181–1186.

Windmann S. (2004). Effects of sentence context and expectation on the McGurk illusion. *Journal of Memory and Language, 50*(2), 212-230.

Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, *160*, 103-109.

Table 1

Ponton-Fusain pair (presented in A or AV modality)	*Pompon-Furet* matched pair (presented in AV or A modality)
Isolated word presentation *Ponton Ponton Ponton Ponton* *Ponton Ponton Ponton Ponton* **Sentence presentation** 1. *Tu apprendras la technique du **fusain**.* 2. *Ils ont construit un ponton sur la plage.* 3. *L'artiste utilise un **fusain** pour dessiner.* 4. *Ce ponton va s'écrouler.* 5. *Ce bateau a détruit le ponton de la plage.* 6. *Ce **fusain** va s'user.* 7. *J'ai taillé le **fusain** de ton père.* 8. *Tu répareras les planches du ponton.*	**Isolated word presentation** *Pompon Pompon Pompon Pompon* *Pompon Pompon Pompon Pompon* **Sentence Presentation** 1. *Elle a cousu un pompon sur son bonnet.* 2. *Il a recueilli le **furet** de la forêt.* 3. *Ce pompon va se décrocher.* 4. *Tu couperas le fil du pompon.* 5. *Il a aperçu un **furet** dans la prairie.* 6. *Ce **furet** va se sauver.* 7. *Elle a choisi le pompon de son écharpe.* 8. *Je nettoierai la cage du **furet**.*

Table *1*: Example of two matched blocks. Each block began with the presentation eight times of an isolated word (=familiarized word, here: *Ponton* and *Pompon*). Then, four sentences containing the familiarized word and four other sentences containing a novel word (=unfamiliarized word, here: *Fusain* and *Furet*) were delivered in a randomized order. Each participant received one of the matched block in auditory-only (A) modality and the other one in audiovisual (AV) modality.

Table 2

Context	Time window	Effect	*df*	*F*	*P*
Isolation	0-80 ms	Lateralization	2,30	5.24	<0.05
		Modality	1,15	9.65	<0.01
	80-150 ms	Lateralization	2,30	8.21	P<0.01
		Position	2,30	11.91	P<0.001
		Lateralization x Position	4,60	3.70	P<0.01
	170-280 ms	Lateralization x Modality	2,30	4.05	P<0.05
		Lateralization x Order	2,30	4.09	P<0.05
		Modality x Order	1,15	7.75	P<0.05
		Order x Position	2,30	6.78	P<0.01
	400-700 ms	Modality	1,15	5.04	P<0.05
		Lateralization x Position	4,60	3.05	P<0.05
		Order x Position	2,30	17.91	P<0.001
Sentence: first vs. second unfamiliarized word	400-700 ms	Order	1,15	10.71	P<0.01
		Order x Position	2,30	5.54	P<0.001
Sentence: first unfamiliarized word vs. first familiarized word in sentence context (ninth overall presentation)	400-700 ms	Order	1,15	11.83	P<0.001

Table *2*: Results of Lateralization x Modality x Order x Position ANOVAs for words in isolation and in sentences. Only significant effects are reported. For words in sentences, the first presentation of the unfamiliarized word was compared a) with the second presentation of the word, b) with the first presentation of the familiarized word, i.e. ninth presentation overall. The Modality x Order interaction was significant over 170-280 ms time window. No other significant interaction involving Modality and Order was found over 0-80 ms, 80-150 ms or 400-700 ms time windows.

Figure caption

Figure 1: Grand-average ERPs for words in isolation in auditory-only (left) and audiovisual (right) modalities. Solid versus dotted lines indicate ERPs to first versus second presentation of words. Time 0 indicates auditory onset of words. Analyses were performed over two time windows of interest (170-280 ms and 400-700 ms), marked by gray rectangles, where word repetition effect was frequently reported in previous studies.

Figure 2: Grand-average ERPs for words in sentence context in auditory-only (left) and audiovisual (right) modalities. Solid versus dotted lines indicate ERPs to first versus second presentation of words. Time 0 indicates auditory onset of words. Analyses were performed over 400-700 ms time window, marked by gray rectangles, where word repetition effect was frequently reported in previous studies.

Appendix: matched pairs of words used in experiment

Block	A (AV) paris		AV (A) paris	
1	Bassin	Panda	Barreau	Pantin
2	Baudet	Siphon	Bottin	Sirop
3	Béret	Ciment	Béton	Ciseau
4	Bijou	Faucon	Bisou	Forain
5	Bison	Panneau	Bidon	Parrain
6	Boulet	Messie	Bouquet	Mérou
7	Burin	Sabbat	Butin	Sapin
8	Félin	Bouchon	Ferry	Boucan
9	Futon	Pâton	Fuseau	Pavot
10	Maton	Bambou	Matou	Bandeau
11	Méfait	Bilan	Mégot	Bidet
12	Mulet	Chameau	Mutin	Chalut
13	Museau	Basson	Mulot	Basset
14	Piment	Bouquin	Pichet	Boudin
15	Pipeau	Manga	Pinot	Manchot
16	Pompon	Furet	Ponton	Fusain
17	Poteau	Muguet	Poney	Muret
18	Satin	Pignon	Sabot	Python

19	Saumon	Poulain	Sauna	Poussin
20	Sureau	Bolet	Sumo	Bonnet