

## Highlights

- Low level linguistic rules impact on the perception of speech sounds
- We investigate when complex rules impact on sound categorization with EEG
- EEG shows that compensation for complex rule modulates an early auditory response
- Speech categorization is thus also sensitive to complex linguistic native rules

**Title:**

Complex linguistic rules modulate early auditory brain responses

**Authors:**

Yue Sun<sup>a,b,c</sup>, Maria Giavazzi<sup>a,d,e</sup>, Martine Adda-Decker<sup>f,g</sup>, Leonardo S. Barbosa<sup>a,b</sup>, Sid Kouider<sup>a,b</sup>, Anne-Catherine Bachoud-Lévi<sup>a,d,e,h</sup>, Charlotte Jacquemot<sup>a,d,e,l</sup>, and Sharon Peperkamp<sup>a,b,l</sup>

**Author affiliations:**

- a. Département d'Etudes Cognitives, Ecole Normale Supérieure - PSL Research University, 29 rue d'Ulm, 75005 Paris, France
- b. Laboratoire de Sciences Cognitives et de Psycholinguistique (EHESS, CNRS, ENS), 29 rue d'Ulm, 75005 Paris, France
- c. Sorbonne Universités, UPMC Univ Paris06, IFD, 4 Place Jussieu, 75252 Paris, cedex 05, France
- d. INSERM U955, team 1 "NeuroPsychologie Interventionnelle", Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris France
- e. Université Paris-Est Créteil, UMR S955, F-94000, Créteil, France
- f. Laboratoire de Phonétique et de Phonologie (CNRS, Sorbonne Nouvelle), 19, rue des Bernardins, F-75005 Paris, France
- g. Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (CNRS, UPMC, Université Paris Sud), rue John von Neumann, bât. 508, B.P.133, F-91403 Orsay, France
- h. AP-HP, Centre de référence maladie de Huntington, Hôpital Henri Mondor and Hôpital Pitié Salpêtrière, Créteil and Paris, France

**Author footnotes:**

1. Joint Last Authors

**Corresponding Author:**

Charlotte Jacquemot

Equipe de NeuroPsychologie Interventionnelle

Département d'Etudes Cognitives

Ecole Normale Supérieure

29 rue d'Ulm, 75005 Paris, France

email: [charlotte.jacquemot@ens.fr](mailto:charlotte.jacquemot@ens.fr)

Tel. +33 (0)1 44 32 26 77

**Abstract:**

During speech perception, listeners compensate for phonological rules of their language. For instance, English place assimilation causes *green boat* to be typically pronounced as *greem boat*; English listeners, however, perceptually compensate for this rule and retrieve the intended sound (*n*). Previous research using EEG has focused on rules with clear phonetic underpinnings, showing that perceptual compensation occurs at an early stage of speech perception. We tested whether this early mechanism also accounts for the compensation for more complex rules. We examined compensation for French voicing assimilation, a rule with abstract phonological restrictions on the contexts in which it applies. Our results reveal that perceptual compensation for this rule by French listeners modulates an early ERP component. This is evidence that early stages of speech sound categorization are sensitive to complex phonological rules of the native language.

**Keywords:**

Electroencephalography, Mismatch negativity, Speech perception, Phonological rules, Assimilation

## 1. Introduction

Perceptual adaptation is a biological requirement to efficiently process the signals of our environment. This adaptation tunes our perception system, making us allocate cognitive resources to relevant information only. Speech processing provides a clear illustration of such adaptation. Exposure to one's native language shapes the auditory system, such that speech perception is tuned to acoustic properties that distinguish native speech sounds and is largely insensitive to those that are not exploited in the native language (Iverson, Kuhl, Akahane-yamada, & Diesch, 2003; Kuhl, Williams, & Lacerda, 1992; Sebastián-Gallés & Soto-Faraco, 1999). For instance, Japanese listeners show persistent perceptual confusion between the English liquid consonants [r] and [l], as Japanese has only one liquid consonant (Flege, Takagi, & Mann, 1996; Goto, 1971). This confusion is likely caused by a low-level mechanism that maps the two non-native sounds onto a single native category, ignoring the acoustic differences between them (Best, McRoberts, & Goodell, 2001; Kuhl, 2000). Studies using electroencephalographic (EEG) recordings report early automatic brain responses that reflect such a mapping mechanism between the speech signal and native speech sounds (Dehaene-Lambertz, 1997; Flagg, Oram Cardy, & Roberts, 2006; Näätänen et al., 1997; Sharma & Dorman, 2000).

However, this mapping mechanism fails to account for the perception of native speech sounds, which can be distorted too. For instance, English listeners frequently misperceive the native consonant [m] as [n] when it is followed by [b]. Since both [m] and [n] exist in English, this perceptual transformation cannot be explained by the low-level mapping mechanism mentioned above. Rather, it arises from listeners' knowledge about a phonological rule of their native language that affects the production of individual speech sounds. Indeed, English has a rule of place assimilation that causes the coronal consonant [n] to be pronounced as its labial counterpart, [m], when it occurs before a labial consonant (e.g. [b]), but not before a non-labial consonant (e.g.

[s]). For instance, the [n] in *green* is typically produced as [m] in *green boat* but not in *green sock*. This [n]-to-[m] transformation in speech production is thus mirrored by a reverse [m]-to-[n] transformation that takes place during speech perception. In other words, listeners compensate for the phonological rule in order to retrieve the sound intended by the speaker (Gaskell & Snoeren, 2008).

Compensation for phonological rules has also been observed in other languages and with other rules (Darcy, Ramus, Christophe, Kinzler, & Dupoux, 2009; Mitterer, Csépe, Honbolygo, & Blomert, 2006). A number of studies have examined the time course of perceptual compensation for phonological rules (Clayards, Niebuhr, & Gaskell, 2015; Mitterer & Blomert, 2003; Mitterer, Csépe, Honbolygo, et al., 2006; Tavabi, Elling, Dobel, Pantev, & Zwitserlood, 2009). For instance, using electroencephalographic (EEG) recordings, Mitterer and Blomert (2003) examined the electrophysiological correlates of perceptual compensation for Dutch place assimilation, which is identical to the English rule mentioned above. They found that perceptual compensation for this rule, by which [m] is perceived as [n] before [b] but not before [s], modulates an early brain response around 100 ms after the onset of the speech sound in question. A similar early brain response related to perceptual compensation was observed in a study which focused on Hungarian liquid assimilation, by which [l] is produced as [r] before [r] (Mitterer, Csépe, Honbolygo, et al., 2006). These findings provide support for the hypothesis that compensation for phonological rules occurs at an early stage of speech sound perception.

So far, research on compensation for phonological rules has focused on rules with a clear phonetic underpinning. For instance, in the case of Dutch and English place assimilation, the coronal consonant [n] can be pronounced as labial [m] whenever it is followed by a labial consonant. Thus, assimilation takes place as long as the target speech sound differs from the following sound in the phonetic feature that is subject to change. Assimilation rules like this can

be considered as extreme cases of coarticulation, by which a sound takes over articulatory properties of an adjacent sound (Church, 1987). Thus, perceptual compensation for these rules may rely on low-level phonetic mechanisms used for compensation for coarticulation (e.g., feature parsing, Gow, 2003). This phonetic mechanism would well explain findings showing that listeners can also compensate for assimilation rules that are not present in their native language. For instance, French listeners compensate for the English place assimilation rule, though to a lesser extent than for their native voicing assimilation rule (Darcy et al., 2009). Likewise, Dutch listeners compensate for Hungarian liquid assimilation, and this compensation also elicits an early brain response (Mitterer, Csépe, Honbolygo, et al., 2006). In these cases, participants necessarily rely on mechanisms that compensate for coarticulation, because they are never exposed to these assimilation rules in their native language.

However, not all phonological rules can be reduced to extreme cases of coarticulation, as there exist rules which are phonetically less straightforward. An example is provided by French. In this language, voiceless consonants may be pronounced as their voiced counterparts before some but not all voiced consonants. For instance, the final voiceless [f] of the word *oeuf* ‘egg’ can be produced as voiced [v] in *oeuf blanc* ‘white egg’ but not in *oeuf noir* ‘black egg’, even though both [b] and [n] are voiced. This rule of voicing assimilation is thus more complex than the rules described above, as not all voiced consonants induce the voicing change. More specifically, voicing assimilation can apply before obstruent consonants (such as [b]) but not before sonorant consonants (such as [n]) (Grammont, 1933). This exceptional behavior of sonorants is language-specific, as in several other languages with voicing assimilation sonorants can induce a voicing change in a preceding voiceless obstruent (Blaho, 2008; Lipski, 1989; Rubach, 1996; Strycharczuk & Simon, 2013; Wheeler, 1986). Thus, in contrast to place assimilation in English and Dutch and to liquid assimilation in Hungarian, voicing assimilation in French is not strictly

phonetically grounded. As a consequence, perceptual compensation for this rule cannot be based on purely phonetic principles but, rather, must involve more complex phonological processing.

An open question is whether the early brain mechanism observed for perceptual compensation of phonetically motivated rules can also explain compensation for more complex phonological rules such as French voicing assimilation. Whereas phonetic compensation takes place very early in speech perception, more abstract phonological compensation could reflect a post-perceptual reinterpretation of the miscategorized sound. Such a late mechanism has been proposed in purely feedforward models (Norris, McQueen, & Cutler, 2000). It reflects the effect of higher order knowledge on the reinterpretation of a low-level sensory output, during the decision making phase of the perceptual judgment. In order to investigate this issue we used high-density EEG to record French participants' electrophysiological responses while they performed a mismatch detection task (see Materials and Methods for more details). Participants listened to a series of four identical precursor stimuli with a vowel-consonant-consonant-vowel structure ( $V_1C_1C_2V_2$ , e.g. *ofbe* or *ofne*), followed by a test stimulus that was either identical to the previous stimuli or differed from them in the voicing of  $C_1$  (*ovbe* or *ovne*).  $C_2$  was either a viable context or an unviable context for voicing assimilation of  $C_1$  (respectively, [b] in *ofbe* and [n] in *ofne*; Figure 1). Since French listeners perceptually compensate for their native voicing assimilation rule (Darcy et al., 2009), they should tend to misperceive *ovbe* as *ofbe* because the former is the assimilated form of the latter. On the contrary, they should rarely misperceive *ovne* as *ofne* because the two forms are not linked by the voicing assimilation rule. In other words, they should have difficulty detecting the voicing change of  $C_1$  between test and precursor stimuli in the viable context but not in the unviable context.

In order to examine when the voicing assimilation rule affects speech perception, we compared the event related potentials (ERPs) elicited by the voicing change of  $C_1$  in the viable and unviable



contexts. Detectable voicing changes introduced by the test stimulus after four identical precursor stimuli should elicit an early ERP called mismatch negativity (MMN), which arises in response to an infrequent change in a repetitive sound sequence and reflects early, preattentive processes (Dehaene-Lambertz, 1997; Näätänen & Alho, 1995; Sams, Paavilainen, Alho, & Näätänen, 1985). Thus, if compensation for voicing assimilation occurs at the early perceptual stage when speech sounds are categorized, an MMN should be observed in the unviable but not in the viable context. If, by contrast, compensation for voicing assimilation reflects a top-down reinterpretation of an already categorized sound, then an MMN should be observed in both the viable and the unviable context. Furthermore, since participants in our paradigm provide a behavioral response, the successful detection of a voicing change should also elicit an ERP at the time window of a late positive component (i.e., P300), which is related to the conscious detection of a deviant item and to the decision-making process involved in giving a response (Polich, 2007). Therefore, this component should arise in the unviable context only, regardless of whether perceptual compensation for voicing assimilation occurs at an early or a later stage.

In addition, we examined whether French participants' expected lowered performance in the detection of a voicing change in the viable context could be due to a general phonetic difficulty to detect a voicing change before an obstruent consonant (such as [b] in *ofbe*) compared to before a sonorant consonant (such as [n] in *ofne*). To this end, we also tested a group of American English listeners on the behavioral task only. If the expected effect in the viable context is merely due to a language-independent phonetic difficulty, American English listeners should behave as French listeners. If, by contrast, it is language-specific, they should behave differently from French listeners, as English does not have a voicing assimilation rule.

## **2. Material and Methods**

### **2.1. Participants**

Thirty right-handed, native French speakers (21 females; mean age: 24 years) participated in the EEG experiment. In addition, twelve right-handed, native speakers of American English (7 females; mean age: 20.1 years), with little or no knowledge of French, participated in the behavioral task only. All participants gave written informed consent to take part in this study and were paid for participation. They reported normal vision, audition and language abilities, and declared having learned no other language other than their native language before the age of ten. Five French participants were excluded from the analyses because of a large amount of artifacts (> 50% of total trials) due to eye blinks, muscle movements, or heartbeats presented in the EEG data.

### **2.3. Paradigm and stimuli**

Participants performed a mismatch detection task on meaningless sound sequences. They listened to a series of four identical precursor stimuli with a vowel-consonant-consonant-vowel structure ( $V_1C_1C_2V_2$ ), e.g. *ofbe* or *ofne*, followed by a test stimulus that was either identical to the previous stimuli e.g. *ofbe-ofbe* or *ofne-ofne* (no-change condition) or differed from them in the voicing of  $C_1$ , e.g. *ofbe-ovbe* or *ofne-ovne* (voicing-change condition).  $C_2$  was either a viable context ([b] in *ofbe*) or an unviable context ([n] in *ofne*) for inducing the perceptual transformation of  $C_1$  (Figure 1). For each condition, half of the trials had precursor items with a voiceless – unassimilated –  $C_1$ , as in the examples above, whereas the other half of the trials had precursor items with a voiced - assimilated -  $C_1$ . Thus, in the voicing-change condition, the change was from voiceless to voiced in half of the trials and from voiced to voiceless in the other half.

We used four  $V_1C_1C_2V_2$  sequences (i.e. *ofbe*, *usda*, *ikdo*, *akvi*) to construct the experimental stimuli (see Appendix for a complete table of stimuli). For each base sequence (e.g., *ofbe*), four items were created as the precursor stimuli (*ofbe*, *ovbe*, *ofne*, *ovne*). The four items differed in that  $C_1$  was either a voiceless or a voiced obstruent (*f* and *v*), and  $C_2$  was either a voiced obstruent

or a voiced sonorant (*b* and *n*). Since obstruents (*b*) but not sonorants (*n*) provide the context for French voice assimilation, the first two items in each quadruplet (*ofbe*, *ovbe*) were used for the viable context and the last two (*ofne*, *ovne*) for the unviable context. For each precursor item (e.g., *ofbe*), three test items were defined, as follows. For the no-change condition, the test item was identical to the precursor item (*ofbe*); for the voicing-change condition, the test item was created by replacing  $C_1$  of the precursor item with its voicing counterpart (*ovbe*). Finally, in order to introduce more variability in the experiment, we added filler items. These items were created by replacing  $C_1$  of the precursor item with another consonant with the same voicing value, but with a change in place of articulation (*osbe*). These filler trials provided clear cases of a change<sup>1</sup>. Both behavioral and electroencephalographic data from these trials were excluded from the analysis.

Precursor stimuli were produced by four female speakers and test stimuli by a male speaker. The presence of multiple speakers requires participants to rely on the stimuli's phonological properties, rather than on purely acoustic information, to do the task (Dehaene-Lambertz et al., 2000; Eulitz & Lahiri, 2004). Recordings were made in a sound-attenuated room with an external audio card (Lexicon Lambda) as digital input and a customized MATLAB program using Psychophysics Toolbox extensions (Brainard, 1997). Tokens were digitized at 44100 Hz / 16 bits. We ensured that  $C_1$  in the unassimilated items (e.g., *f* in *ofbe*) was completely voiceless whereas  $C_1$  in the assimilated items (e.g., *v* in *ovbe*) was completely voiced. Indeed, when voicing assimilation occurs in the viable context in spontaneous speech, it most often induces a categorical transformation from a voiceless sound into a voiced sound rather than a sound with an intermediate voicing (Hallé & Adda-Decker, 2011). For assimilated items, speakers were thus given written forms containing a voiced  $C_1$ . For the unassimilated items, as it is very difficult for

---

<sup>1</sup> As expected, in these trials participants detected the change in both the viable context (e.g. *ofbe-osbe*) and the unviable context (e.g. *ofne-osne*) 94% and 93% of the time, respectively ( $t < 1$ ).

French speakers to naturally produce a voiceless obstruent before a voiced one, we created, for each speaker, the items containing such unassimilated obstruent clusters by splicing the two syllables from two neutral items recorded separately. For instance, *ofbe* was created by combining *of-* from *ofpe* and *-be* from *ovbe*, such that [f] was completely voiceless. This method also allowed us to ensure that the physical difference between the voiceless and the voiced obstruents ([f] and [v]) was identical in the viable and the unviable context. In order to equalize a potential bias that may be introduced as a result of manual editing, all other stimuli were likewise constructed by means of cross-splicing.

All VCCV sequences used for cross-splicing were recorded multiple times by five native French speakers, four women and one man. For each speaker, the best two tokens of each sequence were chosen. After cross-splicing, all stimuli were equalized in length (500 ms) and their intensity was normalized to mean RMS amplitude (60 dB) across all stimuli, using Praat (Boersma, 2002).

Finally, we validated the quality of the stimuli with a trained phonetician and checked for potential sources for misperception of the stimuli other than perceptual compensation<sup>2</sup>.

## 2.4. Procedure

Participants were seated in a quiet testing booth in front of a CRT monitor for instructions and feedback. Stimuli were presented to them binaurally via earphones. Participants' responses were

---

<sup>2</sup> Some of the stimuli contained consonant sequences that could potentially be misidentified for reasons other than perceptual compensation. This is the case for the stimuli with an unassimilated C<sub>1</sub> in the viable context (e.g. *ofbe*). These stimuli are dispreferred by the French phonotactics, as they are usually produced with a voiced C<sub>1</sub> (i.e. *ovbe*), due to the voicing assimilation rule. Although these consonant sequences are phonotactically legal given that voicing assimilation is optional (Dell, 1995; Hallé & Adda-Decker, 2011), one can hypothesize that French listeners perceptually repair these clusters (e.g., *fb*) into phonotactically preferred ones (e.g., *vb* or *fp*), just like they repair phonotactically illegal clusters into legal ones (Hallé & Best, 2007; Hallé, Segui, Frauenfelder, & Meunier, 1998). In order to ensure that French participants correctly perceive these consonant sequences, we asked eight native French listeners, who did not participate in the EEG experiment, to transcribe the stimuli containing these sequences. Their transcriptions confirmed that French listeners are able to correctly perceive these sequences (97.1% correction transcription).

collected with a numeric keypad. The experiment was run using MATLAB Psychophysics Toolbox extensions (Brainard, 1997).

Each trial consisted of four precursor items, produced by the four female speakers, followed by one test item, produced by the male speaker, with an inter-stimulus interval (ISI) of 500 ms. The order of the female voices for the precursor stimuli was counterbalanced across trials. Participants were instructed to respond as quickly and accurately as possible to the test stimulus by pressing one of the two response buttons assigned to the “same” and “different” responses. The side associated with the “same” response was counterbalanced across participants. For each trial, participants had a response time window of 2000 ms, measured from the onset of the test stimulus.

In order to minimize artifacts induced by muscle activity near the EEG recording sensors, French participants were instructed not to move their body and to fixate their gaze on a fixation cross presented in the center of the monitor during each trial. This fixation cross appeared 1500 ms before the auditory stimuli was played and disappeared 1000 ms after the response had been given. Then, a visual cue was presented in the center of the screen for participants to blink their eyes. This visual cue lasted for 2000 ms and was then replaced by the fixation cross for the next trial. Participants were instructed to blink their eyes only during the presentation of the cue.

There were 10 blocks of repeated trials. Each block contained 32 test trials and 16 filler trials (see Appendix for details). Trials were presented in a pseudo-random order, such that no more than two consecutive trials concerned the same radical. The experiment started with a training phase composed of four no-change and four place-change trials. During this phase, participants received visual feedback on their responses (“correct”, “incorrect”, or “too slow”). Results from this phase were excluded from the analysis.

Since the American English participants only performed the behavioral task, they were tested with a slightly adapted procedure. In particular, only 5 blocks of trials were presented to them, no visual cue allowing for an eye blink appeared at the end of each trial, and the interval between trials was shortened from 4500 ms to 2000 ms.

## **2.5. EEG recording and preprocessing**

EEG was recorded using the EGI (Electrical Geodesics Incorporated, Eugene, OR) 256-channel HCGSN sensor net and Net Station 4.1 (EGI software). The EEG signal was referenced to the vertex (Cz) and digitized at a sampling rate of 500 Hz. Impedances for each electrode were kept lower than 50 k $\Omega$  during the experiment.

After acquisition, data were preprocessed using a custom program developed under MATLAB with SPM8 (Litvak et al., 2011) and EEGLab (Delorme & Makeig, 2004) extensions. The continuous EEG data were first band-pass filtered at 0.3 to 20 Hz and then segmented into epochs for each trial. The segmentation was done with respect to the onset of the second phoneme of each test stimulus, i.e. C<sub>1</sub> (in average 130 ms after the stimulus onset), which corresponds to the beginning of the deviation between the no-change and voicing-change items. Epochs were defined from 1000 ms prior to this point to 1000 ms following it, where the 150 ms preceding the onset of the test stimulus (around 280 to 130 ms before the epoch zero point) served as baseline. The segmented epochs were re-referenced against the average of all electrodes and corrected with the 150 ms baseline. Epochs that contained artifacts such as eye blinks, saccades and other excessive muscular activity, as well as disconnected electrodes were rejected (a minimum 50 artifact-free trials was assured for each condition).

## **2.6. EEG data analysis**

EEG analysis was carried out from 0 to 800 ms following the onset of the C<sub>1</sub>. The average ERP was calculated for each participant across Conditions (no-change and voicing-change) and

Contexts (viable and unviable) using all artifact-free trials, independently of whether the behavioral responses were correct or incorrect. To determine the neural correlates of the interaction between the Condition and Context without any *a priori* knowledge on spatial or temporal regions of interest while correcting for multiple comparisons (sensors & time sample), we based our analysis on a cluster-based permutation test (Maris & Oostenveld, 2007) and applied it to a repeated-measures ANOVA. Each sample (one sensor, one time point) of averaged ERP was submitted to an ANOVA with Condition (no-change vs. voicing-change) and Context (viable vs. unviable) as within-participant factors. For the interaction between the two factors, samples whose *F*-value exceeded a threshold (here, *F*-values corresponding to a *p*-value of 0.05) were clustered based on time- and space-adjacency. Each cluster defined in space and time by this procedure was then assigned a cluster-based statistics equal to the sum of the *F*-values of all the samples belonging to the cluster. To test whether this cluster-level statistics could be obtained by chance, the condition labels of the original ERP data of each participant were randomly shuffled. The clustering procedure was then applied on those randomized data, and the *F*-statistic of the biggest cluster from these data was computed. By repeating the random assignment of condition labels to the EEG data 1000 times, we could estimate the distribution of the maximum cluster level *F*-statistic under the null hypothesis. If the original statistic was greater than 95% of the values obtained on randomized data, then the null hypothesis could be rejected with a Monte-Carlo *p*-value < 0.05.

### **3. Results**

#### **3.1. Behavioral detection of voicing change**

##### **3.1.1 French participants**

ANOVAs were performed on both error rates and reaction times with Condition (no-change vs. voicing-change) and Context (viable vs. unviable) as within-participant factors. For error rates we

observed main effects of Condition ( $F(1,24) = 183.7, p < 0.001$ ) and Context ( $F(1,24) = 224.6, p < 0.001$ ) as well as a Condition  $\times$  Context interaction ( $F(1,24) = 235.3, p < 0.001$ , Figure 2A). Post-hoc analyses revealed an effect of Context only in the voicing-change condition ( $F(1,24) = 238.4, p < 0.001$ ), in which participants made more errors for the viable context (65%) than for the unviable context (7.8%). The analysis of reaction times also showed main effects of Condition ( $F(1,24) = 58.61, p < 0.001$ ) and Context ( $F(1,24) = 7.7, p < 0.05$ ) and a Condition  $\times$  Context interaction ( $F(2, 48) = 11.3, p < 0.001$ , Figure 2B). Again, a context effect was found only in the voicing-change condition ( $F(1,24) = 17.1, p < 0.001$ ), in which participants were slower for the viable context (928 ms) than for the unviable context (861 ms). These results confirm that participants had more difficulty detecting the voicing change in the viable context than in the unviable context.

### **3.1.2 American English participants**

Analyses of variance were performed on both error rates and reaction times with the factors Condition (no-change vs. voicing-change) and Context (viable vs. unviable). For error rates we observed a main effect of Condition ( $F(1, 11) = 71.7, p < 0.001$ ) and Context ( $F(1, 11) = 11.9, p < 0.01$ ) as well as a Condition  $\times$  Context interaction ( $F(1, 11) = 13.5, p < 0.01$ ) (Figure 3A). Post-hoc analyses also revealed an effect of Context only in the voicing-change condition ( $F(1, 11) = 13.2, p < 0.01$ ), in which participants made more errors for the viable context (47%) than for the unviable context (33%). The analysis of reaction times revealed a main effect of Condition ( $F(1, 11) = 25.8, p < 0.001$ ) and Context ( $F(1, 11) = 5.17, p < 0.05$ ) but no interaction (Figure 3B).

### **3.1.3 Comparison between French and American English participants**

In order to compare the results of the American English participants with those of the French participants, analyses of variance on both error rates and reaction times were performed with the factors Language (French vs. English), Condition (no-change vs. voicing-change) and Context



(viable vs. unviable). The error rate analysis revealed main effects of Context ( $F(1, 35) = 219, p < 0.001$ ) and Condition ( $F(1, 35) = 251, p < 0.001$ ), as well as interactions between Language and Context ( $F(1, 35) = 46, p < 0.001$ ) and between Context and Condition ( $F(1, 35) = 236, p < 0.001$ ). The Language  $\times$  Context interaction was due to the fact that the effect of Context was larger for French listeners ( $F(1,24) = 224.6, p < 0.001$ ) than for English listeners ( $F(1, 11) = 11.9, p < 0.01$ ). The Context  $\times$  Condition interaction was due to the fact that there was an effect of Context in the voicing-change condition ( $F(1, 36) = 98, p < 0.001$ ) but not in the no-change condition ( $F < 1$ ). Crucially there was also a triple Language  $\times$  Condition  $\times$  Context interaction ( $F(1, 35) = 53.5, p < 0.001$ ). This triple interaction was due to the fact that the phonological context had a larger impact on the detection of a voicing change for French listeners (57.2%) than for English listeners (14.1%) ( $F(1, 35) = 51.6, p < 0.001$ ).

The reaction time analysis revealed a main effect of Condition ( $F(1, 35) = 80, p < 0.001$ ), an interaction between Context and Language ( $F(1, 35) = 9.6, p < 0.01$ ), and an interaction between Context and Condition ( $F(1, 35) = 17, p < 0.001$ ). The Language  $\times$  Context interaction was due to the fact that French listeners were slower in the viable context than in the unviable context (868 ms vs. 841 ms,  $F(1,24) = 7.7, p < 0.05$ ), whereas the opposite holds for English listeners (848 ms vs 869 ms,  $F(1, 11) = 5.17, p < 0.05$ ). The Context  $\times$  Condition interaction was due to the fact that participants were faster in the viable context than in the unviable context for no-change trials (793 ms vs 814 ms,  $F(1, 36) = 9.8, p < 0.01$ ), while they were slower in the viable context than in the unviable context for voicing-change trials (922 ms vs 895 ms,  $F(1, 36) = 7.9, p < 0.01$ ). The triple interaction between Language, Condition and Context, however, did not reach significance ( $F(1, 35) = 3.7, p = 0.06$ ).

### **3.2. ERP responses related to the detection of voicing change**

For French participants, we compared the ERPs in the voicing-change and in the no-change conditions for the viable and the unviable contexts. Since the perceptual transformation occurs only in the viable context, we expected the magnitude of change-induced ERPs to depend on the context. Such a context dependency would be revealed by an interaction between Condition (no-change *vs.* voicing-change) and Context (viable *vs.* unviable).

To determine the electrophysiological correlates of the interaction between the Condition and Context without any *a priori* knowledge on spatial or temporal regions of interest while correcting for multiple comparisons (sensors & time sample), we based our analysis on a cluster-based permutation test (Maris & Oostenveld, 2007) and applied it to a repeated-measures ANOVA (see Materials & Methods 2.6. for more details). This analysis revealed two spatiotemporal clusters for which the interaction between Condition (no-change *vs.* voicing-change) and Context (viable *vs.* unviable) was significant: from 140 to 240 ms after the onset of the deviation above the central electrodes for the first cluster (Monte Carlo  $p < 0.05$ , Figure 4A), and from 360 to 520 ms above the parietal electrodes for the second cluster (Monte Carlo  $p < 0.01$ , Figure 4B).

In order to assess the nature of the two responses, we carried out a series of post-hoc analyses. Specifically, we calculated for each time window the mean voltage over time and sensors within the cluster across condition and context and then conducted *t*-tests on the voltage difference between the voicing-change and no-change conditions for each phonological context. For the first time window (Figure 4A, right panel), the analysis revealed a significant difference in the unviable context ( $t(1,24) = -6.3$ ,  $p < 0.001$ ), reflecting a larger negativity ( $-0.59 \mu\text{v}$ ) in the voicing-change than in the no-change condition. As illustrated in Figure 5, this early response arises at fronto-central electrodes. By contrast, we observed no significant difference in the viable context ( $0.1 \mu\text{v}$ ,  $t(1,24) = 0.42$ ,  $p > 0.1$ ), indicating that the same voicing change does not elicit

an early response in this context. For the second time window (Figure 4B, right panel), the voicing-change minus no-change subtraction over parietal electrodes yielded a marked positivity in the unviable context ( $0.76 \mu\text{V}$ ,  $t(1,24) = 3.76$ ,  $p < 0.001$ ) (Figure 5). Like the first response, this response was absent in the viable context ( $-0.09 \mu\text{V}$ ,  $t(1,24) = -0.51$ ,  $p > 0.1$ ). To further examine whether there was any observable ERP difference in the viable context at all, we conducted the cluster-based permutation test on the difference between voicing-change and no-change conditions in this context within the entire analysis time window (0 – 800 ms). This analysis revealed no significant spatiotemporal cluster (Monte Carlo  $ps > 0.1$ ).

Our behavioral results revealed that although participants tended to fail to perceive the voicing change in the viable context, they correctly discriminated voiced from voiceless obstruents in this context in 35% of the trials. We further examined whether this residual discrimination capacity was reflected in the EEG responses. To do this, we first computed, for each participant, the average ERP difference between the voicing-change and the no-change conditions in the viable context, using all trials (regardless of the behavioral response). We then computed the correlation between the mean amplitude of the ERP difference within the early and the late time-windows on the one hand and participants' detection rate of voicing-change on the other hand. Results from this analysis showed no significant correlation in either time-window (140-240 ms:  $t(1,24) = 0.94$ ,  $p = 0.35$ ; 360-520 ms:  $t(1,24) = -0.98$ ,  $p = 0.36$ ).

#### **4. Discussion**

Context effects are a fundamental feature in speech perception. The categorization of the acoustic signal into speech sounds is affected by the linguistic context in which the sounds occur (Darcy et al., 2009; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; P Hallé, Segui, Frauenfelder, & Meunier, 1998; Mann, 1980; Massaro & Cohen, 1983; Mitterer & Blomert, 2003; Pitt & McQueen, 1998). We investigated the influence of linguistic context on the perception of speech

sounds, through the study of compensation for native phonological rules. Our behavioral results show that French listeners compensate for the voicing assimilation rule of their language, but only in the context which allows the application of this rule. This context-dependent compensation caused detection of a voicing change to be more difficult before a voiced obstruent than before a sonorant.

Our electrophysiological results with French listeners show that the voicing assimilation rule affects speech perception at an early stage of speech sound processing, rather than at a later stage of speech sound reinterpretation. Indeed, in the unviable context, the voicing change elicited both an early and a late ERP response (respectively from 140 to 240 ms and from 360 to 520 ms post-deviance). The topography of the early response is very similar to that of the mismatch negativity (MMN). The MMN is commonly elicited in odd-ball paradigms by the intrusion of a deviant stimulus in a sequence of standard stimuli (Bendixen, SanMiguel, & Schröger, 2012; Dehaene-Lambertz, Dupoux, & Gout, 2000; Dehaene-Lambertz, 1997; Friederici, 2002; Näätänen & Alho, 1995; Näätänen et al., 1997; Paavilainen, 2013; Schröger, 1998). It can be observed without listeners paying attention to the stimuli (Näätänen & Alho, 1995; Näätänen et al., 1997; Sams et al., 1985), and hence reflects an automatic detection of auditory changes (Baldeweg, Williams, & Gruzelier, 1999; Giard, Perrin, Pernier, & Bouchet, 1990). The main difference between our paradigm and standard odd-ball paradigms lies in the fact that we asked participants to make an active comparison between the test item and the precursors. Previous studies have shown that when participants have to focus on the stimuli they perceive, the MMN response is intensified as compared to a classic odd-ball, in which participants' attention is diverted (Woldorff et al., 1998; Pulvermüller & Shtyrov, 2006). The latency and the topography of the MMN we obtained are in line with previous studies conducted with a paradigm in which participants explicitly attend to the auditory input to perform a behavioral task (Bekinschtein et al., 2009; Dehaene-Lambertz et

al., 2000; Dehaene-Lambertz et al., 2005). Also, they are in line with results from the classic odd-ball paradigm with no explicit attention to the auditory stimuli (e.g., Näätänen et al., 1997; Mitterer and Blomert, 2003).

As for the late ERP response elicited by the voicing change in the unviable context, its latency and topography are in accordance with previous descriptions of the P300 component (Polich, 2007). Previous research has shown that the P300 is observed only when participants have to attend to an infrequent auditory change and its latency is aligned with the reaction times of the behavioral response (Jung et al., 2001; O'Connell, Dockree, & Kelly, 2012; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). This component is thus associated with the late, explicit, detection of a task-defined change (for a review see Polich, 2007). Given participants' high detection rate of the voicing change in the unviable context, the presence of both an MMN and a P300 was expected in that context. In particular, the presence of the MMN in the unviable context confirms that French listeners perceive the difference between voiced and voiceless consonants ([v]-[f]) at an early stage of speech perception. In contrast, in the viable context, the same voicing change between [v] and [f] induced neither an MMN nor a P300 response. This result indicates that the [v]-to-[f] transformation due to compensation for the French voicing assimilation rule relies on early mechanisms that apply before the MMN occurs (i.e., earlier than 140 ms after the voicing change). In fact, if compensation relied on a late correction of an already categorized sound, the voicing change should have been detected at early perceptual stages, thus eliciting an MMN not only in the unviable but also in the viable context. The absence of an MMN in the viable context therefore shows that implicit knowledge of a native language rule impacts speech perception at an early stage of signal-to-sound mapping. The absence of the P300 confirms the behavioral results, suggesting that the acoustic differences between [v] and [f] are not even processed at a later stage.

Looking in more detail at the behavioral results with French listeners, we observed a residual capacity to detect the voicing change in the viable context, with 35% of correct responses. This finding is in accordance with Darcy et al. (2009), who showed that in a word detection task, French listeners detected 35% (*sic*) of voicing changes in the viable context. This residual discrimination capacity was not reflected, however, in the EEG data, where there were no ERP differences between the no-change condition and the voicing-change condition in the viable context, and no correlation between individual detection rates and mean amplitude of ERP difference within either the early or the late time-window. This might seem puzzling, as one could expect the presence of a small yet observable MMN and P300 in the viable context, given a non-negligible detection rate of the voicing change in that context. Here we argue that, on the contrary, this discrepancy results from the very fact that compensation is optional. In fact, in order to induce an MMN or a P300 in the voicing-change condition, it is necessary that the precursor stimuli produce enough neural habituation for the perceptual deviation in the test stimulus to trigger a detectable neural response (Näätänen & Alho, 1995; Polich, 2007; Sussman, 2007). Given the nature of the precursor items, this habituation is not always reached. Indeed, when an assimilated item (e.g., *ovbe*) serves as the precursor stimulus, since it is only optionally transformed into its unassimilated counterpart (e.g., *ofbe*) during perception, the four precursor items are not always perceived identically. Thus, although listeners can sometimes detect the voicing change between the last precursor stimulus and the test stimulus, they cannot build an unambiguous representation of the precursor sequence. For this reason, the condition to trigger an observable MMN or P300 is not met (Breen, Kingston, & Sanders, 2013).

We further showed that the difficulty for perceiving the voicing-change between the precursor and test stimuli in the viable context for French listeners cannot be reduced to a universal phonetic difficulty in discriminating voicing differences before obstruents, independently from

language-specific knowledge. Such a universal effect may arise from the phonetic difference between voicing in obstruents and voicing in sonorants: voiced obstruents are commonly produced with an active voicing gesture, whereas sonorants are passively voiced sounds (Stevens, 2000). Behavioral results from the control experiment show that like French listeners, American English listeners have more difficulty discriminating a voicing contrast before obstruents than before sonorants, in spite of not having the voicing assimilation rule in their language. This suggests that voicing is indeed harder to perceive before obstruents than before sonorants. Crucially, however, a comparison between English and French participants revealed that the increased difficulty before obstruents was larger for French than for English listeners, thus confirming our interpretation of the behavioral results obtained with French listeners. That is, French listeners show a language-specific effect of compensation for assimilation over and above a universal phonetic effect.

The time course of perceptual compensation for phonological rules has been previously investigated using EEG measurements (Mitterer & Blomert, 2003; Mitterer, Csépe, Honbolygo, et al., 2006; Tavabi et al., 2009). Results from these studies also showed an absence of an MMN for sound changes in the viable context of phonological rules. However, these studies left an open question regarding the processing mechanism underlying the early brain response, due to the strictly phonetic motivation of the investigated rules. Indeed, the observed early brain response in these studies may reflect processing at a phonetic rather than at a phonological level. The critical aspect of the present study is that we focused on a rule that cannot be caused by a mechanism that analyzes only the phonetic properties of the context, i.e. the presence or absence of voicing in the following consonant. In fact, obstruents followed by a voiced consonant are not always perceptually transformed into their voiceless counterparts. This perceptual transformation applies only when the second consonant is a voiced obstruent. If the second consonant is a sonorant, it

does not trigger perceptual compensation, despite being voiced. This perceptual compensation therefore depends on a more complex, high-level, mechanism that distinguishes among two types of voiced sounds, obstruents and sonorants, in a way that is specific to the French voicing assimilation rule. Our results thus demonstrate that the early stages of the signal-to-sound mapping are not purely governed by low-level phonetic properties of the sound, but are also modulated by high-level phonological knowledge.

In order to ensure that the ERP differences in response to the perceptual compensation for the phonological rule reflected a compensatory mechanism at the phonological level, in the present study we introduced several methodological innovations with respect to previous research. First, we used meaningless stimuli instead of real words in order to minimize lexical biases. Second, we ensured that the compensation mechanism could not be restricted to one specific consonant sequence, by using four different sequences. Finally, our stimuli were produced by multiple speakers (four female speakers for the precursor stimuli and one male speaker for the test stimulus), to induce processing at a phonological rather than an acoustic level (Brunellière, Dufour, Nguyen, & Frauenfelder, 2009; Dehaene-Lambertz et al., 2000; Eulitz & Lahiri, 2004).

The present results open new perspectives to the investigation of how the brain processes phonological rules. In fact, phonological theory differentiates between different types of phonological rules, based on their different origins. Certain rules have a clear phonetic origin (Flemming, 2002; Kirchner, 2013; Steriade, 1997). This holds for instance for place assimilation in Dutch and English, which can be seen as extreme cases of coarticulation between adjacent consonants, as one sound always takes over phonetic properties of its following sound (Gaskell & Snoeren, 2008; Mitterer & Blomert, 2003; Mitterer, Csépe, & Blomert, 2006). French voicing assimilation is an instance of a more complex type of phonological rule. These rules also have phonetic underpinnings but they only apply in a subset of phonetically valid contexts. In spite of



their greater level of complexity, they are typically found across several unrelated languages. For instance, Hungarian has a voicing assimilation rule whose distribution is the same as the one found in French (Blaho, 2004). Finally, there is yet another type of phonological rules, sometimes referred to as “unnatural” because they are phonetically unmotivated (Anderson, 1981; Bach & Harms, 1972; Buckley, 2000). These rules are usually only found in single languages or in a set of related languages, as the result of historical change or of morphological processes.

Our ERP results demonstrate that, just like more straightforwardly phonetic rules of the first type, more complex phonological rules such as French assimilation rule have an impact on early brain responses too, i.e., the MMN. As observed in source reconstruction studies, the MMN is generated by a temporo-frontal network (for a review, see Garrido, Kilner, Stephan, & Friston, 2009). This network includes regions in the temporal lobe – bilateral auditory cortices (Deouell, Bentin, & Giard, 1998; Grau, Fuentemilla, & Marco-Pallarés, 2007; Jemel, Achenbach, Müller, Röpcke, & Oades, 2002; Tiitinen et al., 2006), and prefrontal lobe (e.g., Giard et al., 1990; Opitz, Rinne, Mecklinger, Von Cramon, & Schröger, 2002). Both regions have been shown to contribute to speech sound perception. The temporal auditory cortices play a primary role in the auditory analysis of the speech signal through their fine-grained tuning to one’s native language. Prefrontal regions (e.g., inferior frontal gyrus – IFG) are involved in more complex linguistic analyses of speech information (Hickok & Poeppel, 2007; Myers, 2014). Previous research has also shown functional interactions between the two regions during the perceptual identification of speech sounds (Liebenthal, Sabri, Beardsley, Mangalathu-Arumana, & Desai, 2013) and the detection of auditory deviance (Jemel et al., 2002). In particular, studies using combined fMRI-EEG have suggested that the IFG contributes to a top-down process modulating the deviance detection system in the temporal cortices (Doeller et al., 2003; Opitz et al., 2002). Our finding is in line with this previous research. Specifically, the modulation of the MMN induced by higher-

order linguistic information, such as complex phonological rules, indicates that auditory analysis of speech sounds is directly influenced by complex phonological knowledge. One question left open for future research is whether the last type of the phonological rules that are very abstract and have no phonetic grounding rely on the same early processes or whether they are compensated for by later integrational mechanism.

To conclude, we have demonstrated, using the ERP technique, that a phonological rule that cannot be reduced to phonetic principles has an early impact on the auditory analysis of speech sounds. Thus, the signal-to-sound mapping mechanism is sensitive at an early stage not only to low-level phonetic properties but also to context-sensitive phonological rules.

### **Acknowledgements**

This work was supported by the Agence Nationale pour la Recherche (ANR) under Grant ANR-11-JSH2-0061 awarded to C.J. and M.G., Grant ANR-10-LABX-0087 IEC, Grant ANR-10-IDEX-0001-02 PSL\* and by a doctoral fellowship of the Région Ile-de-France awarded to Y.S..

### **References**

- Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I. P., ... Ilmoniemi, R. J. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology*, 35(2), 211–224.
- Anderson, S. R. (1981). Why phonology isn't "natural." *Linguistic Inquiry*, 12(4), 493–539.
- Bach, E., & Harms, R. T. (1972). How do languages get crazy rules. In R. P. Stockwell & R. K. S. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 1–21). Bloomington: Indiana University Press.

- Baldeweg, T., Williams, J. D., & Gruzelier, J. H. (1999). Differential changes in frontal and sub-temporal components of mismatch negativity. *International Journal of Psychophysiology*, 33(2), 143–148.
- Bekinschtein, T., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences of the United States of America*, 106(5), 1672–1677.  
doi:10.1073/pnas.0809667106
- Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, 83(2), 120–131.  
doi:10.1016/j.ijpsycho.2011.08.003
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109(2), 775–794.
- Blaho, S. (2004). Another look at the misbehaving segments of Hungarian voicing assimilation. In C. Piñón & P. Siptár (Eds.), *Approaches to Hungarian 9: Papers from the Düsseldorf Conference*. Budapest: Akadémiai Kiadó.
- Blaho, S. (2008). The syntax of phonology: A radically substance-free approach. *PhD Dissertation*.
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.

- Breen, M., Kingston, J., & Sanders, L. D. (2013). Perceptual representations of phonotactically illegal syllables. *Attention, Perception & Psychophysics*, 75(1), 101–20.  
doi:10.3758/s13414-012-0376-y
- 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. doi:10.1016/j.cognition.2009.02.013
- Buckley, E. (2000). On the naturalness of unnatural rules. In *Proceedings from the Second Workshop on American Indigenous Languages* (pp. 1–14).
- Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, 25(1-2), 53–69.  
doi:10.1080/01690965.2012.705006
- Clayards, M., Niebuhr, O., & Gaskell, M. G. (2015). The time course of auditory and language-specific mechanisms in compensation for sibilant assimilation. *Attention, Perception & Psychophysics*, 77(1), 311–328. doi:10.3758/s13414-014-0750-z
- Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. van de Vijver (Eds.), *Variation and Gradience in Phonetics and Phonology* (pp. 265–310). Berlin: Mouton De Gruyter.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8(4), 919–924.
- 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. doi:10.1162/089892900562390

- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*(1), 21–33. doi:10.1016/j.neuroimage.2004.09.039
- Dell, F. (1995). Consonant clusters and phonological syllables in French. *Lingua*, *95*(1-3), 5–26. doi:10.1016/0024-3841(95)90099-3
- 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.
- Deouell, L. Y., Bentin, S., & Giard, M. H. (1998). Mismatch negativity in dichotic listening: evidence for interhemispheric differences and multiple generators. *Psychophysiology*, *35*(4), 355–365. doi:10.1017/S0048577298970287
- Doeller, C. F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., & Schröger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: Neuroimaging and electrophysiological evidence. *NeuroImage*, *20*(2), 1270–1282. doi:10.1016/S1053-8119(03)00389-6
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578. doi:10.1037/0096-1523.25.6.1568
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*(4), 577–583.
- Flagg, E. J., Oram Cardy, J. E., & Roberts, T. P. L. (2006). MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs. *Neuroscience Letters*, *397*(3), 263–268. doi:10.1016/j.neulet.2005.12.034

- Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *The Journal of the Acoustical Society of America*, 99(2), 1161–1173.
- Flemming, E. S. (2002). *Auditory representations in phonology*. London: Routledge.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84.
- 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.  
doi:10.1016/j.clinph.2008.11.029
- Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1632–1647.
- 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.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, 9(3), 317–323. doi:10.1016/0028-3932(71)90027-3
- Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65(4), 575–590.
- Grammont, M. (1933). *Traité de phonétique*. Delagrave, Paris.
- Grau, C., Fuentemilla, L., & Marco-Pallarés, J. (2007). Functional neural dynamics underlying auditory event-related N1 and N1 suppression response. *NeuroImage*, 36(3), 522–531.  
doi:10.1016/j.neuroimage.2007.03.027

- Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: a cross-linguistic study of the perception of dental stop+/l/ clusters. *The Journal of the Acoustical Society of America*, *121*(5), 2899–2914. doi:10.1121/1.2534656
- Hallé, P., & Adda-Decker, M. (2011). Voice assimilation in French obstruents: Categorical or gradient. In J. Goldsmith, E. Hume, & L. Wetzels (Eds.), *Tones and features: A Festschrift for Nick Clements* (pp. 149–175). Amsterdam/New York: De Gruyter.
- Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology. Human Perception and Performance*, *24*(2), 592–608.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(May), 393–402. doi:10.1038/nrn2113
- Iverson, P., Kuhl, P. K., Akahane-yamada, R., & Diesch, E. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. doi:10.1016/S0
- Jemel, B., Achenbach, C., Müller, B. W., Röppke, B., & Oades, R. D. (2002). Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. *Brain Topography*, *15*(1), 13–27.
- Jung, T., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, *14*(3), 166–185.
- Kirchner, R. (2013). *An effort based approach to consonant lenition*. London: Routledge.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857.

- Kuhl, P., Williams, K., & Lacerda, F. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608. doi:10.1126/science.1736364
- Liebenthal, E., Sabri, M., Beardsley, S. a., Mangalathu-Arumana, J., & Desai, a. (2013). Neural Dynamics of Phonological Processing in the Dorsal Auditory Stream. *Journal of Neuroscience*, 33(39), 15414–15424. doi:10.1523/JNEUROSCI.1511-13.2013
- Lipski, J. (1989). /S/-Voicing in Ecuadoran Spanish: Patterns and principles of consonantal modification. *Lingua*, 79(1), 49–71.
- Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., ... Friston, K. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, 2011, 852961. doi:10.1155/2011/852961
- Mann, V. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28(5), 407–412. doi:10.3758/BF03204884
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024
- Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, 34(4), 338–348.
- Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: evidence for early compensation. *Perception & Psychophysics*, 65(6), 956–969.
- Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology* (2006), 59(8), 1395–424. doi:10.1080/17470210500198726
- Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, 30(3), 451–479. doi:10.1207/s15516709cog0000\_57



- Myers, E. B. (2014). Emergence of category-level sensitivities in non-native speech sound learning. *Frontiers in Neuroscience*, 8(August), 1–11. doi:10.3389/fnins.2014.00238
- 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., ... Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432–434.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23(03), 299–325.
- O’Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, 15(12), 1729–1735.
- Opitz, B., Rinne, T., Mecklinger, A., Von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage*, 15(1), 167–174.
- Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, 88(2), 109–123. doi:10.1016/j.ijpsycho.2013.03.015
- Pitt, M. a., & McQueen, J. M. (1998). Is Compensation for Coarticulation Mediated by the Lexicon? *Journal of Memory and Language*, 39(3), 347–370. doi:10.1006/jmla.1998.2571
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148. doi:10.1016/j.clinph.2007.04.019

- Rubach, J. (1996). Nonsyllabic analysis of voice assimilation in Polish. *Linguistic Inquiry*, 27(1), 69–110.
- Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62(6), 437–448.
- Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137, 29–39. doi:10.1016/j.bandl.2014.07.010
- Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, & Computers*, 30(1), 131–145.
- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72(2), 111–123.
- Sharma, A., & Dorman, M. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America*, 107(January), 2697–2703.
- Steriade, D. (1997). *Phonetics in phonology: the case of laryngeal neutralization*. Manuscript. Retrieved from <http://www.linguistics.ucla.edu/people/steriade/papers/PhoneticsInPhonology.pdf>
- Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press.
- Strycharczuk, P., & Simon, E. (2013). Obstruent voicing before sonorants. The case of West-Flemish. *Natural Language & Linguistic Theory*, 31(2), 563–588. doi:10.1007/s11049-013-9189-5

Sussman, E. S. (2007). A New View on the MMN and Attention Debate: The Role of Context in Processing Auditory Events. *Journal of Psychophysiology*, 21(3), 164–175.

doi:10.1027/0269-8803.21.3.xxx

Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: a magnetoencephalography study. *PloS One*, 4(2), e4452. doi:10.1371/journal.pone.0004452

Tiitinen, H., Salminen, N. H., Palomäki, K. J., Mäkinen, V. T., Alku, P., & May, P. J. C. (2006). Neuromagnetic recordings reveal the temporal dynamics of auditory spatial processing in the human cortex. *Neuroscience Letters*, 396(1), 17–22. doi:10.1016/j.neulet.2005.11.018

Wheeler, M. W. (1986). Catalan sandhi phenomena. In A. Henning (Ed.), *Sandhi phenomena in the languages of Europe* (pp. 475–488). Berlin: Mouton De Gruyter.

doi:10.1515/9783110858532.475

Woldorff, M. G., Hillyard, S. A., Gallen, C. C., Hampson, S. R., & Bloom, F. E. (1998).

Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, 35(3), 283–292.

## Appendix: Complete list of stimuli used in the study

### A.1) *ofbe*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>ofbe - ofbe - ofbe - ofbe</i>	<i>ofbe</i>
		<i>ovbe - ovbe - ovbe - ovbe</i>	<i>ovbe</i>
	voicing-change	<i>ofbe - ofbe - ofbe - ofbe</i> <i>ovbe - ovbe - ovbe - ovbe</i>	<i>ovbe</i> <i>ofbe</i>
Unviable	no-change	<i>ofne - ofne - ofne - ofne</i>	<i>ofne</i>
		<i>ovne - ovne - ovne - ovne</i>	<i>ovne</i>
	voicing-change	<i>ofne - ofne - ofne - ofne</i> <i>ovne - ovne - ovne - ovne</i>	<i>ovne</i> <i>ofne</i>
	place-change (filler)	<i>ofbe - ofbe - ofbe - ofbe</i>	<i>osbe</i>
		<i>ovbe - ovbe - ovbe - ovbe</i>	<i>ozbe</i>
		place-change (filler)	<i>ofne - ofne - ofne - ofne</i>
<i>ovne - ovne - ovne - ovne</i>			<i>ozne</i>

### A.2) *ikdo*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>ikdo - ikdo - ikdo - ikdo</i>	<i>ikdo</i>
		<i>igdo - igdo - igdo - igdo</i>	<i>igdo</i>
	voicing-change	<i>ikdo - ikdo - ikdo - ikdo</i> <i>igdo - igdo - igdo - igdo</i>	<i>igdo</i> <i>ikdo</i>
Unviable	no-change	<i>ikmo - ikmo - ikmo - ikmo</i>	<i>ikmo</i>
		<i>igmo - igmo - igmo - igmo</i>	<i>igmo</i>
	voicing-change	<i>ikmo - ikmo - ikmo - ikmo</i> <i>igmo - igmo - igmo - igmo</i>	<i>igmo</i> <i>ikmo</i>
	place-change (filler)	<i>ikdo - ikdo - ikdo - ikdo</i>	<i>ipdo</i>
		<i>igdo - igdo - igdo - igdo</i>	<i>ibdo</i>
		place-change (filler)	<i>ikmo - ikmo - ikmo - ikmo</i>
<i>igmo - igmo - igmo - igmo</i>			<i>ibmo</i>

A.3) *usda*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>usda - usda - usda - usda</i>	<i>usda</i>
		<i>uzda - uzda - uzda - uzda</i>	<i>uzda</i>
	voicing-change	<i>usda - usda - usda - usda</i>	<i>uzda</i>
		<i>uzda - uzda - uzda - uzda</i>	<i>usda</i>
	place-change (filler)	<i>usda - usda - usda - usda</i>	<i>ushda</i> [uʃ da]
		<i>uzda - uzda - uzda - uzda</i>	<i>uzhda</i> [uʒ da]
Unviable	no-change	<i>usra - usra - usra - usra</i>	<i>usra</i> [usʁ a]
		<i>uzra - uzra - uzra - uzra</i>	<i>uzra</i>
	voicing-change	<i>usra - usra - usra - usra</i>	<i>uzra</i>
		<i>uzra - uzra - uzra - uzra</i>	<i>usra</i>
	place-change (filler)	<i>usra - usra - usra - usra</i>	<i>ushra</i>
		<i>uzra - uzra - uzra - uzra</i>	<i>uzhra</i>

A.4) *akvi*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>akvi - akvi - akvi - akvi</i>	<i>akvi</i>
		<i>agvi - agvi - agvi - agvi</i>	<i>agvi</i>
	voicing-change	<i>akvi - akvi - akvi - akvi</i>	<i>agvi</i>
		<i>agvi - agvi - agvi - agvi</i>	<i>akvi</i>
	place-change (filler)	<i>akvi - akvi - akvi - akvi</i>	<i>atvi</i>
		<i>agvi - agvi - agvi - agvi</i>	<i>advi</i>
Unviable	no-change	<i>akni - akni - akni - akni</i>	<i>akni</i>
		<i>agni - agni - agni - agni</i>	<i>agni</i>
	voicing-change	<i>akni - akni - akni - akni</i>	<i>agni</i>
		<i>agni - agni - agni - agni</i>	<i>akni</i>
	place-change (filler)	<i>akni - akni - akni - akni</i>	<i>atni</i>
		<i>agni - agni - agni - agni</i>	<i>adni</i>

## Captions to Figures

**Figure 1.** Spectrograms showing representative examples of precursor-test pairs in each context (viable and unviable) and each condition (no-change and voicing-change).

**Figure 2.** Performance of French participants on the mismatch detection task (N = 25). These graphs show mean error rates (A) and mean reaction times (B) as a function of Condition (no-change and voicing-change) and Context (viable and unviable). Error bars give the standard error across participants.

**Figure 3.** Performance of American English participants on the mismatch detection task (N = 12). These graphs show mean error rates (A) and mean reaction times (B) as a function of Condition (no-change and voicing-change) and Context (viable and unviable). Error bars give the standard error across participants.

**Figure 4.** Electrophysiological results. Two spatio-temporal clusters revealing a significant Condition  $\times$  Context interaction. The first cluster consists of left central sensors from 140 to 240 ms after the onset of the deviation (A). The second one consists of centro-posterior sensors from 360 to 520 ms after the onset of the deviation (B). For each cluster, the left panel shows the location of the sensors within the cluster (in red). The middle panel shows evoked potentials averaged across these sensors, for each context and condition. The zero point on the x-axis refers to the onset of the deviation between the no-change and the voicing-change items. Black arrows indicate the onset of the test stimuli. Grey shades mark the time-window of the cluster. The right panel shows the mean evoked potentials across sensors and within the time-window for each condition and context. Error bars give the standard error across participants (N=25).

**Figure 5.** Scalp map of *t*-values over time for the voicing-change *minus* no-change subtraction in viable (*top*) and unviable (*bottom*) contexts. Each scalp map represents mean *t*-values over a 100 ms time-window.

767 **Appendix: Complete list of stimuli used in the study**

768 A.1) *ofbe*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>ofbe - ofbe - ofbe - ofbe</i> <i>ovbe - ovbe - ovbe - ovbe</i>	<i>ofbe</i> <i>ovbe</i>
	voicing-change	<i>ofbe - ofbe - ofbe - ofbe</i> <i>ovbe - ovbe - ovbe - ovbe</i>	<i>ovbe</i> <i>ofbe</i>
	place-change (filler)	<i>ofbe - ofbe - ofbe - ofbe</i> <i>ovbe - ovbe - ovbe - ovbe</i>	<i>osbe</i> <i>ozbe</i>
Unviable	no-change	<i>ofne - ofne - ofne - ofne</i> <i>ovne - ovne - ovne - ovne</i>	<i>ofne</i> <i>ovne</i>
	voicing-change	<i>ofne - ofne - ofne - ofne</i> <i>ovne - ovne - ovne - ovne</i>	<i>ovne</i> <i>ofne</i>
	place-change (filler)	<i>ofne - ofne - ofne - ofne</i> <i>ovne - ovne - ovne - ovne</i>	<i>osne</i> <i>ozne</i>

769 A.2) *ikdo*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>ikdo - ikdo - ikdo - ikdo</i> <i>igdo - igdo - igdo - igdo</i>	<i>ikdo</i> <i>igdo</i>
	voicing-change	<i>ikdo - ikdo - ikdo - ikdo</i> <i>igdo - igdo - igdo - igdo</i>	<i>igdo</i> <i>ikdo</i>
	place-change (filler)	<i>ikdo - ikdo - ikdo - ikdo</i> <i>igdo - igdo - igdo - igdo</i>	<i>ipdo</i> <i>ibdo</i>
Unviable	no-change	<i>ikmo - ikmo - ikmo - ikmo</i> <i>igmo - igmo - igmo - igmo</i>	<i>ikmo</i> <i>igmo</i>
	voicing-change	<i>ikmo - ikmo - ikmo - ikmo</i> <i>igmo - igmo - igmo - igmo</i>	<i>igmo</i> <i>ikmo</i>
	place-change (filler)	<i>ikmo - ikmo - ikmo - ikmo</i> <i>igmo - igmo - igmo - igmo</i>	<i>ipmo</i> <i>ibmo</i>

770

771

772

773 A.3) *usda*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>usda - usda - usda - usda</i>	<i>usda</i>
		<i>uzda - uzda - uzda - uzda</i>	<i>uzda</i>
	voicing-change	<i>usda - usda - usda - usda</i>	<i>uzda</i>
		<i>uzda - uzda - uzda - uzda</i>	<i>usda</i>
	place-change (filler)	<i>usda - usda - usda - usda</i>	<i>ushda</i> [uʃ da]
		<i>uzda - uzda - uzda - uzda</i>	<i>uzhda</i> [uʒ da]
Unviable	no-change	<i>usra - usra - usra - usra</i>	<i>usra</i> [usʁ a]
		<i>uzra - uzra - uzra - uzra</i>	<i>uzra</i>
	voicing-change	<i>usra - usra - usra - usra</i>	<i>uzra</i>
		<i>uzra - uzra - uzra - uzra</i>	<i>usra</i>
	place-change (filler)	<i>usra - usra - usra - usra</i>	<i>ushra</i>
		<i>uzra - uzra - uzra - uzra</i>	<i>uzhra</i>

774 A.4) *akvi*

Context	Condition	Precursor items	Test item
Viable	no-change	<i>akvi - akvi - akvi - akvi</i>	<i>akvi</i>
		<i>agvi - agvi - agvi - agvi</i>	<i>agvi</i>
	voicing-change	<i>akvi - akvi - akvi - akvi</i>	<i>agvi</i>
		<i>agvi - agvi - agvi - agvi</i>	<i>akvi</i>
	place-change (filler)	<i>akvi - akvi - akvi - akvi</i>	<i>atvi</i>
		<i>agvi - agvi - agvi - agvi</i>	<i>advi</i>
Unviable	no-change	<i>akni - akni - akni - akni</i>	<i>akni</i>
		<i>agni - agni - agni - agni</i>	<i>agni</i>
	voicing-change	<i>akni - akni - akni - akni</i>	<i>agni</i>
		<i>agni - agni - agni - agni</i>	<i>akni</i>
	place-change (filler)	<i>akni - akni - akni - akni</i>	<i>atni</i>
		<i>agni - agni - agni - agni</i>	<i>adni</i>

775

776



Figure 1  
[Click here to download high resolution image](#)

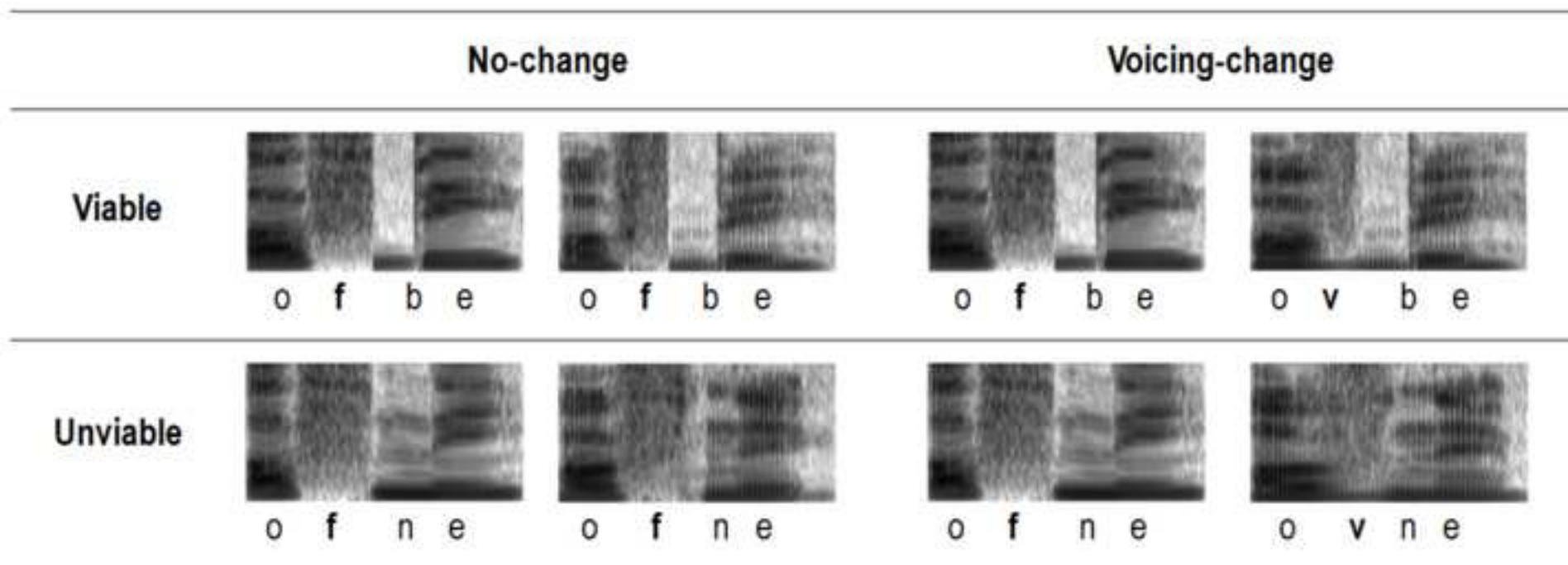


Figure 2  
[Click here to download high resolution image](#)

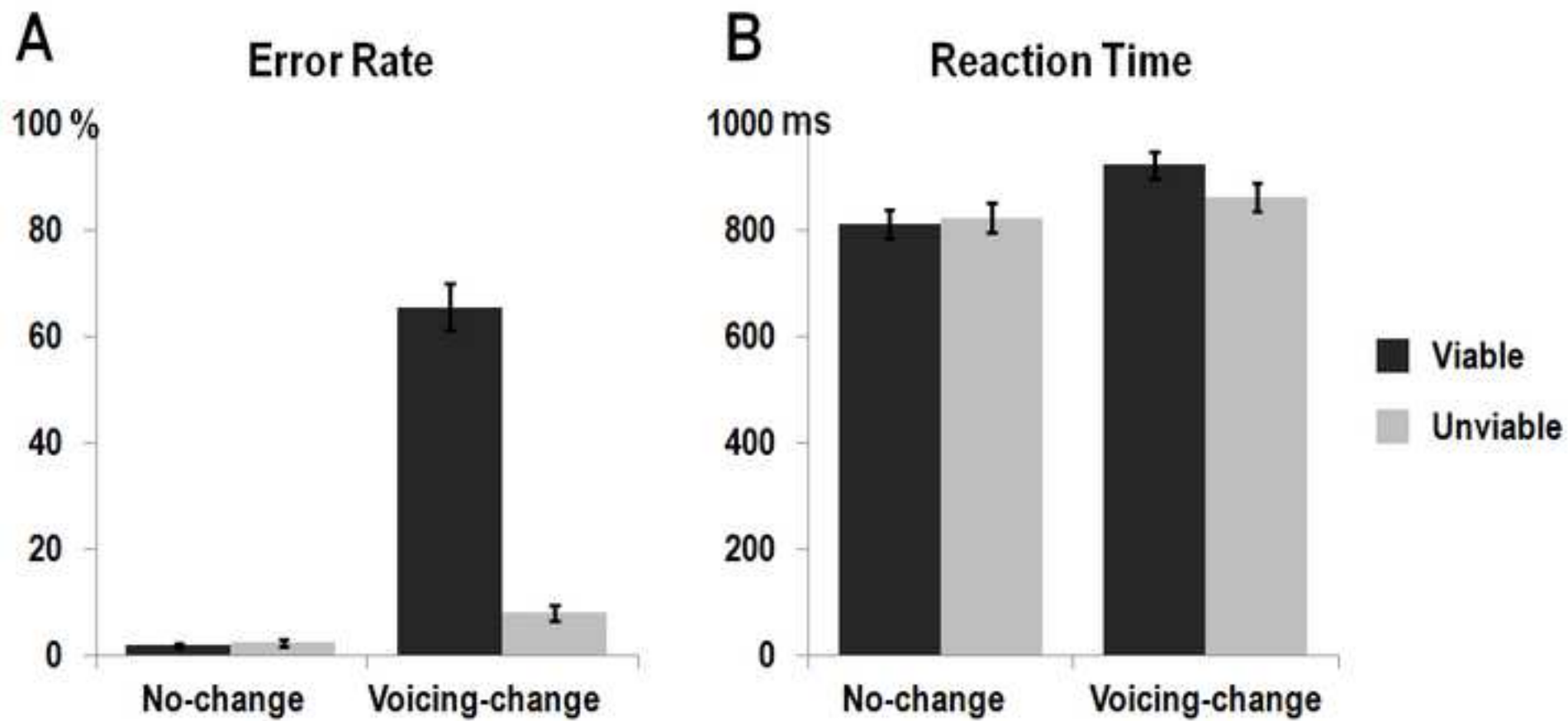


Figure 3  
[Click here to download high resolution image](#)

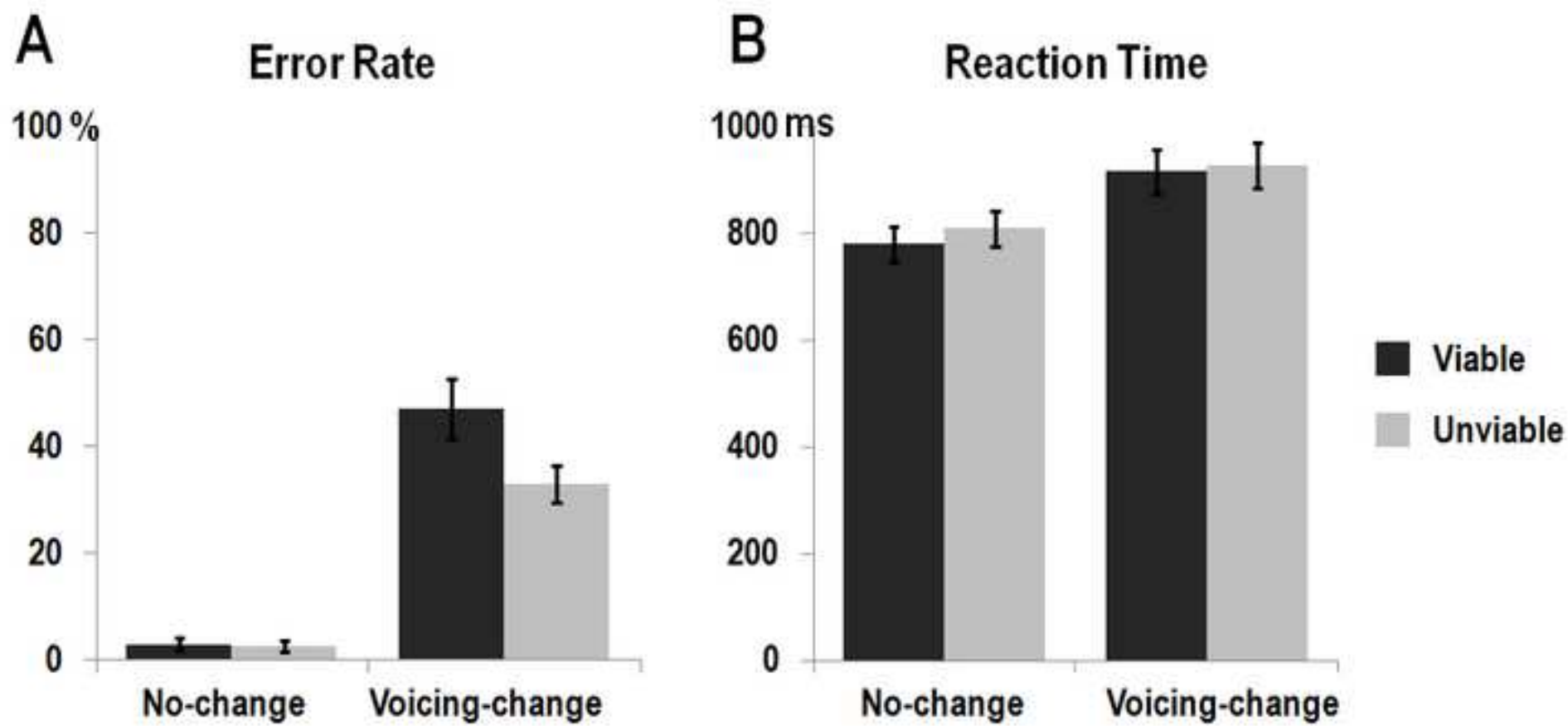


Figure 4  
[Click here to download high resolution image](#)

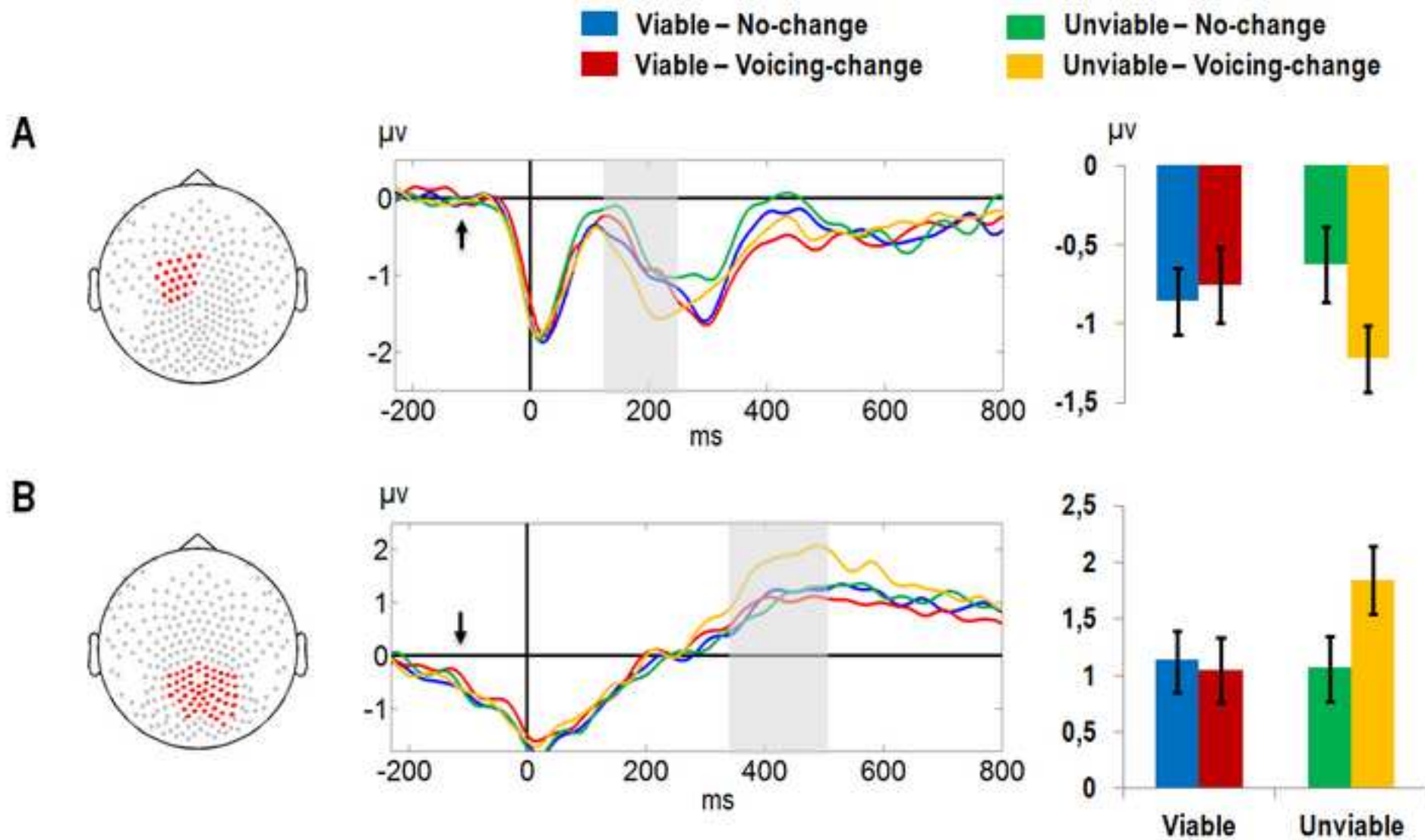


Figure 5  
[Click here to download high resolution image](#)

