



HAL
open science

Adaptive phonemic coding in the listening and speaking brain

Krystyna Grabski, Marc Sato

► **To cite this version:**

Krystyna Grabski, Marc Sato. Adaptive phonemic coding in the listening and speaking brain. *Neuropsychologia*, 2020, 136, pp.107267. 10.1016/j.neuropsychologia.2019.107267 . hal-03096519

HAL Id: hal-03096519

<https://hal.science/hal-03096519>

Submitted on 21 Dec 2021

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

ADAPTIVE PHONEMIC CODING IN THE LISTENING AND SPEAKING BRAIN

Krystyna Grabski & Marc Sato^{CA}

Aix-Marseille Université, CNRS, Laboratoire Parole & Langage, Aix-en-Provence, France

^{CA} Correspondence can be addressed to Marc Sato, Laboratoire Parole & Langage, UMR CNRS 7309, Aix-Marseille Université, 5, avenue Pasteur - 13100 Aix-en-Provence, France. Tel: (+33) (0) 413 552 712 - Email: marc.sato@lpl-aix.fr.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

K.G. and M.S. designed research; K.G. performed research; K.G. and M.S. analyzed data; M.S. wrote the paper.

ACKNOWLEDGMENTS

We thank Laurent Lamalle and Coriandre Vilain for their help in data acquisition. This study was supported by a research grant from Grenoble-INP (BQR 'Modyc: Modélisation dynamique de l'activité cérébrale') to M.S. and K.G.

ABSTRACT

In order to determine the neural substrates of phonemic coding during both listening and speaking, we used a repetition suppression (RS) paradigm in which vowels were repeatedly perceived or produced while measuring BOLD activity with sparse sampling functional magnetic resonance imaging (fMRI). RS refers to the phenomenon that repeated stimuli or actions lead to decreased activity in specific neural populations associated with enhanced neural selectivity and information coding efficiency. Common suppressed BOLD responses during repeated vowel perception and production were observed in the inferior frontal gyri, the posterior part of the left middle temporal gyrus and superior temporal sulcus, the left intraparietal sulcus, as well as in the cingulate gyrus and **presupplementary** motor area. By providing evidence for **common** adaptive neural changes in **premotor and associative auditory and somatosensory brain areas**, the observed RS effects suggest that phonemic coding is partly **driven by** shared sensorimotor regions in the listening and speaking brain.

KEY-WORDS

Speech perception, speech production, dorsal stream, fMRI adaptation, repetition-suppression.

INTRODUCTION

Speech communication requires phonological encoding and execution of the related articulatory plans by the speaker, followed by acoustico-phonetic decoding of the speech signal by the listener. To be efficient, speech communication is thought to imply a principle of parity between the speaker and listener, through common and shared phonemic goals (e.g., Mattingly and Liberman, 1988; Liberman and Whalen, 2000). From this view, theoretical and neurobiological models of speech perception and production argue that phonemic units are coded and processed in multidimensional motor and sensory spaces. In these models, auditory processing and speech motor control partly operate through a cortical dorsal stream that mediate mapping between auditory, somatosensory and articulatory motor speech representations. Motor activity observed during speech perception has been proposed to partly constrain phonetic interpretation of the sensory inputs through the internal generation of candidate articulatory categorizations (e.g., Hickok and Poeppel, 2007; Skipper et al., 2007; Rauschecker and Scott, 2009; Rauschecker, 2011; Schwartz et al., 2012; Bornkessel-Schlesewsky et al., 2015; Skipper et al., 2017; Schomers and Pulvermüller, 2017). During speech production, modulation of neural responses observed within the auditory and somatosensory cortices are thought to reflect feedback control mechanisms in which sensory consequences of the speech motor act are evaluated with actual sensory inputs in order to evaluate accurate phonemic production (e.g., Houde and Nagarajan, 2011; Hickok et al., 2011; Guenther and Vladusich, 2012; Perkell, 2012). Taken together, these models argue for a functional coupling between speech production and perception systems in the speaking and listening brain, with phonetic processing partly driven by internal motor-to-sensory and sensory-to-motor simulation.

From this theoretical perspective, clear support for sensorimotor activity during both acoustico-phonetic processing and speech motor control are provided by fMRI studies that conjointly examined the neural correlates of speech production and perception (e.g., Wilson et al., 2004; Pulvermüller et al., 2006; Skipper et al., 2007; Callan et al., 2010; Zheng, Munhall and Johnsrude, 2010; Tremblay and Small, 2011; Grabski et al., 2013a; Rampinini et al., 2017). In line with the above-mentioned neurobiological models, these studies highlighted common neural activity in sensory and motor neural regions within the auditory cortex, the inferior parietal cortex, as well as the inferior frontal gyrus and adjacent ventral premotor cortex. Focussing on either speech perception or production in isolation, other studies further indicated that sensorimotor activity within the cortical dorsal stream correlates with auditory phonemic categorization (Alho et al., 2012; Chevillet et al., 2013; Alho et al., 2014) in a causal manner (Grabski et al., 2013b; Murakami et al., 2015), can predict the accuracy between the intended and actual phonemic production (Tourville et al., 2007; Golfinoopoulos et al., 2011; Niziolek, Nagarajan and Houde, 2013), and can partly be decoded to reveal produced or perceived phonemic features (Lee et al., 2012; Bouchard et al., 2013; Du et al., 2014; Mesgarani et al., 2014; Rampinini et al., 2017).

In order to determine whether shared sensorimotor brain areas are involved in phonemic adaptive coding during both listening and speaking, we here used an fMRI adaptation paradigm in which vowels, as elementary and universal speech units, were repeatedly perceived and produced by the same set of participants. fMRI adaptation is based on the phenomenon that repeated stimuli or motor acts lead to a reduction in the BOLD signal in specific brain areas that are sensitive to the observed stimuli and/or performed actions (the so-called repetition suppression effect or RS; see Grill-Spector and Malach, 2001; Grill-Spector, Henson and Martin, 2006). Although a number of potential neural/synaptic mechanisms and theoretical models have been proposed to explain RS (for reviews, see Grill-Spector et al., 2006; Friston, 2012; Gotts, Chow and Martin, 2012; Henson, 2012), all are associated with enhanced neural selectivity, increased processing, and information coding efficiency in relation to attributes of the repeated stimulus or motor act. Although debated, RS has also been explained in terms of top-down inference and sensorimotor learning implemented under predictive coding (with reduced prediction errors about the content/precision of a repeated stimulus supposed to be reflected in BOLD suppression; e.g., Friston, 2012; Auztulewicz and Friston, 2016; see also Grotheer and Kovács, 2016).

While a number of fMRI studies examined RS effects during speech perception (Hasson et al., 2007; Joanisse et al., 2007; Vaden et al., 2010; Myers and Swaan, 2012; Lawyer and Corina, 2015) or production (Peeva et al., 2010; Grabski et al., 2012a; Sato et al., 2015; Okada et al., 2018) using vowels, syllables or pseudowords, in our best knowledge no study attempted to identify the neural substrate of adaptive phonemic coding during both listening and speaking. Generally, adaptive phonemic coding here refers to the neurocognitive processes involved in speech perception and production that show enhanced selectivity, increased processing, and information coding efficiency in relation to a repeated phoneme. To this aim, the present fMRI adaptation study involved passive vowel listening and overt vowel production by healthy native French speakers. To minimize possible covert motor simulation, the auditory perception task was first performed, with the participants asked to passively listen to French vowels (/i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/, /ɔ/). In the subsequent production task, participants were asked to overtly produce the same vowels. Crucially, each vowel was perceived or produced in six consecutive trials. This sequence allowed measuring changes in BOLD signal from the first to the sixth repetitions and therefore to precisely determine the time-course of neural adaptation during repeated perceived and produced vowels. Importantly, the selected vowels included distinct phonetic features such as height (close, mid-close and mid-open vowels), backness (front and back vowels) and roundedness (rounded and unrounded vowels). Finally, in order to limit possible noise-induced bias in our results, we used a sparse, clustered acquisition fMRI technique during both vowel perception and production, which allows stimulus presentation and overt responses to occur in relative silence as well as to eliminate the susceptibility artifact due to articulatory-related movement in the production task.

Irrespective of RS, BOLD activity averaged across the six repetitions in the perception and production tasks was first determined. To identify overlapping activation between the two tasks, a conjunction analysis

was subsequently conducted. These analyses allowed us to compare the observed results with previous studies that conjointly examined the common neural substrates of speech perception and production (e.g., Wilson et al., 2004; Callan et al., 2010; Pulvermüller et al., 2006; Skipper et al., 2007; Zheng, Munhall and Johnsrude, 2010; Tremblay and Small, 2011; Grabski et al., 2013a; Rampinini et al., 2017). Critically, for the two tasks, two different time-courses of adaptation across the six repetitions were tested, corresponding either to a linear or a categorical (from the first trial versus the others) decrease of the BOLD response (Noppeney and Penny, 2006; Grabski et al., 2012a). Then, in order to identify common RS effects between the perception and production tasks, a conjunction analysis was performed. Finally, possible linear and categorical repetition enhancements across the six repetitions were also tested.

In keeping with the above-mentioned studies and taking advantage of the fMRI adaptation paradigm, our main hypothesis was that BOLD changes associated with phonemic adaptation might occur in sensorimotor regions during both listening and speaking. Adaptive BOLD responses in shared premotor, auditory and somatosensory brain areas during both repeated perceived and produced vowels would strengthen the theoretical proposal that phonemic coding is partly driven in the same sensorimotor neural regions, through a cortical dorsal stream, in the listening and speaking brain.

METHODS

Participants

Twelve healthy adults (eight males and four females with a mean age of 29 years, ranging from 20 to 42 years), native French speakers, participated in the study after giving their informed consent. All were right-handed according to standard handedness inventory (Oldfield, 1971), had normal or corrected-to-normal vision and reported no history of motor, speaking or hearing disorders. Participants were screened for neurological, psychiatric, other possible medical problems and contraindications to MRI. The protocol was approved by the Grenoble University Ethical Committee and was carried out in accordance with the ethical standards of the 1964 Declaration of Helsinki.

Stimuli

Given the well-known and acknowledged inter-individual differences in the anatomy of the vocal apparatus and its acoustic consequences in speech behaviors (Ladefoged, 2006), the acoustic properties of the perceived and produced vowels were matched for each participant. This procedure allowed determining the neural substrates of vowel perception and production, involving for the two tasks the same acoustic-phonetic dimensions. To this aim, vowels delivered in the perception task were individually recorded by each participant prior to the experiment. Multiple utterances of nine French steady-state vowels (/i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/, /ɔ/) were individually recorded by each participant in a soundproof room (sampling rate of 44.1 kHz with 16-bit quantization recording). The nine selected vowels included a set of distinct phonetic features including height (close, mid-close and mid-open vowels), backness (front and back vowels) and roundedness (rounded and unrounded vowels). With this procedure, six clearly articulated tokens were selected per speaker and per vowel for the perception task. On average, the distribution of formant values for French vowels appeared classical, with some dispersion due to inter-subject and gender differences (see Table S1 in Supplemental Data).

Procedure

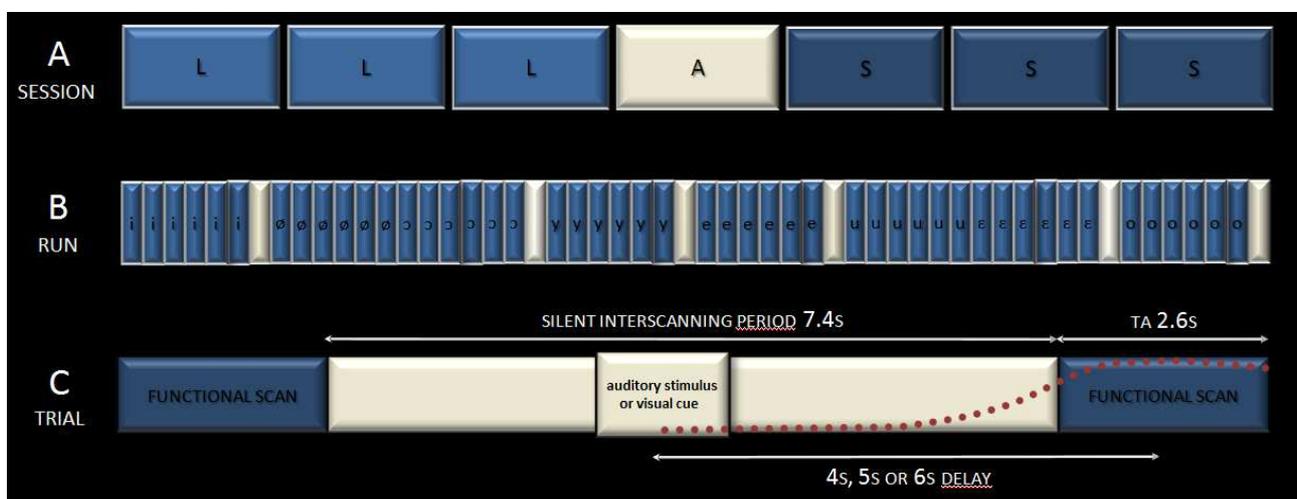


Figure 1. Experimental design. A) The experiment involved passive listening (L) and overt speaking (S), and consisted of six runs, each lasting approximately ten minutes, and one anatomical scan (A). B) Each run consisted of fifty-four trials in which participants were asked to listen to or to overtly produce /i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/ and /ɔ/ vowels, and six baseline trials. In order to determine RS, each vowel was perceived or produced in six consecutive trials. C) For each trial, the time interval between the perceived (auditory stimulus) or produced (visual cue) vowel and the midpoint of the following functional scan acquisition was of 4s, 5s or 6s (the predicted hemodynamic response function is indicated in red, TA: Time of Acquisition).

The experiment consisted of six functional runs: the first three runs involved passive vowel listening and the last three runs involved overt vowel production (see Figure 1). The procedure was similar to the one used in a previous fMRI study (Grabski et al., 2013a), except that, in each functional run, each vowel was here perceived or produced in six consecutive trials. This sequence allowed measuring changes in BOLD signal for repeated compared to novel vowels.

At the beginning of the scanning session, an intensity matching procedure was done for each participant (see Christoffels, Formisano and Schiller, 2007; Grabski et al., 2013a), in order to match as closely as possible the volume of the auditory stimuli in the perception task and that of the auditory feedback in the production task (despite of bone/skull conduction). To this aim, after being placed in the scanner, participants first overtly produced several vowels and then passively listened to few vowels previously recorded from their own voice. Participants were asked whether or not the volume was similar to hearing their own voice and, if necessary, the volume of the stimuli was increased or decreased.

To minimize possible covert motor simulation, the perception task was first performed with the participants just being instructed to pay attention to the auditory stimuli. Participants were asked to passively listen to each of the nine French steady-state vowels previously recorded from their own voice (six distinct occurrences per vowel). A resting condition, without any movement or auditory stimulation, served as baseline. In each of the three functional runs, each vowel (/i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/ or /ɔ/) was randomly presented in six consecutive trials (with six distinct occurrences of the same vowel). Each vowel type differed from the preceding one by at least one or more phonetic features including height, backness and roundedness (e.g., close front unrounded /i/ vowel and mid-open back rounded /ɔ/ vowel). Each trial was 10s in length with a single vowel being presented or without any auditory presentation in the resting condition. In the production task, participants were asked to overtly produce the same nine French steady-state vowels in the same randomized sequence. As in the perception task, a resting condition, without any movement or auditory stimulation, was also added. In each trial (10s), a visual instruction related to the vowel (e.g., "i") or to the resting condition ("--") was displayed for 1000ms. Participants were instructed to produce each vowel from a neutral closed-mouth position as soon as they perceived the visual instruction and to maintain the production until the visual cue disappeared.

In each functional run, each vowel (/i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/ or /ɔ/) was perceived or produced in six consecutive trials in a random sequence. This RS structure allows to measure changes in the BOLD signal for repeated compared to novel stimuli. Altogether, 360 functional scans were therefore acquired (2

tasks x 3 runs x (9 vowels + 1 resting baseline) x 6 repetitions). In addition, three 'dummy' scans at the beginning of each run were added to allow for equilibration of the MRI signal and were removed from the analyses. The total duration of the experiment was around one hour and half.

Data acquisition

Magnetic resonance images were acquired with a 3T whole-body MRI scanner (Bruker Medspec S300) equipped with a transmit/receive quadrature birdcage head coil. Participants laid supine in the scanner with head movements minimized with foam cushions. To reduce auditory exposure to scanner noise, they wore earplugs in addition to protective MRI-compatible headphones, equipped with noise-reducing passive material, through which auditory stimuli were administered to participants during the perception task. In the production task, vowel productions were recorded using an MRI-compatible microphone for offline analysis (see below). Visual instructions were presented using the Presentation software (Neurobehavioral Systems, Albany, USA) and displayed on a screen situated behind the scanner and viewed on a mirror fixed above the subject's eyes.

The fMRI experiment consisted of six functional runs and one anatomical run. Functional images were obtained using a T2*-weighted, echoplanar imaging (EPI) sequence with whole-brain coverage (TR = 10s, acquisition time = 2600ms, TE = 30ms, flip angle = 90°). Each functional scan comprised forty axial slices parallel to the anteroposterior commissural plane acquired in interleaved order (72 x 72 matrix; field of view: 216 x 216 mm²; 3 x 3 mm² in plane resolution with a slice thickness of 3mm without gap). A high-resolution T1-weighted whole-brain structural image was acquired for each participant after the third functional run (MP-RAGE, sagittal volume of 256 x 224 x 176mm³ with a 1 mm isotropic resolution, inversion time = 900ms, two segments, segment repetition time = 2500ms, segment duration = 1795ms, TR/TE = 16/5 in ms with 35% partial echo, flip angle = 30°).

In order to avoid movement artefacts due to vowel production and to minimize scanner noise during both vowel perception and production, a sparse sampling acquisition paradigm was used (e.g., Birn et al., 1999; Hall et al., 1999; Gracco, Tremblay and Pike, 2005). Functional scanning therefore occurred only during a fraction of the TR, alternating with silent interscanning periods, where participants listened to or produced single vowels. Since the rising hemodynamic response is estimated to occur with a 4-6 s delay in case of speech perception and production, the time interval between the vowel onset and the midpoint of the following functional scan acquisition was randomly varied between 4s, 5s or 6s across trials (Grabski et al., 2012a, 2012b, 2013a).

Data Analyses

Data were analyzed using the SPM8 software package (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK) running on Matlab (Mathworks, Natick, MA, USA). The 'Automated Anatomical Labeling' (ALL) atlas (Tzourio-Mazoyer et al., 2002) was used to identify labels for activated

peaks. For visualization, activation maps were superimposed on a standard brain template using the MRICRON software (<http://www.sph.sc.edu/comd/rorden/mricron/>).

Production errors. Participant responses in the production task were analyzed offline for possible error productions. Three types of errors were observed: omission, wrong production and hesitation. In total, 17 functional scans in which an error occurred were removed from the statistical analyses (on average $0.73\% \pm 0.15\%$ of errors per participant), with general performance exceeding 99 %, and at least 90 % correct responses in all runs for all participants.

Data preprocessing. For each participant, the functional series were first realigned by estimating the 6 movement parameters of a rigid-body transformation in order to control for head movements between scans. After segmentation of the T1 structural image (using the unified segmentation model) and coregistration to the mean functional image, all functional images were spatially normalized into standard stereotaxic space of the Montreal Neurological Institute (MNI) using segmentation parameters of the T1 structural image. All functional images were then smoothed using a 6mm full-width at half maximum Gaussian kernel, in order to improve the signal-to-noise ratio and to compensate for the anatomical variability among individual brains.

Individual analyses. Neural activations were analyzed using the General Linear Model, including for each of the six functional runs (three for each task), six regressors of interest (one for each vowel repetition) and the six realignment parameters, with the silent trials forming an implicit baseline. The BOLD response for each event was modeled using a single-bin finite impulse response (FIR) basis function spanning the time of acquisition (2.6s). Before estimation, for each run, a high-pass filtering with a cutoff period of 128 s was applied in order to remove low-frequency drifts (thought to be caused by physiological noise as well as by physical, scanner-related, noise; everything below this frequency was removed from the data, this includes quadratic functions (of time) that are below this frequency). Beta weights associated with the modeled FIR responses were then computed to fit the observed BOLD signal time course in each voxel for each condition. Individual statistical maps were calculated for each repetition in each task with the related baseline and subsequently used for group statistics.

Group analyses. In order to draw population-based inferences, a second-level random effect group analysis was carried-out. A flexible factorial design was used with the task (2 levels: perception, production) and the repetition (6 levels: RS1, RS2, RS3, RS4, RS5, RS6) as within-subject factors and the subjects treated as a random factor (Glascher and Gitelman, 2008).

First, irrespective of RS, t-contrasts were calculated to determine brain activity averaged across the six repetitions, as well as for each repetition individually, in the perception and production tasks (compared to the resting baseline). To identify overlapping activation between the two tasks, conjunction analyses were subsequently conducted on these contrasts.

Second, for the two tasks, two different time-courses of adaptation across the six repetitions were

tested, corresponding to a linear decrease or a categorical decrease (from the first trial versus the others) of the BOLD response. Indeed, in previous fMRI adaption studies, BOLD activity was found to decrease monotonically with the number of stimulus repetitions or to reach a plateau after the first repetition (i.e., with the biggest adaptive changes in the BOLD signal happening after the first repetition; see Grill-Spector and Malach, 2001; Grill-Spector, Henson and Martin, 2006; Noppeney and Penny, 2006). The time courses of adaptation were entered as contrast weights in two parametric t-contrasts (linear decrease: 1 0.6 0.2 - 0.2 -0.6 -1; categorical decrease: 1 -0.2 -0.2 -0.2 -0.2 -0.2) in order to test the predicted parametric patterns of decreasing BOLD signal amplitude. To identify common RS between the perception and production tasks, a conjunction analysis was performed on these parametric contrasts.

Third, in order to identify brain regions that differ between the two tasks with respect to RS effects (i.e., for which significant RS effects were observed in one task but not the other), we recomputed linear and categorical parametric t-contrasts in the listening and the speaking task but with an exclusive mask related to the other task. For concision, results from these exclusive masking contrasts are reported in Supplemental Data (see Figure S1 and Table S4).

Finally, two additional t-contrasts were calculated for each task in order to test possible repetition enhancement across the six repetitions (linear increase: -1 -0.6 -0.2 0.2 0.6 1; categorical increase: -1 0.2 0.2 0.2 0.2 0.2).

All contrasts were calculated with a Family-Wise-Error (FWE) corrected level of $p < .05$ at the cluster level. For the exclusive masking contrasts, the exclusive mask was thresholded at an uncorrected level of $p < .05$ whereas the contrasts of interest was calculated with a FWE corrected level of $p < .05$ at the cluster level (it should be noted that that the more liberal the threshold of an exclusive mask, the more conservative is the masking procedure for the contrast of interest).

RESULTS

Neural correlates of vowel perception and production

Irrespective of RS effects, BOLD activity averaged across the six repetitions in the perception and the production tasks and overlapping activity between the two tasks were first identified (see Figure 2, left). Vowel perception induced large bilateral auditory activity in the superior temporal gyrus/sulcus, including the primary, secondary and associative auditory cortices, extending ventrally to the dorsal part of middle temporal gyrus, dorsally to the ventral part of inferior parietal cortex and parietal operculum, and medially to the thalamus. Vowel production induced bilateral activations of the primary sensorimotor and premotor cortices and of the superior temporal gyrus/sulcus. Additional activity was found in the supplementary motor area and cingulate cortex, the inferior and superior parietal cortices, the parietal operculum, the visual cortex, the anterior insular cortex, the basal ganglia, and the cerebellum. The conjunction analysis revealed common bilateral activation of the primary, secondary and associative auditory cortices, extending dorsally to the ventral part of inferior parietal cortex and parietal operculum. In sum, BOLD activity observed in the perception and production tasks appears fully consistent with previous brain-imaging studies, with large overlapping auditory activity between the tasks as well as specific brain areas classically involved in motor preparation, execution and coordination activated in the production task (for a previous fMRI study on vowel perception and production, see Grabski et al., 2013a).

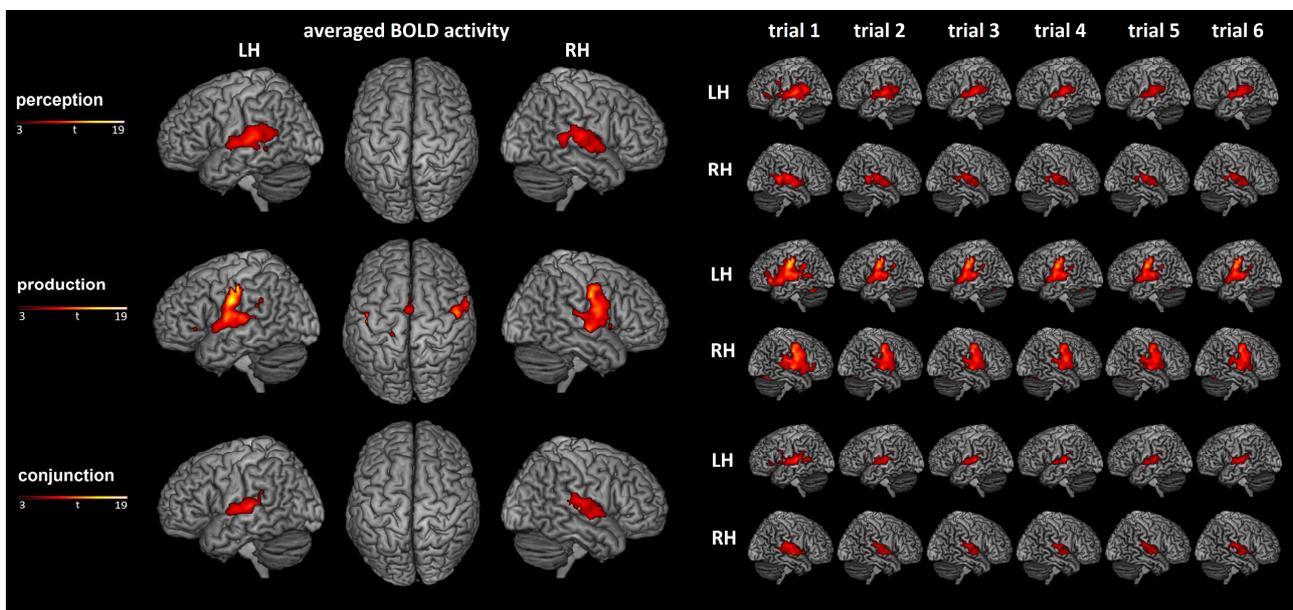


Figure 2. BOLD activity averaged across the six repetitions (left) and for each repetition (right) in the perception and the production tasks and overlapping activity between the two tasks ($p < .05$ FWE corrected at the cluster level).

It should however be noted that, compared to previous studies that conjointly examined speech perception and production (e.g., Wilson et al., 2004; Callan et al., 2010; Skipper et al., 2007; Tremblay and Small, 2011; Grabski et al., 2013a, Rampinini et al., 2017), no common activity was observed in inferior frontal and premotor regions, a result likely due to the adaptation paradigm and decreased activity across

consecutive similar vowels. In line with this hypothesis, common BOLD activity in left inferior frontal and ventral premotor regions was observed for the first vowel, but not for the following ones (see Figure 2, right).

Repetition suppression during vowel perception

Linear and categorical RS effects during vowel perception were observed in largely similar neural clusters (although the location of the reported maximum activation peaks and their corresponding time-course of adaptation across the six consecutive trials slightly differ), with some linear and categorical RS profiles quite similar (see Figure 3 and Tables S2A and S2B for details). While distinct neural populations within the same region/cluster might show specific sensitivity to linear or categorical RS effect, this is likely to be due to the limited sample size and associated inter-subject BOLD variability in evaluating and dissociating linear and categorical RS profiles across participants (see Discussion). These common clusters included neural populations in the superior temporal gyri/sulci, extending ventrally to the middle temporal gyri, in the anterior and posterior parts of the inferior frontal gyrus (pars opercularis and triangularis), extending to the left ventral premotor cortex and anterior insula, and in the cingulate cortex, the supplementary motor area and medial frontal gyrus. In addition, a specific linear RS effect was observed in the left intraparietal sulcus. Conversely, specific categorical RS effects were observed in two clusters located in the left anterior insula, pars orbitalis and pars opercularis of the inferior frontal gyrus as well as in the thalamus.

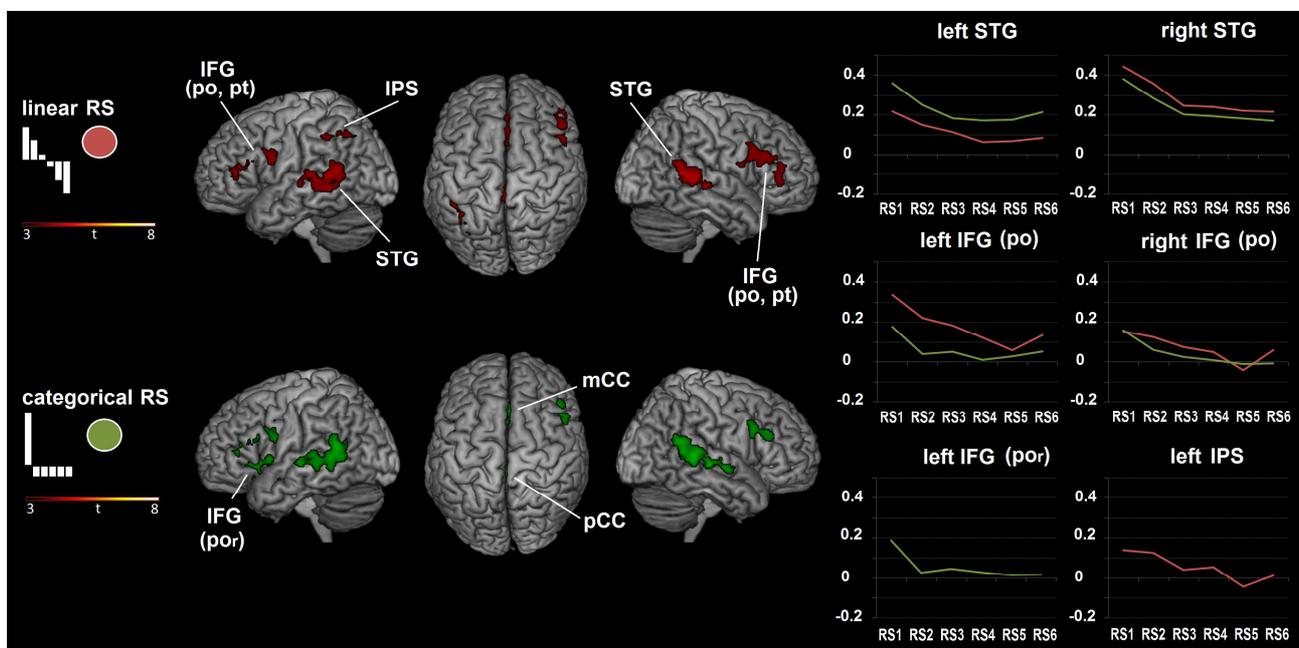


Figure 3. Brain regions showing linear and categorical RS in the perception task (i.e., linear and categorical decreases of BOLD activity across vowel repetition) and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level, see Tables S2A and S2B in Supplemental Data for details).

Repetition suppression during vowel production

Similarly to the vowel perception task, linear and categorical RS effects during vowel production were observed in largely similar neural clusters (see Figure 4 and Tables S3A and S3B for details), including neural populations in the anterior and posterior parts of the inferior frontal gyrus (pars opercularis, triangularis and orbitalis), extending to the premotor and anterior insular cortices, in the posterior part of the left middle temporal gyrus extending to the superior temporal sulcus, in the basal ganglia and thalamus, in the supplementary motor area and cingulate cortex, in the left inferior parietal lobule extending to the precuneus and cuneus, and in the extrastriate cortex. In addition to these effects, categorical RS effects were specifically observed in the left posterior part of the inferior temporal gyrus, in the right inferior parietal lobule extending to the precuneus and cuneus, and in the superior part of the cerebellum.

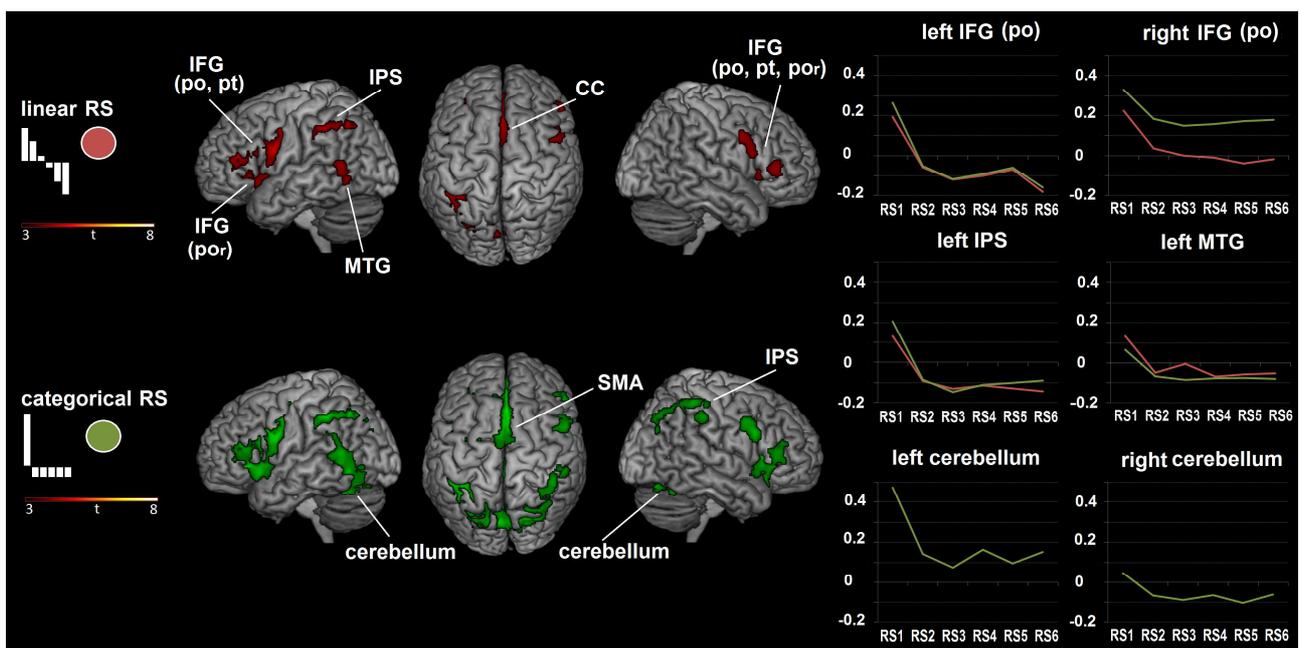


Figure 4. Brain regions showing linear and categorical RS in the production task (i.e., linear and categorical decreases of BOLD activity across vowel repetition) and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level, see Tables S3A and S3B in Supplemental Data for details).

Common repetition suppression during vowel perception and production

Brain regions showing common RS effects to the tasks are provided in Figure 5 and Tables 1A and 1B. Both linear and categorical BOLD decreases were observed in the pars opercularis and pars triangularis of the left inferior frontal gyrus extending to the ventral premotor cortex, and in the cingulate cortex, the supplementary motor area (pre-SMA) and medial frontal gyrus. In addition, linear RS was specifically observed in the left intraparietal sulcus, as well as categorical RS in the pars orbitalis of the left inferior frontal gyrus and anterior insula, in the pars opercularis and pars triangularis of the right inferior frontal gyrus, and in the posterior part of the left middle temporal gyrus extending to the superior temporal sulcus.

Importantly, although RS effects were observed in these brain areas during both listening and speaking, the magnitude of BOLD activity across vowel repetition was similar across tasks in several brain areas but it differed in other regions (see the contrast estimates in Tables 1A and 1B). For example, RS observed in the left middle temporal gyrus was similar in both tasks, suggesting similar adaptive changes in this region. On

the contrary, the magnitude of BOLD activity observed in the pre-SMA was similar in both tasks but with a stronger BOLD activity in the first vowel during the speaking task. This later result argues for qualitatively similar but quantitatively distinct adaptive changes across tasks in this region.

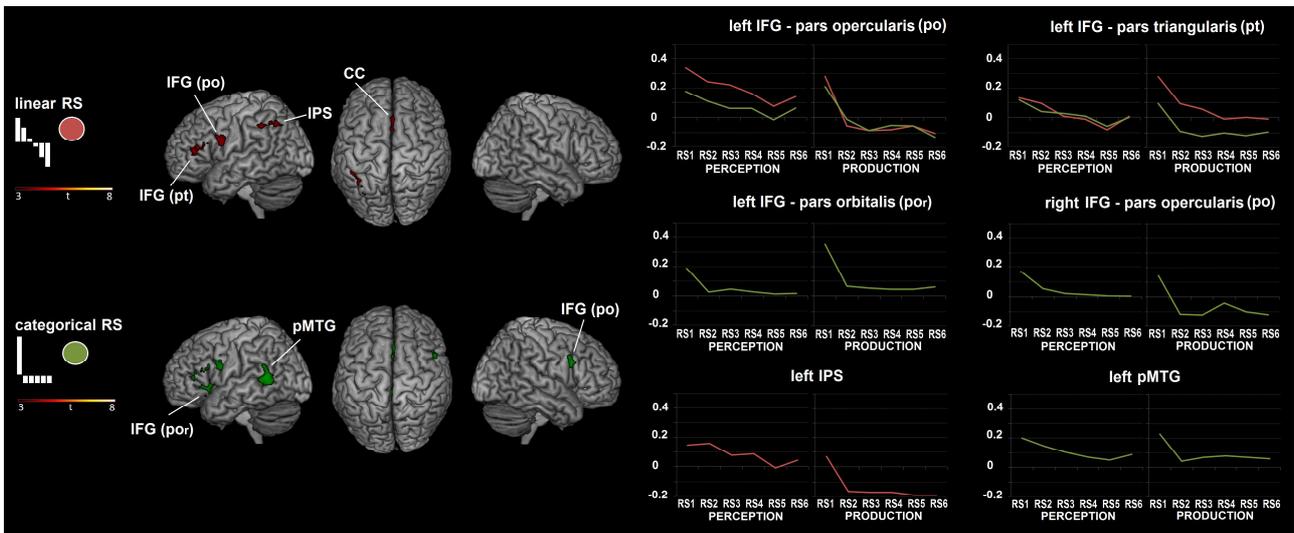


Figure 5. Brain regions showing overlapping linear and categorical RS in the perception and production tasks (i.e., linear and categorical decreases of BOLD activity across vowel repetition) and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level, see Tables S1A and S1B for details).

Repetition enhancement

No significant linear or categorical increase of BOLD activity was observed neither during vowel perception nor during vowel production.

Clusters & Regions	H	MNI coordinates			T	Perception - Contrast estimates						Production - contrast estimates					
		x	y	z		R1	R2	R3	R4	R5	R6	R1	R2	R3	R4	R5	R6
cluster 1 (256 voxels)																	
middle cingulate cortex	L/R	4	18	44	4.60	0.10	0.01	-0.02	-0.09	-0.09	-0.08	0.37	-0.03	0.00	0.03	0.00	0.04
medial frontal gyrus	L/R	-2	22	40	4.15	0.12	0.01	0.00	-0.06	-0.10	-0.05	0.40	-0.02	-0.01	0.08	0.05	0.02
cluster 2 (198 voxels)																	
inferior frontal gyrus (pars opercularis)	L	-48	6	28	4.67	0.34	0.24	0.22	0.16	0.07	0.14	0.28	-0.06	-0.09	-0.08	-0.05	-0.11
cluster 3 (117 voxels)																	
inferior parietal lobule	L	-36	-58	44	4.57	0.15	0.16	0.08	0.09	-0.01	0.04	0.07	-0.16	-0.17	-0.17	-0.19	-0.19
cluster 4 (116 voxels)																	
inferior frontal gyrus (pars triangularis)	L	-40	34	16	4.54	0.13	0.10	0.01	-0.01	-0.08	0.01	0.28	0.10	0.06	-0.01	0.00	-0.01

Table 1A. Maximum activation peak summary of brain regions showing overlapping linear RS between the perception and production tasks, and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level).

Clusters & Regions	H	MNI coordinates			T	Perception - Contrast estimates						Production- Contrast estimates					
		x	y	z		R1	R2	R3	R4	R5	R6	R1	R2	R3	R4	R5	R6
cluster 1 (417 voxels)																	
middle cingulate cortex	L/R	8	20	40	4.84	0.07	-0.03	-0.06	-0.05	-0.04	-0.07	0.20	-0.06	-0.05	0.02	0.00	0.00
supplementary motor area (pre-SMA)	L/R	0	14	48	3.66	0.13	0.00	0.03	-0.10	-0.14	-0.06	0.48	-0.04	0.00	0.05	0.03	0.02
cluster 2 (287 voxels)																	
precentral gyrus	L	-44	8	32	4.79	0.34	0.22	0.18	0.13	0.08	0.16	0.10	-0.20	-0.17	-0.17	-0.10	-0.23
inferior frontal gyrus (pars opercularis)	L	-38	8	26	4.08	0.17	0.10	0.05	0.00	0.01	0.02	0.19	-0.06	0.01	-0.03	0.02	-0.06
inferior frontal gyrus (pars triangularis)	L	-40	32	18	3.90	0.10	0.05	0.00	-0.02	-0.08	-0.01	0.23	0.07	0.02	-0.01	-0.01	-0.01
cluster 3 (280 voxels)																	
middle temporal gyrus	L	-48	-50	10	4.76	0.14	0.10	0.03	0.03	0.00	0.04	0.14	-0.02	-0.01	-0.01	0.00	-0.02
cluster 4 (236 voxels)																	
insula	L	-34	26	2	5.27	0.19	0.03	0.05	0.03	0.01	0.02	0.35	0.07	0.05	0.05	0.05	0.06
inferior frontal gyrus (pars orbitalis)	L	-32	24	-12	3.53	0.14	0.02	-0.03	0.00	-0.01	0.02	0.15	-0.09	-0.11	-0.10	-0.03	-0.04
cluster 5 (150 voxels)																	
inferior frontal gyrus (pars opercularis)	R	44	14	30	4.24	0.17	0.05	0.02	0.01	0.00	0.00	0.14	-0.12	-0.12	-0.04	-0.10	-0.12
inferior frontal gyrus (pars triangularis)	R	40	18	24	4.20	0.12	0.01	0.03	-0.03	-0.08	0.00	0.10	-0.15	-0.13	-0.08	-0.10	-0.06
cluster 6 (87 voxels)																	
middle cingulate cortex	L/R	0	-30	32	4.01	0.10	0.02	-0.04	-0.01	0.00	-0.10	-0.01	-0.22	-0.25	-0.24	-0.16	-0.18
posterior cingulate cortex	L/R	6	-40	24	3.58	0.07	-0.03	-0.05	-0.02	-0.02	-0.03	0.06	-0.11	-0.13	-0.05	-0.06	-0.07

Table 1B. Maximum activation peak summary of brain regions showing overlapping categorical RS between the perception and production tasks, and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level).

DISCUSSION

Using a well-established fMRI adaptation paradigm, we examined whether there exist common neural regions sensitive to phonemic adaptation during repeated vowel production and perception. Our results provide the first demonstration of common adaptive neural changes in the listening and speaking brain and show that they take place in premotor and associative auditory and somatosensory regions, including the inferior frontal gyri, the posterior part of the left middle temporal gyrus and superior temporal sulcus, the left intraparietal sulcus, as well as the cingulate gyrus and presupplementary motor area. By providing evidence for adaptive changes in the same premotor and associative auditory and somatosensory brain areas, these results suggest that phonemic representations are coded in shared sensorimotor regions in the listening and speaking brain.

Specificities and limitations

Before we discuss our results, it is important to highlight several specificities and limitations of the present study. Regarding the tasks, we tried to carefully minimize the use of an articulatory rehearsal strategy during speech perception in order to limit motor process as a possible confounding factor between the two tasks. To do so, a passive perception task was used that was devoid of any explicitly defined phonemic judgment. This was done to limit verbal working memory strategies and subvocal rehearsal of the auditory material that are usually associated with meta-phonological tasks (e.g., Sato et al., 2009). In addition, the perception task was performed before the production task. In our view, mixing the perception and production tasks from trial to trial would have possibly induced some covert motor production of the perceived vowel during auditory speech perception. The experimental design was therefore unbalanced between the two tasks in terms of attentional resources and fatigue. Although this may have impacted in some way BOLD activity in the production task, it is to note that a 15 minutes long anatomical scan was acquired between the two tasks, during which participants stayed at rest. As a matter of fact, the averaged BOLD activity appears largely in line with previous fMRI studies of speech perception and production (e.g., Wilson et al., 2004; Callan et al., 2010; Skipper et al., 2007; Tremblay and Small, 2011; Grabski et al., 2013a; Rampinini et al., 2017).

As for the stimuli, we used vowels, considered universally in the world's languages as the most basic phonemic units poorly contaminated by complex coarticulation effects (Schwartz et al., 1997), hence thought to reveal the core perceptuo-motor network of speech perception and production. The selected vowels include a set of distinct phonetic features including height (close, mid-close and mid-open vowels), backness (front and back vowels) and roundedness (rounded and unrounded vowels), with each repeatedly perceived or produced vowel differing from the preceding one by at least one or more of these phonetic features. In order to match the acoustic properties of the perceived and produced vowels, vowels delivered in the perception task were individually recorded by each participant prior to the experiment. This procedure contrasts with previous fMRI studies of speech perception that did not use self-related speech

stimuli. However, it appears from the literature that behavioral and neurocognitive differences between the perception of self vs. other speech stimuli are limited, with increased right inferior frontal activity when comparing self-voices to familiar-voices (Nakamura et al., 2001; Kaplan et al., 2008) but no clear behavioral differences (Rosa et al., 2008; Aruffo and Shore, 2012; Treille et al., 2017). Finally, at the beginning of the scanning session, an intensity matching procedure was done for each participant (see Christoffels, Formisano and Schiller, 2007; Grabski et al., 2013a), in order to match as closely as possible the volume of the auditory stimuli in the perception task and that of the auditory feedback in the production task (despite bone/skull conduction).

Regarding the RS paradigm, each vowel was perceived or produced in six consecutive trials in each run (with six distinct occurrences of the same vowel). This sequence allowed measuring changes in BOLD signal from the first to the sixth repetitions and to determine linear and categorical BOLD decreases during repeated perceived and produced vowels (Noppeney and Penny, 2006; Grabski et al., 2012a). Due to the sparse sampling acquisition procedure, the time-lag between each repeated vowel was of 10 seconds. Although RS has been previously observed for a wide range of time-lags (from few hundred milliseconds to hours), it has been shown that different brain regions show varying sensitivity to different repetition time-lags (Grill-Spector and Malach, 2001; Grill-Spector, Henson and Martin, 2006; Barron et al., 2017). The present results therefore appear obviously related to the selected number of repetitions and time-lag across trials. Finally, two classical limits of fMRI adaptation effects to be mentioned are related to adaptation cascades across cortical processing hierarchies (i.e., adaptation can alter the signals sent from early to subsequent areas in the hierarchical processing stages, thereby altering the downstream responses) and to attention and expectation with related top-down effects known to interact with adaptation (for a review, see Larsson et al., 2016). From these possible confounding effects, the passive listening and active speaking tasks obviously differed in terms of attentional demands, therefore limiting a major role of attention in the common adaptive changes observed in the two tasks. Despite the exogenous and endogenous nature of the two tasks, since the time-lag and number of vowel repetitions was kept constant through the experiment, expectation effects cannot however be fully excluded from our results. Finally, the use of two distinct perception and production tasks minimizes but does not exclude the possibility of adaptation cascades, given the well-acknowledged auditory-to-motor and motor-to-auditory processing hierarchies, as well as sensorimotor feedback loops, during speech perception and production (see below).

Finally, our findings have to be interpreted with caution due to the rather limited sample size. Reduced statistical power and increased probability of false-negative findings due to limited sample size is a central problem in neuroimaging research (Thirion et al., 2007; Hupé, 2015). From that point, several linear and categorical RS effects during vowel perception and/or production were observed in largely similar neural clusters with some linear and categorical profiles quite similar. In our view, this is likely due to the limited sample size and associated inter-subject BOLD variability in evaluating linear and categorical RS profiles

across participants. Despite clear BOLD decrease across the six trials in all reported clusters, a clear dissociation between linear and categorical RS thus appears difficult. From this clear limitation however, the observed results on the first vowel repetition are in line with previous fMRI studies of speech perception and production (e.g., Wilson et al., 2004; Callan et al., 2010; Skipper et al., 2007; Tremblay and Small, 2011; Grabski et al., 2013a; Rampinini et al., 2017), and the observed RS effects appear largely compatible with previous fMRI adaptation studies that examined adaptive coding during either speech perception or speech production (see below).

Common repetition suppression during vowel perception and production

Notwithstanding the above-mentioned specificities and limitations of the present study, it is important to first note that, irrespective of the observed RS effects, BOLD activity averaged across trials observed in both the perception and production tasks appears consistent with previous brain-imaging studies, with large overlapping auditory activity between the tasks as well as specific brain areas classically involved in motor preparation, execution and coordination activated in the production task (for a review on the neural correlates of vowel perception and production, see Grabski et al., 2013a).

Regarding RS effects, our results demonstrate common suppressed BOLD responses during repeated vowel perception and production in the inferior frontal gyri, the posterior part of the left middle temporal gyrus and superior temporal sulcus, the left intraparietal sulcus, as well as in the cingulate gyrus and presupplementary motor area. As previously mentioned, while no past study attempted to identify the neural substrates of adaptive phonemic coding during both listening and speaking, a few examined RS effects during speech perception (Hasson et al., 2007; Joanisse et al., 2007; Vaden et al., 2010; Myers and Swaan, 2012; Lawyer and Corina, 2015) or production (Peeva et al., 2010; Grabski et al., 2012a; Sato et al., 2015; Okada et al., 2018) using vowels, syllables or pseudowords. As discussed above, RS effects critically depend on several factors, including the task, the stimuli, the number of repetitions and the time-lag between repetitions. Despite clear differences on these experimental factors, these studies also revealed RS effects in the left and/or right inferior frontal gyrus and adjacent ventral premotor cortex (Hasson et al., 2007; Vaden et al., 2010; Myers and Swaan, 2012; Grabski et al., 2012a; Sato et al., 2015; Sato et al., 2015; Okada et al., 2018), in the posterior part of the left middle temporal gyrus and/or superior temporal sulcus (Joanisse et al., 2007; Vaden et al., 2010; Myers and Swaan, 2012; Lawyer and Corina, 2015; Peeva et al., 2010), in the left intraparietal sulcus and adjacent dorsal inferior parietal cortex (Joanisse et al., 2007; Grabski et al., 2012a; Sato et al., 2015), and in the supplementary motor area and/or cingulate cortex (Vaden et al., 2010; Peeva et al., 2010; Sato et al., 2015).

Although both a linear and a categorical BOLD decrease was observed within the pars opercularis and triangularis of the left inferior frontal gyrus and in the cingulate cortex, specific linear RS effect was observed in the left intraparietal sulcus as well as specific categorical RS effect in the pars orbitalis of the left inferior frontal gyrus, the left posterior middle temporal gyrus and superior temporal sulcus, and in the

pars opercularis and pars triangularis of the right inferior frontal gyrus. Although these results might suggest that the time-course of neural adaptation **differs** across these regions, they should be interpreted with caution due to the limited sample size (see above).

Finally, apart from common adaptive neural changes, **several** neural regions showed sensitivity to RS only during the listening or speaking task (see Table S4 and Figure S1). RS observed in visual brain areas during speaking, including the extrastriate cortex and left occipito-temporal brain areas, can obviously be linked to the repeated visual presentation of the vowel to be produced (see Grabski et al., 2012a and Sato et al., 2015 for similar results using fMRI adaptation on repeated orofacial movements, vowels and/or syllables). Adaptive neural changes in the basal ganglia, **insular cortex** and superior cerebellum were also restricted to the production task. The observed RS effects in these regions in the speaking but not the listening task are likely explained by their traditionally assigned role in the speech domain to articulatory motor initiation/planning and online regulation of motor commands via thalamo-motor projections (Jürgens, 2002; Riecker et al., 2005; for similar RS effects during articulatory speech and non-speech movements, see **Grabski et al., 2013b**). **In addition, exclusive RS effects during the speaking task were also observed in subregions of the left and right inferior and middle frontal gyri, inferior parietal lobules as well as supplementary motor area and cingulate gyrus. Conversely, subregions of the right pars opercularis of the right inferior frontal gyrus and of the left and right middle and superior temporal gyri, including the transverse temporal gyri, only showed RS in the listening task.** Note that these auditory regions were found to be activated in both the listening and speaking tasks when RS was not **taken** into account (i.e., averaging BOLD activity across trials). This contrastive result appears in line with a number of fMRI adaptation studies which reported RS effects in the superior temporal gyrus during listening (Hasson et al., 2007; Vaden et al., 2010; Lawyer and Corina, 2015) but not during speaking (Peeva et al., 2010; Okada et al., 2018; but see Sato et al., 2015 for adaptive changes in the right transverse temporal gyrus during syllable production). In a recent EEG study on N1/P2 auditory evoked potentials, we also found evidence for such exogenous but not endogenous auditory adaptive processes during listening and speaking (Sato and Shiller, 2018). Although speculative, the fact that no auditory adaptation in the superior temporal gyrus occurred during speaking might reflect a floor effect on auditory neural responses or, rather, a “resetting” of auditory processing during speech motor planning and control in this region (Sato and Shiller, 2018).

From a broader perspective, the implication of the above-mentioned neural regions showing common suppressed BOLD responses is well acknowledged in speech perception and motor control. Densely connected with hierarchically earlier auditory core areas as well as with the inferior frontal and parietal cortices (Cattani et al., 2005; Dick and Tremblay, 2012), selectivity for phoneme processing in the posterior part of the left middle temporal gyrus and superior temporal sulcus has been repeatedly observed in past multivariate fMRI and intracranial EEG studies (e.g., Bouchard et al., 2013; Du et al., 2014; Mesgarani et al., 2014; Rampinini et al., 2017). While the role of the opercular part of the inferior frontal gyrus and of the adjacent ventral premotor cortex is classically assigned to motor preparation and planning of articulatory

and speech movements (Riecker et al., 2005; Bohland and Guenther, 2006), their involvement have also been demonstrated in the course of phonemic processing (e.g., Myers et al., 2009; Lee et al., 2012; Niziolek and Guenther, 2013). In addition, the sensitivity to phoneme categories of the posterior inferior frontal gyrus (pars opercularis) may also be related to the role of more anterior inferior frontal areas (pars orbitalis and triangularis) in higher order decision-making processes (Neubert et al., 2014). Neurons in the inferior parietal cortex and intraparietal sulcus have also been shown to be involved in phonemic categorization (e.g., Celsis et al., 1999; Jacquemot et al., 2003; Raizada and Poldrack, 2007; Desai et al., 2008). Finally, the involvement of the supplementary motor area and cingulate gyrus in both speech perception and production is also widely attested (Hertrich et al., 2016; Lima et al., 2016). During speaking, these regions are thought to participate in the initiation of speech motor programs, internally specified action selection, inhibition and in higher superordinate planning functions (Alario et al., 2006; Tremblay and Gracco, 2006; Tourville and Guenther, 2011; for a review, see Hertich et al., 2016). Their recruitment during auditory speech processing also suggests their contribution to auditory imagery and higher-order sensorimotor control and predictive functions (for a review, see Lima et al., 2016).

In conclusion, adaptive BOLD responses observed in common premotor and associative auditory and somatosensory brain areas during repeated perceived and produced vowels add new empirical evidence that phonemic coding are driven in common sensorimotor neural regions, through a cortical dorsal stream, in the listening and speaking brain. These results appear in line with a number of neurobiological models (e.g., Hickok and Poeppel, 2007; Skipper et al., 2007; Rauschecker and Scott, 2009; Houde and Nagarajan, 2011; Hickok et al., 2011; Rauschecker, 2011; Guenther and Vladusich, 2012; Bornkessel-Schlesewsky et al., 2015; Skipper et al., 2017; Schomers and Pulvermüller, 2017) that support a functional coupling between speech production and perception systems, with phonetic processing partly driven by internal motor-to-sensory and sensory-to-motor simulation. From this view, our results also appear largely compliant with a number of studies showing that both sensory and motor activity correlate with auditory phonemic categorization (Alho et al., 2012; Chevillet et al., 2013; Alho et al., 2014), can predict the accuracy between the intended and actual phonemic production (Tourville et al., 2007; Golfinopoulos et al., 2011; Niziolek, Nagarajan and Houde, 2013), and can partly be decoded to reveal produced or perceived phonemic features (Lee et al., 2012; Bouchard et al., 2013; Du et al., 2014; Mesgarani et al., 2014; Rampinini et al., 2017). From a broader perspective, previous fMRI adaptation studies on non-communicative and silent orofacial and manual actions also revealed that repeated perceived and/or produced actions cause RS in the inferior frontal gyrus and intraparietal sulcus (e.g., Dinstein et al., 2007; Chong et al., 2008; Hamilton and Grafton, 2009; Kilner et al., 2009; Grabski et al., 2012; Sato et al., 2015). These results and the present ones therefore argue for a role of premotor and sensory regions in action goal (de)coding during perception and motor control (Jeannerod, 1994; Rizzolatti, Fogassi and Gallese, 2001; Friston, 2011).

REFERENCES

- Alario FX, Chinay H, Lehericy S, Cohen L (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, 1076(1): 129-143.
- Alho J, Lin FH, Sato M, Tiitinen H, Sams M, Jääskeläinen IP (2014). Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. *Frontiers in Psychology*, 5: 394.
- Alho J, Sato M, Sams M, Schwartz JL, Tiitinen H, Jaaskelainen IP (2012). Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. *NeuroImage*, 60(4): 1937-1946.
- Aruffo C, Shore DI (2012). Can you McGurk yourself? Self-face and self-voice in audiovisual speech. *Psychonomic Bulletin & Review*, 19: 66–72.
- Auksztulewicz R, Friston K (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, 80: 125-140.
- Barron HC, Garvert MM, Behrens TEJ (2017). Repetition suppression: a means to index neural representations using BOLD? *Phil. Trans. R. Soc. B* 371: 20150355.
- Birn RM, Bandettini PA, Cox RW, Shaker R (1999). Event-related fMRI of tasks involving brief motion, *Human Brain Mapping*, 7(2): 106-114.
- Bohland JW, Guenther FH (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, 32(2), 821–841.
- Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP (2015). Neurobiological roots of language in primate audition: common computational properties. *Trends in Cognitive Science*, 19(3): 142–150.
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, 495(7441): 327-332.
- Callan D, Callan A, Gamez M, Sato MA, Kawato M (2010). Premotor cortex mediates perceptual performance. *Neuroimage*, 51(2): 844–858.
- Catani M, Jones DK, Ffytche DH (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57: 8–16.
- Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, 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: 135–144.
- Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Journal of Neuroscience*, 33: 5208-5215.

- Chong TT, Cunnington R, Williams MA, Kanwisher N, Mattingley JB (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology*, 18: 1576–1580.
- Christoffels IK, Formisano E, Schiller NO (2007). Neural Correlates of Verbal Feedback Processing: an fMRI Study Employing Overt Speech. *Human Brain Mapping*, 28(9): 868-879.
- Desai R, Liebenthal E, Waldron E, Binder JR (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience*, 20: 1174–1188.
- Dick AS, Tremblay P (2012). Beyond the arcuate fasciculus: consensus and controversy in the connective anatomy of language. *Brain*, 135(12): 3529–3550.
- Dinstein I, Hasson U, Rubin N, Heeger DJ (2007). Brain areas selective for both observed and executed movements. *Journal of Neurophysiology*, 98: 1415–1427.
- Du Y, Buchsbaum BR, Grady CL, Alain C (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci USA*, 111: 7126–7131.
- Friston K (2011). What is optimal about motor control? *Neuron*, 72: 488–498.
- Friston K (2012). Predictive coding, precision and synchrony. *Cognitive Neuroscience*, 3: 238–239.
- Glascher J, Gitelman D (2008). Contrast weights in flexible factorial design with multiple groups of subjects. Unpublished tutorial. (available at http://www.sbirc.ed.ac.uk/cyril/download/Contrast_Weighting_Glascher_Gitelman_2008.pdf [assessed 18.05.2019]).
- Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Nieto-Castanon A, Guenther FH (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage*, 55: 1324–1338.
- Gotts SJ, Chow CC, Martin A (2012). Repetition priming and repetition suppression: A case for enhanced efficiency through neural synchronization. *Cognitive Neuroscience*: 3, 227–259.
- Grabski K, Lamalle L, Sato M (2012a). Somatosensory-motor adaptation of orofacial actions in posterior parietal and ventral premotor cortices. *Plos One*, 7(11): e49117.
- Grabski K, Lamalle L, Vilain C, Schwartz JL, Vallée N, Troprès I, Baciú M, Le Bas JF, Sato M (2012b). Functional MRI assessment of orofacial articulators: neural correlates of lip, jaw, larynx and tongue movements. *Human Brain Mapping*, 33(10): 2306-2321.
- Grabski K, Schwartz JL, Lamalle L, Vilain C, Vallée N, Baciú M, Le Bas JF, Sato M (2013a). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics*, 26(3): 384-408.
- Grabski K, Tremblay P, Gracco V, Girin L., Sato M (2013b). 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.
- Gracco VL, Tremblay P, Pike GB (2005). Imaging speech production using fMRI. *NeuroImage*, 26: 294-301.

- Grill-Spector K, Henson R, Martin A (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Science*, 10: 14–23.
- Grill-Spector K, Malach R (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica (Amsterdam)*, 107: 293–321.
- Grotheer M, Kovács G (2016). Can predictive coding explain repetition suppression? *Cortex*, 80: 113-124.
- Guenther FH, Vladusich T (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25: 408–422.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999). Sparse temporal sampling in auditory fMRI. *Human Brain Mapping*, 7(3): 213-23.
- Hamilton AF, Grafton ST (2009). Repetition suppression for performed hand gestures revealed by fMRI. *Human Brain Mapping*, 30: 2898–2906.
- Hasson U, Skipper JI, Nusbaum HC, Small SL (2007). Abstract coding of audiovisual speech: beyond sensory representation. *Neuron*, 56 : 1116-1126.
- Henson RN (2012). Repetition accelerates neural dynamics: in defense of facilitation models. *Cognitive Neuroscience*, 3: 240–241.
- Hertrich I, Dietrich S, Ackermann H (2016). The role of the supplementary motor area for speech and language processing. *Neuroscience & Biobehavioral Reviews*, 68: 602–610.
- Hickok G, Poeppel D (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8: 393–402.
- Houde JF, Nagarajan SS (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5: 82.
- Hupé JM (2015). Statistical inferences under the Null hypothesis: common mistakes and pitfalls in neuroimaging studies. *Frontiers in Neuroscience*, 9: 18.
- Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience*, 23: 9541–9546.
- Jeannerod M (1994). The representing brain. Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17: 187–245.
- Joanisse MF, Zevin JD, McCandliss BD (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, 17(9): 2084-2093.
- Jürgens U (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, 26(2): 235–258.
- Kaplan JT, Aziz-Zadeh L, Uddin L, Iacoboni M (2008). The self across the senses: The neural response to

one's own face and voice. *Social, Cognitive, and Affective Neuroscience*, 3: 218-223.

Kilner JM, Neal A, Weiskopf N, Friston KJ, Frith CD (2009). Evidence of mirror neurons in human inferior frontal gyrus. *The Journal of Neuroscience*, 29: 10153–10159.

Ladefoged P (2006). *A course in phonetics* (5th ed.). Boston, MA: Thomson Wadsworth.

Larsson J, Solomon SG, Kohn A (2016). fMRI adaptation revisited. *Cortex*, 80: 154-160.

Lawyer L, Corina D (2014). An investigation of place and voice features using fMRI-adaptation. *J. Neurolinguistics*, 27: 18–30.

Lee YS, Turkeltaub P, Granger R, Raizada RDS (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J. Neurosci.*, 32: 3942-3948.

Lieberman AM, Whalen DH (2000). On the relation of speech to language. *Trends in Cognitive Science*, 3(7): 254–264.

Lima CF, Krishnan S, Scott SK (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neuroscience*, 39(8): 527-542.

Mattingly IG, Liberman AM (1988). Specialized perceiving systems for speech and other biologically significant sounds. In GMG Edelman, WE Gall, WM Cowan (Eds.), *Auditory function: neurological bases of hearing* (pp. 775-793). New York: Wiley.

Mesgarani N, Cheung C, Johnson K, Chang EF (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343: 1006–1010.

Murakami T, Kell CA, Restle J, Ugawa Y, Ziemann U (2015). Left dorsal speech stream components and their contribution to phonological processing. *Journal of Neuroscience*, 35(4): 1411-1422.

Myers EB, Blumstein SE, Walsh E, Eliassen J (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, 20: 895–903.

Myers EB, Swan K (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *Journal of Cognitive Neuroscience*, 24(8): 1695-1708.

Nakamura K, Kawashima R, Sugiura M, Kato T, Nakamura A, Hatano K, Nagumo S, Kubota K, Fukuda H, Ito K, Kojima S (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia*, 39 : 1047-1054.

Neubert FX, Mars RB, Thomas AG, Sallet J, Rushworth MF (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* 81(3): 700–713

Niziolek CA, Nagarajan SS, Houde JF (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience*, 33(41): 16110–16116.

Noppeney U, Penny W (2006). Two approaches to repetition suppression. *Human Brain Mapping* 27(5):

411-416.

- Okada K, Matchin W, Hickok G (2018). Phonological feature repetition suppression in the left inferior frontal gyrus. *Journal of Cognitive Neuroscience*, 30(10): 1549-1557.
- Oldfield RC (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9: 97-114
- Peeva MG, Guenther FH, Tourville JA, Nieto-Castanon A, Anton JL, Nazarian B, Alario FX (2009). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *NeuroImage*, 50(2): 626-38.
- Perkell JS (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25: 382–407.
- Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the USA*, 103(20): 7865-70.
- Raizada RD, Poldrack RA (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, 56(4): 726-40.
- Rampinini AC, 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. *Scientific reports*, 7(1): 17029.
- Rauschecker JP (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, 271: 16–25.
- Rauschecker JP, Scott, SK (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Riecker A, Mathiak K, Wildgruber D, Erb M, Hertrich I, Grodd W, Ackermann H (2005). fMRI reveals two distinct cerebral networkssubserving speech motor control. *Neurology*, 64(4), 700–706.
- Rizzolatti G, Fogassi L, Gallese V (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Review Neuroscience*, 2: 661–670.
- Rosa C, Lassonde M, Pinard C, Keenan JP, Belin P (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain Cogn.*, 68: 204-214.
- 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, Vilain C, Lamalle L, Grabski K (2015). Adaptive coding of orofacial and speech actions in motor and somatosensory spaces with and without overt motor behavior. *Journal of Cognitive Neuroscience*, 27(2): 334-51.

- Sato M, Shiller DM (2018). Auditory prediction during speaking and listening. *Brain and Language*, 187: 92-103.
- Schomers MR, Pulvermüller F (2017). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Front Hum Neurosci.*, 10: 435.
- Schwartz JL, Boë LJ, Vallée N, Abry C (1997). Major trends in vowel system inventories. *Journal of Phonetics*, 25: 233–254.
- Schwartz JL, Ménard L, Basirat A, Sato M (2012). The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5): 336–354.
- Skipper JI, Devlin JT, Lametti DR (2017). The hearing ear is always found close to the speaking tongue: review of the role of the motor system in speech perception. *Brain and Language*, 164: 77–105.
- Skipper JI, Van Wassenhove V, Nusbaum HC, Small SL (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10): 2387–2399.
- Thirion B, Pinel P, Meriaux S, Roche A, Dehaene S, Poline JB (2007). Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. *Neuroimage*, 35: 105-120.
- Tourville JA, Reilly KJ, Guenther FH (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39: 1429–1443.
- Tourville JA, Guenther FH (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7): 952–981.
- Treille A, Vilain C, Kandel S, Sato M (2017). Electrophysiological evidence for a self-processing advantage during audiovisual speech integration. *Experimental Brain Research*, 235(9): 2867-2876.
- Tremblay P, Gracco VL (2006). Contribution of the frontal lobe to externally and internally specified verbal responses: fMRI evidence. *NeuroImage*, 33(3): 947-957.
- Tremblay P, Small SL (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage*, 57: 1561–1571.
- 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: 273-289.
- Vaden KI Jr, Muftuler LT, Hickok G (2010). Phonological repetition-suppression in bilateral superior temporal sulci. *NeuroImage*, 49(1): 1018-1023.
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7: 701–702.
- Zheng ZZ, Munhall KG, Johnsrude IS (2010). Functional overlap between regions involved in speech

perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, 22(8): 1770–1781.

LEGENDS

Figure 1. Experimental design. A) The experiment involved passive listening (L) and overt speaking (S), and consisted of six runs, each lasting approximately ten minutes, and one anatomical scan (A). B) Each run consisted of fifty-four trials in which participants were asked to listen to or to overtly produce /i/, /y/, /u/, /e/, /ø/, /o/, /ɛ/, /œ/ and /ɔ/ vowels, and six baseline trials. In order to determine RS, each vowel was perceived or produced in six consecutive trials in a pseudorandom sequence. C) For each trial, the time interval between the perceived (auditory stimulus) or produced (visual cue) vowel and the midpoint of the following functional scan acquisition was of 4s, 5s or 6s (the predicted hemodynamic response function is indicated in red).

Figure 2. Brain activity averaged across the six repetitions (left) and for each repetition (right) in the perception and the production tasks (i.e., irrespective of RS) and overlapping activity between the two tasks ($p < .05$ FWE corrected at the cluster level).

Figure 3. Brain regions showing linear and categorical RS in the perception task (i.e., linear and categorical decreases of BOLD activity across vowel repetition) and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level, see Tables S2A and S2B in Supplemental Data for details).

Figure 4. Brain regions showing linear and categorical RS in the production task (i.e., linear and categorical decreases of BOLD activity across vowel repetition) and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level, see Tables S3A and S3B in Supplemental Data for details).

Figure 5. Brain regions showing overlapping linear and categorical RS in the perception and production tasks (i.e., linear and categorical decreases of BOLD activity across vowel repetition) and contrast estimates for the six repetitions ($p < .05$ FWE corrected at the cluster level, see Tables 1A and 1B for details).