Comparing the selectivity of vowel representations in cortical auditory vs. motor areas: A repetition-suppression study

Marjorie Dole, Coriandre Emmanuel Vilain, Célise Haldin, Monica Baciu, Emilie Cousin, Laurent Lamalle, Hélène Loevenbruck, Anne Vilain, Jean-Luc Schwartz

Comparing the selectivity of vowel representations in cortical auditory vs. motor areas: A repetition-suppression study

Marjorie Dole[(1)], Coriandre Vilain[(1)], Célise Haldin[(2)], Monica Baciu[(2)], Emilie Cousin[(2, 3)], Laurent Lamalle[(3)], Hélène Lœvenbruck[(2)], Anne Vilain[(1)], Jean-Luc Schwartz[(1)]

(1) Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France
(2) Univ. Grenoble Alpes, CNRS, LPNC, Grenoble, France
(3) Univ. Grenoble Alpes, Inserm, CHU Grenoble Alpes, IRMaGe, Grenoble, France

Authors' e-mail addresses:

marjorie.dole@cea.fr
coriandre.vilain@gipsa-lab.grenoble-inp.fr
celise.haldin@univ-grenoble-alpes.fr
monica.baciu@univ-grenoble-alpes.fr
emilie.cousin@univ-grenoble-alpes.fr
laurent.lamalle@univ-grenoble-alpes.fr
helene.loevenbruck@univ-grenoble-alpes.fr
anne.vilain@gipsa-lab.grenoble-inp.fr
jean-luc.schwartz@gipsa-lab.grenoble-inp.fr

Corresponding author: Marjorie Dole (marjorie.dole@cea.fr)

Abstract

A computational model of speech perception, COSMO (Laurent et al., 2017), predicts that speech sounds should evoke both auditory representations in temporal areas and motor representations mainly in inferior frontal areas. Importantly, the model also predicts that auditory representations should be narrower, i.e. more focused on typical stimuli, than motor representations which would be more tolerant of atypical stimuli. Based on these assumptions, in a repetition-suppression study with functional magnetic resonance imaging data, we show that a sequence of 4 identical vowel sounds produces lower cortical activity (i.e. larger suppression effects) than if the last sound in the sequence is slightly varied. Crucially, temporal regions display an increase in cortical activity even for small acoustic variations, indicating a release of the suppression effect even for stimuli acoustically close to the first stimulus. In contrast, inferior frontal, premotor, insular and cerebellar regions show a release of suppression for larger acoustic variations. This "auditory-narrow motor-wide" pattern for vowel stimuli adds to a number of similar findings on consonant stimuli, confirming that the selectivity of speech sound representations in temporal auditory areas is narrower than in frontal motor areas in the human cortex.

Keywords: Repetition-suppression, selectivity, auditory representations, motor representations, vowel processing

1. Introduction

1.1. Questioning accounts of sensory-motor interactions in speech perception

The hypothesis of a role for sensory-motor interactions in speech perception was introduced with the Motor Theory of Speech Perception by Liberman and his colleagues in the Haskins Labs in the 1960s (Liberman et al., 1967), as a way of explaining how the perceptual system resolves the variability in acoustic realizations associated with a given phonological code. The proposal of the Motor Theory of Speech Perception is that the gesture rather than the sound characterizes the phoneme. Consequently, proponents of the Motor Theory of Speech Perception have argued that listeners are able to recover the articulatory (Liberman et al., 1967) or motor (Liberman & Mattingly, 1985) cause of the speech sound during the decoding process and that the articulatory-motor gesture would provide the underlying invariant structure, thus bringing order to the acoustic disorder.

However, this proposal has been hotly debated and criticized, and "auditory theories of speech perception", in which speech decoding does not rely on speech production knowledge or articulatory-motor representations, have been proposed and defended (e.g. Diehl et al., 2004) with a number of experimental and functional arguments (e.g. Kingston & Diehl, 1994; Kluender 1994; Lotto, 2000; Massaro & Oden 1980; Nearey, 1990). This has led to the introduction of perceptuo-motor theories of speech perception, which attempt to integrate and reconcile these different sets of arguments combining auditory processing and motor knowledge into a coherent framework (Schwartz et al., 2012; Skipper et al., 2007).

Over the last twenty years, a large number of experimental data provided by neuroimaging tools have demonstrated the existence of sensory-motor links in the human brain during speech perception tasks (e.g. Benson et al., 2001; Fadiga et al., 2002; Pulvermüller et al., 2006; Watkins et al., 2003; see a review in Skipper et al., 2017), and confirmed that these links do have a potentially causal role in speech perception (d'Ausilio et al., 2009, 2011; Möttönen et al., 2013, 2014; Sato et al., 2009, 2011; see a review in Schomers & Pulvermüller, 2016; and a caveat on the importance of this causal role in Stokes et al., 2019). A striking finding, however, is that motor areas are more activated in noisy (Binder et al., 2004; Du et al., 2014; Zekveld et al., 2006) or in atypical listening conditions (Callan et al., 2004, 2014; Wilson & Iacoboni, 2006), and that their modulatory role in speech perception is more apparent for ambiguous or noisy stimuli (d'Ausilio et al., 2009, 2011; Sato et al., 2011). The reason for this phenomenon, namely the recruitment of motor regions mainly under adverse listening conditions, remains to be understood. Importantly, the Motor Theory of Speech Perception, which states that the motor system would systematically be recruited during the processing of acoustic stimuli for the extraction of invariant decoding cues, does not predict that there should be an increase in motor recruitment for acoustic stimuli presented under adverse conditions.

1.2. The Auditory-Narrow Motor-Wide property in the COSMO computational model

In recent years, we have developed a computational Bayesian model of speech communication, COSMO (Moulin-Frier et al., 2012, 2015), which provides interesting insights into the possible role of the motor system in speech perception under adverse conditions. COSMO (for "Communicating Objects using Sensory-Motor Operations") explores the hypothesis that there are two possible accesses to the phonological code from the incoming sound, namely an "auditory pathway" and a "motor pathway".

The auditory pathway is based on a direct relationship between sounds and phonological categories. This is expressed, in the Bayesian probabilistic framework in which COSMO was designed, by the probability distribution $P_A(S/O)$ where S represents the attributes of the sound (the Sensory input, auditory cues in the following) and O is the object of the communication

between the speaker and the listener (object being conceived in a broad sense, from phonemes to concepts – in the present paper, O stands for the phoneme category). The probability distribution $P_A(S/O)$ is learnt by the child (or by the COSMO computational agent) from speech input provided by the child's caregivers or peers (or the agent's tutors). The learning process for the auditory pathway is direct and therefore simple. It requires the child to associate stimuli and objects – here sounds and phonemes – directly, in a supervised way. In the COSMO framework, the auditory pathway is an implementation of aspects of speech perception that are best explained by auditory theories (Moulin-Frier et al., 2012, 2015). It provides an optimal representation of the stimuli in the environment, with a straightforward relationship from sound to object (Kleinschmidt & Jaeger, 2015).

Conversely, the link between stimuli and objects in the motor pathway summarized in the probability distribution $P_M(S/O)$, is indirect. It is mediated by the recovery of motor commands (or articulatory gestures) M from the sound stimulus, generally referred to as the "acoustic-to-articulatory inversion process". Because of the versatility of the speech apparatus, several combinations of articulatory gestures can result in the same acoustic product. Therefore, the articulatory-to-acoustic relation is many-to-one: there are many possible articulatory/motor solutions for a single acoustic outcome. Recovering a gesture or a motor command from the input sound is therefore an ill-posed problem. COSMO solves this issue within the Bayesian formalism (Moulin-Frier et al., 2012, 2015). Bayesian inversion rules result in what is called a marginalization process, in which the link between stimuli and objects includes all possible values of the motor variable M, according to the following formula:

$$P_M(S/O) \propto \Sigma_M P(M/O) P(S/M) \qquad \textit{(Eq. 1).}$$

In this equation (where α means "proportional to", that is equal modulo a normalization factor), the distribution P(M/O) relates motor gestures to objects and is called the motor repertoire, whereas P(S/M) predicts sensory outputs from motor commands and is called the internal forward model. These two distributions must also be learned by the child. However, the learning process for the motor pathway is more complex than for the auditory pathway and cannot be conceived as supervised, since the motor gestures associated with the incoming sounds are not provided by tutors. It is hypothesized that infants gradually combine endogenous exploration and exogenous tuning (Warlaumount, 2020). Endogenous exploration consists in playing with their vocal tract and learning the relationship between motor commands and acoustic outputs, i.e. the internal forward motor model (or P(S/M)). Exogenous tuning consists in learning to pick the appropriate motor commands to best imitate the acoustic targets in their environment, i.e. in tuning the motor repertoire or P(M/O)). Importantly, the summation over the variable M in Eq. (1) means that the Bayesian resolution of the inversion process (finding an adequate gesture M for a given sound S) consists in actually exploiting all the possible gestures M on the basis of their likelihood.

This learning process has been shown to be computationally tractable (Laurent et al., 2017). However, the complexity of the motor pathway, and particularly of the probabilistic inversion expressed through the marginalization operation in Eq. (1), blurs the relationship between sensory inputs and phonetic categories provided by the $P_M(S/O)$ distribution. As a consequence, Laurent et al. (2017) revealed that, in comparison with the optimal probability distribution in the auditory pathway $P_A(S/O)$, $P_M(S/O)$ is less peaked and less well tuned to the acoustic stimuli provided in the learning process. As illustrated in Figure 1A, a typical sound stimulus in the environment (S1) is associated with a high $P_A(S/O)$ probability in the auditory pathway, whereas it will be associated with a weaker $P_M(S/O)$ probability in the motor pathway.

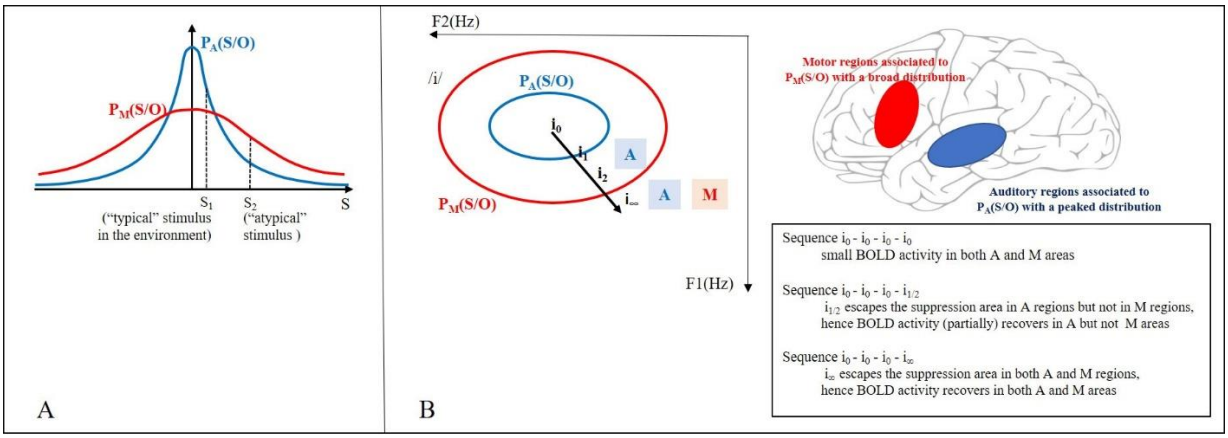

Figure 1. The probability distributions $P_A(S/O)$ and $P_M(S/O)$ and the "Auditory-Narrow Motor-Wide" property in COSMO. A: A "typical" stimulus S_1 is better recognized by the auditory pathway while an "atypical" stimulus S_2 is better recognized by the motor pathway. B: Rationale and hypotheses of the present study.

Laurent et al. (2017) suggested that the objects (O) associated with typical stimuli (S1) are therefore optimally recognized in the auditory pathway. Conversely, the objects related to atypical input stimuli (e.g. noisy or pronounced with an accent never experienced by the agent) should be better identified in the motor pathway, due to the wider distribution $P_M(S/O)$ (stimulus S_2 in Figure 1A). Simulations of identification scores using the COSMO model confirm the presence of these two distinct profiles (Moulin-Frier et al., 2012, Laurent et al., 2017). This pattern has been coined the "Auditory-Narrow Motor-Wide property" (ANMW). To our knowledge, this is the first explanatory account for the increased role of the motor system in noise or under atypical conditions (Barnaud et al., 2018). Importantly, COSMO also considers a perceptuo-motor decoding process based on the fusion of the auditory and motor decoding pathways, in line with the perceptuo-motor theory proposed by Schwartz et al. (2012). This perceptuo-motor system takes the best of the properties of its two components, allowing optimal recognition performance for both typical stimuli (through the auditory branch) and atypical stimuli (though the motor branch).

1.3. The repetition-suppression paradigm adapted to the ANMW property

The objective of the present study is to provide neurocognitive experimental evidence for the ANMW property. For this aim, a repetition-suppression (RS) paradigm was chosen. This paradigm is based on the robust finding that when a given stimulation is repeated several times, the neural response to the stimulus decreases. This phenomenon operates over a wide range of temporal scales and experimental paradigms, and has been reported at the level of single-cell recordings as well as EEG/MEG or fMRI data (Grill-Spector et al., 2006). Importantly, recording repetition-suppression effects in fMRI or MEG provides a way to test which variations in the stimulation lead to a decrease in the cerebral response. In the visual system this has led to the characterization of sensitivity to orientation, color or motion (Engel, 2005; Huk & Heeger, 2002).

The RS paradigm, originally developed in EEG studies, has been transferred to fMRI and adapted to speech for over 20 years (see first studies in Celsis et al., 1999; Zevin & McCandliss, 2005). To describe a typical experiment, Chevillet et al. (2013) investigated fMRI responses to sequences of 2 syllables from a continuum of synthetic stimuli varying between /ga/ and /da/. The presented sequence consisted of either 2 identical syllables, or of a first syllable followed by a slightly different syllable, which either belonged to the same category or to another

category. The RS paradigm was expected to lead to a low neural response when the same stimulus was repeated twice, due to suppression of neuronal activity, whereas the variation in the last stimulus was expected to lead to a higher response (less suppression). The fMRI analysis revealed a significant increase in cortical activity in the left anterior and posterior auditory cortex when the last stimulus was varied but remained within the same category, and a further increase in activity in the left premotor cortex when the second stimulus was varied and changed category.

Similar results with a few variations were reported by Raizada & Poldrack (2007), Myers et al. (2009), Joanisse et al. (2007), Myers & Swan (2012), Altmann et al. (2014), Lawyer & Corina, (2014), Alho et al. (2016). They all concerned plosives in a consonant-vowel sequence, focusing on the categorization of place of articulation or voicing. They showed differences in the set of regions activated for within-category stimulus variation contrasted to between-category variations. Overall, within-category variations mainly induce activity in the temporal cortex (left posterior Superior Temporal Gyrus pSTG and anterior Middle Temporal Gyrus aMTG in Chevillet et al., 2013; left Superior Temporal Gyrus STG in Alho et al., 2016; and right STG in Myers et al., 2009). Only Myers et al. (2009) also found activity in the parietal and frontal cortex under this condition (right Supramarginal Gyrus SMG and bilateral Inferior Frontal Gyrus IFG). Between-category variations induce additional activity in temporal (left posterior STG in Zevin & McCandliss, 2005; left pSTS/STG in Altmann et al., 2014; left STS-MTG in Joanisse et al., 2007; bilateral STG in Lawyer & Corina, 2014), parietal (left SMG in Celsis et al., 1999; Raizada and Poldrack, 2007; left Inferior Parietal Cortex IPC in Joanisse et al., 2007) and frontal regions (left Pre-Motor Cortex, PMC in Chevillet et al., 2013; and left IFG in Myers et al., 2009 and Alho et al., 2016).

One single study has explored vowel stimuli using this paradigm. Altmann et al. (2014) compared MEG responses to pairs of consonant-vowel stimuli which were either identical or differed by the consonant (e.g. /ba/ vs. /da/) or by the vowel (e.g. /ba/ vs. /bo/). Pairs of different stimuli elicited greater activation than pairs of identical stimuli in the left STG. However, while this increased activity in the left STG was only observed when the difference involved crossing a categorical boundary in the case of consonants (i.e. in the between-category condition), it was shown in both the within- and between-category cases for vowels.

Therefore, overall, there is a trend to find increased activity in frontal areas in the between-category condition compared to the intra-category condition, but this trend is variable between studies, and it is not reported in the one study concerning vowels. Most of these data are interpreted in terms of pre-categorial vs. categorial processes, with regions specifically activated during a stimulus variation associated with a category change being interpreted as playing a specific role in the categorization process *per se*. Authors diverge on the anatomical localization of these regions, however (e.g. frontal for Myers, 2009 or Chevillet et al., 2013; vs. temporal or temporo-parietal in Joanisse et al., 2007, or Altmann et al., 2014). The lack of effect of category change for vowels is interpreted by Altmann et al. (2014) as related to the less categorial perception of vowels (Schouten et al., 2003).

In the present study, we specifically attempt to explore the ANMW hypothesis using the RS paradigm on a vowel perception task. To operationalize the ANMW hypothesis in neurocognitive terms, we start from the architecture proposed in COSMO (Barnaud et al., 2018), according to which the auditory knowledge stored in $P_A(S/O)$ is represented in temporal regions (superior temporal gyrus and sulcus) while the motor knowledge leading to the distribution $P_M(S/O)$ is represented in frontal regions including the motor and premotor cortices and the inferior frontal gyrus. We then further assume that the distributions can be translated into patterns of neural activity, with differential neural selectivity within auditory and motor regions. Selectivity is related to the set of stimuli that provide a given response in a neural

channel Ch. It can be described by a probability distribution P(S/Ch). Our hypothesis is that the selectivity distributions are sharp in temporal auditory regions, whereas they are wide in frontal motor regions, because of the marginalization process described in Eq. (1). As a result, a small variation in the acoustic stimulus leads to a change in the neural channel in auditory regions, while a larger variation would be required for a change in motor regions. Therefore, a small acoustic change is expected to result in release of suppression in auditory but not in motor regions.

In the present study, we use stimuli varying from prototypical vowels (i.e. vowels that will be unambiguously identified by listeners, e.g. i_0), to vowels progressively deviating from the prototype (i_1, i_2, i_3, etc), to vowels clearly distant from the prototype (i_∞). The design and assumptions of the study are presented in Figure 1B. In the top left corner of the plot, we present the standard deviation ellipses of the hypothetical sensory distributions $P_A(S/O)$ and $P_M(S/O)$ associated with the distribution of responses to sensory inputs around a prototypical /i/ (i_0), with a broader ellipse for $P_M(S/O)$ than for $P_A(S/O)$, in line with the ANMW hypothesis. The upper right part of the plot provides the neuroanatomical translation of the ANMW hypothesis, with a sharp distribution $P_A(S/O)$ in auditory regions and a wide distribution $P_M(S/O)$ in motor regions. The text block at the bottom right of the plot provides the experimental RS hypothesis. In a sequence of 4 identical vowels such as i_0-i_0-i_0-i_0, the fMRI-BOLD activity is expected to be weak due to suppression. When the last stimulus is slightly modified (e.g. i_1 or i_2) BOLD activity is expected to partially recover, but recovery would likely occur for smaller acoustic modifications in auditory regions compared to motor regions. Therefore, the prediction is that activity could be enhanced in auditory regions for sequences such as i_0-i_0-i_0-$i_{1/2}$ where i_1 or i_2 are close to i_0. In motor regions, activity should only increase for sequences such as i_0-i_0-i_0-i_∞, where i_∞ is sufficiently distant from i_0. In other words, this study aims at revealing a differential involvement of temporal auditory regions and frontal motor regions in the RS paradigm for vowel stimuli.

2. Material and methods

2.1. Participants

Three groups of participants were recruited for this study, two groups for pilot behavioral tests to select suitable stimuli for the repetition-suppression task, and a third group for the fMRI repetition-suppression task itself. Participants in the three groups were all different. The study was approved by the local ethics committee (CPP Sud Est V, ID RCB 2019-A00293-48) and registered on ClinicalTrials.gov (NCT number 03102983). Written informed consent was obtained from all participants before the study in accordance with Helsinki guidelines.

For the first set of pilot behavioral tests, a first group of 12 French listeners (G1 in the following, age between 23 and 58, 9 female subjects) and a second group of 7 participants (G2 in the following, age between 21 and 52, 5 female subjects), all from Université Grenoble Alpes, participated in the experiments. All participants were native speakers of French with normal audition and normal or corrected-to-normal vision, and reported no history of psychiatric or neurological disorders.

For the fMRI repetition-suppression task, 19 healthy participants (10 females) were initially included but 18 were retained for the analyses, aged 18 to 39 years (G3 in the following; mean age: 24.22, S.D.: 6.29, 9 females). One female participant was excluded from the analyses for methodological reasons. The participants received compensatory retribution for their participation. All were native speakers of French with normal or corrected-to-normal vision. All were right-handed (Edinburgh Laterality Inventory; Oldfield, 1971) and had audiometric

pure-tone thresholds not exceeding 25 dB HL at frequencies 250 Hz, 500 Hz, 750 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 3000 Hz, 4000 Hz and 6000 Hz. Participants reported no history of psychiatric or neurological disorders.

2.2. Stimuli

2.2.1. Synthetic vowel sounds

The experiment included vowel stimuli in the /i/ (front unrounded) and /u/ (back rounded) regions. Stimuli were synthetic, obtained by a Klatt formant synthesizer (Klatt & Klatt, 1990) available in Praat (Boersma & Weenink, 2021). All stimuli had the same 200 ms duration and the same pitch (decreasing linearly from 147 Hz at the beginning of the stimulus to 122 Hz at the end), and only the formant values were varied. The formant values were adjusted so as to start at the top of the vowel space, with a minimum F1 value around 250 Hz (stimuli i_0 and u_0), and then moving along the boundaries of the vowel space in regular F1, F2 and F3 steps, in a perceptual Bark scale (Schroeder et al., 1979). This provided two sets of stimuli in the front unrounded and in the back rounded region respectively. In the first set, stimuli i_0 to i_{10} had F1 increasing from 2.6 to 4.6 Bark in 0.2 Bark steps, and F2 and F3 decreasing in 0.1 Bark steps from 13.3 to 12.3 and from 15.6 to 14.6 Barks, respectively. In the second set, stimuli u_0 to u_{10} had F1 increasing from 2.6 to 4.6 Bark in 0.2 Bark steps, and F2 and F3 increasing in 0.1 Bark steps from 5.8 to 6.8 and from 13.0 to 14.0 Barks, respectively. Three additional stimuli were prepared so as to provide stimuli that were significantly different from i_0 and u_0: namely stimuli i_∞ and u_∞, with an F1 value at 5.8 Barks, and a control stimulus a. All formant values are provided in Table 1.

2.2.2. Selection of suitable individual vowel sounds

Given the way the sounds were synthesized, the expectation was that stimuli i_0 and u_0 would be respectively perceived as /i/ and /u/. Then, it was expected that front vowel stimuli from i_1 to i_{10} would gradually be perceived as mid-high and mid-open vowels /e/ and /ɛ/. Back vowel stimuli from u_1 to u_{10}, were expected to be more and more identified as mid-high /o/ and mid-open /ɔ/. Stimuli i_∞, u_∞ and a, were expected to be respectively perceived as /ɛ/, /ɔ/ and /a/. To assess how the synthetic stimuli were actually perceived, a categorization test was performed by the first group of 12 listeners, G1, with a forced choice procedure, among the following answers: /i/, /e/, /ɛ/, /u/, /o/, /ɔ/ and /a/. The participants were presented with 9 repetitions of each of the 25 stimuli (i_0 to i_{10}, u_0 to u_{10}, i_∞, u_∞ and a), with all stimuli presented in a random order. The average categorization percentages for the stimuli i_0 to i_{10} and u_0 to u_{10} are displayed in Figure 2. They show that i_0 and u_0 are indeed good prototypical exemplars of the high vowels /i/ and /u/, with a switch to the mid-high vowels /e/ and /o/ around i_4 and u_4, whereas the last stimuli i_{10} and u_{10} were categorized as mid-open vowels /ɛ/ and /ɔ/ respectively. As expected, i_∞ and a were 100% categorized as /ɛ/ and /a/, whereas u_∞ was actually categorized halfway between /ɔ/ and /a/.

Since we had no clear predictions about the optimal distance from the prototypical i0 and u0 stimuli that would allow partial recovery from repetition-suppression, we decided to further evaluate three pairs of stimuli. The first pair consists of i_1 and u_1, with an F1 distance with prototypes equal to 0.2 Bark, thus quite close to prototypes. The sounds in the second pair, i_2 and u_2, correspond to an F1 distance with prototypes equal to 0.4 Bark, which is likely to allow clear discrimination already, but should still remain within the high /i/ or /u/ category, given the categorization responses displayed in Figure 2. Finally, the third pair involves two stimuli even further away from the prototypes, i.e. i_4 and u_4. These sounds belong to the mid-high category, given that i_4 was perceived as /e/ and that u_4 was perceived at the boundary between /u/ and /o/.

To characterize the perceptual distance between the selected stimuli {i_0, i_1, i_2, i_4} on the front side and {u_0, u_1, u_2, u_4} on the back side, we performed two additional perceptual tests on these two sets. The first test consisted in an ABX discrimination test with A or B the i_0 (resp. u_0) prototype, B or A one stimulus in the {i_0, i_1, i_2, i_4} (resp. {u_0, u_1, u_2, u_4}) set, and X one of the two stimuli in the pair. The task was to decide whether X was closer to A or B. There were 10 occurrences of each pair with X randomly selected within the pair. All stimuli in the test were grouped in a single block and presented in a random order, different for each participant. The second additional test was a quantitative assessment of the perceptual distance to the prototype. For this test we added the stimuli i_∞, u_∞. The task consisted in listening to pairs such as i_0-i_X or u_0-u_X, X being a value within the {0, 1, 2, 4, ∞} set, and in providing a subjective evaluation of the perceptual distance of the two stimuli in the pair on a scale varying between 1 (no audible difference) and 7 (maximal possible perceptual distance). For each pair, the two orders of stimuli in the pair were presented 5 times, with all pairs mixed in a single block in a random order different for each participant.

The 7 participants in group G2 participated in the two additional tests. The results are displayed in Figure 3. They show that stimuli i_1 and u_1 are poorly discriminated from the prototypes in the ABX test and perceptually quite close (especially for u_1, with discrimination from u_0 close to chance and distance close to $_1$, meaning no audible difference). Stimuli i_2 and u_2 are rather well discriminated, although far from perfectly, but display a rather small perceptual distance (around 2). Stimuli i_4 and u_4 are clearly discriminated (although not perfectly) and display rather large distances (around 5), though still far from the maximal distance obtained for stimuli i_∞ or u_∞.

This set of perceptual evaluations led us to discard stimuli i_1 and u_1, which might not lead to any recovery from the repetition-suppression effect, and to select the set of stimuli {i_0, i_2, i_4, i_∞} on one side and {u_0, u_2, u_4, u_∞} on the other side of the vowel (F1, F2) space for the repetition-suppression fMRI experiment to be presented in the next section.

Name	F1 (Hz)	F2 (Hz)	F3 (Hz)	F1 (Bk)	F2 (Bk)	F3 (Bk)
i_0	247	2125	2980	2.6	13.3	15.6
i_1	267	2090	2940	2.8	13.2	15.5
i_2	287	2060	2900	3	13.1	15.4
i_3	307	2030	2855	3.2	13,0	15.3
i_4	330	2000	2810	3.4	12.9	15.2
i_5	350	1970	2770	3.6	12.8	15.1
i_6	370	1940	2730	3.8	12.7	15,0
i_7	392	1910	2690	4	12.6	14.9
i_8	414	1880	2650	4.2	12.5	14.8
i_9	436	1855	2615	4.4	12.4	14.7
i_{10}	459	1830	2575	4.6	12.3	14.6
u_0	247	602	2030	2.6	5.8	13,0
u_1	267	615	2060	2.8	5.9	13.1
u_2	287	628	2090	3	6,0	13.2
u_3	307	641	2125	3.2	6.1	13.3
u_4	330	654	2155	3.4	6.2	13.4
u_5	350	667	2190	3.6	6.3	13.5
u_6	370	681	2220	3.8	6.4	13.6
u_7	392	694	2255	4	6.5	13.7
u_8	414	708	2290	4.2	6.6	13.8
u_9	436	721	2320	4.4	6.7	13.9

u_{10}	459	735	2355	4.6	6.8	14,0
i_{∞}	602	1660	2500	5.8	11.2	14.2
u_{∞}	602	838	2500	5.8	7.5	14.2
a	735	1086	2500	6.8	9,0	14.2

Table 1 – (F1, F2, F3) values in Hz and in Bark for the 25 stimuli. F4 was fixed at 3560 Hz (16.8 Bark). All Bark values are computed from Hz values by the formula in Schroeder et al. (1979).

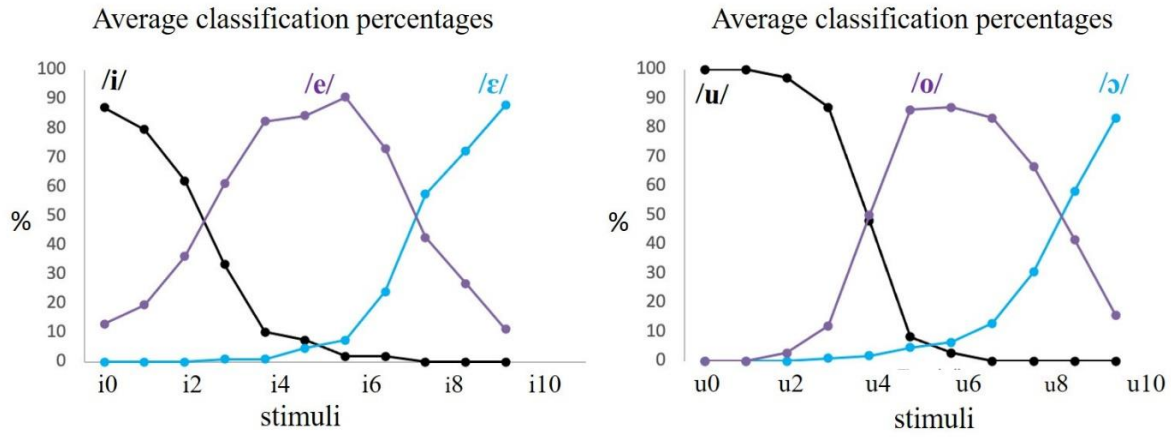

Figure 2. Categorization responses for front unrounded stimuli i0 to i10 (left) and back rounded stimuli u0 to u10 (right) averaged on the 12 listeners. Error bars correspond to standard error of the mean.

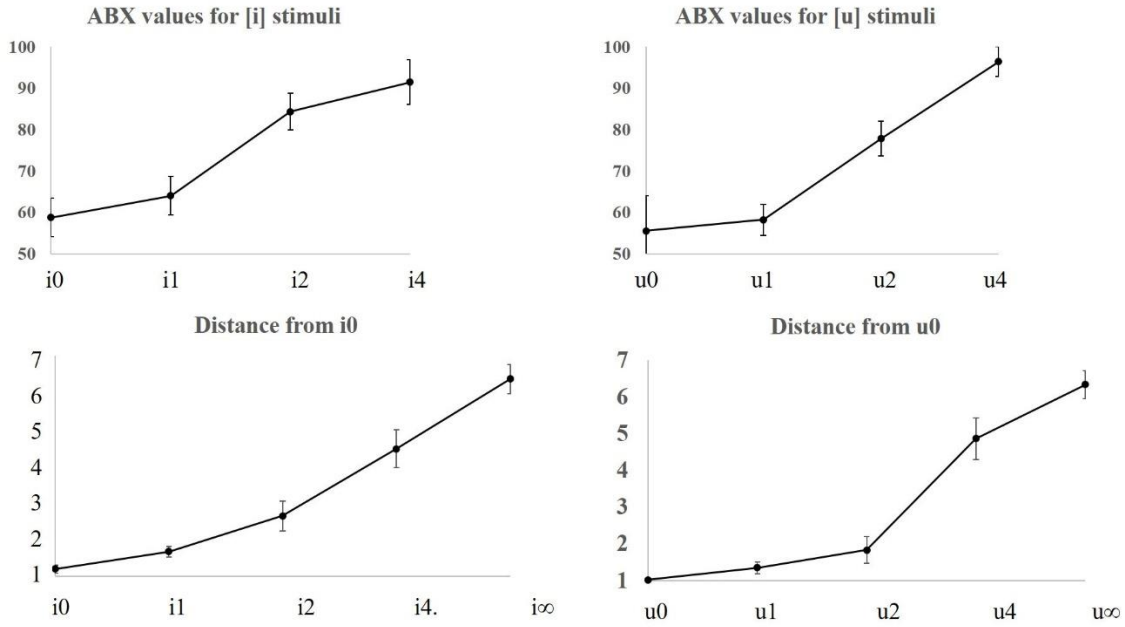

Figure 3. Top: Discrimination from i_0 (left) and u_0 (right) in an ABX experiment, 50% corresponds to chance and 100% to perfect discrimination. Bottom: Normalized perceptual distance from i_0 (left) and u_0 (right), from1 (no audible difference) to 7 (maximal distance). Error bars correspond to standard error of the mean.

2.2.3. Selected trains of stimuli

Finally, the stimuli for the RS task were composed of trains of 4 vowels consisting in 3 identical prototypical vowels (i_0 or u_0), followed by a fourth vowel that could be either the same prototypical stimulus (i_0 / u_0) or a different stimulus (i_2, i_4, i_∞ / u_2, u_4, u_∞), with increasing distance between the first 3 vowels and the last one, resulting in the following four conditions: Repetition (REP): i_0-i_0-i_0-i_0 / u_0-u_0-u_0-u_0, Non-Repetition 1 (NREP1): i_0-i_0-i_0-i_2 / u_0-u_0-u_0-u_2, Non-Repetition 2 (NREP2): i_0-i_0-i_0-i_4 / u_0-u_0-u_0-u_4, and Non-Repetition ∞ (NREP∞): i_0-i_0-i_0-i_∞ / u_0-u_0-u_0-u_∞.

Catch trials consisting of trains of i_0-i_0-i_0-a and u_0-u_0-u_0-a stimuli were also added to maintain participants' attention on the auditory stimuli. In each train of vowels, the vowel duration was 200 ms and the interstimulus interval was 50 ms, for a total stimulus duration of 950 ms.

2.3. fMRI repetition-suppression protocol

2.3.1. Task and procedure

2.3.1.1. Vowel Perception experiment

Functional MRI experiments underwent by the G3 participants were performed at the Centre for neuroimaging of the University Hospital CHU Grenoble Alpes (IRMaGe, Grenoble, France). The participants lying in the MRI scanner listened to auditory stimuli via an audio system compatible with high magnetic fields (MR Confon). Visual instructions were displayed through a mirror located in front of their eyes, reflecting a screen which was positioned behind the MRI scanner. The task consisted in carefully listening to the 4-vowel sequences and pressing a key when the last vowel was an /a/ (catch trials). There were altogether 4 conditions of interest plus one condition for catch trials and a baseline consisting of a silent condition, with 54 trials per condition and a total of 324 trials (see Table 2).

Condition	**Vowel: /i/**	**Vowel: /u/**
REP (54 trials)	i_0-i_0-i_0-i_0 (27 trials)	u_0-u_0-u_0-u_0 (27 trials)
NREP1 (54 trials)	i_0-i_0-i_0-i_1 (27 trials)	u_0-u_0-u_0-u_1 (27 trials)
NREP2 (54 trials)	i_0-i_0-i_0-i_2 (27 trials)	u_0-u_0-u_0-u_2 (27 trials)
NREP∞ (54 trials)	i_0-i_0-i_0-i_∞ (27 trials)	u_0-u_0-u_0-u_∞ (27 trials)
Catch Trials (54 trials)	i_0-i_0-i_0-a (27 trials)	u_0-u_0-u_0-a (27 trials)
Silent condition (54 trials)	-	-

Table 2 – Summary of the conditions used in the fMRI experiment

The 324 stimuli were divided into three runs of 108 items each, separated by a short pause; the order of sessions was counterbalanced between participants. In each session, the order of conditions was counterbalanced, with the constraint of always alternating /i/- and /u/- stimuli. A motor localizer session was added at the end of the fMRI recording session. The total experiment duration for the three sessions was about 45 min.

2.3.1.2. Post-Scan behavioral test

Importantly, given that the fMRI task involved perceptual detection of catch-trials /a/ stimuli, there was no control of the perception of i_0-u_0, i_2-u_2, i_4-u_4 or i_∞-u_∞ stimuli. Therefore, right after the fMRI experiment, a post-scan check was run outside the magnet. Specifically, participants performed a behavioural test in a quiet room in order to evaluate their subjective perception of

between-vowel distances. To this aim, they performed the quantitative perceptual distance evaluation task applied for the preparation of the stimuli with the G2 group (see Section 2.2). They were presented with the following pairs of vowels: i_0-i_0, u_0-u_0, i_0-i_2, u_0-u_2, i_0-i_4, u_0-u_4, i_0-i_∞, u_0-u_∞, in both orders, e.g. i_0-i_2 or i_2-i_0. Seventy pairs of stimuli (5 repetitions for each pair and each order) were thus presented in a random order. Participants were asked to subjectively evaluate the distance between the vowels in the pair, on a 1 to 7 scale. Due to technical problems, 2 of the 18 participants were unable to complete this post-scan test. For the other 16 participants, the responses were similar to those provided in the pilot study (displayed in Figure 3).

2.3.2 fMRI data acquisition

MR images were acquired with a whole-body 3T MR Philips imager (Achieva 3.0T TX Philips, Philips Medical Systems, Best, NL) with a 32-channel head coil for all of the participants. The chronology of fMRI sequences was as follows: vowel perception fMRI run #1, vowel perception fMRI run #2, anatomical MRI, vowel perception fMRI run #3, motor localizer fMRI and B0 fieldmap. A T2*weighted (Gradient Echo, GE) echo-planar imaging (EPI) sequence sensitive to blood oxygen level dependent (BOLD) contrast was used for the functional scans.

For the vowel perception experiment, the acquisition parameters were: 9 s repetition time (TR), 30 ms echo time, 90° flip angle, 240 mm x 240 mm in-plane field of view, 80 x 78 acquisition matrix size, 2.5 SENSE factor. 53 axial slices (2.80 mm thickness, separated by a 0.20 mm gap) covering the entire brain and parallel to the anterior commissure-posterior commissure plane were acquired in a sequential mode (ordered from Head to Feet); reconstructed voxel size was 3x3x3 mm^3. A total of 108 dynamic volumes per run were acquired in a sparse imaging procedure, which introduces a silent gap between subsequent/successive volume acquisition. Functional scanning therefore occurred over a fraction of the TR (2720 ms per volume, over the 9000 ms TR), alternating with silent inter-scanning periods during which the auditory stimuli were presented. The time interval between the onset of auditory stimuli and the midpoint of the following acquisition scan was varied randomly between 4s, 5s, or 6s to cover the typical delay range where the maximum of the BOLD hemodynamic response function occurs (the sparse sampling acquisition technique used in this study was similar to the one described in Grabski et al., 2012). Figure 4 summarizes the detail of the acquisition sequence.

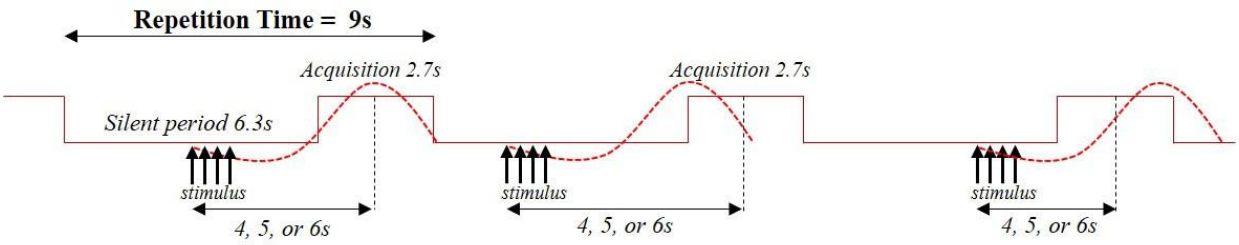

Figure 4. Timeline of the fMRI acquisition experiment.

Complementary to this vowel perception experiment, participants completed a motor localizer test. This test was run in order to draw functional regions of interest which served for further analyses. However, since the analyses focused on these regions of interest did not bring any additional information, the motor localizer and the related analyses are not described in the present paper.

A 3D T1-weighted high-resolution three-dimensional anatomical volume was also acquired using a MP-RAGE sequence: 1170 ms TI, 10.5 ms TR, 4.9 ms TE, 8° flip angle, 220 mm x 220 mm in-plane field of view with 175 mm H-F coverage, 316 x 315 x 250 acquisition matrix, 1.5 x 1.5 SENSE factors, isotropic 0.7 mm^3 reconstructed voxel size.

To correct images for geometric distortions, a B0 fieldmap was obtained from two gradient echo data sets acquired with a FLASH sequence. The fieldmap was subsequently used during data pre-processing.

2.4. fMRI Data analyses

Statistical analyses were performed using SPM12 statistical parametric mapping software (Wellcome Department of Cognitive Neurology, UK, www.fil.ion.ucl.ac.uk/219 spm/software/spm8/) running under Matlab 7.9 (The Mathworks Inc., Natick, USA). In addition, we used the SPM extension Anatomical Automatic Labelling (AAL, Tzourio-Mazoyer et al., 2002) for effect localizations, and, when necessary, the Yale BioImage suite package (mni2tal mapping from Lacadie et al., 2008, http://www.bioimagesuite.org).

The first 5 volumes of each scanning session, during which MR signal reaches steady-state, were discarded. fMRI data underwent two categories of analyses: spatial pre-processing and statistical analyses on the spatial pre-processed data.

2.4.1. Spatial pre-processing

Data pre-processing included realignment on the mean volume of each session, unwarping, co-registration on the anatomical volume and normalisation using DARTEL Tool from SPM12. On the latter point, structural T1-weighted scans of the 18 participants were segmented into different tissue types. Intensity average of the grey and white matter images were generated to use as an initial template for DARTEL registration (6 iterations). This template was aligned with the MNI Template using affine transform, and each functional scan was then aligned with this template. Finally, images were smoothed with an 8-mm FWHM Gaussian kernel. No high-pass filtering was applied.

2.4.2. Statistical analyses

Data analysis was performed using the general linear model (GLM, Friston et al., 1995) as implemented in SPM12.

First-level (individual) analyses

For each participant, the GLM modelled each experimental condition as five regressors (REP, NREP1, NREP2, NREP∞, Baseline) depending on the stimulus onset time and convolved with a canonical hemodynamic response function (HRF). Stimulus duration was set in the model at 1s. Six regressors were added for the six realignment parameters. The GLM was then used to generate parameter estimates of activity at each voxel and for each condition. Statistical parametric maps were generated from linear contrasts between the HRF parameter estimates for the different conditions and for between-conditions contrasts.

Firstly, to globally evaluate the brain regions associated with each condition of interest, each condition was contrasted against the silent baseline: REP > Baseline, NREP1 > Baseline, NREP2 > Baseline and NREP∞ > Baseline. Secondly, the repetition-suppression effect was evaluated as a function of increasing acoustic distance. Following our hypotheses, neural activity should be enhanced in auditory regions for the NREP1 condition in comparison to the REP condition, and in auditory and motor regions for the NREP∞ condition compared to the REP condition. The status of the NREP2 condition between these two patterns was not predictable *a priori*. To assess this set of hypotheses, each non-repetition condition was thus

contrasted to repetition: NREP1 > REP, NREP2 > REP, NREP∞ > REP. To evaluate the effect of the increasing acoustical distance, the non-repetition consecutive conditions (in terms of increasing distance) were also contrasted to each other, i.e. NREP2 > NREP1 and NREP∞ > NREP2. Finally, the (exploratory) contrast NREP∞ > NREP1 was also included for the sake of completeness.

Second-level (group) analyses

For both analyses, resulting images from each subject were entered in a second-level (random effect) model (Friston et al., 1999). A one-sample t-test was used; resulting contrasts were thresholded at a whole-brain $p<0.05$ family-wise error (FWE) corrected at the voxel level ($T > 6.91$) and at $p<0.05$ FWE corrected at the cluster-level for the first part of the analyses (section 3.1). For the second part of the analyses, contrasts were thresholded at $p<0.0001$ uncorrected at the voxel-level ($T > 4.79$), and $p<0.05$ FWE corrected at the cluster-level (section 3.2).

Local maxima are reported in the MNI space. Localisations were obtained using AAL (Tzourio-Mazoyer et al., 2002) and, when necessary, the Yale BioImage suite package (mni2tal mapping from Lancadie et al., 2008). The resulting tables present extended local maxima for each cluster of activation.

3. Results

3.1. REP, NREP1, NREP2 and NREP∞ compared to Baseline

The results of the first analysis contrasting (each condition REP, NREP1, NREP2 and NREP∞ versus Baseline) are provided in Figure 5 and Figure 6 and Table 3.

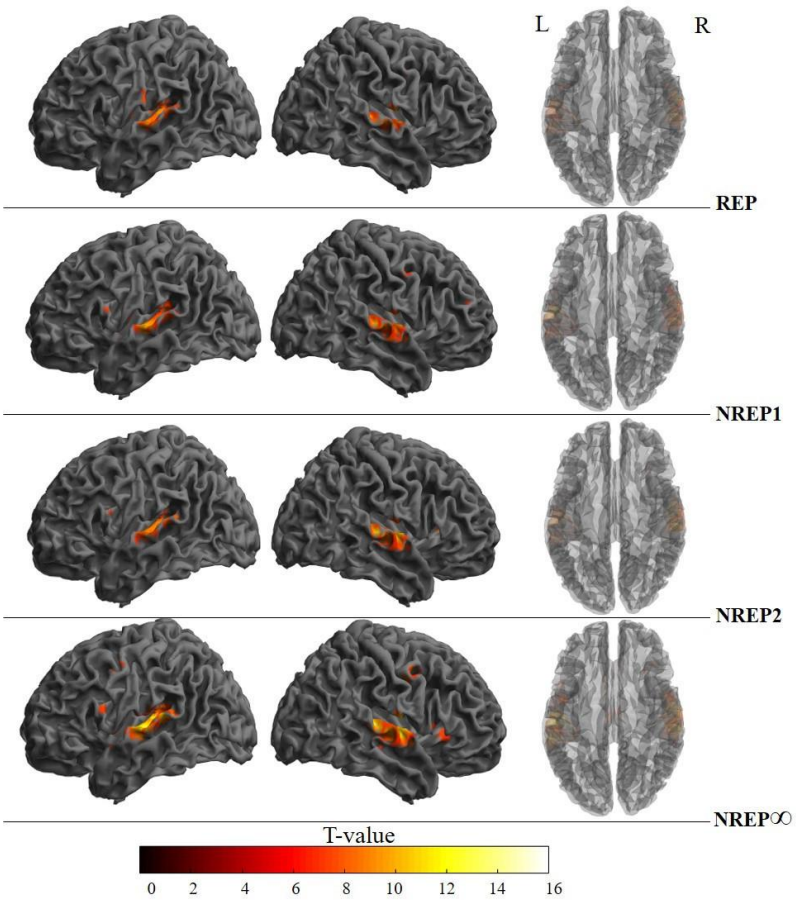

Figure 5. Maps of brain regions significantly activated for conditions REP, NREP1, NREP2 and NREP∞ (in successive rows). Maps are thresholded at p<0.05 FWE corrected at the cluster and voxel-levels (t(17)>6.91)**.** Activations are superposed on the canonical SPM12 MNI template.

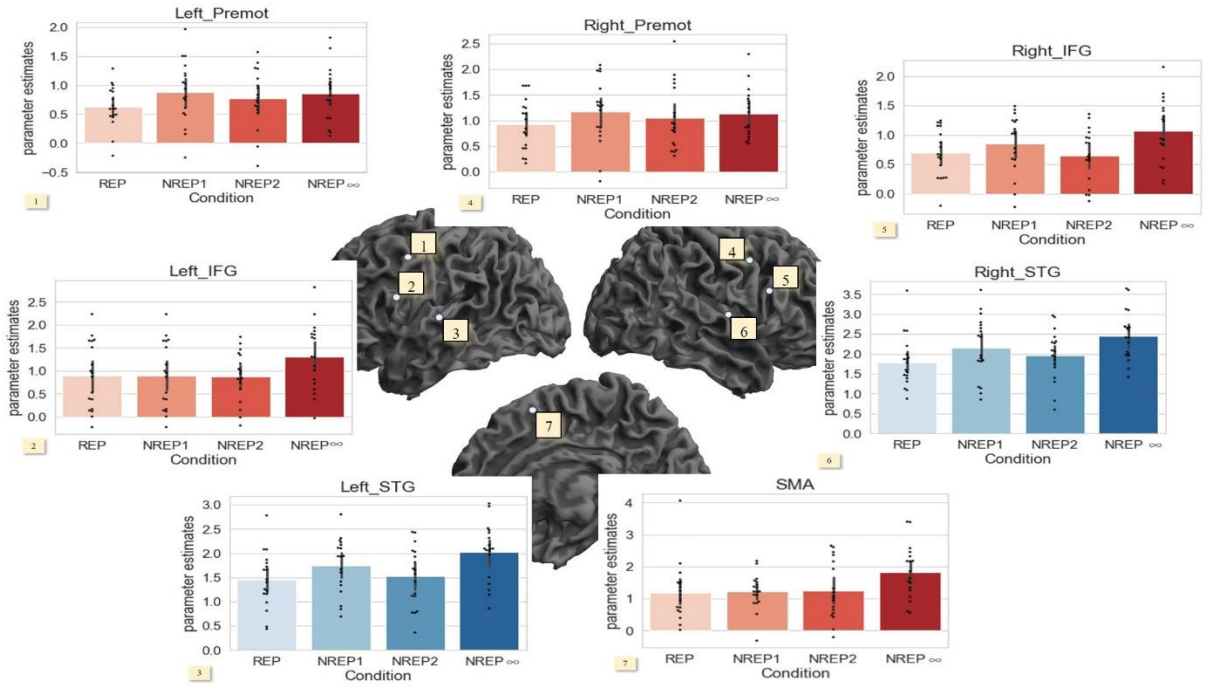

Figure 6. Parameter estimates (group means and individual data points) extracted from:
1: [-48 0 42] (Left Premotor Cortex); 2: [-51 6 18] (Left Inferior Frontal Gyrus); 3: [-57 -21 3] (Left Superior Temporal Gyrus); 4: [51 -3 39] (Right Premotor Cortex); 5: [45 15 21] (Right Inferior Frontal Gyrus); 6 [60 -9 9] (Right Superior Temporal Gyrus); 7: [-3 12 54] (Supplementary Motor Area). Error bars correspond to 95% confidence interval. Red colors indicate frontal and motor regions, and blue colors indicate auditory regions.

Compared to Baseline, the perception of vowels in the REP condition yielded bilateral activation in the auditory cortex (BA 41), superior temporal gyrus (STG, BA 22) and left precentral gyrus (BA6). When the distance between the first three vowels and the last one increased, i.e. in the NREP1 and NREP2 conditions compared to Baseline, the bilateral STG remained activated, and activity in frontal areas seemed to extend, including the right premotor region, right medial cingulate gyrus, right inferior frontal gyrus (BA45) and left cerebellum-Crus1 (see Table 3). The NREP∞ condition activated a larger network, including supplementary motor area (SMA, medial BA6), left and right precentral (M1, BA4) gyri, left mid-cingulate gyrus, left IFG (BA 44), right insula, and left and right cerebellum (pars 6).

To better visualize the location of activations for the four different contrasts, they are projected onto the same template in Figure 7. As can be seen on the figure, when progressively increasing the acoustic distance with REP, activation in the right STG seems to progress anteriorly, from y = [1; -38] in the REP > Baseline contrast, to y = [5; -34] in the NREP1 > Baseline contrast, and y = [11; -40] in the NREP∞ > Baseline contrast. In the left STG, activation diffuses both anteriorly and posteriorly, from y = [-6; -43] in the REP > Baseline contrast, to y= [5; -41] in the NREP1 > Baseline contrast and y = [7; -44] in the NREP∞ > Baseline contrast. Concerning motor regions (premotor and supplementary motor area), while there is almost no increase of activation in the REP > Baseline contrast, activation slightly increases in the NREP1 > Baseline contrast and is higher in the NREP∞ > Baseline contrast. No additional activation was observable for the NREP2 > Baseline contrast in comparison to the REP > Baseline contrast at a p<0.05 FWE-corrected level.

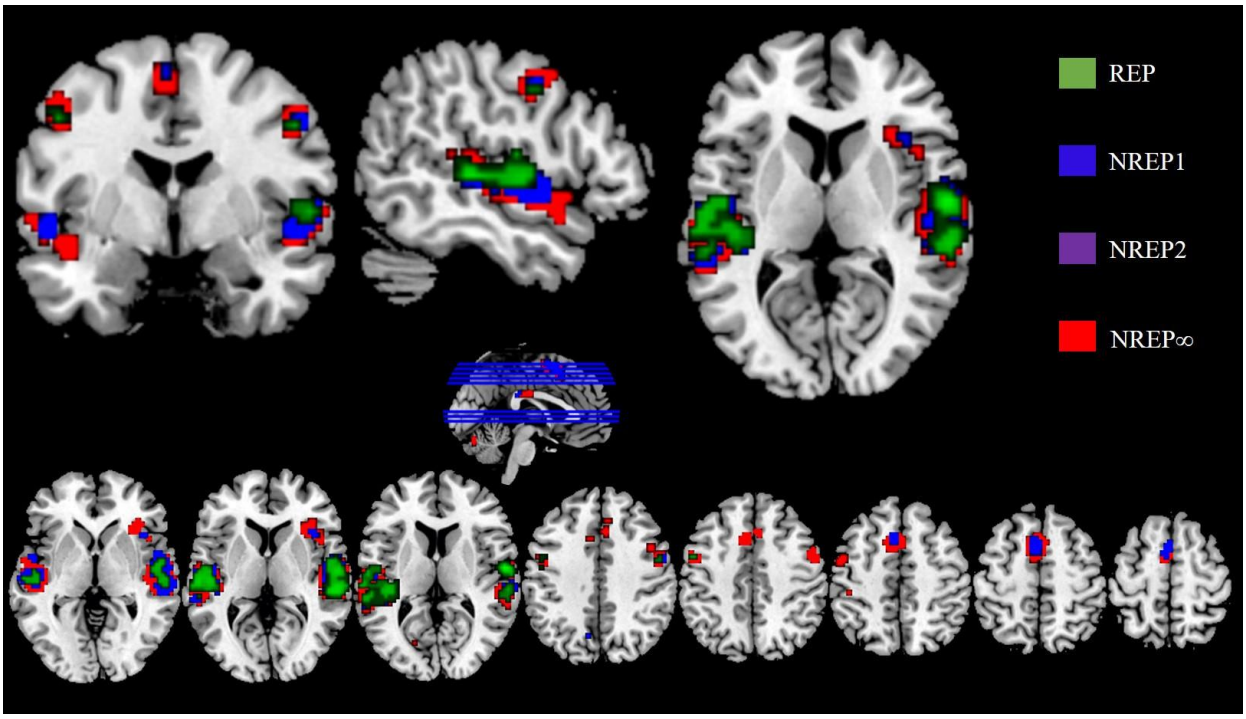

Figure 7. Maps of brain regions significantly activated for REP, NREP1, NREP2, and NREP∞ conditions relative to Baseline, superposed on the same ch2bet template. Maps are thresholded at $p<0.05$ FWE corrected at both the cluster and the voxel levels.

Condition	Region Label	BA	Peak Voxel (MNI coordinates)			Cluster Size (voxels)	T score
			X	y	z		
REP	Right superior temporal gyrus *(extending in rolandic operculum and right Heschl)*	41, 22, 1	57	-9	9	164	11.78
	Left superior temporal gyrus *(extending in rolandic operculum, Heschl gyrus and postcentral gyrus)*	41, 22, 1	-57	-21	3	269	10.43
	Right angular gyrus*	39	33	-54	33	3	7.78
	Right precentral gyrus (premotor and Supplementary motor area)	6	51	-3	39	3	7.73
	Left precentral gyrus (premotor and Supplementary motor area)	6	-48	0	45	5	7.35
	Left precentral gyrus (premotor and Supplementary motor area)	6	-48	-9	42	1	7.10
	Right Inferior frontal gyrus – pars triangularis	44	45	15	21	1	7.03
	Left Inferior frontal gyrus – pars triangularis	44	-48	3	24	1	6.94
NREP1	Left superior temporal gyrus *(extending in middle temporal, Heschl and rolandic operculum)*	41, 22	-57	-21	3	256	12.59
	Right superior temporal gyrus *(extending in rolandic operculum and right Heschl)*	22, 41, 1	57	-27	3	277	11.77
	Supplementary Motor Area	6	-3	12	57	41	10.00
	Right precentral gyrus (right premotor)	6	51	-3	39	7	7.58
	Left superior temporal gyrus	22	-51	-3	-3	14	7.54
	Left inferior frontal gyrus – pars opercularis	44	-51	9	18	4	7.45
	Left precuneus	7	-9	-66	39	1	7.30
	Right middle frontal gyrus	10	39	45	18	1	7.16
	Right inferior frontal gyrus – pars opercularis	44	45	15	21	2	7.09
	Right medial cingulate	-	0	-30	27	5	7.08
	Right inferior frontal gyrus (pars triangularis) *extending in right insula*	45	39	21	6	9	7.07
	Right precuneus	7	12	-69	30	2	6.90
NREP2	Right superior temporal gyrus *(extending in Heschl and Rolandic operculum)*	22, 41	60	-24	3	262	12.59
	Left superior temporal gyrus *(extending in Heschl gyrus and rolandic operculum)*	41	-48	-18	3	207	10.67
	Right insula	-	27	24	6	5	8.82
	Left cerebellum – Crus1	-	-45	-60	-33	6	8.32
	Left Supplementary Motor Area	6	0	12	54	10	7.85
	Right cerebellum – pars 6	-	12	-72	-21	3	7.81
	Right precuneus *(extending in left)*	7	6	-72	48	6	7.53

	Left inferior frontal gyrus – pars triangularis *	45	-21	27	12	1	7.41
	Right precentral gyrus (premotor and Supplementary motor area)	6	51	0	42	3	7.28
	Left inferior frontal – pars opercularis	44	-51	6	18	3	7.26
	Left insula *	13	-21	27	3	1	7.20
	Left precentral gyrus (premotor and Supplementary motor area)	6	-48	0	39	2	7.10
	Right mid- cingulate *	23	6	-21	27	1	7.04
	Left mid- cingulate *	23	-3	-27	27	2	7.02
NREP∞	Right superior temporal gyrus *(extending in rolandic operculum, Helschl, and insula)*	1, 41	60	-9	9	402	16.29
	Left superior temporal gyrus *(extending in Heschl gyrus, rolandic operculum, middle temporal gyrus)*	41	-48	-18	3	481	14.69
	Right precentral gyrus (right premotor and supplementary motor) *(extending in right medial frontal gyrus)*	6	51	-3	42	32	10.39
	Supplementary motor area *(left, extending in right)*	6	0	12	54	113	9.99
	Left precentral gyrus (premotor and supplementary motor)	6	-48	0	45	27	8.91
	Left mid-cingulate gyrus* *(extending in right cingulate gyrus)*	23	-6	-18	27	23	8.58
	Right insula *(extending in inferior frontal gyrus – pars opercularis and triangularis)*	13, 44	33	24	3	47	8.51
	Right cerebellum – lob. 6	-	27	-63	-24	22	8.48
	Right inferior frontal gyrus – pars opercularis *(extending in pars triangularis)*	44	45	15	21	8	8.33
	Left cerebellum -lob. 6 *(extending in right)*	-	-6	-75	-21	26	8.27
	Left precentral gyrus *(extending in left inferior frontal gyrus pars opercularis and triangularis)*	6,44	-54	6	15	16	7.92
	Right medial superior frontal gyrus *(extending in left and right cingulate and supplementary motor area)*	8	6	21	42	8	7.64
	Left postcentral gyrus	1	-42	-30	51	1	7.18
	Left cerebellum – pars 6	-	-15	-69	-21	1	7.11
	Right medial superior frontal gyrus	8	6	30	42	2	7.10
	Left superior temporal gyrus	22	-51	6	-6	1	7.04
	Left calcarine sulcus	17	-18	-72	12	1	7.01
	Left cerebellum – pars 6	-	-30	-63	-27	1	6.94

Table 3. MNI coordinates (extended local maxima) for each condition contrasted against baseline. We used the SPM AAL extension for effects localization, and when necessary, Yale BioImage suite package for Brodman's areas identification. Threshold is fixed at $p<0.05$ FWE corrected at the voxel level and cluster level ($t(17)>6.91$). T-values represent the highest t-value at the voxel level. *nearest grey matter

3.2. Between-condition contrasts

The results of the second part of the analysis assessing the difference in the repetition-suppression effect as a function of acoustic distance are provided in Table 4, and the networks of brain areas significantly activated in the comparisons between the 3 NREP conditions and the REP condition are displayed in Figure 8.

The NREP1 > REP contrast displayed significant increase in activation in the right superior temporal gyrus (BA22) extending in the middle temporal gyrus (BA 21), and the right hippocampus. At the threshold of p<0.0.5 FWE corrected at the voxel level, there was no significant increase in activation when NREP2 was contrasted to REP or when NREP2 was compared to NREP1. As Figure 8 shows, at a less stringent threshold (p<0.001 uncorrected), we can observe a significant activation in the right STG.

The NREP∞ > REP contrast displayed significant differences in a larger network of brain regions, including bilateral superior temporal gyrus extending to the bilateral inferior frontal gyrus (BA 47) and insula, left and right cerebellum, left medial superior frontal gyrus (BA 8) extending to the left mid-cingulate, and left and right calcarine sulcus (BA 17). The NREP∞ > NREP2 contrast was associated to a significant increase in activation in the left and right insula (BA13), bilateral superior temporal gyrus (BA 41 and 22), right inferior frontal (BA45) gyrus and supplementary motor area (BA6). The NREP∞ > NREP1 contrast provided significant differences only in the left and right supplementary motor area.

The reverse contrasts (REP > NREP1, REP > NREP2, and REP > NREP∞, NREP1 > NREP2, NREP1 > NREP∞ and NREP2 > NREP∞) yielded no significant increase in activation.

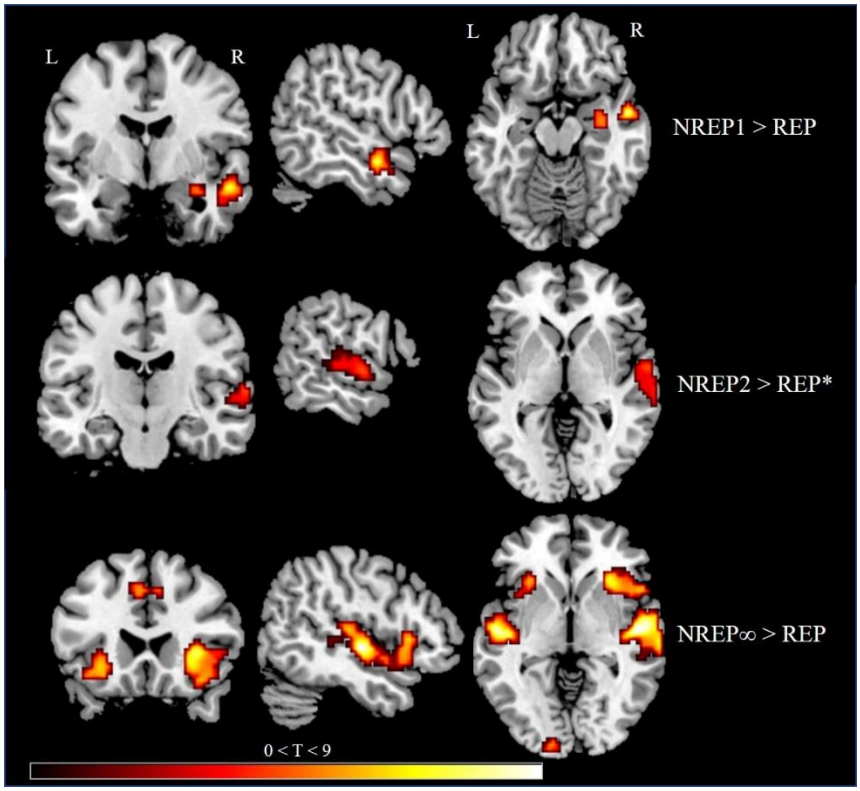

Figure 8. Maps of brain regions significantly activated for NREP1 > REP (upper panel), NREP2 > REP (middle panel), and NREP∞ > NREP (lower panel) contrasts. Maps are thresholded at p<0.0001 uncorrected at the voxel level and p<0.05 FWE corrected at the cluster level (t(17)>4,71). Activations are projected on the ch2bet template. * For representation purpose, NREP2>REP is thresholded at p<0.001 uncorrected at the voxel level and p<0,05 FWE corrected at the cluster level.

Condition	Region Label	BA	Peak Voxel (MNI coordinates)			Cluster Size (voxels)	T score
			x	y	z		
NREP1 > REP	Right superior temporal gyrus *(extending in middle temporal gyrus and temporal pole)*	22	54	-3	-15	54	9.80
	Right hippocampus *(extending in right amygdala and putamen)*	-	33	-9	-15	25	5.77
	Right superior temporal gyrus *(extending in middle temporal gyrus)*	41	66	-21	3	28	5.73
NREP2 > REP	*No suprathreshold cluster*						
NREP∞ > REP	Right superior temporal gyrus *(extending in right insula, Heschl gyrus and right inferior frontal gyrus)*	22, 41, 13	51	-12	-3	760	11.56
	Left superior temporal gyrus *(extending in left insula and inferior frontal gyrus – pars opercularis)*	22	-54	-12	0	441	10.97
	Right calcarine sulcus	17	15	-69	9	57	7.01
	Right cerebellum – pars 6	-	27	-63	-24	36	6.90
	Left and right medial superior frontal gyrus *(extending in supplementary motor area and cingulate)*	8	-6	15	42	87	6.73
	Left cerebellum – crus1 *(extending in pars 6)*	-	-45	-60	-30	55	6.37
	Left calcarine *(extending in left middle occipital gyrus)*	18	-12	-99	-3	19	6.06
NREP∞ > REP1	Left and right supplementary motor area	8	6	21	45	65	6.04
NREP2 > NREP1	*No suprathreshold cluster*						
NREP∞ > NREP2	Right insula *(extending in inferior frontal gyrus – pars opercularis)*	13, 45	30	21	-3	184	8.49
	Left superior temporal gyrus *(extending in left middle temporal gyrus)*	41, 22	-54	-18	3	124	7.61
	Left insula *(extending in inferior frontal gyrus – pars orbitalis)*	13	-30	21	-6	57	6.58
	Right supplementary motor area *(extending left supplementary motor area and in left and right cingulate*)	6, 8	6	12	45	83	6.53
	Right Inferior frontal gyrus – pars opercularis *(extending in pars triangularis)*	44	42	6	21	21	6.37
	Right superior temporal gyrus *(extending in Heschl gyrus)*	22	51	-9	-9	26	6.30
	Left calcarine *(extending in medial occipital gyrus)*	18	-9	-96	0	17	5.90

Table 4. MNI coordinates (extended local maxima) for between condition contrasts. We used the SPM AAL extension for effects localization, and when necessary, Yale BioImage suite package for Brodman's areas identification. Threshold is fixed at $p<0.0001$ uncorrected at the voxel level and $p<0.05$ FWE corrected at the cluster level ($t(17)>4.71$). T-values represent the highest t-value at the voxel level.

4. Discussion

4.1. The basic network of vowel perception

As a preliminary note, the auditory speech conditions REP, NREP1, NREP2 and NREP∞, yielded a significant increase in BOLD response compared with the silent baseline, in a network that includes the auditory cortex and superior temporal gyrus bilaterally, the inferior frontal gyrus (BA44/45) bilaterally, the bilateral premotor cortex, SMA, insula and cerebellum. This sensory-motor network linking auditory regions with somatosensory-motor regions through the dorsal pathway has been described in a number of previous studies (Hickok & Poeppel, 2007; Skipper et al., 2005, 2017; Schomers & Pulvermüller, 2016). It is also consistent with all other studies on the RS paradigm using phonetic stimuli, as mentioned in Section 1.3, and with a number of studies focused on vowel perception (e.g. Arnaud et al., 2013; Grabski et al., 2013; Grabski & Sato, 2020; Husain et al., 2006; Joanisse & Gati, 2003; Rampini et al., 2017).

4.2. Repetition-suppression and the ANMW hypothesis

The novelty of the present study concerns the distinct patterns of activation in the temporal and frontal regions associated with a repetition-suppression paradigm involving vowel stimuli. Our major finding is that a small acoustic variation between the final stimulus and the three preceding ones in a vowel sequence (condition NREP1 compared with REP) produces a significant increase in activity (i.e. release of suppression) in a set of regions associated with auditory processing, namely the right superior temporal gyrus, Heschl's gyrus and the right hippocampus. Larger acoustic distances (NREP∞) provide an additional significant increase in activity in regions associated with speech-related planning and motor programming (right insula and right pars orbitalis and pars opercularis, left insula, right medial superior frontal gyrus, supplementary motor area, as well as bilateral cerebellum), in addition to those considered to be involved in auditory decoding (left STG).

The distinct patterns of repetition-suppression recovery associated with small and large acoustic distances have never been described previously with vowel stimuli. As presented in Section 1.3, the only previous study using a similar paradigm with vowels is the MEG study by Altmann et al. (2014). Their study revealed a region within the left superior temporal cortex that was differentially activated for pairs of stimuli consisting of the same or different consonant-vowel (CV) syllables, at around 430 to 500 ms after the onset of the second stimulus of the pair. Yet, although this difference in activation occurred only for categorical variations in the consonant of the CV stimuli, it was observed both for within- and for between-category variations of the vowel. The authors interpreted this result as evidence for a categorical effect at the neural level for consonant perception, but not for vowel perception. In their view, their finding is consistent with observed behavioral differences in the categorization of consonants and vowels, consonants being "represented in a more categorical-like manner than vowels".

In contrast, effects similar to those obtained in the present study have been found in previous studies with consonants (e.g. in Chevillet et al., 2013, or Alho et al., 2016, see Section 1.3). Yet, the results seem to differ largely from one experiment to another, as shown in Section 1.3. To get a clearer picture of the similarities and differences between the studies, in Table 5, we present a summary of all available fMRI or MEG repetition-suppression studies which involve sequences of phonetic stimuli varying around a prototype. Strikingly, in spite of the large variations in the pattern of activations in these experiments, it appears that all experimental data can be summarized by 3 rules which apply systematically to all these studies.

Study	Technique	Phonetic categories	Conditions	Paradigms	Results	Temporal/Motor selectivity
Celsis 1999	fMRI	Consonants [t] vs. [d] in Ca stimuli (ta-pa)	4-stimuli trains s-s-s-d, d either same (S) or different from s (D), s being always [ta] and d being either [ta] (S) or [pa] (D)	Passive listening (with no task)	D > S : left SMG	No temporal or motor area
Zevin & McCandliss, 2005	fMRI	Consonants [ɹ] vs. [l] in Ca stimuli (ɹa-la)	4-stimuli trains s-s-s-d, d either same (S) or different from s (D), meaning that s is ɹa and d is la or conversely	Passive listening (with no task)	D > S: left pSTG/SMG, plus a broad network with less consistent pattern (Note: small number of subjects (8), p<0.005 uncorrected)	AT
Joanisse et al., 2007	fMRI	Consonants [d] vs. [g] in Ca stimuli (da-ga)	4-stimuli trains s-s-s-d, the last one (d) either same (S) or different from the first 3 (s), within the same category (W) or changing category (B)	Pre-attentive passive listening, stimuli not attended and presented together with a movie displayed silently with subtitles	W vs. S: nothing B > W: left (STS-MTG-SMG)	0 then AT
Myers et al., 2009	fMRI	Consonants [t] vs. [d] in Ca stimuli (ta-da)	5-stimuli pairs s-s-s-s-d, d either same (S) or different from s, within the same category (W) or changing category (B)	Active listening without categorization (with a decision task on pitch in catch trials)	W > S: bilateral IFG, right STG - SMG, left INS B > W: left IFG B > S: bilateral IFG, left pSTG	AT-MP then MP
Chevillet et al., 2013	fMRI	Consonants [d] vs. [g] in Ca stimuli (da-ga)	2-stimuli pairs s-d, d either same (S) or different from s, within the same category (W) or changing category (B)	Active dichotic listening with a distracting task on the same stimuli (deciding on the compared duration of the stimuli between the two ears)	W > S: left aMTG and left pSTG B > W : left PMC B > S: left (aMTG, pSTG, PMC)	AT then MP
Altman et al., 2014	MEG	Consonants [b] vs. [d] or Vowels [a] vs. [o] in CV stimuli	2-stimuli pairs s-d, d either same (S) or different from s, within the same category (W) or changing category (B)	Active listening (same-different task)	B > (W=S) in left ST for consonants (B=W)> S in left ST for vowels	0 then AT (consonants) AT then 0 (vowels)
Lawyer & Corina, 2014	fMRI	CV syllables varying in C category by 1, 2 or 3 phonetic features (place, voice, manner)	10-to-16 repetitions of s followed by d, d either same (S) or different from S varying by 1, 2 or 3 phonetic features (D1, D2, D3)	Passive listening (high vs. low frequency decision on a noise band presented 10s after the stimuli)	D > S: R STG, bilateral STS + cingulate, anterior cingulate, MFG and several areas in the cerebellum.	AT-MP
Alho et al., 2016	MEG	Consonants [d] vs. [g] in Ca stimuli (da-ga)	2-stimuli pairs s-d, d either same (S) or different from s, with s and d within the same category (W) or from two different categories (B)	ATTEND condition without categorization task (active dichotic listening with a distracting task on the same stimuli deciding on the compared duration of the stimuli between the two ears)	W > S: left aSTG B > W: left IFG (POp) + left aINS	AT then MP

Alho et al., 2016	MEG	Consonants [d] vs. [g] in Ca stimuli (da-ga)	2-stimuli pairs s-d, d either same (S) or different from s, with s and d within the same category (W) or from two different categories (B)	IGNORE condition with pre-attentive passive listening, stimuli not attended and presented together with a movie displayed silently with subtitles	W vs. S: nothing B vs. W: nothing B vs. S: left MTC + left pTC	0 then 0
Present study	fMRI	Vowels [i] or [u]	4-stimuli trains s-s-s-d, d either same (S) or different from s, with variable distances between s and d, close (D-close) or far (D-far)	Active listening (with a decision task on phonetic identiity on catch trials)	D-close > S: right STG D-far > D-close: bilateral IFG + SMA D-far > S: bilateral STG + bilateral IFG + SMA	AT then MP

Table 5 – Summary of main findings in previous RS studies exploring same-different paradigms with phonetic stimuli.
Column 2 – fMRI: functional Magnetic Resonance Imaging; MEG: Magnetoencephalography.
Column 3 – CV: Consonant-Vowel; Ca: Consonant-Vowel with vowel [a].
Column 4 – S: Same; D: Different; D-close: Different with close stimuli; D-far: Different with far stimuli; W: Within; B: Between.
Column 6 – SMG: Supramarginal Gyrus; STG: Superior Temporal Gyrus (a-anterior, p-posterior); MTG: Middle Temporal Gyrus (a-anterior); IFG: Inferior Frontal Gyrus; INS: insula; PMC: Premotor Cortex; ST: Superior Temporal gyrus/sulcus; STS: Superior Temporal Sulcus; MFG: Middle Frontal Gyrus; POp: Pars Opercularis; MTC: Middle Temporal Cortex; pTC: posterior Temporal Cortex; SMA: Supplementary Motor Area. X>Y: larger response in a given region for condition X compared with condition Y; X=Y: no significant difference in response in a given region for condition X compared with condition Y.
Column 7 – AT: difference between conditions in Auditory Temporal regions; MP: difference between conditions in speech planning – Motor Programming regions. 0: no difference between same (S) and different (W, B, D) conditions. X then Y: difference between same and close conditions in regions X, difference between close and far conditions in regions Y.

First, all data converge on the fact that when the last stimulus in a sequence differs from the previous one(s), cortical activity increases. As can be observed in Table 5 (column "Results"), activation is larger in conditions D (last stimulus Different), W (last stimulus different but Within the same category), or B (last stimulus different with Between category distinction) than S (all stimuli are the Same). Furthermore, when two types of deviations for the last stimulus are compared, larger deviation leads to an additional increase in cortical activity. There is increased cortical activity in conditions D-far compared to D-close (two "Different" conditions with larger differences in one case, D-far, than in the other, D-close) and in condition B compared to condition W. This is actually the precise manifestation of the RS effect, and it is systematic in all these studies. Therefore, this first observation could be summarized as the "repetition-suppression rule".

Second, in the last column of Table 5, we consider the selectivity in the "auditory temporal" (AT) regions, grouping the superior temporal gyrus and sulcus, and in the "speech planning - motor programming" (MP) regions, grouping the inferior frontal gyrus, premotor cortex, supplementary motor area, insula and cerebellum. More precisely, we compare in this last column the activity increase in these two sets of regions, when the last stimulus in the sequence shifts from Same to Different and, when possible, from Close to Far (or from Within to Between). Strikingly, a systematic regularity emerges which provides the second summarizing rule: RS recovery systematically happens in the auditory temporal regions before the motor regions. In other words, RS recovery is triggered for smaller acoustic variations in the auditory temporal regions than in the motor regions. Indeed, when only two conditions are compared (Same and Different), the increase in activity in the Different condition occurs in Auditory Temporal regions (for Zevin & McCandliss, 2005, or Lawyer & Corina, 2014, adding MP regions in the cerebellum in the second case). When three conditions are examined, the observed pattern is either "no change for Within and a change in Auditory Temporal regions for Between conditions" (Joanisse et al., 2007; Altmann et al., 2014, for plosives) or "change in Auditory Temporal regions for Within/D-close and a change in Motor regions for Between/D-far" (Myers et al., 2009 – although in this case there was synchronous change in Auditory Temporal and Motor regions for Within; Chevillet et al., 2013; Alho et al., 2016; and the present study).

Crucially, this selectivity pattern is the exact application of the ANMW hypothesis at the heart of the present study, so we will call it the "ANMW rule". This rule applies to all 10 studies in Table 5, including the present one. The present study adds one key point to this pattern, that is the fact that the ANMW-rule also applies to vowel stimuli, while previous evidence for this rule only concerned consonants.

Third, attention seems to modulate the pattern of responses in a consistent way, with less activity overall and therefore less difference between conditions when attention is decreased. Although the way in which attention is controlled is highly variable across these studies (see column "Paradigms" in Table 5), the trend is clear. It applies to the comparison of conditions "ATTEND" and "IGNORE" in the study by Alho et al. (2016). It also emerges when comparing the studies by Joanisse et al. (2007) and Chevillet et al. (2013) who used similar stimuli, but differed in attention involvement. In Joanisse et al.'s (2007) study, where participants were instructed not to pay attention to the sound, there were no difference between conditions W and S, but differences emerged in the Auditory Temporal region when comparing B and W. In contrast, Chevillet al. (2013) proposed a distraction task which actually led participants to listen carefully to the acoustic stimuli although the focus was not on categorization per se. They found a difference between conditions S and W in Auditory Temporal, and then between W and B in Motor Programming regions. The attentional modulation of cortical activity, globally and across conditions, is consistent with general findings on the role of attention in modulating

activity and selectivity in neural processes (Spitzer et al., 1988; Hillyard et al., 1998; Murray & Wojciulik, 2004). It is also in line with numerous reports of fMRI speech perception experiments showing a modulation of cortical responses with selective attention in the superior temporal and inferior frontal cortex (e.g. Hugdahl et al., 2003; Sabri et al., 2008; Wild et al., 2012). This third rule can therefore be called the "attention rule".

Three additional elements in the current data deserve to be discussed. The first one concerns the lack of increase in activation in the NREP2 > REP contrast while there is a significant increase in activation in the NREP1 > REP contrast (see Table 4). This is unlikely to be due to the stimuli, given the increase in acoustic and perceptual distances from stimuli i_0-u_0 to i_2-u_2 (used in NREP1) and then to i_4-u_4 (used in NREP2, see Figures 2 and 3). A possible interpretation could be that in conditions REP and NREP1, where the acoustic difference between the first three stimuli and the last one is either null (REP) or small (NREP1), the listeners search more actively for a perceptual difference at the end of the sequence. This would lead to an increase in attentional processing in these two conditions compared with the other two conditions NREP2 and NREP∞, where the last stimulus clearly differs from the first three ones. This could explain an increase in activation in conditions REP and NREP1 compared with condition NREP2 (and NREP∞). This increased processing would mask the increase in neural activity in NREP2 that should result from the release in RS. Two opposite mechanisms (less attention but more release in RS in NREP2 relative to REP and NREP1) would be at play which could explain the observed activation contrasts.

This suggests that the effects of the hypothetical ANMW principle and of attentional processes could interact in a possibly complex way. Moreover, it must be noticed that the pattern of beta estimates within each set of voxels (see Figure 6) does not display smooth monotonically increasing patterns as could be assumed based on the selectivity patterns in Figure 1A – even if we take into account the potential role of attentional processes possibly differing in conditions REP and NREP1 compared with conditions NREP2 and NREP∞. This is likely due to the complex and partly unknown relationships relating neural individual and collective activity to BOLD response in a given cortical region (Vanzetta & Slovin, 2010; Zhang et al., 2020). Future studies explicitly monitoring or evaluating the amount of cognitive effort and attention in relation to the ANMW principle should be set up to further analyse these interactions and the underlying pattern of cortical activation in temporal vs. frontal areas.

Second, it is not without interest that, in addition to the observed increased activity in the frontal regions for between-category vowel processing (the NREP∞ condition), increased activation was also observed in the bilateral cerebellum. In line with propositions made by Grandchamp et al. (2019), this cerebellar recruitment can be interpreted as the involvement of an internal model when processing a new vowel category. The internal model would provide cortical frontal regions with a motor specification inverted from the new sound.

Finally, there is a trend in the present data to observe greater differences in cortical activity in the right hemisphere compared with the left one, since a difference between NREP1 and REP in the STG was found only in the right hemisphere, while activation in the IFG in the NREP∞ condition was increased bilaterally, as compared with the two other conditions. This pattern is actually not really surprising, since the stimuli used in the present study are synthetic and based on small variations around the prototypical i_0/u_0 stimuli. They are therefore atypical speech sounds likely to induce larger activity in the right hemisphere. Numerous studies have found a shift in the lateralization of cerebral processing from left to right hemisphere for degraded (noisy) speech stimuli (Bidelman & Howell, 2016). Other studies have found right hemisphere predominance for the processing of speaker indexical information included in speech sounds (McGettigan & Scott, 2012) or the processing of phonetic variability between talkers (Luthra,

2021). A right hemisphere shift is also observed in some of the studies listed in Table 5 (e.g. for Zevin & McCandliss, 2005; Myers et al., 2009; Lawyer & Corina, 2004).

4.3. Alternative interpretations

The set of experimental data on phonetic repetition-suppression described in Table 5 can be and has been interpreted with other systems of explanations than the present set of three rules including the "ANMW rule".

First, a number of authors proposed a dichotomy between pre-categorical auditory/phonetic processing in posterior areas (typically STG/STS/SMG), and categorical processing and decision in anterior areas (including IFG/PMC). This is the argument developed in particular by Myers et al. (2009), Chevillet et al. (2013) or Alho et al. (2016). The underlying rationale is that phonological categorisation occurs in the inferior frontal regions, where "integration of auditory and motor information" takes place (Chevillet et al., 2013). This view is in line with a number of data on the role of anterior regions in successful phonological categorization, for both auditory (Alho et al., 2012, 2014) and audiovisual stimuli (Hasson et al., 2007). However, a number of other studies also provide evidence for categorisation processes in the STG/SMG complex (e.g. Chang et al., 2010; DeWitt & Rauschecker, 2012; Jacquemot et al., 2003), and specifically converge on the role of the SMG in the representation and manipulation of phonological units in speech perception (Paulesu et al., 1993; Caplan et al., 1995; Dehaene-Lambertz et al., 2005; Jacquemot & Scott, 2006). As a matter of fact, a number of studies listed in Table 5 show temporal rather than frontal regions to be specifically associated with Between-category sequences (e.g. Zevin and McCandliss, 2005; Joanisse et al., 2007; Altmann et al., 2014; Lawyer & Corina, 2014).

A second line of interpretation would rather relate frontal activity in the Between-category conditions to active decision-making and executive processes (e.g. Binder et al., 2004; Blumstein et al., 2005; Joanisse et al., 2007), possibly including access to the phonological loop in working memory (Baddeley et al., 1984) and access to the output phonological buffer in the IFG (Jacquemot & Scott, 2006). This account could be viewed as consistent with the pattern of results in Table 5, in which most conditions that included an explicit auditory task prompting the listener to pay attention to the stimuli yielded activity in IFG/PMC in the Between condition (apart from the MEG data in Altmann et al., 2014). However, all of these studies, apart from the present one, involved an auditory task in which the participants were indeed asked to listen carefully to the stimuli, not for categorization or discrimination, but rather for a decision on some other component of the sound (pitch or duration). This led Chevillet et al. (2013) to argue that IFG activity in their study could not be conceived as attention-related, since the task at play did not involve phonological categorization, and to conclude that the PMC activity was related to "automatic sensorimotor integration of speech".

Overall, the rather complex pattern of data in Table 5 seems difficult to reconcile with either of these two interpretations. It is instead more compatible with the gradient view that underlies the two ground rules proposed to operate in addition to the "RS rule", namely a difference in selectivity in temporal vs. frontal regions at play in the ANMW hypothesis, and a gradient role of attention. Yet, of course, the different arguments are not mutually exclusive, and it is likely that differential selectivity (as in ANMW), auditory-motor integration mechanisms and phonological executive processes are all at play across the RS studies reviewed in this discussion.

5. Conclusion

In this work, we have designed an original fMRI RS experiment based on the predictions from a computational model. The prediction is related to the so-called Auditory-Narrow Motor-Wide property, according to which auditory processing of speech sounds – presumably related to activations in temporal regions in the human cortex – would be narrower, i.e. more tuned to prototypical sounds, than motor processing – presumably related to frontal regions. We designed the study on this basis, to explore whether fMRI responses to vowel sounds in a repetition-suppression paradigm would provide distinct recovery patterns in temporal and frontal regions. The data we obtained generally confirm this prediction. A small modification of the acoustic input led to a significant increase in the BOLD response in temporal regions (in the right superior temporal gyrus and Heschl's gyrus) whereas a significant increase in regions likely associated to speech planning and motor programming (right insula and right pars orbitalis and pars opercularis, left insula, right medial superior frontal gyrus, supplementary motor area and cerebellum) occurred for a larger modification of the vowel input. Interestingly, our analysis of previous neuroimaging studies reveals that this pattern, which we have termed the ANMW rule, seems to apply to all other previous studies.

These data are in line with the behavior of the COSMO model and suggest that the interpretation of greater cortical activity in motor regions for atypical stimuli – e.g. noisy or accented – may be related to the ANMW property. Regardless of the modelling aspects, these data add to the overall picture of repetition-suppression fMRI experiments, adding data for vowel stimuli to the large amount of data already available for consonants.

Acknowledgement:

This work was performed on the IRMaGe platform member of France Life Imaging network (grant ANR-11-INBS-0006). This work was supported by the European Research Council under the 7th European Community Program (FP7 / 2007-2013 Grant Agreement No.339152 – "Speech Unit(e)s").

References

Alho, K., Salonen, J., Rinne, T., Medvedev, S.V., Hugdahl, K., Hämäläinen, H., 2012. Attention-related modulation of auditory-cortex responses to speech sounds during dichotic listening. Brain Res. 1442, 47-54.

Alho, K., Rinne, T., Herron, T.J., Woods, D.L., 2014. Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. Hear Res. 307, 29-41.

Alho, J., Green, B.M., May, P.J.C., Sams, M., Tiitinen, H., Rauschecker, J.P., Jääskeläinen, I.P., 2016. Early-latency categorical speech sound representations in the left inferior frontal gyrus. Neuroimage. 129, 214-223.

Altmann, C.F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., Fukuyama, H., 2014. Categorical speech perception during active discrimination of consonants and vowels. Neuropsychologia. 64, 13-23.

Arnaud, L., Sato, M., Ménard, L., Gracco, V.L., 2013. Repetition-suppression for speech processing in the associative occipital and parietal cortex of congenitally blind adults. PLOS ONE 8(5): e64553.

Baddeley, A., Lewis, V., Vallar, G., 1984. Exploring the Articulatory Loop. The Quarterly Journal of Experimental Psychology Section A. 36(2), 233–252.

Barnaud, M.L., Bessière, P., Diard, J., Schwartz, J.L., 2018. Reanalyzing neurocognitive data on the role of the motor system in speech perception within COSMO, a Bayesian perceptuo-motor model of speech communication. Brain Lang., 187, 19-32.

Benson, R.R., Whalen, D.H., Richardson, M., Swainson, B., Clark, V.P., Lai, S., Liberman, A.M., 2001. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain Lang. 78(3), 364-96.

Bidelman, G.M., Howell, M., 2016. Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception. NeuroImage. 124, 581-590.

Binder, J., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci. 7, 295–301.

Blumstein, S.E., Myers, E.B., Rissman, J., 2005. The perception of voice onset time: an fMRI investigation of phonetic category structure. J Cogn Neurosci. 17(9), 1353-66.

Boersma, P., Weenink, D., 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.56, retrieved 3 November 2021 from http://www.praat.org/.

Callan, D.E., Jones, J.A., Callan, A.M., Akahane-Yamada, R., 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. Neuroimage. 22(3), 1182-94.

Callan, D.E., Callan, A.M., Jones, J.A., 2014. Speech motor brain regions are differentially recruited during perception of native and foreign-accented phonemes for first and second language listeners. Front Neurosci. 8; 275.

Caplan, D., Gow, D., Makris, N., 1995. Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. Neurology. 45(2), 293–298.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I., Nespoulous, J.L., Chollet, F.,1999. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. Neuroimage. 9(1), 135-44.

Chang, E., Rieger, J., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T, 2010. Categorical speech representation in human superior temporal gyrus. Nat Neurosci. 13, 1428–1432.

Chevillet, M.A., Jiang, X., Rauschecker, J.P., Riesenhuber, M., 2013. Automatic Phoneme Category Selectivity in the Dorsal Auditory Stream. Journal of Neuroscience. 33 (12), 5208-5215.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., Fadiga, L., 2009. The motor somatotopy of speech perception. Curr Biol. 19(5), 381-385.

D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I., Craighero, L., 2011. Tongue corticospinal modulation during attended verbal stimuli: Priming and coarticulation effects. Neuropsychologia. 49 (13), 3670-3676.

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.

DeWitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci USA. 109(8), 505-514.

Diehl, R., Lotto, A., Holt, L., 2004. Speech perception. Annual Review of Psychology, 55, 149–179.

Du, Y., Buchsbaum, B.R., Grady, C.L., Alain, C., 2014. Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc Natl Acad Sci USA. 111(19), 7126-7131.

Engel, S.A., 2005. Adaptation of oriented and unoriented color-selective neurons in human visual areas. Neuron. 45(4), 613-23.

Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. European Journal of Neuroscience. 15, 399-402.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R.S., Turner, R., 1995. Analysis of fMRI time-series revisited. Neuroimage. 2(1), 45-53.

Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. How many subjects constitute a study? Neuroimage. 10(1), 1-5.

Grabski, K., Lamalle, L., Vilain, C., Schwartz, J.L., Troprès, I., Vallée, N., Baciu, M., Le Bas, J.F., Sato, M., 2012. Functional MRI assessment of orofacial articulators: neural correlates of lip, jaw, larynx and tongue movements. Human Brain Mapping. 33(10), 2306-2321.

Grabski, K., Tremblay, P., Gracco, V.L., Girin, L., Sato, M., 2013. A mediating role of the auditory dorsal pathway in selective adaptation to speech: A state-dependent transcranial magnetic stimulation study. Brain Research. 1515, 55-65.

Grabski, K., Sato, M., 2020. Adaptive phonemic coding in the listening and speaking brain. Neuropsychologia. 136, 107267.

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., Lachaux, J.P., Dohen, M., Perrier, P., Garnier, M., Baciu, M., Lœvenbruck, H., 2019. The ConDialInt Model: Condensation, Dialogality, and Intentionality Dimensions of Inner Speech Within a Hierarchical Predictive Control Framework. Front. Psychol. 10:2019.

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci. 10(1), 14-23.

Hasson, U., Skipper, J. I., Nusbaum, H. C., & Small, S. L., 2007. Abstract coding of audiovisual speech: beyond sensory representation. Neuron. 56(6), 1116–1126.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8(5), 393-402.

Hillyard, S.A., Vogel, E.K., Luck S.J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Phil. Trans. R. Soc. Lond. B.Biol. Sci. 353, 1257–1270.

Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L.M., Niemi, J., 2003. The effects of attention on speech perception: An fMRI study, Brain and Language. 85 (1), 37-48.

Huk, A.C., Heeger, D.J., 2002. Pattern-motion responses in human visual cortex. Nat Neurosci. 5(1), 72-75.

Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., Horwitz, B., 2006. Neural bases of categorization of simple speech and nonspeech sounds. Hum Brain Mapp. 27(8), 636-651.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., Dupoux, E., 2003. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. J. Neurosci, 23(29), 9541-9546.

Jacquemot, C., Scott, S.K., 2006. What is the relationship between phonological short-term memory and speech processing? Trends Cogn. Sci. 10, 480-486.

Joanisse, M.F., Gati, J.S., 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage. 19(1), 64-79.

Joanisse, M.F., Zevin, J.D., McCandliss, B.D., 2007. Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm. Cereb Cortex. 17(9), 2084-2093.

Kingston, J., Diehl, R.L. 1994. Phonetic knowledge. Language, 70, 419–54.

Klatt, D.H., Klatt, L.C., 1990. Analysis, synthesis and perception of voice quality variations among male and female talkers. Journal of the Acoustical Society of America. 87, 820–856.

Kleinschmidt, D. Jaeger, T. F., 2015. Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. Psychological Review, 122, 148–203.

Kluender K.R., 1994. Speech perception as a tractable problem in cognitive science. In M.A. Gernsbacher (Ed.) Handbook of Psycholinguistics (pp. 173–217). San Diego, CA: Academic.

Lacadie, C.M., Fulbright, R.K., Constable, R.T., Papademetris, X. 2008. More accurate Talairach coordinates for neuroimaging using nonlinear registration. NeuroImage, 42(2), 717-725.

Laurent, R., Barnaud, M.L., Schwartz, J.L., Bessière, P., Diard, J., 2017. The complementary roles of auditory and motor information evaluated in a Bayesian perceptuo-motor model of speech perception. Psychological Review. 124(5), 572-602.

Lawyer, L., Corina, D., 2014. An Investigation of Place and Voice Features Using fMRI-Adaptation. J Neurolinguistics. 27(1).

Liberman, A.M., Cooper, F.S., Shankweiler, D., Studdert-Kennedy, M., 1967. Perception of the Speech Code. Psychological review. 74, 431-461.

Liberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. Cognition. 21(1), 1–36.

Lotto, A.J., 2000. Language acquisition as complex category formation. Phonetica, 57, 189–96.

Luthra, S., 2021. The Role of the Right Hemisphere in Processing Phonetic Variability Between Talkers. Neurobiology of Language. 2, 138-151.

Massaro, D.W., Oden, G.C., 1980. Evaluation and integration of acoustic features in speech perception. Journal of the Acoustical Society of America, 67, 996–1013.

McGettigan, C., Scott, S. K., 2012. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? Trends in Cognitive Sciences. 16, 269-276.

Möttönen, R., Dutton, R., Watkins, K.E., 2013. Auditory-motor processing of speech sounds Cerebral Cortex. 23 (5), 1190-1197.

Möttönen, R., van de Ven, G.M., Watkins, K.E., 2014. Journal of Neuroscience. 34 (11), 4064-4069.

Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.L., Diard, J., 2012. Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory Bayesian modeling study. Language and Cognitive Processes, Taylor & Francis (Routledge). 27 (7-8), 1240-1263.

Moulin-Frier, C., Diard, J., Schwartz, J.L., Bessière, P., 2015. COSMO ("Communicating about Objects using Sensory–Motor Operations"): A Bayesian modeling framework for studying speech communication and the emergence of phonological systems. Journal of Phonetics. 53, 5-41.

Murray, S., Wojciulik, E., 2004. Attention increases neural selectivity in the human lateral occipital complex. Nat Neurosci. 7, 70–74.

Myers, E.B., Blumstein, S.E., Walsh, E., Eliassen, J., 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychological science. 20(7), 895–903.

Myers, E.B., Swan, K., 2012. Effects of Category Learning on Neural Sensitivity to Non-native Phonetic Categories. J Cog Neurosci. 24 (8), 1695–1708.

Nearey, T.M., 1990. The segment as a unit of speech perception. Journal of Phonetics, 18, 347–73

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9, 97–113.

Paulesu, E., Frith, C.D., Frackowiak, R.S., 1993. The neural correlates of the verbal component of working memory. Nature. 362(6418), 342-345.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci USA. 103(20),7865-70.

Raizada, R.D., Poldrack, R.A., 2007. Selective amplification of stimulus differences during categorical processing of speech. Neuron. 56(4), 726-40.

Rampinini, A., Handjaras, G., Leo, A., Cecchetti, L., Ricciardi, E., Marotta, G., Pietrini, P., 2017. Functional and spatial segregation within the inferior frontal and superior temporal cortices during listening, articulation imagery, and production of vowels. Sci Rep 7, 17029.

Sabri, M.., Binder, J.R., Desai, R., Medler, D.A., Leitl, M.D., Liebenthal, E., 2008. Attentional and linguistic interactions in speech perception. NeuroImage. 39(3), 1444-1456.

Sato, M., Tremblay, P., Gracco, V., 2009. A mediating role of the premotor cortex in phoneme segmentation. Brain and Language. 111(1), 1-7.

Sato, M., Grabski, K., Glenberg, A., Brisebois, A., Basirat, A., Ménard, L., Cattaneo, L., 2011. Articulatory bias in speech categorization: evidence from use-induced motor plasticity. Cortex. 47(8), 1001-1003.

Schomers, M.R., Pulvermüller, F., 2016. Is the Sensorimotor Cortex Relevant for Speech Perception and Understanding? An Integrative Review. Front Hum Neurosci. 10, 435.

Schouten, B., Gerrits, E., van Hessen, A., 2003. The end of categorical perception as we know it. Speech Commun. 41(1), 71-80.

Schroeder, M.R., Atal, B.S., Hall, J.L., 1979. Objective measure of certain speech signal degradations based on masking properties of human auditory perception, in: Lindblom, B., Öhman, S. (Eds.), Frontiers of Speech Communication Research, Academic Press, London. 217-229.

Schwartz, J.-L., Basirat, A., Ménard, L., Sato, M., 2012. The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. Journal of Neurolinguistics, 25, 336–354.

Skipper, J.I., Nusbaum, H.C., Small S.L., 2005. Listening to talking faces: motor cortical activation during speech perception. NeuroImage, 25, 76–89.

Skipper J.I., van Wassenhove, V., Nusbaum, H.C., Smal,l S.L., 2007. Hearing lips and seeing voices : how cortical areas supporting speech production mediate audiovisual speech perception. Cereb Cortex, 17(10), 2387-2399.

Skipper, J.I., Devlin, J.T., Lametti, D.R., 2017. The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. Brain Lang. 164, 77-105.

Spitzer, H., Desimone, R., Moran, J., 1988. Increased attention enhances both behavioral and neuronal performance. Science. 240(4850), 338-40.

Stokes, R.C., Venezia, J.H. Hickok, G., 2019. The motor system's [modest] contribution to speech perception. Psychonomic Bulletin & Review, 26, 1354–1366.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15(1), 273-89.

Vanzetta, I., Slovin, H., 2010. A BOLD Assumption. Frontiers in Neuroenergetics, 2(24), 1-4.

Warlaumont, A., 2020. Infant Vocal Learning and Speech Production. In J. Lockman & C. Tamis-LeMonda (Eds.), The Cambridge Handbook of Infant Development: Brain, Behavior, and Cultural Context (Cambridge Handbooks in Psychology, pp. 602-631). Cambridge: Cambridge University Press.

Watkins, K.E., Strafella, A.P., Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 41(8), 989-94.

Wild, C.J., Yusuf, A, Wilson, D.E., Peelle, J.E., Davis, M.H., Johnsrude, I.,S., 2012. Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention. Journal of Neuroscience. 32 (40), 14010-14021.

Wilson, S.M., Iacoboni, M., 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. Neuroimage. 33(1), 316-25.

Zekveld, A., Heslenfeld, D., Festen, J., Schoonhoven, R., 2006. Top-down and bottom-up processes in speech comprehension. NeuroImage. 32, 1826-36.

Zevin, J.D., McCandliss, B.D., 2005. Dishabituation of the BOLD response to speech sounds. Behav Brain Funct 1.

Zhang, X., Pan, W-J., Dawn Keilholz, S., 2020. The relationship between BOLD and neural activity arises from temporally sparse events, NeuroImage, 207, 11639