Journal of Neurolinguistics 52 (2019) 100849

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

Processing of non-contrastive subphonemic features in French homophonous utterances: An MMN study

Noelia Do Carmo-Blanco[a,*], Michel Hoen[b], Stéphane Pota[c], Elsa Spinelli[c], Fanny Meunier[a,*]

[a] *Université Côte d'Azur, CNRS, BCL, Nice, France*
[b] *Oticon Medical, Vallauris, France*
[c] *LPNC, Université Grenoble Alpes, CNRS, UMR5105, Grenoble, France*

ARTICLE INFO

Keywords:
Acoustic cues
Mismatch negativity (MMN)
Intra-speaker variability
Speech segmentation
Homophones
Fundamental frequency (f0)

ABSTRACT

Native listeners process and understand homophones, such as *la locution* 'the phrase' vs. *l'allocution* 'the speech', both [lalɔkysjɔ̃], without much semantical ambiguity in connected speech. Yet, behavioral experiments show that disambiguation is partial under intra-speaker variability without semantical context. To investigate electrophysiological correlates of perception of non-contrastive subphonemic features in French homophonous sequences, we examined the event-related potential Mismatch Negativity (MMN) using a multitoken stimuli oddball paradigm. Stimuli were taken from multiple natural productions of nominal homophonous utterances. In the first experiment, we used the first syllables, while in the second experiment, the whole utterances.

The homophonous sequence elicited an MMN response in both experiments. This suggests that non-contrastive acoustic features that differentiate homophones, such as pitch and duration, are robust enough despite intra-speaker variability to allow listeners to automatically extract regularities associated with each utterance. This ability of the perception system might contribute to correct segmentation and comprehension of ambiguous utterances.

1. Introduction

Spoken word recognition consists of mapping a complex signal made of numerous combinations of acoustic features onto lexical representations stored in memory. Given the high complexity and variability of natural speech, only certain acoustic features are used for this mapping. These features are language-specific: during their first year of life, infants become indeed attuned to the phonetic distinctions that are used phonemically in their native language (Kuhl et al., 2006; Werker & Tees, 1984), i.e. contrastive cues. During development, the infant's ability to discriminate foreign speech sounds decreases, while at the same time the ability to discriminate native speech sounds improves (Cheour et al., 1998; Kuhl et al., 2006; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005). At adult age, the system has the ability to focus on pertinent cues of the native language and to lose sensitivity to non-contrastive cues. However, some non-contrastive cues could be pertinent for speech processing in a given language. For example, some non-semantically nor phonemically contrastive acoustic features in French, such as pitch or duration, could be relevant for segmenting connected speech. In this paper, we examined electrophysiological correlates of the processing of non-contrastive cues associated with word boundaries in

[*] Corresponding authors. University of Nice Sophia Antipolis, Laboratoire Bases, Corpus, Langage UMR CNRS 7320, Campus Saint Jean d'Angély - SJA3/MSHS-SE, 24 Avenue des Diables Bleus, 06357, Nice Cedex 4, France.

E-mail addresses: ndocarmo@unice.fr (N. Do Carmo-Blanco), fanny.meunier@unice.fr (F. Meunier).

https://doi.org/10.1016/j.jneuroling.2019.05.001
Received 12 December 2018; Received in revised form 18 March 2019; Accepted 14 May 2019
Available online 25 June 2019

French in an adult population.

In contrast to written language, where words are separated by blank spaces, there are no clear word boundaries in spoken language. Speech continuity is thus one major challenge for spoken word recognition. However, an abundance of converging data coming from analyses of production show that despite the apparent continuity, there are subtle acoustic cues, albeit very variable from one production to the other, that are correlated with word-initial boundaries (Ito & Strange, 2009; Turk & Shattuck-Hufnagel, 2014). This is the case for example for durational differences, as in *one spade* vs. *once paid* where the phonetic realization of [s] is shorter in the latter (Shatzman & McQueen, 2006). Because of speech continuity, listeners are often confronted with transient ambiguities. In some sequences, these ambiguities can be total, as for homophonic sequences (i.e. phonemically identical sequences) such as *la fiche* 'the sheet' vs. *l'affiche* 'the poster' both [lafiʃ] in French. Despite these ambiguities, in everyday life, listeners easily segment words in the speech chain and they are rarely misled in the presence of sentential context. Yet, correct segmentation of such two-word utterances drops to 75% when presented in isolation (Schaegis, Spinelli, & Welby, 2005; Spinelli, Welby & Shaegis, 2007; see also, Quené, 1992). Importantly, when allowing for natural speech variability (intra- and inter-speakers), correct discrimination drops to 66%. Hence, while being significantly above chance, complete disambiguation of such homophonic sequences is not systematically achieved. This demonstrates that despite the presence of acoustic cues correlated to words boundaries in the speech signal, ambiguity increases when natural speech variability is taken into account. In this paper, we took into account intra-speaker variability and we used multiple productions of homophonous French utterances to study electrophysiological correlates of the processing of non-contrastive features.

In English, there are numerous allophonic cues associated with word boundaries that listeners can exploit to achieve correct segmentation. For example, Altenberg (2005) and Ito and Strange (2009), by comparing sequences such as *chief school* and *chief's cool,* showed a shorter voice onset times (VOTs) of voiceless stops and a shorter duration of [s] in *chief's cool* (see also Turk & Shattuck-Hufnagel, 2000). Nevertheless, it has been shown that non-native speakers do not exploit such allophonic cues the same way and that segmentation strategies do not depend only on the characteristics of the acoustic signal but also on the native language of the listener. French (Shoemaker, 2014b), Spanish (Altenberg, 2005) and Japanese (Ito & Strange, 2009) speakers underperform when segmentation relies on VOTs. On the same note and of particular interest for us, Tyler and Cutler (2009) compared the use of suprasegmental cues (vowel length and pitch) across speakers of English, French, and Dutch using an artificial-language learning approach. It is noteworthy that the French language is characterized by right-edge (iambic) while English and Dutch by more left-edge (trochaic) boundary phenomena. Apart from a primary stress to the last full syllable of a phrase-final word, French has no systematic stress, although a rise in fundamental frequency (f0) at the beginning of the first content-word syllable can often be found (Spinelli, Welby, & Schaegis, 2007; Welby, 2007), which is known as secondary stress (Di Cristo, 1998, pp. 195–218). Yet, most models in the frame of the autosegmental-metrical approach consider it as a loose boundary marker. Primary and secondary stress are also known as initial (IA) and final accents (FA) (Astésano, Bard, & Turk, 2007). The role of IA is not fully understood in French. Although there is no agreement on its use in prosodic structure, it has been shown that it can contribute to segment lexical units (Welby, 2007). This has led to question the claim that French listeners are "deaf" to accentuation and to claim that IA might have a more important role than frequently described in prosodic models. Yet, stress is not contrastive in French and pitch accents variations in words do not signal differences in meaning. When it comes to both English and Dutch, most content words start with a stressed syllable with acoustic reflections on pitch movement and duration. Stress is contrastive in both languages. Results of the artificial-language learning experiments ran by Tyler and Cutler (2009) showed a different use of acoustic cues depending on the native language of the listeners. French speakers exploited vowel length and pitch movement only when they appeared in word-final position but not in word-initial position. In contrast, English speakers used pitch movement only in word-initial position, whereas Dutch speakers were sensitive to pitch in both initial and final positions. Moreover, Dutch and English listeners exploited durational cues to a greater extent than French listeners, who only benefited from vowel lengthening when it was a right-edge cue but not left. Sensitivity to pitch (i.e. direction, slope and height) and duration seems to be largely determined by the language background of the listener (Chandrasekaran, Krishnan, & Gandour, 2007; Jongman, Qin, Zhang, & Sereno, 2017; Zora, Schwarz, & Heldner, 2015). Overall, these results highlighted the importance of the characteristics of the speaker's native language in the use of certain acoustic cues.

A study on French by Spinelli et al. (2007) examined the acoustic cues linked to elision in French by using homophonic pairs such as *l'affiche* and *la fiche*. Elision corresponds to the case in which the vowel of clitics, here the definite article *la,* is elided before vowel-initial words thus giving *l'affiche*. Their analyses revealed multiple differences in formant and f0 values, as well as in segmental and syllabic durations between the two [la] (i.e. l#a and la#). However, correlational analyses revealed only a few significant effects. Using percentages of identification, they showed that the segmentation into *la#fiche* by participants was linked to higher F_2 values of the first vowel [a] and marginally linked to longer duration of [a]. On the other side, segmentation into *l#affiche* was correlated with higher f0 values of [a] (see also Welby, 2003). Yet, if a conservative correction such as Bonferroni was applied, only the latter intonational effect remained significant. Furthermore, no effect was found when using reaction time to reflect online segmentation. Interestingly, duration and f0 values, two cues often studied and contrastive in many languages, are not contrastive in French, which leaves open the question of whether French speakers are able to automatically extract this information from natural speech, in particular in a context of intra-speaker variability.

Behavioral studies on specific non-contrastive cues in French have not allowed drawing strong conclusions on the processing of non-contrastive acoustic features. For example, a recent study tested to what extent duration is used as an online cue for comprehension in French (Shoemaker, 2014a). In the experiment, consonant duration was manipulated while all other factors were held constant. Stimuli were ambiguous phrases (i.e., *un air* or *un nerf*) in which the pivotal consonants (i.e. in the example [n]) were instrumentally shortened or lengthened (overall mean $_{\text{natural productions}}$ = 92 ms, mean $_{\text{shortened}}$ = 63 ms and mean $_{\text{lengthened}}$ = 125 ms). In an AX discrimination task, listeners were sensitive to durational differences only between extreme tokens (mean duration difference of 62 ms), in other words, only when “the difference was greatly exaggerated with respect to what would be a normal distribution of allophonic variation”. Although French production data have revealed significant differences in duration between liaison consonants (e.g., [n] in *un air*) and initial consonants (e.g., [n] in *un nerf*), the differences are rather small when compared to those in non-natural speech experiments. According to Spinelli, McQueen, and Cutler (2003), liaison consonants were on average 8 ms shorter than initial consonants (15 ms in Gaskell, Spinelli, & Meunier, 2002) and pivotal consonants were on average 10 ms longer than word-final consonants. The discrepancy between the range of natural duration differences and the ones tested experimentally is such that it is still unclear whether French speakers make use of durational cues during natural speech processing.

In the present study, we focused on the encoding of such non-contrastive features by French native speakers in the case of variable natural productions. The exploitation of these cues in natural speech remains understudied, and particularly with online tasks that do not involve discrimination between utterances. Since the majority of behavioral studies have used word-form identification tasks as opposed to word comprehension and meaning, the role of acoustic cues in segmentation might be largely overestimated, as suggested by Mattys and Melhorn (2007). These researchers asked their participants to choose between near homophonous utterances in context. Results showed that when the context information was incongruent, segmentation was led by the semantical context. This indicates a strong role of higher order information in segmentation. However, in one of the experiments, participants were instructed to direct their attention to the fine acoustic details of the utterances. This caused participants to give more credence to acoustic than to lexical information. In this line, one might ask how relevant such acoustic cues are when processing speech without having a specific task and without focused attention. One might also ask if stimulus complexity affects such processing. In our experiments, we used two types of stimuli: syllables in the first and nominal sequences in the second. This allowed investigating the effect of stimulus complexity on the processing of fine-grained acoustic features without changing task or attentional focus.

The question arisen in this paper is whether non-contrastive features prove relevant to segmentation in French. We examined whether these fine acoustic variations, related to word segmentation, are sufficiently robust for the speech perception system to be processed without focused attention within a multiple token context. To address this question, we used the event-related potential (ERP) Mismatch negativity (MMN; Näätänen, Paavilainen, Rinne, & Alho, 2007; Näätänen & Alho, 1995), which allows us to study speech sounds perception without asking participants to focus on stimuli and without a behavioral task related to the stimuli. The auditory MMN is elicited by unexpected changes in some aspects of a regular continuous auditory stream, such as pitch, intensity or duration (Näätänen et al., 2007). This component is observed in the oddball paradigm, in which one rare sound (deviant) occurs in a series of frequent stimuli (standards). This fronto-central negative wave, which peaks between 100 and 200 ms after the deviance onset, reflects the formation of sensory memory traces from statistical regularities in the input signal. In the case of spoken words, the memory trace for standard stimuli contains a sound representation close to that stored in long-term memory (Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993), i.e., close to the mental lexicon's representation. The MMN is also sensitive to long-term memory traces resulting from the listener's experience with spoken language (Näätänen et al., 2001, 1997; Shtyrov, Kujala, Palva, Ilmoniemi, & Näätänen, 2000).

The MMN is extremely useful to study the processing of acoustic regularities within variable speech productions, such as with multiple tokens of the same linguistic unit. For instance, this component was elicited when deviant vowel phonemes came from productions by different speakers (for example three productions of [a], i.e./a_1/,/a_2/and/a_3/) and the standard vowel was another phoneme ([u] for example; Shestakova et al., 2002; Eulitz & Lahiri, 2004), thus suggesting that the invariant relevant information for vowel recognition is extracted and encoded. It is to note that the MMN amplitude does not only reflect physical acoustical differences (i.e. Euclidean distances in the formant space). Deguchi et al. (2010) showed that the MMN elicited in the context of vowels produced by different speakers was greater in active than in passive listening condition. Moreover, the MMN was sensitive to the processing of different vowels as a function of the acoustic distance only in the active condition. A question left open is whether non-phonemically contrastive acoustic features can be processed without focused attention. The MMN is also sensitive to stimulus familiarity, thus being larger for familiar than for non-familiar speech sounds. Similarly, the processing of an acoustic vowel contrast elicited a greater MMN in native than in non-native speakers (Peltola, Kujala, Tuomainen, & Ek, 2003). Dehaene-Lambertz, Dupoux, and Gout (2000) showed that while French phonemes elicited an MMN in native speakers (French), the MNN was reduced or absent in non-native speakers (Japanese) who are unable to discriminate these phonemes. Finnish speakers seem to be more sensitive to vowel duration than German speakers (Kirmse et al., 2008). Overall, the neural system is tuned to the native language and shaped to be efficient in processing phonemes that belong to the native language. The relevance of the acoustical variation in a given language is, therefore, a key factor.

Brunellière, Dufour, and Nguyen (2011) exposed Southern French speakers and standard French speakers to the words *épée* “sword” and *épais* “thick” against *épi* [epi] as the standard stimulus in an oddball paradigm. In standard French, *épée* and *épais* have different phonological realizations (i.e. [epe] vs. [epè]). However, in southern French, they are produced as homophones (both [epè]). Despite the divergent phonemic classification, the elicited MMNs were similar in both groups of participants. This suggests that the auditory perception system in southern French speakers also differentiates between both productions. Although this contrast is not part of their speech production, southern speakers have been exposed to both phonemes from an early age through TV, for example, or national media. On the other hand, differences in topography were found between conditions in standard French speakers but not in southern ones, which was explained by differences in semantical access. If this is the case, in our experiments and in particular in our second experiment in which we used nominal sequences, differences in topography might be found between homophonous utterances.

Although some EEG studies have looked at specific acoustical cues, very few studies have been conducted in French language. Regarding pitch, Zora et al. (2015) studied the processing of stress contrastive homonyms in English (*upsét-úpset*). EEG results revealed the encoding of intensity and f0 prosodic features. In Hungarian, where there is a systematic and contrastive stress on the first syllable of the word, it has been shown that f0 and consonant duration features elicit a larger MMN than other features (Honbolygó, Kolozsvári, & Csépe, 2017). When it comes to phoneme length and speech segmentation, Menning, Imaizumi, Zwitserlood, and Pantev (2002) investigated the perception of the same phoneme as a function of its duration ([o] [o:]). In certain languages, such as Japanese, this feature is contrastive for semantic processing but also crucial to divide words into segments. In the experiment, non-native participants were trained to discriminate phonemes differing in small durational differences. The native group showed better performance as well as higher MMNm amplitude (Näätänen, 2001). Yet, both parameters increased in the non-native group as a result of training. It is important to note that in the language of the non-native group, German in this case, vowel duration can also contribute to semantic processing. In French language, Aguilera, El Yagoubi, Espesser, and Astésano (2014) used the MMN component to study whether listeners could discriminate IA. To that aim, the same words presenting IA were resynthesized without IA. They found that French listeners could perceive IA. Yet, those resynthesized token are unfamiliar to the speech perception system so their stress patterns might not be stored as templates in long-term memory. It remains to be determined whether the perception system is sensitive to such subtle prosodic cues in natural speech production when they are non-contrastive for the speakers.

In the present study, we focused on the perception of subphonemic differences in the homophonous French sequence: *la locutio*n [la#lɔkysjɔ̃] vs. *l'allocution* [l#alɔkysjɔ̃], where the # marks word boundaries. A modified version of the oddball paradigm was used, in which each stimulus came from different productions of the same speaker. Acoustic measurements were carried out to characterize the differences between the two initial syllables of the sequences ([la]). In a first experiment (Syllable experiment), we used the homophonous syllables [la#] vs. [l#a], with [la]s excised from the natural productions of *la locution* vs. *l'allocution*. In this experiment, the length of the tokens was controlled without corrupting the signal in order to isolate the effect of other prosodic features such as pitch. In a second experiment (Word experiment), we used different productions of the same linguistic unit (i.e. homophone determiner + noun sequences such as la#locution) to study natural language processing. The aim was not to isolate a precise feature but to test whether the perception system can distinguish between homophones differing in non-contrastive subphonemic cues in French language. This has never been investigated with online methods such as EEG, which do not need the participants to be engaged in a task that requires focusing on stimuli. ERPs were recorded while French speakers were presented with four standards, which corresponded to four different tokens of the same linguistic unit. The four standards were followed by a test stimulus that could be another token of the standard (Identical condition), a phonemically identical sequence (Homophonous condition) or, as a control, a phonemically different sequence (Dissimilar condition). In the Dissimilar condition, the syllables were phonemically different ([l#a] or [la#] vs. [l#i]).

We hypothesized that: a) since the stimuli of the Different condition ([li]) are phonemically different from the standards ([la]), they should elicit an MMN. b) if fine-grained acoustic cues are encoded by the system, the homophonous deviants ([la#] and [l#a]) should map onto different linguistics representations and, therefore, elicit an MMN. c) if only perceptually relevant regularities in the input are taken into account, no MMN should be found under the Identical segmentation condition because standards should activate a unique linguistic representation despite the intra-standard variability. d) if an MMN is observed in the Homophonous condition, it could present an asymmetric pattern in accordance with the nature of the standard (consonant- or vowel-initial, i.e. [la#] or [l#a]) as observed in previous behavioral studies. Indeed, behavioral priming experiments in ambiguous utterances have shown that one word segmentation such as l#af (from *l'affiche*) led to the activation of both target and non-target, while the two-word segmentation such as la#f (from *la fiche*) led to the activation of only the target (Spinelli et al., 2007). Moreover, since the MMN latency in acoustic experiments is sensitive to stimulus duration, differences in latency could be expected in the Word experiment. e) if the processing of allophonic features in French homophones allows different semantic activations, we might find differences in MMN topography (Brunellière et al., 2011). Lastly, apart from topography and latency measures, in the Word experiment we expected to replicate results from the Syllable experiment.

2. Materials and methods

2.1. Participants

Eighteen native French volunteers participated in the Syllable experiment (10 females, M = 22 years, SD = 3) and 19 others in the Word experiment (9 females, M = 21 years, SD = 3). They were all right-handed, had normal hearing (pure-tone thresholds not exceeding 20 dB HL over a 125 Hz to 8 kHz range) and no documented history of language impairment nor neurological disorders. The two experiments were conducted in accordance with the Declaration of Helsinki. After being informed of the experimental procedure, all the participants provided written consent to take part in the study and were paid for their participation. The research protocol was approved by the Ethical Committee of the Grenoble Hospital (CHU of Grenoble, France; ID RCB: 2012-A01653-40).

2.2. Stimuli

Three French utterances, *la locution* ([lalokysiɔ̃], ‘the locution’), *l'allocution* ([lalokysiɔ̃], ‘the speech’) and *l'illocution* ([lilokysiɔ̃], ‘the illocution’) were excised from sentences recorded by the same native French female speaker who was unaware of any experimental details. The experimental utterances were extracted from 5 recordings of carrier sentences, thus resulting in 5 versions of each nominal sequence (mean sequence durations were 823 ms, 889 ms, and 872 ms respectively amplitude level normalization at 65 dB-A) that were used in the Word experiment. The use of 5 different tokens for each condition allowed the study of natural speech variation processing. As stated in the introduction, variability between productions coming from different speakers can impair

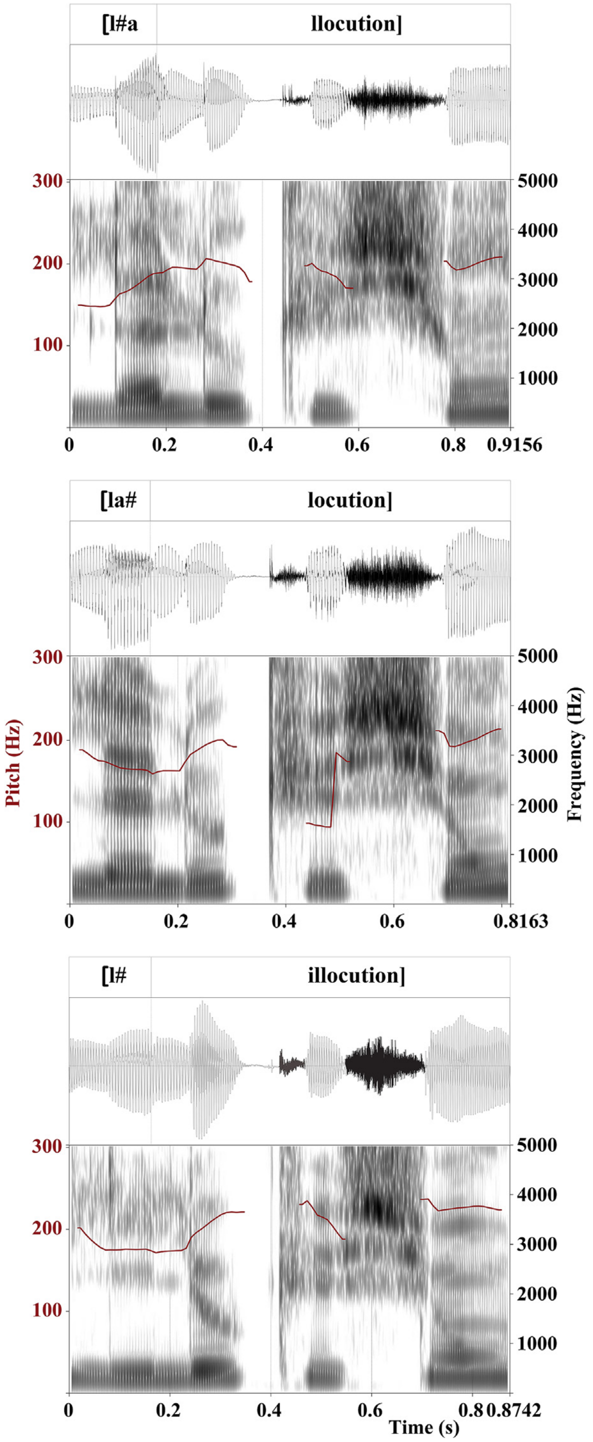

Fig. 1. Waveform and spectrogram of one token of each condition in the Word experiment.

performance in an ABX task. Indeed, the acoustic distance between different tokens of the same linguistic unit coming from different speakers might be high, and could be equivalent to a homophone contrast. For this reason, only intraspeaker variability was considered in these experiments.

Fig. 1 depicts the intensity envelope and the spectrogram of one token from each condition in the Word experiment. Stimulus intensity was normalized in both experiments. The acoustic characteristics (Tables 1 and 2) confirmed both variability in the different productions of the same token as well as subphonemic regularities, in accordance to what has been described in the literature

Table 1

Summary of acoustic measurements for all the stimuli used in the Syllable experiment. Measures: Duration of the segment, duration of the vowel, f0, F_1 and F_2 values of the first vowel.

	Production 1	Production 2	Production 3	Production 4	Production 5	Mean	SD
Duration of the segment (ms)							
[la#]	140.00	140.50	140.00	140.50	140.30	**140.26**	**0.25**
[l#a]	139.52	139.59	139.30	140.00	139.84	**139.65**	**0.27**
[l#i]	140.40	140.40	138.96	139.73	140.90	**140.08**	**0.75**
Duration of the vowel (ms)							
[la#]	70.00	75.38	74.15	80.00	62.30	**72.37**	**6.66**
[l#a]	64.00	75.38	80.44	64.68	83.54	**73.61**	**8.95**
[l#i]	82.82	77.44	78.80	87.24	90.52	**83.36**	**5.53**
f_0 value of the first vowel (Hz)							
[la#]	165.55	174.27	169.94	169.31	168.42	**169.50**	**3.15**
[l#a]	201.12	203.19	202.18	205.73	202.96	**203.04**	**1.71**
[l#i]	183.17	178.35	184.17	177.58	180.25	**180.70**	**2.90**
F_1 value of the first vowel (Hz)							
[la#]	498.46	537.32	560.42	559.33	565.91	**544.29**	**27.85**
[l#a]	660.07	607.25	676.78	673.33	651.44	**653.77**	**27.93**
[l#i]	376.06	325.99	330.30	325.27	326.62	**336.85**	**22.01**
F_2 value of the first vowel (Hz)							
[la#]	2019.54	1932.79	1959.04	1838.30	1821.11	**1914.16**	**83.49**
[l#a]	1881.72	1824.66	1835.17	1818.95	1804.7	**1833.04**	**29.35**
[l#i]	2488.90	1403.42	2395.83	2596.99	2517.84	**2280.60**	**495.60**

Table 2

Summary of acoustic measurements for each of the stimulus used in the Word experiment. Measures: Duration of the first syllable, duration of the vowel in the 1st syllable, f0, F_1 and F_2 values of the first vowel.

	Production 1	Production 2	Production 3	Production 4	Production 5	Mean	SD
Duration of the 1st syllable (ms)							
La locution	149.52	134.12	131.79	122.84	146.95	**137.04**	**11.09**
L'allocution	200.90	184.70	186.13	201.27	198.17	**194.23**	**8.15**
L'illocution	161.83	156.36	157.74	152.84	140	**153.75**	**8.34**
Duration of the vowel in the 1st syllable (ms)							
La locution	81.19	81.49	70.62	76.55	84.39	**78.85**	**5.39**
L'allocution	86.6	80.30	85.03	87.01	80.03	**83.79**	**3.40**
L'illocution	83.05	88.31	77.74	83.99	6200	**79.02**	**10.23**
f_0 value of the first vowel (Hz)							
La locution	164.31	171.35	164.59	159.71	158.74	**163.74**	**5.01**
L'allocution	189.27	189.93	174.62	185.92	183.06	**184.56**	**6.20**
L'illocution	173.57	177.77	172.21	167.98	171.60	**172.63**	**3.54**
F_1 value of the first vowel (Hz)							
La locution	483.60	490.56	486.15	542.46	499.64	**500.48**	**24.25**
L'allocution	648.72	614.17	571.82	571.47	617.40	**604.72**	**33.07**
L'illocution	316.06	322.28	283.38	259.14	312.08	**298.59**	**26.63**
F_2 value of the first vowel (Hz)							
La locution	2012.83	1994.19	2021.48	2090.76	1976.84	**2019.22**	**43.55**
L'allocution	2037.45	2025.75	1956.02	2042.05	2012.44	**2014.74**	**34.77**
L'illocution	2392.84	2363.16	2387.42	2578.83	2408.50	**2426.15**	**86.89**

(Spinelli et al., 2007). The content-word initial syllables [l#a] in *l'allocution* had longer durations than [la#] in *la locution* (respectively M = 194.23 ms, SD = 8.15 and M = 137.04 ms, SD = 11.09. This difference was statistically significant, as assessed with a *t-test*, $p < 0.001$). Intonational differences between the two homophonous conditions were also identified, as shown by mean f0s of the first vowel ([a]). Higher f0 values were observed for content-word initial vowels ([a] in *l'allocution*) than in the [a] of *la locution* (+33.54 Hz in the Syllable experiment and +20.82 Hz in the Word experiment). In our sample of stimuli, this difference was statistically significant, as assessed by a *t-test* ($p < 0.001$). Finally, the [a] in *l'allocution* had higher F_1 values, thus indicating a lower tongue position (hence, a more canonical pronunciation of a low-vowel), than its homophonous counterpart, [a] in *la locution*, (+109 and + 104.23 Hz in Experiments 1 and 2 respectively, $p < 0.001$). No differences were found in our sample regarding F_2 in the stimuli of any of the experiments ($p > 0.05$). It is also worth noting that the [l#a] productions elicited less variability than the [la#] productions in two measures: duration of the first syllable and f0 values of the first vowel, which suggests a more systematic pronunciation. For further details concerning the different values of each production, see Tables 1 and 2 The acoustic feature extractions, amplitude envelopes, and spectrograms were performed with the software Praat (Boersma & Weenink, Phonetic Sciences, University of Amsterdam, The Netherlands). It is to note that measurements are based on mean values over the vowel duration. Since the stimuli were excised in the first experiment to match 140 ms duration, the time window selected for the measurement differed between experiments. Therefore, the mean f0 and F1 values also differ slightly.

For the Syllable experiment, the first syllable [la] ([la#] from *la locution* or [l#a] from *l'allocution*) and the first syllable [li] ([l#i] from *l'illocution*) were then excised from their respective nominal sequences. All the extractions were performed at the closest zero-crossing point of stimuli boundaries in the acoustic signal. Syllable duration was equalized without corrupting the signal (i.e. by cutting the syllables at the same length). The acoustic properties of the stimuli used in the Syllable experiment are reported in Table 1, including duration of the sequence, duration of the first vowel, f0 and first and second vowel formant values. The table reports the subphonemic differences within tokens from the same condition and between conditions. In the Word experiment, the homophonous nominal sequences were used as stimuli. Fig. 1 shows the spectrogram and waveform of one stimulus of each conditions in the Word experiment: [l#allocution] (top panel), [la#locution] (middle panel), and deviant condition [l#illocution] (bottom panel).

2.3. Procedure

Both experiments took place in an electrically and acoustically shielded chamber. Participants sat comfortably in front of an LCD display. They were instructed to watch a movie of their choice without sound and to ignore the auditory stimuli presented in stereo binaurally via headphones at a comfortable listening level (65 dB SPL). Sounds were presented in a modified version of the oddball paradigm, in which a series of four standards (coming from four different productions) was followed by a stimulus in the test position.

The test item could be

- another production of the standard (Identical condition, same segmentation with intraspeaker variability)
- a stimulus from 5 different productions of a homophonous sequence (Homophonous condition), i.e., phonemically identical to the standards, but with different segmentation
- or a stimulus from 5 different productions of a non-homophonous [li] sequence (Dissimilar condition), i.e., phonemically different from the standards.

As a result, the homophonous or dissimilar stimulus could be presented after 4, 9 or 14 standard stimulus, thus making unpredictable which stimulus was going to be presented in the test position. Prediction was also hindered by the fact of having different tokens of the standard and test stimulus.

In the Syllable experiment, the initial syllables of the utterances (i.e. la#, l#a, l#i) were used as stimuli. In the Word experiment, the nominal sequences were used (i.e. la#locution, l#allocution and l#illocution). The procedure was the same for both experiments. The presentation order for the standards and test items was pseudo-randomized, and a 500 and 250 ms inter-stimulus interval (ISI) was applied in the Syllable and the Word experiment respectively. This allowed a good trade-off between the number of trials and experiment length. Moreover, to compare results from both experiments, the total length should be equivalent. Each experiment was divided into two consecutive blocks where the standard stimuli in one block were the homophonous stimuli in the other block, and vice-versa. The two blocks were separated by a 5-min break. The order of blocks was counterbalanced across subjects. In the Syllable experiment, which lasted 45 min, 1800 stimuli were presented. In the Word experiment, 1125 stimuli were presented, and therefore the experiment lasted 55 min.

2.4. EEG recording, pre-processing and analyses

EEG recording was performed using the Biosemi system with 32 active electrodes (Electro-Cap International, Inc., Ohio; Biosemi, ActiveTwo, version 5.36) positioned according to the International 10–20 system. The EEG data were collected at a sampling rate of 2 kHz over a [0.1–400 Hz] bandwidth and referenced to a common reference (CMS) and ground (DRL) directly integrated into the cap.

Offline analyses were performed using EEGLAB toolbox (Delorme & Makeig, 2004) and customed routines. Bad channels were spherically interpolated. Continuous data were downsampled to 1000 Hz and visually inspected to discard artifactual segments. Data were re-referenced to the average of the mastoid electrodes and a 0.5–30 Hz bandpass filter was applied. Regarding ocular artifacts, an independent component analysis allowed the identification and subtraction of this type of artifacts. Epochs were generated related to stimulus onset over a time window from −100 ms to 800 ms. They were baseline-corrected by subtracting the average signal activity across the 100 ms pre-stimulus period. Finally, epochs were separately averaged for each test and standard stimuli in each of the two blocks for each participant. We did not collapse data across blocks to be able to study potential asymmetries depending on the nature of the standard stimulus. As mentioned at the end of the introduction, the presence of one of the homophones could allow discriminating the other one more easily.

The MMN component was computed by subtracting the ERP waveform elicited by the standard from the ERP elicited by the deviant. This procedure minimizes the influence of the physical differences between the two types of stimuli in the early ERP responses. Grand-averages ERPs were submitted to statistical analysis. The grand-grand average MMN difference wave was maximal at Fz, followed by Cz, as already described in the literature (see Bishop, 2007 for a review). The Fz electrode was therefore selected for statistical analysis. The Identical condition, which was not expected to elicit an MMN response, was used as a control condition. For each subject, the amplitude of each MMN component was measured within a 50-ms window centered on the peak of the grand-grand average difference wave (i.e. component mean amplitude for each condition; see Zora et al., 2015, for another example of MMN mean amplitude measurement centered at the peak latency in a study of lexical stress).

Differences in mean amplitude between conditions were assessed using a multilevel mixed model approach, including data from both experiments. This allowed the study of between and within subjects' effects in the same statistical test, which is regarded as a hierarchical or nested mixed model. Fixed effects corresponded to the variables of interest, namely condition (Identical, Homophonous and Dissimilar), type of standard ([la#] and [l#a]) and experiment (syllable or word). The interaction between condition and experiment was also considered in the model. Type of standard was included as random slope for the random effect of subjects. These analyses were run in R, using the lmer function from the lme4 package (Bates & Maechler, 2009). Since we compared models with different nested fixed effects, the maximum likelihood (ML) was used to estimate variance components (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). The lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015) allowed to determine *p*-values (for more methodological detail see Luke, 2017).

The other statistical analyses were run between only homophonous conditions for each experiment. With the aim of studying differences in MMN topography between the homophonous conditions of each experiment, we ran paired *t*-tests fdr-corrected including data from the 32 electrodes. Topographical analyses were performed for the same time window that was used for the mean amplitude analysis. Four region of interest (ROI) were defined: frontal (Fz, AF3, AF4), right central (FC2, F4, C2, FC6), left central (FC1, F3, C3, FC5) and central (Cz, CP1, CP2). We run a 2 × 4 rANOVA with factors (type of standard) and ROI (four levels corresponding to each ROI). In addition, we also compared the time course of the processing of homophones. For each experiment, paired *t*-tests were computed on MMN latency measures at Fz. Yet, MMN peak latency values should be interpreted with caution, since they are not as reliable as mean amplitude measurements (Bishop, 2007).

3. Results

3.1. MMN Syllable experiment: descriptive statistics

Fig. 2a depicts the grand average standard minus test difference ERPs at Fz, for the three conditions, and that when l#a (top) and la# (bottom) were used as standards. As expected for both standard contexts, only the Homophonous and Dissimilar conditions revealed an MMN component. This negative difference wave peaked at 265 ms and 225 ms when elicited in the Homophonous and the Dissimilar conditions respectively, which corresponds to a mismatch response. No negativity was observed in the Identical condition (i.e., the other token of the standard), thus confirming that the intra-standard variability was not considered as deviant. Fig. 2b shows the component topography for the three conditions during the time window that was used for statistical analysis (240–290 ms for the Homophonous and 200–250 ms for the Dissimilar condition).

3.2. MMN Word experiment: descriptive statistics

Fig. 3a depicts the grand average standard minus test difference ERPs at Fz, for the three conditions in the word experiment, and

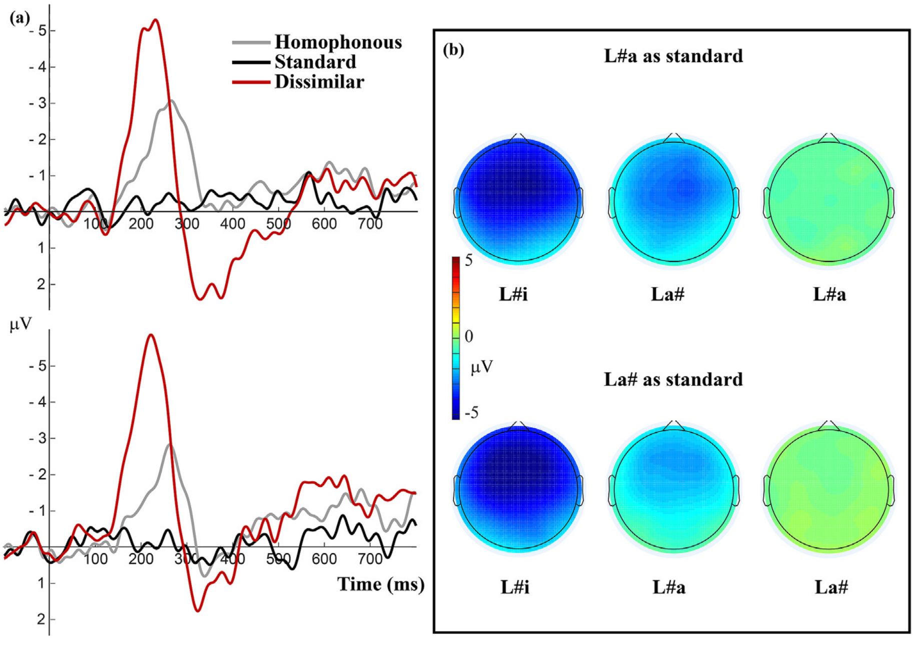

Fig. 2. EEG results in the Syllable Experiment. a) Grand-average standard minus test waveforms at Fz for the three conditions. The ERPs elicited when l#a was the standard are depicted at the top. The ERPs elicited when la# was the standard are depicted at the bottom. b) Grand-average MMN topography for the three conditions during the 50 ms time window around each MMN peak.

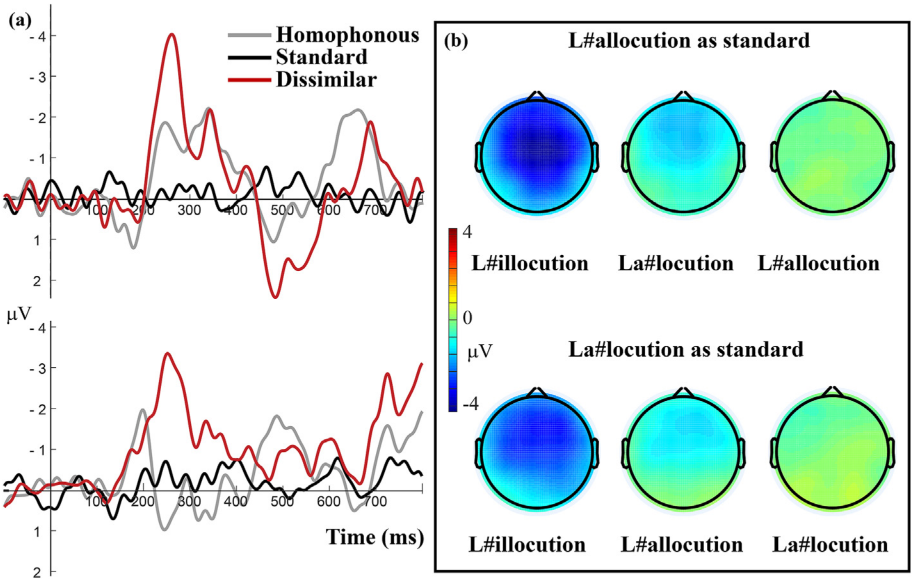

Fig. 3. EEG results in the Word Experiment. a) Grand-average standard minus test waveforms at Fz for the three conditions. The ERPs elicited when l#allocution was the standard are depicted at the top. The ERPs elicited when la#locution was the standard are depicted at the bottom. b) Grand-average MMN topography for the three conditions during the 50 ms time window around each MMN peak.

that when l#allocution (top) and when la#locution (bottom) were used as standards. Like in the previous experiment, only the Homophonous and Dissimilar conditions revealed an MMN component. This negative difference wave peaked at 250 ms when elicited by the Dissimilar condition. The latency of the MMN peak in the Homophonous condition depended on the standard. When la#locution was the standard, the MMN elicited by l#allocution peaked at 200 ms, whereas the homophonous MMN peaked at 235 ms when elicited by l#allocution as standard. Fig. 3b shows the component topography for the three conditions during the time window that was used for statistical analysis.

3.3. Statistical analysis

A multilevel mixed linear model analysis on component mean amplitude including data from the two experiments revealed a main effect of Condition. When comparing to the Identical segmentation condition, mean amplitude was increased by an estimate factor of 2.38 μV in the Homophonous condition ($t = 6.55$, $p < 0.001$) and by 4.94 μV in the Dissimilar condition ($t = 13.60$, $p < 0.001$). The main effects of Experiment was not significant ($t = 0.04$, $p > 0.05$). Yet, the interaction between Condition and Experiment was significant as follows: the mean voltage in the Homophonous condition was higher in the Syllable experiment ($t = 2.33$, $p < 0.05$) than in the Word experiment, with an estimate of 1.18 μV difference in amplitude. Similarly, the mean voltage in the Dissimilar condition was also greater in the Syllable experiment ($t = 3.84$, $p < 0.001$) than in the Word experiment, with an estimate of 1.94 μV. Despite the greater MMN response in the Syllable experiment, the main results were replicated in the Word experiment (i.e. an MMN response in the Homophonous condition and a larger MMN response in the Dissimilar condition). The main effect of Type of standard was not significant ($t = 0.25$, $p > 0.05$), thus suggesting that the MMN amplitude did not differ as a function of the type of standard. Lastly, for each experiment, we assessed differences in MMN mean amplitude between the Homophonous and Dissimilar conditions. Two-way anova repeated measures (rANOVA) with factors Condition (Homophonous and Dissimilar) and type of standard (la# or l#a) revealed a significant main effect of Condition in the Syllable experiment ($F(1,17) =$, $p < 0.0.001$, $\eta = 0.766$). This result indicated that the MMN mean amplitude was greater in the Dissimilar conditions. The other main effect and interaction were non-significant (F < 2.67, $\eta < 0.14$). rANOVA yielded the same results in the Word experiment. The main effect of Condition was significant ($F(1,18) = 61.67$, $p < 0.0.001$, $\eta = 0.517$), thus indicating that mean amplitude was greater in the Dissimilar conditions. The other main effect and interaction were non-significant ($F < 2.1$, $\eta < 0.03$).

The MMN topography and latency were compared between the Homophonous conditions of each experiment (e.g. the MMN topography and latency elicited in the Homophonous condition when la# was the standard vs. the MMN elicited when l#a was the standard in the Syllable experiment). To assess differences in topography between the homophonous MMNs, we ran fdr-corrected *t-tests* for each of the 32 electrodes during the time window of interest. None of the test was significant ($p > 0.05$, $t < 1.8$, $d < 0.6$) in either of the experiments. Neither the main effect of type of homophone, F(1, 17) = 0.91, p > 0.05, ŋ = 0.05, nor the main effect of ROI, F(3, 51) = 0.87, p > 0.05, ŋ = 0.05, was significant. Their interaction was not significant either, F(3, 51) = 0.91, p > 0.05, ŋ = 0.05. Similar results were obtained in the word experiment (F(1, 17) = 0.31, p > 0.05, ŋ = 0.02 for the main effect of type of standard, F(3, 51) = 1.02, p > 0.05, ŋ = 0.05, for the main effect of ROI and F(3, 51) = 0.13, p > 0.05, ŋ = 0.01 for their interaction). In the Syllable experiment, when comparing latency between Homophonous conditions, no differences were found in a

100 ms time-window around the MMN peak at 265 ms ($t < 1, p > 0.05$). In the Word experiment, latency measures were computed in a time window between 175 and 275 ms after stimulus onset. The difference in latency between Homophonous conditions was significant ($t = 4.49$, $p < 0.001$, $d = 1.69$). The MMN peaked 35 ms earlier when la#locution was the standard.

4. Discussion

The present study aimed to test whether non-contrastive acoustic features differentiating homophonous sequences are relevant for the speech perception system. This was examined under the condition of intra-speaker variability and without sentential context. To the best of our knowledge, the MMN studies that have used multiple linguistic tokens as standard stimuli, used a deviant stimulus that was phonemically different. Those studies support the processing of vowel sounds pre-attentively regardless of speaker variability. Yet, here we test subphonemic differences that are not contrastive in a particular language. Most of the studies addressing the processing of prosodic cues have compared native vs non-native speakers of a language, and so native vs non-native allophonic contrasts (Altenberg, 2005; Jongman et al., 2017; Rojczyk, 2018; Shoemaker, 2014b; Tyler & Cutler, 2009). The present study focused on the processing of homophonous utterances (i.e. same phonological representation) by native speakers in natural settings. In two different experiments, we examined the automatic processing of allophonic features in a non-stress-timed language: French (e.g. duration and pitch do not encode meaning). To study the processing of these prosodic cues without focused attention, we measured the MMN component, which was elicited in a modified version of the oddball paradigm. To get close to natural language conditions, each type of stimulus, both standard and test, were different tokens of the same linguistic unit produced by the same speaker.

As expected, the Identical condition, where different productions of the same token were used, did not elicit an MMN response in any experiment. Thus, the slight changes in the acoustic properties of speech, corresponding to the variations between different productions of the same word, were not considered as deviant by the neural system. These results add to and extend a growing literature showing that the MMN response does not merely represent discrimination of repetition of identical stimuli or a sequence of stimuli, but it is a more complex comparator (Eulitz & Lahiri, 2004; Shestakova et al., 2002). The neural system would be able to extract the statistical characteristics of the physical properties of speech utterances and use them to discriminate speech contrasts.

The most important result of the study was the presence of the MMN in the Homophonous conditions, and that regardless of the type of standard stimuli (i.e. one of the two homophonous utterances) and the type of stimuli (i.e. syllable or word). It is well established that the MMN can reflect the activation of speech elements and phoneme representations stored in long-term memory (Pulvermüller & Shtyrov, 2006). Here, we show that the MMN is also sensitive to meaningful subphonemic differences even when they are non-contrastive in a particular language. Indeed, the neural system encodes relevant fine prosodic cues such as slight changes in duration and pitch that allow differentiating between homophonous utterances even when such cues are non-contrastive in the language of the speaker. This result suggests that the representations of speech sounds in memory include subphonemic information, and therefore they are not reduced to the phoneme level. More research is needed on allophonic cues before conclusions are drawn on speech representation units. What we can conclude from the presence of the MMN response in the Homophonous but its absence in the Identical condition is that the MMN is a powerful tool for tracking relevant changes in the speech signal. This confirms that the MMN component is able to extract out what is common in a stream of stimuli to assess without focused attention whether a subsequent stimulus falls within the range of relevant acoustic variance while remaining unresponsive to intra-speaker variations of the same speech sequence. The MMN response would therefore reflect the measure of deviance from a sensory prediction. This response discriminates between subtle relevant and irrelevant differences that are not used to differentiate phonemes in a language.

The MMN response had already been described in the context of multi-token variability. Yet, standard and test stimuli were different vowels (Deguchi et al., 2010; Eulitz & Lahiri, 2004) or differed in the sex of the speaker (Brunellière, Dufour, Nguyen, & Frauenfelder, 2009). Here, we show for the first time an MMN response to subphonetic differences in the context of multitoken stimuli from the same speaker, which approaches the natural use of language. The underlying representations of phonemes in the mental lexicon might be abstract enough to deal with the numerous possibilities of acoustic features in spoken language.

As predicted, an MMN response was observed in the Dissimilar conditions in each experiment. Their amplitude was greater than the MMN amplitude in the Homophonous conditions. The stimuli *l'i* and *l'illocution* are phonemically distinct from the standards and differ mainly in F_1 and F_2. Indeed, despite an unclear contrast in duration or pitch, each dissimilar stimulus was acoustically further from the standard than its homophonous counterpart. Accordingly, Michelas, Frauenfelder, Schön, and Dufour (2016) found a stronger EEG response to phonemic than to stress information for French speakers. In contrast to our results, Honbolygó et al. (2017) found differences in MMN amplitude when comparing phonemic with contrastive acoustic deviations. F0 and consonant duration changes elicited a larger MMN component than phonemic changes. This suggests an enhanced sensitivity to prosodic changes for Hungarian speakers. Whereas the acoustic variations they studied are contrastive in Hungarian, the acoustic deviations studied here are not. The difference in results as a function of the language confirms the importance of the MMN as a tool to measure the relevance of a deviation in a specific language.

The asymmetry of the time course of the component in the Word experiment was another major finding. In the Syllable experiment, where duration was equalized, no differences were found in MMN latency regardless of the standard stimuli. The MMN response can be ascribed to other acoustic differences, such as pitch variations. In the Word experiment, the Homophonous test stimulus *l'allocution* (*la locution* being the standard) elicited an earlier MMN peak latency than its homophonous counterpart *la locution*. However, in our sample, the first syllable of *la locution* was in average 57 ms shorter in duration than the first syllable *l'allocution*. Stimulus durational differences are well known to cause differences in MMN latency (Tervaniemi et al., 1999), and indeed a 35 ms difference in latency was found in the Word experiment. Another interpretation could be that since *la* in *la locution* is one of the most common definite articles in French (the$_{\text{feminine}}$) with a production timely effective, the speech perception system is highly trained in its processing. Thus, a deviation from this norm might be discriminated earlier. Further experiments are needed to clarify this point.

We did not find differences in the topographical distribution of the component between homophonous conditions in any of the experiments. This might indicate that there was no access to lexical characteristics of the speech. This result is in accordance with previous findings observed for southern French speakers in the processing of French phonemes (Brunellière et al., 2011). When speakers did not produce a phonemic contrast between two words in a particular regional variety of French, and so the utterances were treated as homophones, no differences were found in MMN topography between conditions when the two homophones were presented as deviant stimuli. On the other hand, speakers of standard French who differentiate phonemically between utterances showed differences in MMN distribution. Differences in MMN topography were also found in Kirmse et al. (2008). This study investigated pre-attentive processing of vowel duration in Finnish vs. German speakers. They found differences in MMN latency (i.e. shorter latency in the Finnish group) as well as differences in topography. They concluded that Finnish speakers were more sensitive to duration contrasts. Importantly, differences in topography were also explained by long-term language experience. Yet, differences in exposure to the stimuli could not account for our results, and moreover, the frequency of both homophonous words was similar. Taken together, non-contrastive subphonemic cues might be processed automatically by the speech perception system, without activating distinctive semantical networks. Accordingly, at the behavioral level, disambiguation of homophonous sequences without context is not fully achieved, and performance decreases when multiple tokens of the same unit are presented (Spinelli et al., 2007). As already mentioned, it has also been shown that the MMN elicited during active listening is larger and more sensitive to with-in category acoustic difference than in passive listening (Deguchi et al., 2010). It is possible that topography differences were not found because there is no semantical access in the perception of homophones without focused attention, and particularly for low-frequency words such as the ones used in this research). This latter hypothesis could be tested with an MMN protocol where participants are performing an active perception task.

Lastly, MMN amplitude was greater in the Syllable experiment than in the Word experiment. This could be explained by differences in the complexity of the stimuli that were used in each experiment. The source of the MMN generators depends on the cognitive information that elicits the response. The more complex, the more variable the brain response might have been. Moreover, stimuli were shorter in the Syllable experiment than in the Word experiment. The speech perception system could be more sensitive to acoustic details in shorter stimuli than in longer ones. This interpretation is in agreement with the higher MMN amplitude in the Homophonous and the Dissimilar conditions in the Syllable experiment. On the other hand, the duration of each standard and test stimulus differed in the word experiment, thus resulting in differences in MMN latency. This might have caused a reduction in amplitude during the averaging process. Finally, it is to be considered that although no evidence of semantical processing of homophones was found, differences in meaning between stimuli in the Syllable and the Word experiment could also explain the differences in the MMN amplitude.

By looking at the acoustic analyses, it is not clear which of the features accounted for the overall MMN effect under the Homophonous conditions. Despite the absence of lexical stress in French, acoustic analysis has shown that pitch accent can convey speaker commitment to the content of the proposition (Michelas, Portes, & Champagne-Lavau, 2015). Differences in prosodic features are also found in discourse markers such as *alors* (then) and *et* (and) in French, as a function of the meaning that they convey in the sentence (Didirková, Christodoulides, & Simon, 2018). When it comes to nominal homophonous utterances, allophonic feature differences have never been characterized in detail. In our sample, multiple fine acoustic differences were found between the natural production of the first syllable from the homophonous *l'allocution* and *la locution*, as shown by the acoustic properties of the stimuli. The first syllable of *l'allocution* has a longer duration than the first syllable of *la locution.* Interestingly, the vowel length in [l#a] was not systematically longer than the vowel in [la#]. Yet, the syllabic segment [la#] was longer in duration. This brings additional support in favor of the syllable as a prosodic unit. Another acoustic difference between the two segments was a slightly higher F_1 value in [a] from *l'allocution*. There are also prosodic differences in pitch height, such as a higher f0 value for the vowel [a] in *l'allocution*. Therefore, it is unclear which of the sound properties accounted for the overall MMN effect under the Homophonous conditions. Since differences in MMN response were found in the Syllable experiment despite the equal duration of the stimuli, f0 seems a particularly good candidate. Moreover, Honbolygó et al. (2017) identified f0 and consonant length as more important features than intensity or vowel duration, as indicated by larger MMN components elicited by deviations in these features. Future research should isolate the effect of each f0 dimension (e.g. height, slope and direction) on automatic segmentation. Since our stimulus sample only included a few tokens from a single speaker, the acoustical analysis should be conducted on an extended sample to reach conclusions on acoustic features. Moreover, to reach a deeper understanding of the neural representations of subphonemic categories in French, ERP should be recorded during attentive perception of homophonous segments.

In the two experiments, we investigated automatic discrimination of French homophonous utterances without focused attention. We found that the MMN was sensitive to such nominal segments, which differ in non-contrastive features. The most likely candidate to account for the result is pitch. There is a growing body of behavioral research on listeners' perceptions of intonation patterns (Dilley & McAuley, 2008; Mattys, 2004). Regarding word boundaries, our findings would be in line with previous studies that have identified f0 as a modulator of the segmentation of homophonic sequences in French. Indeed, Spinelli, Grimault, Meunier, & Welby (2010) reported that increasing f0 by resynthesizing the vowel [a] of *c'est la mie* [se#la#mi] (it is the crumb) increased incorrect identification (i.e. more *c'est l'amie* [se#l#ami] (it is the friend) responses) in a forced choice task. In non-lexical-stress languages, intonation patterns have been mostly investigated at the sentence level, for instance, in rhetorical relations or irony. It has been shown that sarcasm in French is acoustically characterized by a higher pitch, which can also help to correctly identify sarcasm even in the absence of context (Lœvenbruck, Jannet, D'Imperio, Spini, & Champagne-Lavau, 2013; González Fuente, Prieto Vives, & Noveck, 2016). In order to draw stronger conclusions on the role of pitch in discrimination of ambiguous utterances, such as homophones, this acoustic feature should be the target of further online studies using a larger sample of stimuli. Our results are in line with previous MMN data on stress processing that support the encoding of IA in French (Aguilera et al., 2014). IA seems to be processed at the level of the syllable representation, and it might consequently contribute to word recognition in French. Our findings extend previous research showing that French listeners are not deaf to stress patterns (Michelas, Esteve-Gibert, & Dufour, 2018).

A large body of literature supports the processing of lexical (Pettigrew et al., 2004), semantic (Shtyrov, Hauk, & Pulvermüller, 2004) and syntactic speech information such as grammar (Pulvermüller & Shtyrov, 2003) without focused attention. Here we add evidence in support of an automatic processing of non-contrastive fine prosodic information. Furthermore, our study supports the contribution of the MMN as a promising tool to understand the brain's response to speech. The MMN not only reflects mere identity comparison, but also a higher order perceptual process where the statistical acoustic regularities of speech signals are extracted.

5. Conclusions

Overall, our study shows that the speech perception system automatically processes fine subphonemic features, such as duration and pitch, even when they are not lexically contrastive in a language. This processing does not require the subject's focused attention, which contrasts with the incomplete disambiguation of homophonous utterances found in behavioral studies. The fine allophonic cues that differentiate homophonous sequences are sufficiently robust despite intra-speakers' variability to allow listeners to extract regularities associated with each production. The extraction of these regularities by the speech perception system may serve the final goal of automatic and quick guidance toward correct segmentation and comprehension.

If the speech perception system is sensitive to the fine-grained acoustic information of initial syllables in French, such subphonemic cues could play an important role in word segmentation in other languages. Future research should confirm these findings in speakers of other languages where duration is not lexically contrastive (e.g. Spanish) or speakers of non-stressed languages (e.g. Finnish). Our results are also a foundation for further research in French language. The potential effect of the different allophonic features should be examined. For instance, f0 height and direction should be manipulated to better understand the contribution of the major f0 dimensions. Moreover, our results should be replicated in the context of inter-speaker variability. The understanding and identification of the prosodic cues that are relevant to speech segmentation could help to improve machine-learning classification and have applications in the field of automatic speech recognition.

Acknowledgements

This research was supported by a European Research Council grant to the SpiN Project (No. 209234).

References

Aguilera, M., El Yagoubi, R., Espesser, R., & Astésano, C. (2014). Event-related potential investigation of initial accent processing in French. *Proceedings of Speech Prosody,* 383–387.

Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research, 21*(4), 325–358. https://doi.org/10.1191/0267658305sr250oa.

Astésano, C., Bard, E. G., & Turk, A. E. (2007). Structural influences on initial accent placement in French. *Language and Speech, 50*(3), 423–446. https://doi.org/10.1177/00238309070500030501.

Bates, D., & Maechler, M. (2009). Package 'lme4'(Version 0.999375-32): Linear mixed-effects models using S4 classes. Available (April 2011) at http://Cran.r-Project.Org/Web/Packages/Lme4/Lme4.Pdf.

Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychological Bulletin, 133*(4), 651–672. https://doi.org/10.1037/0033-2909.133.4.651.

Brunellière, A., Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2009). Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition, 111*(3), 390–396. https://doi.org/10.1016/j.cognition.2009.02.013.

Brunellière, A., Dufour, S., & Nguyen, N. (2011). Regional differences in the listener's phonemic inventory affect semantic processing: A mismatch negativity (MMN) study. *Brain and Language, 117*(1), 45–51. https://doi.org/10.1016/j.bandl.2010.12.004.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*(1), 148–156. https://doi.org/10.1016/j.brainres.2006.10.064.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351.

Deguchi, C., Chobert, J., Brunellière, A., Nguyen, N., Colombo, L., & Besson, M. (2010). Pre-attentive and attentive processing of French vowels. *Brain Research, 1366*, 149–161. https://doi.org/10.1016/j.brainres.2010.09.104.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*(4), 635–647.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Di Cristo, A. (1998). *Intonation in French. Intonation systems.* A Survey of Twenty Languages.

Didirková, I., Christodoulides, G., & Simon, A. C. (2018). The Prosody of Discourse Markers alors and et in French A Speech Production Study. In: *Proc. 9th International Conference on Speech Prosody* (pp. 503–507). 2018.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*(3), 294–311.

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in speech recognition. *Journal of Cognitive Neuroscience, 16*(4), 577–583.

Gaskell, M. G., Spinelli, E., & Meunier, F. (2002). Perception of resyllabification in French. *Memory & Cognition, 30*(5), 798–810.

González Fuente, S., Prieto Vives, P., & Noveck, I. A. (2016). A fine-grained analysis of the acoustic cues involved in verbal irony recognition in French. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.). *Speech prosody 2016; 2016 may 31-june 3; boston, United States of America* (pp. 902–906). International Speech Communication Association. https://doi.org/10.21437/SpeechProsody. 2016-185 [Place Unknown], 2016.

Honbolygó, F., Kolozsvári, O., & Csépe, V. (2017). Processing of word stress related acoustic information: A multi-feature MMN study. *International Journal of*

*Psychophysiology, 118*https://doi.org/10.1016/j.ijpsycho.2017.05.009.

Ito, K., & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *Journal of the Acoustical Society of America, 125*(4), 2348–2360. https://doi.org/10.1121/1.3082103.

Jongman, A., Qin, Z., Zhang, J., & Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *Journal of the Acoustical Society of America, 142*(2), EL163–EL169. https://doi.org/10.1121/1.4995526.

Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., & Jacobsen, T. (2008). Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *International Journal of Psychophysiology, 67*(2), 131–143.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: tests in linear mixed effects models. R package version 2.0-20.* Vienna: R Foundation for Statistical Computing.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y.

Lœvenbruck, H., Jannet, M. A. B., D'Imperio, M., Spini, M., & Champagne-Lavau, M. (2013). Prosodic cues of sarcastic speech in French: Slower, higher, wider. *14th annual conference of the international speech communication association (interspeech 2013)* (pp. 3537–3541). .

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 397.

Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America, 122*(1), 554–567.

Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning & Memory, 9*(5), 253–267. https://doi.org/10.1101/lm.49402.

Michelas, A., Esteve-Gibert, N., & Dufour, S. (2018). On French listeners' ability to use stress during spoken word processing. *Journal of Cognitive Psychology, 30*(2), 198–206. https://doi.org/10.1080/20445911.2017.1394862.

Michelas, A., Frauenfelder, U. H., Schön, D., & Dufour, S. (2016). How deaf are French speakers to stress? *Journal of the Acoustical Society of America, 139*(3), 1333–1342.

Michelas, A., Portes, C., & Champagne-Lavau, M. (2015). When pitch accents encode speaker commitment: Evidence from French intonation. *Language and Speech, 59*(2), 266–293. https://doi.org/10.1177/0023830915587337.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1–21.

Näätänen, R., & Alho, K. (1995). Mismatch negativity-a unique measure of sensory processing in audition. *International Journal of Neuroscience, 80*(1–4), 317–337.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport, 4*, 503–506.

Peltola, M. S., Kujala, T., Tuomainen, J., & Ek, M. (2003). *Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. 352*, 25–28. https://doi.org/10.1016/S0304-3940(03)00997-2.

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Finnigan, S., Alku, P., Kei, J., et al. (2004). Automatic auditory processing of English words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear and Hearing, 25*(3), 284–301.

Pulvermüller, F., & Shtyrov, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *NeuroImage, 20*(1), 159–172.

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology, 79*(1), 49–71. https://doi.org/10.1016/j.pneurobio.2006.04.004.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7-and 11-month-old American infants. *Developmental Science, 8*(2), 162–172.

Rojczyk, A. (2018). Nonnative perception of allophonic cues to word boundaries: Lou spills versus loose pills for speakers of Polish. *Language Acquisition, 00*(00), 1–9. https://doi.org/10.1080/10489223.2018.1433672.

Schaegis, A., Spinelli, E., & Welby, P. S. (2005). Perception of phonemically ambiguous spoken sequences in French. *Proceedings of 27th annual conference of the cognitive*. Science Society.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68*(1), 1–16.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., et al. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *NeuroReport, 13*(14), 1813–1816. https://doi.org/10.1097/00001756-200210070-00025.

Shoemaker, E. (2014a). Durational cues to word recognition in spoken French. *Applied PsychoLinguistics, 35*(2), 243–273.

Shoemaker, E. (2014b). The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Studies in Second Language Acquisition, 36*(4), 709–731. https://doi.org/10.1017/S027226311400014X.

Shtyrov, Y., Hauk, O., & Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: The mismatch negativity to action words. *European Journal of Neuroscience, 19*(4), 1083–1092.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage, 12*(6), 657–663.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language, 48*(2), 233–254.

Spinelli, E., Welby, P. S., & Schaegis, A. L. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: The case of French elision. *Language & Cognitive Processes, 22*(6), 828–859. https://doi.org/10.1080/01690960601076472.

Spinelli, E., Grimault, N., Meunier, F., & Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Attention, Perception, and Psychophysics, 72*(3), 775–787. https://doi.org/10.3758/APP.72.3.775.

Tervaniemi, M., Lehtokoski, A., Sinkkonen, J., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Test–retest reliability of mismatch negativity for duration, frequency and intensity changes. *Clinical Neurophysiology, 110*(8), 1388–1393.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28*(4), 397–440.

Turk, A. E., & Shattuck-Hufnagel, S. (2014). Timing in talking: What is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society of London B Biological Sciences, 369*(1658), 20130395.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America, 126*(1), 367–376. https://doi.org/10.1121/1.3129127.

Welby, P. S. (2003). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation.* The Ohio State University.

Welby, P. S. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication, 49*(1), 28–48.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.

Zora, H., Schwarz, I.-C., & Heldner, M. (2015). Neural correlates of lexical stress. *NeuroReport, 26*(13), 791–796. https://doi.org/10.1097/WNR.0000000000000426.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R.* Springer Science & Business Media.