HAL
open science

Theta activity and phase resetting during perception of French homophonous utterances

Noelia Do Carmo-Blanco, Michel Hoen, Elsa Spinelli, Fanny Meunier

To cite this version:

Noelia Do Carmo-Blanco, Michel Hoen, Elsa Spinelli, Fanny Meunier. Theta activity and phase resetting during perception of French homophonous utterances. Language, Cognition and Neuroscience, 2022, 37 (2), pp.154-164. 10.1080/23273798.2021.1950786 . hal-03348382

HAL Id: hal-03348382

https://hal.univ-grenoble-alpes.fr/hal-03348382

Submitted on 3 Jan 2022

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

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

Theta activity and phase resetting during perception of French homophonous utterances

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

[a]*Université Côte d'Azur, BCL, CNRS UMR7320, France*

[b]*Oticon medical, Vallauris, France*

[c]*Université Grenoble Alpes, LPNC, CNRS, UMR5105, Grenoble, France*

[*] Corresponding author: 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. Tel.: + 33 4 89 15 22 83, email: fanny.meunier@unice.fr, ndocarmo@unice.fr

Theta activity and phase resetting during perception of French homophonous utterances

Speech perception involves segmenting a continuous stream of speech into its word components. This can be challenging in the case of homophonous utterances only differing in non-contrastive subphonemic features. Yet, the speech perception system seems able to discriminate subphonemic deviation in homophonous utterances, since it has been shown to elicit a mismatch response (MMN). Here, we focused on the oscillatory correlates, namely phase resetting and power, of non-contrastive subphonemic deviation processing in language. An oddball task that considered natural intraspeaker variability was used.

Subphonemic deviance elicited intertrial phase coherence (ITC) differences in the theta band at Fz during the time window of the MMN. No differences in power were found. This suggests that the processing of subphonemic deviation in speech signals, reflected by the MMN, might rely on mechanisms of phase resetting. ITC might facilitate the synchronous firing of functional networks involved in the processing of subphonemic deviance.

Keywords: homophone; time-frequency; theta activity; ITC; oddball

Introduction

Understanding speech requires detecting acoustic characteristics of speech sounds and matching them to the phonemic representations stored in memory, i.e. encoding and mapping (Mcqueen, 2005). Thus, a complex stream of speech signal has to be segmented into meaningful units to achieve speech perception. This process can be particularly challenging in the case of homophonous sequences in natural speech, where no clear word boundaries are present. Homophonous utterances do not show phonemic differences. Such utterances are

common in spoken language (such as *two lips* and *tulips*) and in particular in French as a consequence of the liaison phenomenon. The sequence [lafiʃ], for example, depending on the segmentation, can be understood as *la fiche* ‘the card’ or *l’affiche* ‘the poster’ and the sequence [lami] as *l’amie* ‘the friend’ or *la mie* ‘the crumb’. The only possible features to access to distinct lexical categories are subphonemic differences, such as changes in pitch and vowel duration. Listeners are able to distinguish between nominal homophonous sequences with semantical context without much ambiguity. Yet, performance drops to 75 % of correct discrimination when homophonous utterances are presented in isolation (Quené, 1992) and to 66% when using multiple tokens of the same linguistic unit, i.e. with intra and interspeaker variability (Spinelli et al., 2007). Indeed, small changes in articulation patterns can elicit relevant differences at the acoustic level (Junqua & Haton, 1996).

The event-related potential (ERP) mismatch negativity (MMN) has proven revealing to investigate speech processing, and particularly to study the phonological mapping. This component can be elicited by unexpected changes in a regular stream of auditory stimuli (Näätänen et al., 2007), for instance by differences in pitch, intensity, or duration, and thus it allows studying discrimination between speech sounds (Bishop, 2007). Moreover, the MMN can occur even without focused attention on the stimuli. All this makes it a suitable tool to understand how language comprehension occurs. This component has been explained as an “error response” within a predictive coding mechanism (Garrido et al., 2009). In such model, perception is based on inference, as the brain would anticipate future events. The MMN would reflect a prediction error: a mismatch between expectations and current sensory input. Since

the discovery of the MMN, a myriad of studies has focused on the conditions under which the component is elicited. When it comes to homophone sequences, it has been shown that such utterances differing in non-contrastive features can elicit an MMN component (Do Carmo-Blanco et al., 2019). This was shown using a variant of the oddball paradigm that used multiple productions of the same linguistic unit (i.e. the standard stimuli were different tokens from the same speaker and the deviant stimuli either a homophonous token or a phonemically different token). As expected, the MMN component was greater for a phonemically different deviant than for a homophonous test stimulus differing in subphonemic features.

ERPs only reveal time-locked neural activity. Time-frequency analysis shows other dimensions of processing, including non-time-locked activity, and provides information about speech processing at multiple frequency bands. In particular, the auditory MMN component has been related to changes in theta activity (Ko et al., 2012; Koerner et al., 2016; Yordanova & Kolev, 1998). In the present article, we aimed at investigating oscillatory correlates of subphonemic deviance in homophonous sequences, as indexed by the MMN. This study can contribute to improving our understanding of the dynamics of language processing and speech perception. At the same time, it can further our understanding of the mechanisms underlying the MMN component.

Although it has been posited that theta activity underlies the MMN response (Fuentemilla et al., 2008; Herrmann et al., 2014), the oscillatory correlates of speech sound deviation are not completely elucidated. An increase in theta band seems to be elicited by deviant but also by standard stimuli in an oddball paradigm. Apart from changes in theta

power, some studies have shown greater Inter-Trial Coherence (ITC) elicited by deviant stimuli (Fuentemilla et al., 2008; Hsiao et al., 2009; Ko et al., 2012). ITC, also known as inter-trial phase coherence (ITPC), represents time-locked oscillatory activity. It measures phase synchrony or degree of phase-locking between single trials in a particular frequency band, i.e. phase consistency over trials (Tallon-Baudry et al., 1996). It is defined within a single electrode and it ranges between 0 and 1 (Cohen, 2014), with 1 indicating perfect synchrony between trials. The few previous oddball studies that addressed the oscillatory correlates of the MMN used simple tones differing either in duration (Fuentemilla et al., 2008, Hsiao et al., 2009) or frequency (Ko et al., 2012). In one of the pioneering time-frequency studies, Fuentemilla et al. (2008) used an oddball paradigm where the deviant tone differed in duration. They found differences in ITC between standard and deviant tones responses associated with the temporal source of the MMN, as measured in the mastoid electrodes, whereas not in power. Yet, they were not able to show that such a reset mechanism was involved in the frontal component source of the MMN (see Giard et al., 1990 for a detailed description of each subcomponent). They concluded that each subcomponent could have different generation mechanisms. By also using an oddball paradigm, Hsiao et al. (2009) showed differences in both power and ITC in auditory change detection of tone duration deviance in temporal regions. Differences in ITC were also found in frontal regions, which were not accompanied by differences in power. Yet, such stimuli contrast with the complexity of speech sounds, which contain multiple harmonics.

One of the theories of speech perception, the oscillatory model (Poeppel et al., 2007), posits that neural oscillations are involved in speech processing. On this basis, it has been argued that different oscillatory responses not only reflect sound processing but also linguistic processing. Pérez et al. (2017) recorded the EEG activity of pairs of participants, speaker and listener, engaged in oral narratives. They found an inter-brain coupling (i.e. entrainment) of neural oscillations. Thus, the pattern of brain activity synchronizes between speaker and listener. Apart from cortical coupling to the speech envelope, they found that some of the synchronization patterns seemed to occur without being mediated by the physical properties of speech. Kösem & van Wassenhove (2017) suggested that phase entrainment (i.e. neural oscillatory activity phase alignment with oscillatory signals such as speech) could be the mechanism supporting segmentation and speech processing. Indeed, theta frequency corresponds to syllable rate in a continuous speech stream. Yet, Kösem & van Wassenhove (2017) also suggested that theta entrainment to the speech envelope could be involved in the processing of phonetic features and phonological information. Such entrainment mechanism seems necessary, albeit not sufficient, to the processing of intelligible speech and speech comprehension, as certainly, no study has shown otherwise. Neural oscillations have also been proposed to support speech processing by phase alignment during phoneme onset or offset (Meyer, 2018). As to theta ITC, it has also been linked to an attentional shift given relevant auditory information, such as one’s name (Tamura et al., 2015).

Whereas some studies have investigated ERP correlates of homophone perception and deviation in different languages, such as English (Conwell, 2015) or French (Aguilera et al.,

2014; Brunellière et al., 2009), no study has investigated their oscillatory correlates. The characteristics of brain oscillations, including power and phase, have been described during detection of tone deviance. Interestingly, it has been proposed that theta band ITC might be more sensitive to deviance of simple tones than power and typical ERPs. Theta band ITC changes have been found in subjects who did not show an MMN response in the time domain (Bishop & Hardiman, 2010).

To the best of our knowledge, only one study has reported oscillatory correlates of pitch deviance in speech with a passive oddball task. Li & Chen (2018) used a single syllable word whose meaning is dependent on the pitch contour in Chinese. They found an increase in theta power associated with unattended processing of lexical pitch contour variations in speech. It is to note that pitch contour is lexically contrastive in Chinese. This study, like most of MMN studies, used one single production of the standard stimulus, which contrasts with the variability found in natural speech production. Each production is subjected to variations in timing and vocal tract movements (Junqua & Haton, 1996), which make every production of the same unit different. Despite differences in every production, we identified them as the same unit. To date, the oscillatory signature of an oddball task has never been investigated when multiples tokens of the same speech sequence have been used as standard stimuli.

In this paper, we asked whether theta activity reflects subphonemic deviance in homophonous sequences (non- contrastive) as reflected in the MMN. We also aimed to contribute to disentangling the role of theta power and ITC in the discrimination of speech

sounds. These questions were studied by using stimuli taken from natural speech. To approach ecological conditions, we used multiple productions of the standard and test stimuli.

We used the first syllable of the French homophonous nominal utterances, *la locutio*n [la#lɔkysjɔ̃] vs. *l'allocution* [l#alɔkysjɔ̃]. These syllables differ in non-contrastive subphonemic features, such as pitch and duration. Importantly, sequence duration was kept constant, since most of the previous studies show theta effects elicited by stimulus duration differences. A phonemically different syllable was used as the control stimulus ([l#a], [la#], [l#i]). Theta band response has been posited to reflect the MMN (see review of studies using time-frequency analysis for MMN paradigms by Herrmann et al., 2014). Given that theta power was found sensitive to pitch contrasts in tonal languages such as Chinese without focused attention, one might expect a similar effect of subphonemic deviance. As part of the prosodic information, suprasegmental information can convey meaning. Such features might be relevant and processed at the low-level acoustic brain areas without focused attention. Moreover, an MMN component has been evoked by a subphonemic mismatch. Although phase reset is not the only mechanism in the generation of ERPs (Klimesch et al., 2007), it has been shown that phase resetting of ongoing oscillations might contribute to the generation of ERPs (Makeig, 2002). As a result, differences in ITC were also expected.

Method

Participants

18 native monolingual French healthy participants (10 females, M = 22 years, SD = 3) gave

informed consent to participate in this study. They all reported normal hearing. None of them reported having history of language or neurological disorder. The experimental protocol was approved by the Ethical Committee of the Grenoble Hospital (CHU of Grenoble, France; ID RCB: 2012-A01653-40).

Stimuli and procedure

The initial syllables la#, l#a, and l#i were excised from the homophonous utterances *La locution* ([lalokysiɔ̃], 'the locution'), *L'allocution* ([lalokysiɔ̃], 'the speech'), and from *L'illocution* ([lilokysiɔ̃], 'the illocution') from recorded sentences that were read by a native French female speaker who was unaware of the experimental purpose. The duration of the syllables was equalized to 140 ms without corrupting the signal, i.e. just by excising the syllable waveforms at the same length. Extractions were performed at the closest zero-crossing point in the acoustic signal. In order to investigate natural speech variation processing, for each syllable (i.e. la#, l#a and l#i) five different tokens coming from five different recordings were used.

The experimental design consisted in a modified version of the oddball paradigm where trains of four different standards (i.e. la# or l#a) were followed by a test stimulus, which could be another token of the standard stimulus (Identical condition), a homophone syllable, i.e. one of the five recordings (Homophonous condition) or the phonemically different syllable l#i, i.e. one of the five recordings (Dissimilar condition). The homophonous or the dissimilar test stimulus was unpredictable, as it could appear after 4, 9 or 14 standard stimuli. The

stimulus display was pseudorandomized and a 500 ms interstimulus interval was set. Two blocks of 1800 stimuli were presented in the experiment, which lasted 45 minutes, that including a break between blocks. The syllables la# and l#a were used as standards in each block, respectively, in random order. Table 1 shows the acoustic characteristics of the stimuli used in the experiment, as measured with the software Praat. Mean pitch in the content-word initial vowel ([a] in *l'allocution*) was higher than in the article vowel ([a] in *la locution*), as assessed by a t-test (mean difference of + 33.54 Hz, $p < 0.001$). Similarly, F_1 was also higher in the [a] in *l'allocution* (+ 109 Hz, $p < 0.001$). The waveform and spectrogram of one token of each condition can be found in Do Carmo-Blanco et al. (2019).

Participants watched a silent movie of their choice during the experiment and were instructed to ignore the auditory stimuli over the headphone. Sound stimuli were delivered binaurally at 65dB SPL. For detailed information about the experimental procedure, see Do Carmo-Blanco et al. (2019).

EEG recording and analysis

EEG was collected with a 32 electrodes Biosemi system at 2 kHz and filtered between 0.1 and 400 Hz. Electrode offset was kept under 20 mV. (i.e. the average of voltage at each electrode relative to the common mode or average potential of the subject determined by the CMS, or the common mode sense, active electrode, and the passive DRL electrode, or driven right leg, which is characteristic of the Biosemi system). The electrode offset measures how far-off from the common mode an electrode is. Data were visually inspected for artifacts. Bad channels

were splined interpolated and artifactual segments removed before ICA decomposition. Independent components representing blinks and lateral eye movements were subtracted from the signal. The recording of 32 electrodes EEG allowed representing the topography of the MMN component. The expected MMN response elicited by the stimuli (waveform and topography) can be found in Do Carmo-Blanco et al. (2019).

Time-frequency analysis

The continuous EEG data were epoched from -1500 to 2000 ms around stimulus onset to avoid edge artifacts during time-frequency decomposition and later truncated. The signal was bandpass filtered (0.3-180 Hz) and digitized at a sampling rate of 256 Hz. All the offline processing and analysis were performed with the EEGLAB toolbox (Delorme & Makeig, 2004) and custom routines.

For each EEG epoch and each frequency, power (i.e. square amplitude) was calculated by complex Morlet wavelet-based spectral decomposition, by using 25 frequencies logarithmically spaced from 2 to 40 Hz. The number of cycles in the wavelet was increasing from 3 to 10 according to frequency. Stimulus-related changes in power were computed as dB change with respect to a 200 ms baseline between 250 ms and 50 ms before stimulus onset. For each subject, changes in power with respect to the baseline period were averaged for each of the 3 conditions (Standard, Homophone and Dissimilar conditions).

Phase coherence across trials was measured through ITC, which reflects the extent to which oscillation phase values in a particular frequency and time point are consistent over trials, according to the following formula:

$$ITC(t,f) = 1/n \sum_{k=1}^{n} e^{i\phi^{k}(t,\,f)}$$

Where ITC indexes the average phase angle across all epochs, t and f depict each time point and frequency, the complex number i, and ϕ^k the phase angle in radians (for details on the code see Cohen, 2014). As ITC is sensitive to the number of trials, the number of trials was matched across conditions (i.e. a total of 110 epochs randomly chosen for each condition for each block per participant). Those same trials were used to compute power. Time-points and frequencies were the same as those used to compute power. Data from the two blocks were collapsed. As such, the homophonous deviants were pooled together whether they were *l'a* or *la* (first or second block depending on the participant), thus avoiding any confounding effect due to physical differences. Epochs were separated for each condition and participant.

Statistical analysis

The statistical analysis sought to assess whether significant differences in power and ITC at Fz could be found between conditions (i.e. Identical vs Homophonous, Identical vs Dissimilar and Homophonous vs Dissimilar). To solve the problem of multiple comparisons in multidimensional EEG data, such differences in mean power and ITC were assessed with cluster permutation tests (Maris & Oostenveld, 2007). One of the advantages of this method is the

possibility to use any test statistic. The null hypothesis we are testing is that the probability distributions of the EEG data in the different conditions are equal.

Data from each condition were used to simulate a null hypothesis test through 2000 permutations of conditions. In this procedure, data from each condition are pooled together and shuffled. Data are then randomly assigned to conditions, which is used to compute a test statistic (here a two-sided *t-test*) following each permutation. Adjacent time-frequency bins with a test statistic value over a threshold are grouped together and the sum of the values in each cluster is computed. This allows drawing a histogram of the test statistics, here of the largest significant clusters for each permutation under the random partition. Thus, this method considers the size of the clusters to create the null distribution that is used to determine the significance threshold value. In our case, at each permutation, the sum of the t-values in the largest cluster was stored. These values are compared to those observed when comparing conditions in the actual data. The clusters in the experimental data with a value over the significance threshold value determined by permutation (95-percentile value) remain significant. The size of the clusters was measured with the Matlab function bwconncomp, with a connectivity value of 8 on the basis of temporal and spectral adjacency. The *t* threshold value was set to 2.11 (the 99.5 quantile of a *t*-distribution, corresponding to a 2-sided t-test with 17 degrees of freedom at $p = 0.05$). The *p*-value cut-off was set to $p = 0.05$ to achieve a good trade-off between false positives and false negatives. A lower p-value can prevent from including points in a cluster where the null hypothesis is true and reduce sensitivity for smaller clusters.

It is to note that cluster-based permutation tests are not meant to give a precise temporal, spatial or frequency band estimation (Sassenhagen & Draschkow, 2019). Moreover, the precise timing of the effect also depends on the filter used. Power and ITC were measured at Fz, where the MMN component reached its maximum. Fz is the electrode most commonly used for the study of this component (Näätänen et al., 2007). The aim was to relate the time-frequency results to the MMN results reported in Do Carmo-Blanco et al. (2019).

Cluster-based permutations has become one of the most popular approaches to deal with the multiple comparison correction problem in multidimensional EEG data. Yet, since the significance level threshold was set a priori to $p < .05$, it did not provide p-values. Thus, for each significant cluster determined by cluster-based permutations, we performed a two-way repeated-measures analysis of variance (rANOVA), with factors Frequency (if more than one frequency range) and Condition (Identical, Homophonous, and Dissimilar), using the Statistical Package for the Social Science (SPSS). Clusters which showed a main effect of condition by rANOVA (α level $< .05$) were submitted to subsequent 2-tailed *t* test for post hoc comparisons. We also applied the Greenhouse-Geisser correction for *p* values when the sphericity assumption was not met. The corrected *p* value was reported with the uncorrected degrees of freedom (Vasey & Thayer, 1987). Partial eta squared (η_p^2) was the measure of effect size reported with ANOVA.

Results

Power

In all conditions, stimuli elicited an increase in power with respect to the baseline period in

fronto-central regions, which was greatest at 6 Hz between 200 and 300 ms approximately. Figure 1 shows power at Fz for the 3 conditions, time-locked to stimulus onset between 4 and 40 Hz. The figure shows an increase in theta power with respect to baseline in the three conditions. Theta reaches its maximum between 200 and 400 ms after stimulus onset, with maximum values of 0.56 dB of change with respect to the baseline in the Identical condition, and 0.84 and 1.28 dB in the Homophonous and Dissimilar condition. Non-parametrical Cluster-based permutation testing on power did not yield any significant cluster ($p > 0.05$ for a two-sided *t*-test) at the time-window of interest (from -100 to 600 ms after stimulus onset) for any of the 3 pairwise comparisons (i.e. Identical vs Homophonous, Identical vs Dissimilar and Homophonous vs Dissimilar).

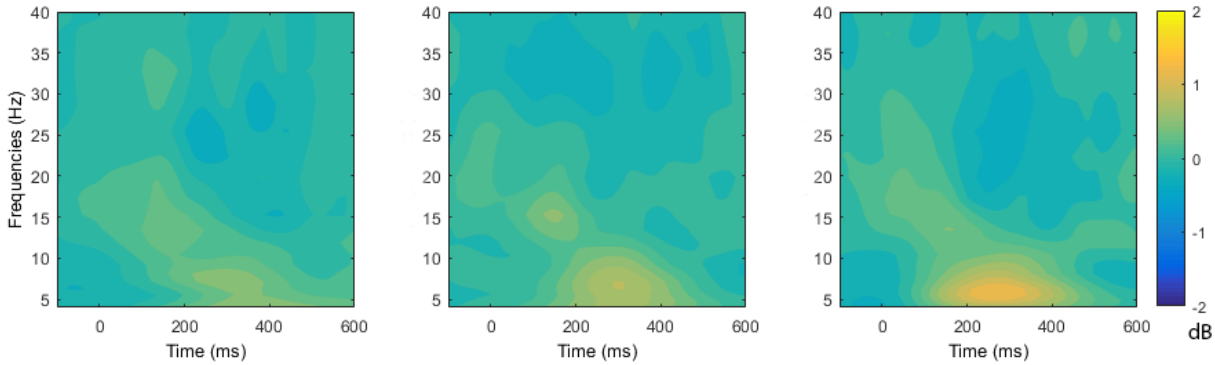

Figure 1. Power at Fz. Changes in power with respect to baseline across time and frequencies for the Standard (left panel), Homophone (middle panel) and Dissimilar (right panel) conditions. No significant differences were found between conditions.

ITC

Figure 2A shows ITC at Fz for the 3 conditions during the time-window between -100 and 600 ms after stimulus onset. ITC was highest in the theta band. The Dissimilar condition showed the highest phase coherence values in the theta band, reaching an ITC of 0.32 around 200 ms après stimulus onset. Theta ITC reached its maximum of 0.29 between 275 et 300 ms in the Homophonous condition and around 300 ms in the Identical condition, with a value of 0.26. Figure 2B depicts significant clusters after cluster-based permutation correction for multiple comparisons. The dark blue blobs depict significant clusters ($p < 0.05$). Non-significant differences in means are masked in light blue. Mean ITC was lower in the Identical than in the Homophonous condition for a cluster in the low theta range between 75 and 200 ms after stimulus onset. It is to remember that time-frequency analyses lack the temporal resolution of ERPs (Sassenhagen & Draschkow, 2019). When comparing to the deviant Dissimilar condition, significant differences expanded the whole theta band (from 4 to 8 Hz). ITC was higher for the Homophonous condition than for the Identical condition for the low-theta band between 75 and 200 ms. Differences in ITC between the Dissimilar and the other two conditions expanded the whole theta band. ITC was higher for the Dissimilar condition between 75 and 200 ms.

We run a two-way rANOVA on theta ITC between 100 and 175 ms after stimulus onset. This time-window was chosen considering that time-frequency analysis does not have

high temporal resolution. rANOVA with factors Frequency (low and mid-range theta) and Condition (Identical, Homophonous and Dissimilar) showed a main effect of Condition ($F(2,34) = 10.27$, $p = .004$, $\eta_p^2 = .377$). ITC was lower in the Identical condition (mean ITC in the time-frequency bins was 0.143). ITC reached 0.189 in the Homophonous condition and 0.264 in the Dissimilar condition. All pairwise comparisons were significant (Identical vs Homophonous, p =.001, Identical vs Dissimilar p =.003, and Homophonous vs Dissimilar p =.017). The main effect of Frequency ($F(1,17) = 7.25$, $p = .015$, $\eta_p^2 = .299$) was also significant, indicating that ITC was lower in low than in mid-range theta (0.187 vs 0.211). Finally, the interaction was also significant ($F(2,34) = 4.16$, $p = .027$, $\eta_p^2 = .197$).

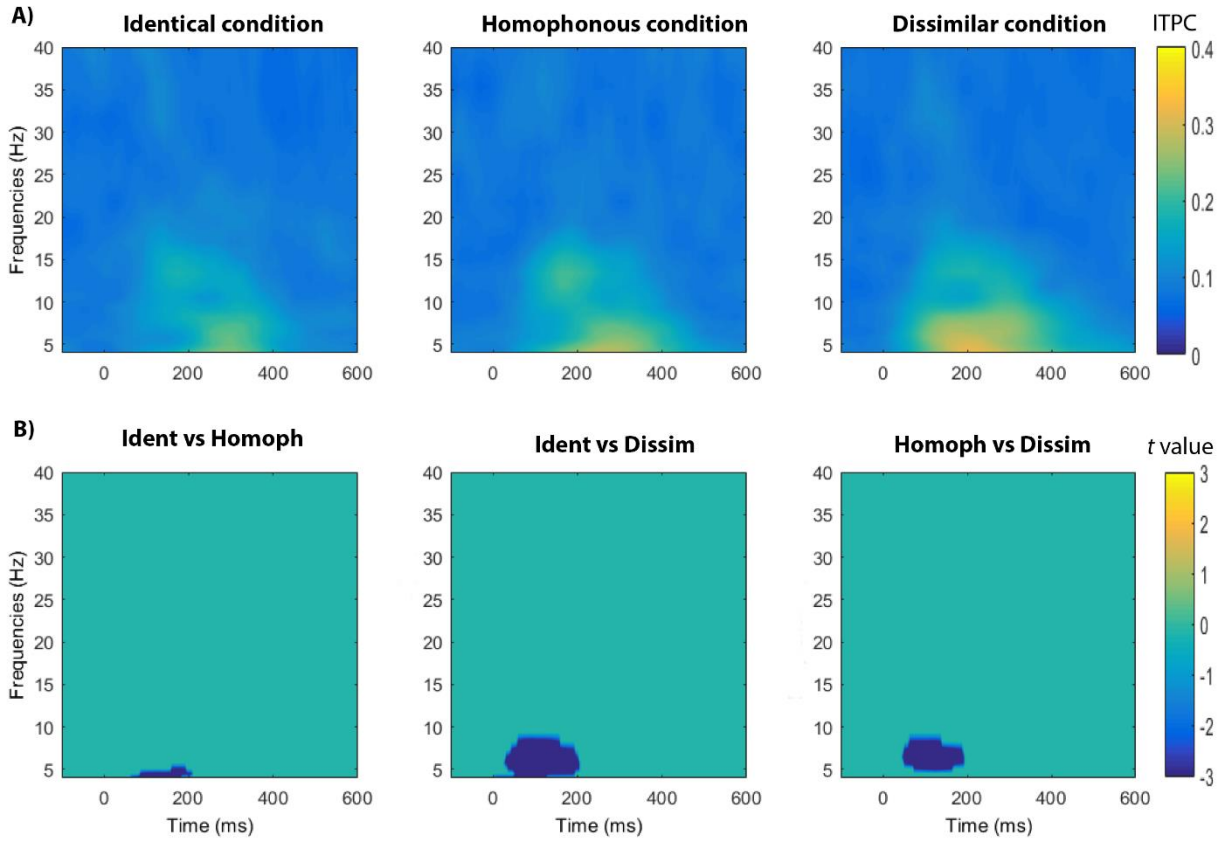

Figure 2. ITC at Fz (A) and statistical analysis (B). (A) ITC for the 3 conditions across time and frequencies. ITC increases in the theta band in each of the conditions. The peak of this increase was reached earlier in the Dissimilar condition than in the other two conditions, and it was latest for the identical condition. (B) Cluster-corrected t values. ITC was higher in the Homophonous condition than in the Identical condition in the low theta band between 75 and 200 ms (bottom left panel). ITC was higher in the Dissimilar condition than in the Identical condition, spanning the whole theta frequency band between 75 and 200 ms (bottom middle panel). Cluster-based permutation tests also revealed a significant cluster in the theta band when comparing the Dissimilar to the Homophonous condition. ITCP was higher in the Dissimilar condition.

Discussion

The MMN component has been crucial to the study of phoneme perception and discrimination and also non-phonemically contrastive acoustic features. The aim of this study was to further elucidate the oscillatory correlates of subphonemic deviation detection in homophonous sequences, while also investigating the relation between the MMN and theta activity. For this reason, we focused our analysis on Fz since the MMN elicited by the stimuli of this experiment reached its maximum at this electrode (Do Carmo-Blanco et al., 2019). We were interested in the contribution of oscillatory components of EEG, namely power and ITC. ITC is a measure of cortical synchrony that, like power, is not reflected in the grand average ERP (Makeig et al., 2004). Such oscillatory responses have mostly been characterized in the processing of deviation of simple tones.

Although some articles have investigated the underlying oscillatory mechanisms of sound deviation processing, very few studies have used speech signals, which are far more complex and contain multiple harmonics. Here, stimuli were taken from natural linguistic utterances, and more precisely from multiple productions of the same speaker. Thus, intraspeaker variability was taken into account. French homophone syllables differing in non-contrastive acoustic features in French, such as pitch and F_1, were used as stimuli in an oddball task. Unlike some previous studies where the deviant tone differed in duration (Fuentemilla et al., 2008; Hsiao et al., 2009), stimulus duration was kept constant, and that without corrupting the signal. The results did not show differences in theta power between any of the conditions (Identical, Homophonous and Dissimilar). However, we could find differences in ITC between

each pairwise comparison. Theta ITC results mirror the MMN findings reported in Do Carmo-Blanco et al. (2019) in the same task. As further described below, the homophonous deviant did elicit an MMN component.

Our study was not able to show differences in theta power between the homophonous and the identical condition, at the time and location of the MMN component. Power might not explain such fine-grained deviation detection. Although this contradicts Hsiao et al. (2009) results' in tone duration deviation, where the MMN was accompanied by increased power, it could be mitigated by the fact that their MMN peaked at temporal regions. In frontal regions power was not increased. Differences could also be explained by differences in the stimuli used. The standard stimuli used in our tasks were complex syllables coming from different productions of the same speaker, thus differing slightly between each other. This contrasts with a simple single tone. As a result, the processing of linguistic units' deviation might be different.

In contrast to the above-mentioned studies on the processing of pure tones, Li & Chen (2018) used speech stimuli. They found an increase in theta power associated with the processing of semantical pitch contour variations. The difference in results between theirs and this study could be explained by the contrastive character of pitch in Chinese. The syllable-level pitch variations have lexical implications, whereas the subphonemic cues in the homophonous sequences studied here are not semantically contrastive in French although necessary to access the intended meaning via the appropriate segmentation. Their study did not report ITC values to compare with.

We were not able to show differences in power between the identical and the dissimilar condition. Theta oscillations have been related to speech processing, particularly to sound analysis and phonological processing (see Kösem & van Wassenhove, 2017 for a review), and also to cognitive effort (Bosseler et al., 2013). Accordingly, theta power is higher when elicited by non-native than by native phonetic categories in an oddball paradigm. Since this effect is not present in young infants, it has been explained as an increased effort, which would be the result of the decline in the processing of non-native sounds in adults. In our study, power was not sensitive to phonemic deviation, either (la# ou l#a vs li#), at the time and location of the MMN component. Although Kösem & van Wassenhove (2017) suggest that theta oscillations are sensitive to phonetic features, they describe stages such as the identification of consonants in syllables. The stimuli here did not show differences in consonants. Our results do not contradict Bosseler et al. (2013), either. Auditory stimuli in our task were not attended to. Power could index cognitive effort given focused attention. To better understand the role of theta power, it would be needed to compare the results of this experiment with data from attended stimuli in the same MMN task.

Our results showed significant differences in ITC values between each condition, where the homophonous condition exhibited higher phase coherence values than the identical condition. This suggests that the processing of subphonemic deviation in speech signals relies on mechanisms of phase resetting. At the same time, our results also support that theta ITC may be the preceding oscillatory mechanism related to the MMN component. It is relevant to note that the homophonous MMN peaked at 265 ms after stimulus onset (Do Carmo-Blanco et

al., 2019), whereas the ITC effect spanned from 75 to 200 ms. Thus, ITC effects lasted longer and preceded the MMN. As phase resetting is linked to temporal synchrony of populations of neurons, ITC could have contributed to the MMN. An increase in ITC could be a requisite to elicit the mismatch response to subphonemic deviation. This result is in accordance with the proposal that ITC could be as good predictor of deviation as the MMN (Bishop & Hardiman, 2010).

It is to note that the phase of high-frequency oscillations (beta and gamma band) has been related to word parsing of monosyllabic words in ambiguous utterances (Kösem et al., 2016), thus suggesting a conscious representation of speech. Low-frequency oscillations have been linked to pre-lexical segmentation. ITC could contribute to the processing of fine-grained stress patterns that are involved in speech segmentation. This hypothesis could be tested in future studies by comparing the neural response to stress patterns related to segmentation and related to word meaning.

The results from the ERP and time-frequency analysis in this study support that subphonemic characteristics of speech are processed in a bottom-up fashion, independent of lexical and semantic context and without focused attention. The phonemic representation of a speech sound has to be flexible enough to include subphonemic categories, within the range of intraspeaker variability. Thus, the processing of subphonemic differences has to tolerate small natural speech fluctuations, such as those present in different productions from the same speaker in the same contextual conditions. Such flexible categorical boundaries would contribute to the understanding of productions coming from individuals with speech

production disorders and diverse accents. Indeed, the phonetic representation of a speech sound from someone suffering from dysarthria or having an accent could be notably distant (i.e. acoustic distance in the formant space) to the phonetic representation from a healthy native speaker.

One of the proposed features to account for the ITC results in homophone deviation (i.e. non-contrastive subphonemic deviation) is pitch. Indeed, acoustic analyses showed differences in pitch between the identical and the homophonous conditions (see Do Carmo-Blanco et al., 2019). Pitch discrimination is not only relevant in tonal languages, but also for linguistic processing, such as prosodic information, and for music processing. A growing body of literature suggests a link between pitch processing and theta power. For instance, successful detection of pitch changes has been related to theta power in Chinese without attention (Li & Chen, 2018). Pure tone discrimination can be predicted from theta amplitude (Florin et al., 2017). Regarding other factors that could affect the theta response, Ko et al. (2012) showed that probability does not account for the results. Our results also support that pitch deviance might be an important factor in ITC response in the theta band in an oddball paradigm. Moreover, we were able to confirm that differences in duration are not necessary to elicit such ITC effect.

Taken together, we showed that ITC is linked to the processing of non-contrastive fine acoustic deviations that support word segmentation. In this line, low-frequency oscillations have also been related to parsing and speech segmentation (Ding et al., 2016; Kösem et al., 2016), and particularly theta phase information (Ghitza et al., 2013). The oscillatory frequency

information allows neurons to fire synchronously thus being key for the timing of neural activity. Synchronous repetitive firing of neurons has been linked to the coupling of functional networks (Varela et al., 2001). Indeed, it facilitates the activation of functional networks by increasing the likelihood that neurons entrain one another (Bastiaansen & Hagoort, 2006). This preliminary study adds more evidence to the contribution of synchronous neural activity in the theta band to acoustic and phonological processing. Our results also support that the MMN component could arise through reorganization of the phase of ongoing brain activity. Since no differences in power were elicited by homophonous deviation, it is possible to posit that a phase resetting mechanism occurred (Cohen, 2014). A mechanism of phase resetting of the low theta oscillations might underlie the MMN response and the processing of subphonemic deviation.

Although the precise contribution of the different acoustic features to the results was not investigated, this work is a foundation for future speech studies in natural settings. Apart from the contribution of these features, such as pitch or F_1, future research should consider inter-speaker variability. This would contribute to understanding the processing of language in use. The design of such experimental procedure might be challenging. It would have to consider that the acoustic distance between the same productions coming from different speakers can be as high as the acoustical distances between homophone phonemes from the same speaker. Regardless of the precise feature that could account for the results, the present experiment further our understanding of speech sounds discrimination and the neural

representation of subphonemic categories in French language. Here, it is shown that French listeners are not “deaf” to accentuation.

No differences were found in power elicited by subphonemic deviance. It could be argued that this study did not have enough power to reveal such an effect. The lack of significant effects could also be due to higher inter-individual variability in theta power. Finally, in order to draw strong conclusions about the role of power, future research should compare changes in power under attentive and pre-attentive processing of the same stimuli. Since speech perception is modulated by prior expectations, the neural processing might depend on the level of attention. This is in accordance with the studies showing that theta is modulated by attentional control of behaviour (e.g. see Başar et al., 2001; Orekhova et al., 2006) As so, frontal theta power has been linked to cognitive control.

The findings presented in this study support that theta activity is associated with non-contrastive speech deviation processing. ITC in the theta band might promote information processing by, for instance, facilitating long-scale coupling. Subphonemic deviation detection in MMN is reflected in theta ITC.

Our results support that speech change detection reflects the prediction error between the comparison of an auditory speech input and a previous memory trace. The prediction error is not only reflected on the MMN component, but also in theta phase resetting. The measure of the prediction error takes into account natural intra-speaker variability, which is in accordance with the use of previous knowledge and the integration of top-down prediction in speech perception (Sohoglu et al., 2012). If prior knowledge and expectations are influenced by the

level of attention of the subject, the underlying processing to achieve speech perception might also depend on the attentional state of the subject. Here we bring more evidence on the role of phase resetting on the processing of natural speech deviation detection.

Conclusion

This study provides additional electrophysiological data suggesting that theta phase resetting is involved in MMN generation and can be as good predictor of deviation detection as the ERP component. More specifically, we showed that ITC supports the processing of subphonemic deviation in speech even in the context of intraspeaker variability. Speech perception combines prior knowledge, such as subtle variations coming from different productions of the same speech unit, and expectations with sensory information to achieve discrimination of speech sounds even in the absence of focused attention.

Acknowledgments: We acknowledge Stéphane Pota for data collection.

Funding sources: This work was supported by a European Research Council grant to the SpiN Project (No. 209234).

Disclosure of interest: The authors report no conflict of interest.

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 2014*, 383–387.

Başar, E. ro., Schürmann, M. arti., & Sakowitz, O. live. (2001). The selectively distributed theta system: functions. *International Journal of Psychophysiology*, *39*(2–3), 197–212. https://doi.org/10.1016/S0167-8760(00)00141-0

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, *159*(06), 179–196. https://doi.org/10.1016/S0079-6123(06)59012-0

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

Bishop, D. V. M., & Hardiman, M. J. (2010). Measurement of mismatch negativity in individuals: A study using single-trial analysis. *Psychophysiology*, *47*(4), 697–705. https://doi.org/10.1111/j.1469-8986.2009.00970.x

Bosseler, A. N., Taulu, S., Pihko, E., Mäkelä, J. P., Imada, T., Ahonen, A., & Kuhl, P. K. (2013). Theta brain rhythms index perceptual narrowing in infant speech perception. *Frontiers in Psychology*, *4*(October), 1–12. https://doi.org/10.3389/fpsyg.2013.00690

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

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press.

Conwell, E. (2015). Neural responses to category ambiguous words. *Neuropsychologia*, *69*, 85–92. https://doi.org/10.1016/j.neuropsychologia.2015.01.036

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

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158.

Do Carmo-Blanco, N., Hoen, M., Pota, S., Spinelli, E., & Meunier, F. (2019). Processing of non-contrastive subphonemic features in French homophonous utterances: An MMN study. *Journal of Neurolinguistics*, *52*. https://doi.org/10.1016/j.jneuroling.2019.05.001

Florin, E., Vuvan, D., Peretz, I., & Baillet, S. (2017). Pre-target neural oscillations predict variability in the detection of small pitch changes. *PLoS ONE*, *12*(5), 1–19. https://doi.org/10.1371/journal.pone.0177836

Fuentemilla, L., Marco-Pallarés, J., Münte, T. F., & Grau, C. (2008). Theta EEG oscillatory activity and auditory change detection. *Brain Research*, *1220*, 93–101. https://doi.org/10.1016/j.brainres.2007.07.079

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Ghitza, O., Giraud, A.-L., & Poeppel, D. (2013). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience*, *6*, 340.

Giard, M., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology*, *27*(6), 627–640. https://doi.org/10.1080/23273798.2016.1238495

Herrmann, C. S., Rach, S., Vosskuhl, J., & Strüber, D. (2014a). Time-frequency analysis of event-related potentials: A brief tutorial. *Brain Topography*, *27*(4), 438–450. https://doi.org/10.1007/s10548-013-0327-5

Herrmann, C. S., Rach, S., Vosskuhl, J., & Strüber, D. (2014b). Time–frequency analysis of event-related potentials: a brief tutorial. *Brain Topography*, *27*(4), 438–450.

Hsiao, F. J., Wu, Z. A., Ho, L. T., & Lin, Y. Y. (2009). Theta oscillation during auditory change detection: An MEG study. *Biological Psychology*, *81*(1), 58–66. https://doi.org/10.1016/j.biopsycho.2009.01.007

Junqua, J.-C., & Haton, J.-P. (1996). *Speaker Variability and Specificity* (pp. 127–153). https://doi.org/10.1007/978-1-4613-1297-0_4

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews*, *31*(7), 1003–1016. https://doi.org/10.1016/j.neubiorev.2007.03.005

Ko, D., Kwon, S., Lee, G.-T., Im, C. H., Kim, K. H., & Jung, K.-Y. (2012). Theta Oscillation Related to the Auditory Discrimination Process in Mismatch Negativity: Oddball versus Control Paradigm. *Journal of Clinical Neurology*, *8*(1), 35. https://doi.org/10.3988/jcn.2012.8.1.35

Koerner, T. K., Zhang, Y., Nelson, P. B., Wang, B., & Zou, H. (2016). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A mismatch negativity study. *Hearing Research*, *339*, 40–49.

Kösem, A., Basirat, A., Azizi, L., & Van Wassenhove, V. (2016). High-frequency neural activity predicts word parsing in ambiguous speech streams. *Journal of Neurophysiology*, *116*(6), 2497–2512.

Kösem, A., & van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition and Neuroscience*, *32*(5), 536–544. https://doi.org/10.1080/23273798.2016.1238495

Li, X., & Chen, Y. (2018). Unattended processing of hierarchical pitch variations in spoken sentences. *Brain and Language*, *183*(16), 21–31. https://doi.org/10.1016/j.bandl.2018.05.004

Makeig, S. (2002). Response: Event-related brain dynamics – unifying brain electrophysiology. *Trends in Neurosciences*, *25*(8), 390. https://doi.org/10.1016/s0166-2236(02)02198-7

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, *8*(5), 204–210. https://doi.org/10.1016/j.tics.2004.03.008

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Mcqueen, J. M. (2005). Speech Perception. In *Handbook of Cognition* (pp. 256–276). SAGE Publications Ltd. https://doi.org/10.4135/9781848608177.n11

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621. https://doi.org/10.1111/ejn.13748

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

Orekhova, E. V, Stroganova, T. A., Posikera, I. N., & Elam, M. (2006). EEG theta rhythm in infants and preschool children. *Clinical Neurophysiology*, *117*(5), 1047–1062.

Pérez, A., Carreiras, M., & Duñabeitia, J. A. (2017). Brain-To-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-04464-4

Poeppel, D., Idsardi, W. J., & Van Wassenhove, V. (2007). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1071–1086.

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

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

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

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *16*(13), 4240–4249. https://doi.org/10.1016/j.neuropsychologia.2011.02.038

Tamura, K., Karube, C., Mizuba, T., Matsufuji, M., Takashima, S., & Iramina, K. (2015). Phase-locked theta activity evoked in patients with severe motor and intellectual disabilities upon hearing own names. *Brain and Development*, *37*(8), 764–772.

Varela, F., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature Reviews. Neuroscience*, *2*(4), 229–239. https://doi.org/10.1038/35067550

Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: a multivariate solution. *Psychophysiology*, *24*(4), 479–486. https://doi.org/10.1111/j.1469-8986.1987.tb00324.x

Yordanova, J., & Kolev, V. (1998). Single-sweep analysis of the theta frequency band during an auditory oddball task. *Psychophysiology*, *35*(1), 116–126.

Tables

Table 1. Acoustic measurements for each of the stimuli used in the experiment. Summary of 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	1838.3	1821.11	**1914.16**	**83.49**
[l#a]	1881.72	1824.66	1835.17	1818.95	1804.7	**1 833.04**	**29.35**
[l#i]	2488.9	1403.42	2395.83	2596.99	2517.84	**2 280.60**	**495.60**

Figures

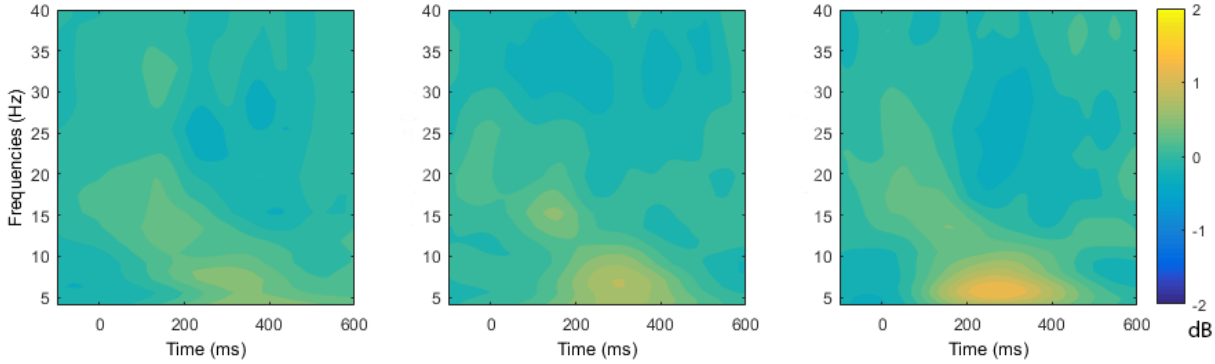

Figure 1

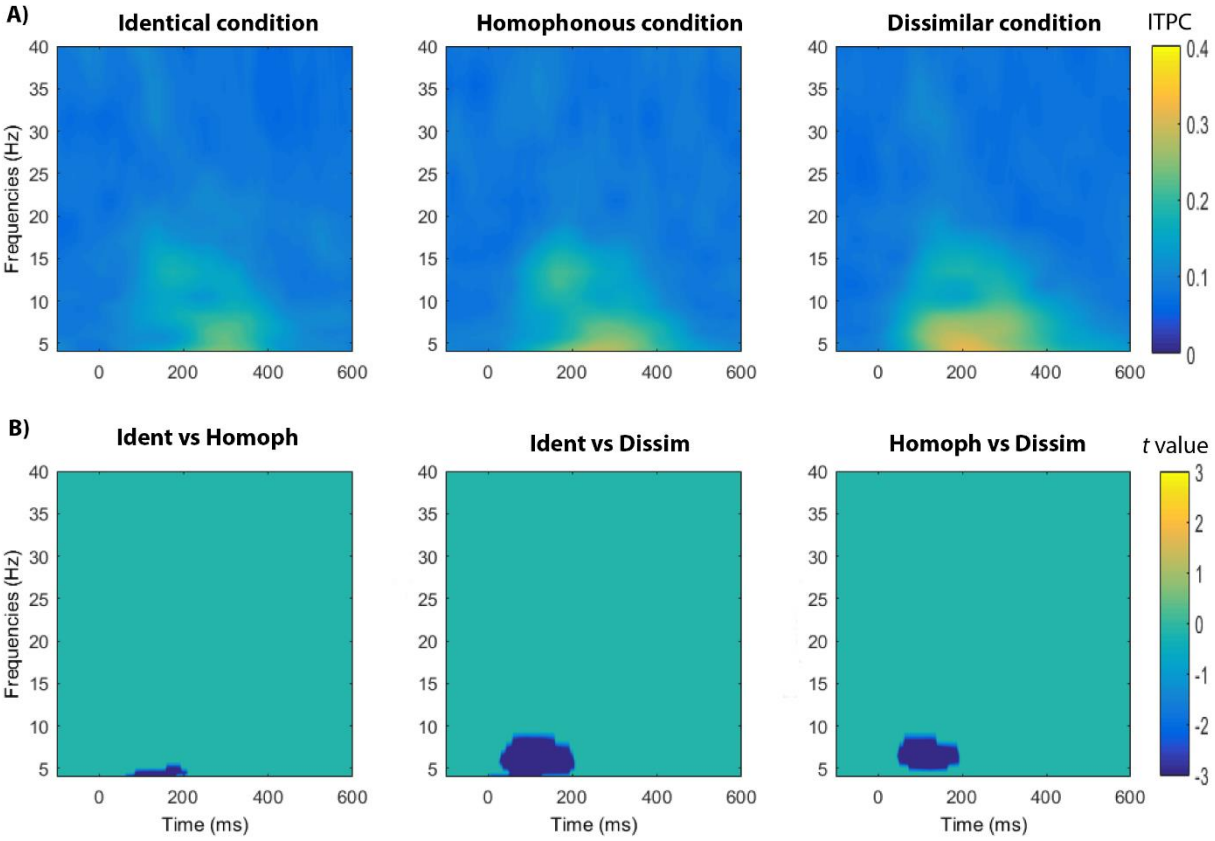

Figure 2

Figures Captions

Figure 1. Power at Fz. Changes in power with respect to baseline across time and frequencies for the Standard (left panel), Homophone (middle panel) and Dissimilar (right panel) conditions. No significant differences were found between conditions.

Figure 2. ITC at Fz (A) and statistical analysis (B). (A) ITC for the 3 conditions across time and frequencies. ITC increases in the theta band in each of the conditions. The peak of this increase was reached earlier in the dissimilar condition than in the other two conditions, and it was latest for the identical condition. (B) Cluster-corrected t values. ITC was higher in the homophonous condition than in the Identical condition in the low theta band between 75 and 200 ms (bottom left panel). ITC was higher in the Dissimilar condition than in the identical condition, spanning the whole theta frequency band between 75 and 200 ms (bottom middle panel). Cluster-based permutation tests also revealed a significant cluster in the theta band when comparing the Dissimilar to the Homophonous condition. ITCP was higher in the Dissimilar condition.