Neural correlates of phonetic convergence and speech imitation

Maëva Garnier[1]*, Laurent Lamalle[2,3,4,5] and Marc Sato[1]

[1] GIPSA-LAB, Département Parole et Cognition, CNRS et Grenoble Université, UMR5216, Grenoble, France
[2] Inserm US 017, Grenoble, France
[3] IRMaGe, Université Grenoble Alpes, Grenoble, France
[4] CHU de Grenoble, UMS IRMaGe, Grenoble, France
[5] CNRS, UMS 3552, Grenoble, France

Speakers unconsciously tend to mimic their interlocutor's speech during communicative interaction. This study aims at examining the neural correlates of phonetic convergence and deliberate imitation, in order to explore whether imitation of phonetic features, deliberate, or unconscious, might reflect a sensory-motor recalibration process. Sixteen participants listened to vowels with pitch varying around the average pitch of their own voice, and then produced the identified vowels, while their speech was recorded and their brain activity was imaged using fMRI. Three degrees and types of imitation were compared (unconscious, deliberate, and inhibited) using a go-nogo paradigm, which enabled the comparison of brain activations during the whole imitation process, its active perception step, and its production. Speakers followed the pitch of voices they were exposed to, even unconsciously, without being instructed to do so. After being informed about this phenomenon, 14 participants were able to inhibit it, at least partially. The results of whole brain and ROI analyses support the fact that both deliberate and unconscious imitations are based on similar neural mechanisms and networks, involving regions of the dorsal stream, during both perception and production steps of the imitation process. While no significant difference in brain activation was found between unconscious and deliberate imitations, the degree of imitation, however, appears to be determined by processes occurring during the perception step. Four regions of the dorsal stream: bilateral auditory cortex, bilateral supramarginal gyrus (SMG), and left Wernicke's area, indeed showed an activity that correlated significantly with the degree of imitation during the perception step.

Keywords: phonetic convergence, imitation, speech production, speech perception, sensory-motor interactions, internal models

INTRODUCTION

When they interact, speakers tend to imitate their interlocutor's posture (Shockley et al., 2003), gestures, facial expressions, and breathing (Chartrand and Bargh, 1999; Estow and Jamieson, 2007; Sato and Yoshikawa, 2007). Regarding their interlocutor's speech, such convergence effects also occur at the phonetic, lexical, and syntactic levels (Natale, 1975; Pardo, 2006; Delvaux and Soquet, 2007; Kappes et al., 2009; Aubanel and Nguyen, 2010; Bailly and Lelong, 2010; Miller et al., 2010; Babel, 2012; Babel and Bulatov, 2012). The phenomenon of "phonetic convergence," also referred to as "accommodation," "entrainment," "alignment," or "chameleon effect," not only concerns supra-segmental parameters such as vocal intensity (Natale, 1975), fundamental frequency (f_0) (Gregory et al., 1993, 2000; Bosshardt et al., 1997; Goldinger, 1997; Babel and Bulatov, 2012) and long-term average spectrum (Gregory et al., 1993, 1997, 2000; Gregory and Webster, 1996) but also temporal and spectral cues to phonemes like voice onset time (VOT) of stop consonants (Sancier and Fowler, 1997; Shockley et al., 2004; Nielsen, 2011) and the first two formants of vowels (F_1, F_2; (Babel and Bulatov, 2012; Pardo, 2010; Sato et al., 2013). This phenomenon appears to be quite subtle, f_0 and speech rate

showing the greatest sensitivity to phonetic convergence (Pardo, 2010; Sato et al., 2013).

Most of the literature considers this convergence phenomenon as primarily driven by social or communicative motivations. Convergence behaviors may aim at placing the interaction on a "common ground" of sounds and gestures, which is hypothesized to improve communication at the social level and/or at the intelligibility level.

Several theories predict that speakers converge more toward people they like, and from whom they want to be liked in return (Byrne, 1997; Chartrand and Bargh, 1999; Babel, 2009), toward people they are acquainted with (Lelong and Bailly, 2012) or toward people who exert a leadership (Pardo, 2006) or any kind of social dominance/hierarchy on them (Gregory and Hoyt, 1982; Street and Giles, 1982; Gregory, 1986; Giles et al., 1991; Gregory and Webster, 1996). More generally, the Communication Accommodation Theory (CAT; Giles et al., 1991) considers phonetic convergence and divergence as a social tool to mark the desire to belong to or to distinguish oneself from a social group (Giles et al., 1973; Giles, 1973; Bourhis and Giles, 1977; Tajfel and Turner, 1979; Giles et al., 1991). Work by Krashen (1981)

and Pardo (2006) supports the idea that phonetic convergence is driven by empathy, rather than by the desire to be liked. In any case, no evidence has been provided yet, supporting the idea that we like people more if they converge toward us [although, on the other side, previous studies showed that we like people more after imitating them (Adank et al., 2013)].

Phonetic convergence is also believed to improve communication at the intelligibility level. Producing speech sounds and lexical forms that are more similar to the own repertoire of the interlocutor may facilitate phonetic decoding and lexical access. However, no study has shown such intelligibility benefits yet [although, on the other side, previous studies showed it is easier to understand an accent after imitating it (Adank et al., 2010)]. In fact, this idea appears contradicted by the fact that our own speech—that cannot be more similar to our own sound repertoire—is not more intelligible to us than speech produced by others (Hawks, 1985).

Several additional observations lead us to partly reconsider the idea that phonetic convergence may be primarily driven by social and communicative motivations. First, phonetic convergence was also observed in non-interactive tasks of speech production (Goldinger, 1998; Namy et al., 2002; Shockley et al., 2004; Vallabha and Tuller, 2004), even at the basic level of vowel production (Sato et al., 2013). Partial imitation of lip gestures and vocal sounds was also observed in newborns and small children (Heimann et al., 1989; Meltzoff and Moore, 1997). Such imitation processes appear to be involuntary (Garrod and Clark, 1993) and are believed to play a key role in cognitive development, in particular for language acquisition (Chen et al., 2004; Serkhane et al., 2005; Nagy, 2006). Some authors support the idea that these automatic imitation processes still exist in adults, but that they may be neutralized by inhibition processes. They formulate this hypothesis from the observation of patients with fontal brain lesions and a loss of social inhibition, who systematically repeat and imitate their interlocutor (Brass et al., 2003, 2005; Spengler et al., 2010). These studies suggest that imitation would be innate and involuntary while inhibiting imitation, and controlling the degree of this inhibition, is what we may learn with age.

Rizzolatti and colleagues (e.g., Iacoboni et al., 1999; Rizzolatti et al., 2001; Gallese, 2003) have suggested the idea of a "direct matching" between perception and action, as the basis for imitation of motor tasks. Main empirical support of this theoretical proposal comes from the discovery of mirror neurons in the macaque brain (Rizzolatti et al., 1996, 2002). Mirror neurons are a small subset of neurons, found in the macaque ventral premotor cortex and the anterior inferior parietal lobule that fire both during the production of goal-directed actions and during the observation of a similar action made by another individual. In humans, homologous brain regions were also found to be involved in both action perception and production (notably, the pars opercularis of Broca's area, located in the posterior part of the inferior frontal gyrus; Rizzolatti and Arbib, 1998). Such a "motor resonance" was observed not only for finger, hand and arm movements (Tanaka and Inui, 2002; Buccino et al., 2004; Molnar-Szakacs et al., 2005), but also for mouth and lip movements (Fadiga et al., 2002; Wilson, 2004; Skipper et al., 2007; D'Ausilio et al., 2011). This overlapping network appears to be

hard wired, or at least to function from the very beginning of life (Sommerville et al., 2005; Nyström, 2008).

Regarding speech, a number of models also support the idea of a direct matching between perception and motor systems (for reviews, see Galantucci et al., 2006; Schwartz et al., 2012). Motor theories of speech perception argue that speech is primarily perceived as articulatory gestures (Liberman and Mattingly, 1985; Fowler, 1986) and sensory-motor theories postulate that phonetic coding/decoding and representations are shared by speech production and perception systems (Skipper et al., 2007; Rauschecker and Scott, 2009; Schwartz et al., 2012). Brain imaging studies provide evidence for an involvement of the motor system in speech perception (Fadiga et al., 2002; Wilson, 2004; Skipper et al., 2007). Anatomical connections between posterior superior temporal regions, the inferior parietal lobule, and the posterior ventrolateral frontal lobe in the premotor cortex were attested using diffusion tensor magnetic resonance imaging (Catani and Jones, 2005). Recent neurobiological models of speech perception and production postulate the existence of a dorsal sensory-motor stream (Hickok and Poeppel, 2000, 2004, 2007; Poeppel and Hickok, 2004) mapping acoustic representations onto articulatory representations, including the posterior inferior frontal gyrus, the premotor cortex, the posterior superior temporal gyrus/sulcus, and the inferior parietal lobule (Callan et al., 2004; Guenther, 2006; Skipper et al., 2007; Dick et al., 2010).

To sum up, all these observations and models support the idea that humans have shared representations of the motor commands of an action and of its sensory consequences. This functional coupling between perception and action systems, through these shared representations, argues for perception not only consisting in information decoding but also contributing to the automatic and involuntary "update" or "recalibration" of these shared sensory-motor representations.

This brings us to reconsider the mechanisms underlying the phenomenon of phonetic convergence and to explore the hypothesis that automatic and involuntary imitation of phonetic features might reflect a sensory-motor learning, taking place as soon as speech is perceived. In favor of this hypothesis is the fact that speakers modify their way of speaking not only during the interaction with their interlocutor, but also after the interaction. This "after-effect" concerns not only speech production, but also speech perception: vowel categorization was found to be modified after repeated exposure to someone else's speech (Sato et al., 2013). Furthermore, passive listening, without any motor involvement, appears to be sufficient to observe these after-effects (Sato et al., 2013).

The present study aimed at determining the neural substrates of phonetic convergence and more particularly at: (1) understanding whether phonetic convergence and deliberate imitation of speech are underpinned by the same neurocognitive mechanisms, (2) examining to what extent sensory-motor brain areas are involved during deliberate and unconscious imitations of speech, and (3) better understanding the degree of control and consciousness that one can have on imitation and its inhibition.

On the basis of previous studies, showing the involvement of the dorsal stream in voluntary imitation of speech (Damasio and Damasio, 1980; Caramazza et al., 1981; Bartha and Benke, 2003;

Molenberghs et al., 2009; Irwin et al., 2011; Reiterer et al., 2011) and fast overt repetition (Peschke et al., 2009), we assumed the dorsal stream to be also involved in phonetic convergence. We expected deliberate imitation and unconscious convergence to be based on the same mechanisms but to rely on the modulated activation of the dorsal stream, particularly during the perception step of the perception-action loop. Finally, we also hypothesized that phonetic convergence can be inhibited to some extent, and that this inhibition also relies on activity changes of the dorsal stream.

To explore these hypotheses, we simultaneously recorded speech signals and neural responses of 16 participants, in three tasks of speech imitation with varying degrees of will and consciousness: voluntary imitation, phonetic convergence, intended inhibition of phonetic convergence. In these tasks, we focused on one phonetic feature particularly sensitive to that phenomenon: f_0, which was varied specifically for each participant, from -20 to $+20\%$ around his/her own average pitch. We used a go-nogo paradigm in order to compare brain activations during the whole imitative process or during its perception and production steps only. In addition, two other speech control tasks (passive perception and production) were included in order to compare brain activations during perception and motor steps of the imitative process, with brain activations during usual perception and non-imitative production of vowels.

METHODS

PARTICIPANTS

Sixteen right-handed and healthy participants (11 males and 4 females of 27 ± 5 years old), French native speakers, volunteered to participate in the experiment. None of them had any speaking or hearing disorders. None of them had previously received explicit information about phonetic convergence phenomena. The study received the ethic approval from the Centre Hospitalier Universitaire de Grenoble, from the Comité de Protection des Personnes pour la Recherche Biomédicale de Grenoble and from the Agence Française de Sécurité Sanitaire des Produits.

PROCEDURE

The experiment consisted of three tasks of interest and two reference tasks.

T1. Reference task: passive auditory perception of vowels.
T2. Vowel production task. The vowels to be produced were played to the participant through headphones. Participants were expected to partly and unconsciously imitate these stimuli (convergence effect).

T3. Vowel production reference task. The vowels to be produced were displayed on a screen viewed by the participant. Participants were expected to produce vowels according to their own speech representations.
T4. Vowel imitation task. Like in T2, vowels were played to the participant through headphones. Participants were asked to produce these vowels and to ≪ imitate the voice heard ≫.
T5. Vowel production and convergence inhibition task. Participants were briefly informed about the existence of convergence phenomena. Like in T2 and T3, vowels were played to the participant through headphones. They were asked to produce these vowels as close as they could from their habitual production, trying not to follow the stimuli.

Participants were simply informed that the experiment would consist in the production and perception of vowels. The two first tasks were presented as such to the participants, in order for them not to suspect the audio stimuli to influence their own production. The voluntary imitation and inhibition tasks were thus left for the end of the experiment. These five tasks were followed by a brain anatomical scan. The whole procedure was completed in one and an half hour.

The audio stimuli used in the conditions T1, T2, T4, and T5 consisted of 27 different vowels, specifically selected for each participant. First, a vowel database with modified pitches was created from 3 French vowels ([e], [oe], [o]) produced by a reference male speaker and a reference female speaker. Pitches were artificially shifted by steps of 5 Hz from 80 to 180 Hz for the male vowels, and from 150 to 350 Hz for the female vowels. This pitch manipulation was performed using the PSOLA module integrated in Praat, which enables to modify pitch without affecting formants or speech rate. Before the experiment, each participant was also recorded while producing a series of vowels, in order to determine his/her habitual pitch (see **Table 1**). Finally, for each participant, 27 stimuli were selected from the vowel database, with the 9 quantified frequencies closest to 80, 85, 90, 95, 100, 105, 110, 115, and 120% of his/her habitual pitch. The visual stimuli used in the condition T3 consisted of the 3 symbols ≪éé≫, ≪eu≫, and ≪oo≫.

The two reference tasks (T1 and T3) consisted in 54 trials. The 27 audio or visual stimuli described above were presented in a pseudo-random order and in alternation with 27 ≪ void ≫ stimuli (i.e., no sound in T1 or no displayed vowel in T3). These ≪ void ≫ stimuli were used as a baseline for the comparison of neural activations. Each trial lasted 10 s. Stimuli were played

Table 1 | Participants' information.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
Gender	M	F	F	M	M	M	M	F	M	F	M	M	M	M	F	M
Age	22	30	27	23	25	39	24	30	24	25	38	28	25	27	27	23
Average habitual f_0 (Hz)	129 ± 3	219 ± 6	198 ± 5	127 ± 3	111 ± 1	130 ± 5	131 ± 3	298 ± 6	131 ± 5	239 ± 12	125 ± 6	135 ± 9	126 ± 4	146 ± 4	237 ± 18	121 ± 8

(T1) or displayed (T3) during the first 500 ms. One second later, a fixation cross was displayed during 500 ms which indicated when the participant had to produce the vowel for the T3 condition.

The three tasks of interest (T2, T4, and T5) consisted of 81 trials. In these tasks, the 27 audio stimuli were presented twice, in a pseudo-random order and in alternation with the 27 ≪ void ≫ stimuli (i.e., no sound). Concretely, one third of the time, audio stimuli were followed by a green cross, indicating to the participant that he/she should produce the vowel (≪ Go ≫). One other third of the time, audio stimuli were followed by a red cross, meaning that the participant should remain quiet (≪ No Go ≫). The last third of the time, no stimulus was played and a red cross was displayed (Baseline). This go/no-go paradigm enables to compare the neural activations in a double task of speech production and perception, with those in a task of ≪ active ≫ perception, i.e., when participants perceive vowels with the goal of producing them afterwards, but finally without carrying out any motor action.

MATERIAL AND DATA ACQUISITION
Visual instructions were displayed on a screen located behind the participant, using a video projector and the Presentation software (Neurobehavioral Systems, Albany, EU). Participants could read them by reflection, thanks to a mirror placed above their eyes. Audio stimuli were played though MRI-compatible headphones. The audio level was set to a sufficient intensity so that participants could hear the stimuli correctly, despite the earplugs they wore to protect them from the scanner noise. The production of vowels was recorded thanks to a microphone placed 1 m away from their mouth.

Anatomic and functional images were acquired with a whole body 3T scanner (Bruker MedSpec S300) equipped with a transmit/receive quadrature volume head coil. The fMRI experiment consisted of five functional runs and one anatomical run. Functional images were obtained using a T2*-weighted, echoplanar imaging (EPI) sequence with whole-brain coverage ($TR = 10$ s, acquisition time = 2600 ms, $TE = 30$ ms, flip angle = 90°). Each functional scan comprised forty axial slices parallel to the anteroposterior commissural plane acquired in interleaved order (72×72 matrix; field of view: 216×216 mm^2; 3×3 mm^2 in plane resolution with a slice thickness of 3 mm without gap). A ≪ sparse sampling ≫ acquisition paradigm was used in order to minimize potential artifacts articulatory movements could induce on functional images. This acquisition technique stems from the time delay existing between the neural activity linked to a motor or perceptual task and the associated hemodynamic response. Based on the estimation of this delay in tasks of speech production and perception by previous studies (Grabski et al., 2013), the functional scan was chosen to start 4.7 s after stimulus perception, thus 3.7 s after vowel production (in T2–T5). A high-resolution T1-weighted whole-brain structural image was acquired for each participant after the third functional run (MP-RAGE, sagittal volume of $256 \times 224 \times 176$ mm^3 with a 1 mm isotropic resolution, inversion time = 900 ms, two segments, segment repetition time = 2500 ms, segment duration = 1795 ms, $TR/TE = 16/5$ in ms with 35% partial echo, flip angle = 8°).

ACOUSTIC ANALYSIS
The acoustic analyses were performed using Praat software. A semi-automatic procedure was used to segment vowels on the basis of intensity and duration criteria. Hesitations and mispronunciations were removed from the analyses. f_0 values were calculated, using an autocorrelation method, from a time interval defined as ±25 ms of the maximum peak intensity of the sound file.

The stimuli were specific to each participant, with f_0 values varying between approximately −20 and +20% of the participant's average habitual f_0 (see **Table 1**). Consequently, the f_0 values of the produced vowels were also converted to the participant's range, expressed as the percentage of deviation from his/her average habitual f_0 : $\Delta f_0 = (f_{0produced} - f_{0habitual})/f_{0habitual}$.

fMRI DATA PREPROCESSING AND STATISTICAL ANALYSIS
Data were analyzed with the software SPM5 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London, UK). The fMRI data of one participant (S3) were artifacted by a metalic pin and could therefore not be included in the analysis. The results reported in the fMRI data section of this article thus concern the 15 remaining participants.

For each participant, functional images were realigned, normalized in the reference space of the Montreal Neurological Institute (MNI) and smoothed with a 6 mm width Gaussian low-pass filter.

The hemodynamic responses corresponding to the experimental conditions were then estimated with a general linear model, including the characterization of a unique impulse response for each functional scan and taking body movements into account through regressors of non-interest.

Whole brain statistical analysis
Eight T-contrasts were tested (see **Table 2**), in order to identify brain regions specifically involved in vowel perception or production (when listening passively or actively, and with different degrees of imitation), compared to a resting condition.

Table 2 | Detail of the eight individual t-contrasts tested, and of the corresponding cognitive tasks explored.

		T-Contrasts	Cognitive tasks	
PERCEPTION	1	T1: Perception – Rest		Passive perception
	2	T2: Task NoGo – Rest		-Production (unconscious imitation)
	3	T4: Task NoGo – Rest	Perception in preparation of	-Deliberate imitation
	4	T5: Task NoGo – Rest		-Inhibited imitation
PRODUCTION	5	T3: Production – Rest		Production from visual instructions
	6	T2: Task Go – NoGo		Production (unconscious imitation)
	7	T4: Task Go – NoGo		Deliberate imitation
	8	T5: Task Go – NoGo		Inhibited imitation

Using SPM, a flexible factorial group analysis was conducted from these individual contrasts, corresponding to a One-Way repeated measures ANOVA (one factor TASK with 8 levels).

Eight T-contrasts were tested in order to identify brain regions specifically involved in each task of vowel perception and/or production, compared to a resting condition. Two conjunctions were calculated from the first four contrasts examining neural correlates of vowel perception, as well as from the four following contrasts examining neural correlates of vowel production. Two F-contrasts tested the main effect between the vowel perception conditions (1,2,3,4) and between the vowel production conditions (5,6,7,8).

For these contrasts, statistical significance was considered for $p < 0.05$, corrected for multiple comparisons («false discovery rate» test for perception tasks and « family-wise error » test for production tasks), with activation clusters wider than 25 voxels.

The 3D coordinates of the center of gravity of the activated clusters, normalized in the MNI reference space were assigned to functional areas of the brain thanks to the SPM Anatomy toolbox and on the basis of cytoarchitectonic probabilities. When not assigned in the SPM Anatomy toolbox, brain regions were labeled using Talairach Daemon (Lancaster et al., 2000).

Regions of interest analysis

This study hypothesizes that the dorsal stream would be involved in speech imitation and phonetic convergence. Particular attention was therefore paid to neural activations in regions of the dorsal stream. With the SPM Anatomy toolbox, 7 ROIs were defined in both hemispheres, from the cytoarchitectonic probability of

– Region TE (including TE1.0, TE1.1, and TE1.2)
– Region TE3 (Wernicke's area, including the Spt area),
– Supramarginal Gyrus (IPC PF, PFm, PFcm)
– Region BA6 (premotor cortex and supplementary motor area)
– Regions BA44 and BA45 (Broca's area)
– The Insula

Using Marsbar, eight T-contrasts (similar to **Table 2**) were tested from individual fRMI data, in order to determine the difference of neural activations in the ROIs previously defined, between tasks of vowel perception and/or production (when listening passively or actively, and with different degrees of imitation), and a resting condition.

Using SPSS software, a One-Way repeated measures ANOVA was then conducted on these individual differences of neural activation observed in each ROI. Statistical significance was considered for $p < 0.001$, *post-hoc* analyses being corrected for multiple comparisons (Bonferroni).

Finally, we performed a Pearson correlation analysis to determine the correlation between the average activation of each ROI, for each participant, in the deliberate and unconscious imitation tasks, and their demonstrated degree of imitation (defined from the behavioral data, as the slope coefficient between their produced f_0, and that of the followed stimuli).

RESULTS

BEHAVIORAL RESULTS

Figure 1 summarizes the average behaviors observed in the deliberate imitation (T3), unconscious imitation (T2), and inhibited imitation (T5) tasks. On average, the observed tendencies confirmed our expectations:

– participants were able to imitate almost perfectly the pitch of the audio stimuli (T3; slope coefficient of 0.87, $r = 0.900$, $p < 0.001$).
– participants unconsciously followed the pitch of the audio stimuli in the production task when vowels were presented auditorily (T2; $r = 0.635$, $p < 0.001$). This unconscious imitation was, however, not as strong as voluntary imitation (Slope coefficient of 0.57). It is worthwhile noting that this order of magnitude is much higher than the convergence effects usually reported in behavioral studies (slope coefficient of 0.08 in Sato et al., 2013, for instance).
– participants were able to inhibit almost completely this convergence effect when informed about its existence (T5; Slope coefficient of 0.08, $r = 0.067$, $p = 0.17$).

At the individual level, however, varying behaviors were observed. **Figure 2** gives an overview of these different individual behaviors. **Table 3** summarizes the results of the Pearson correlation analysis.

Some participants demonstrated better imitation abilities than others but all of them were able to follow variations of pitch (slope coefficients from 0.62 to 1.00).

Five speakers (see bottom panel of **Figure 2**) did not show any significant behavioral difference in the variation of f_0 between deliberate and unconscious imitations: they completely followed the pitch of the audio stimuli, even in task T3 for which they were not told of or conscious about convergence effects (slope coefficients from 0.78 to 0.98).

Eight speakers (see top panel of **Figure 2**) showed a significantly weaker degree of imitation in the unconscious imitation task. The slope of the convergence effect showed a great inter-speaker variability, from 0.11 to 0.69.
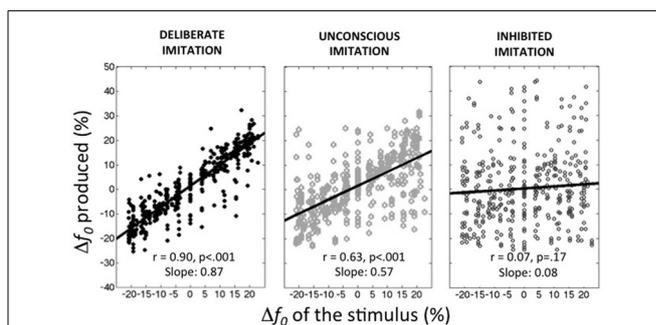

FIGURE 1 | Average behaviors observed in the three tasks of deliberate, unconscious, and inhibited imitation. Vowel stimuli were presented with 9 f_0 values, varying around the habitual average pitch of each participant. The y-axis represents how participants modified their produced f_0 from their habitual average f_0 (see **Table 1**).
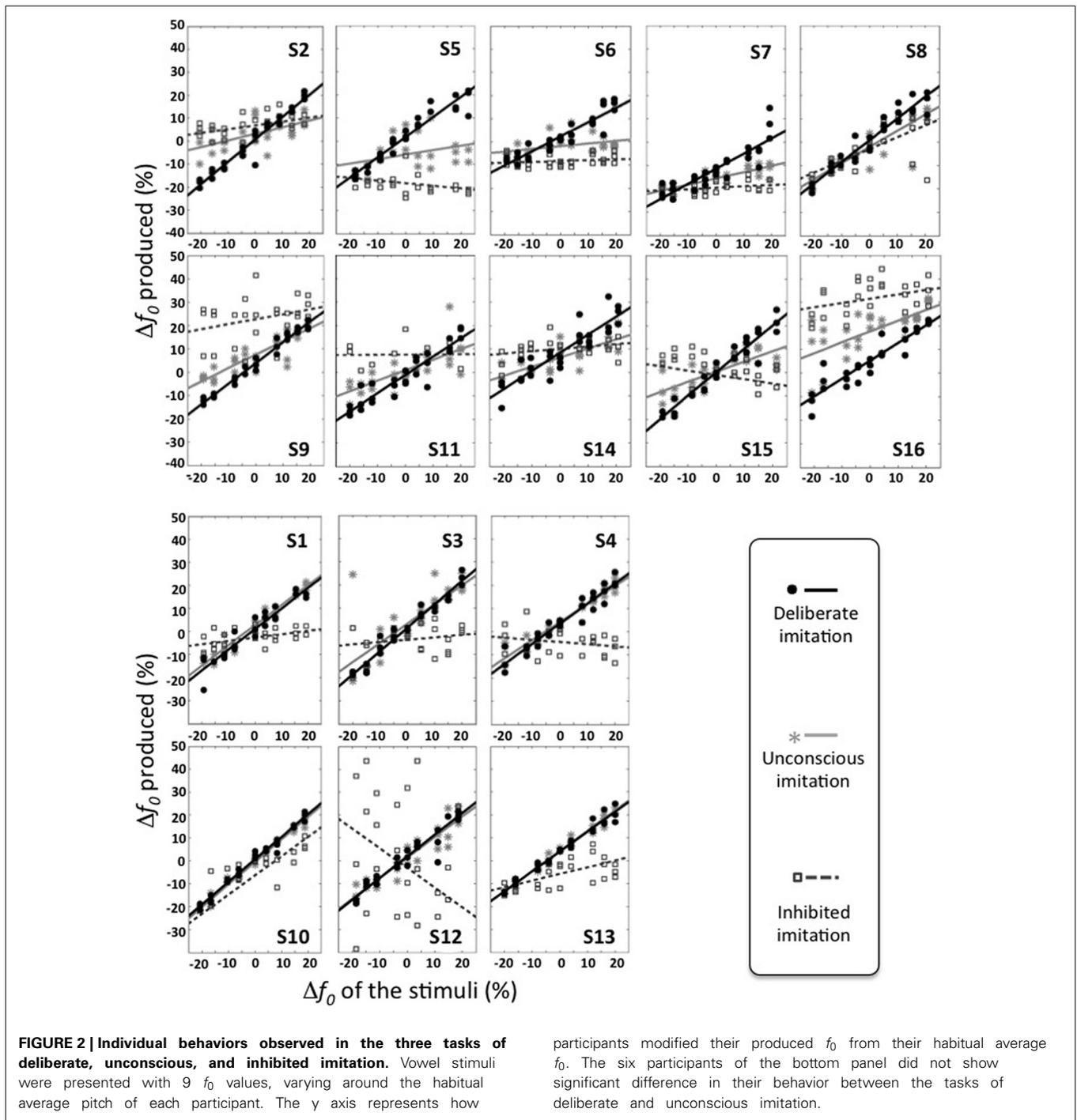
FIGURE 2 | Individual behaviors observed in the three tasks of deliberate, unconscious, and inhibited imitation. Vowel stimuli were presented with 9 f_0 values, varying around the habitual average pitch of each participant. The y axis represents how participants modified their produced f_0 from their habitual average f_0. The six participants of the bottom panel did not show significant difference in their behavior between the tasks of deliberate and unconscious imitation.

Great inter-speaker variability was observed in the inhibition task too. Ten out of 16 speakers (S3, S4, S5, S6, S7, S9, S11, S12, S14, S16) did not show a significant correlation between their produced f_0 and that of the stimuli in this task, which supports the idea that they were able to inhibit the convergence effect.

Three speakers (S1, S2, S13) showed a significant and positive correlation between their produced f_0 and that of the stimuli, with a significantly weaker linear regression slope than in the unconscious imitation task (respectively 0.15, 0.16, and 0.29).

These speakers were thus able to partially compensate for the convergence effect.

Two speakers (S8 and S10) also showed a significant correlation between their produced f_0 and that of the stimuli, but with a linear regression slope significantly almost as high (respectively 0.51 and 0.65) as in the unconscious imitation task. Inhibiting the convergence effect was therefore very hard for these participants.

Finally, one of the speakers (S15) even showed a significant but negative correlation between her produced f_0 and that of

Table 3 | Individual results of Pearson's correlation test, examining the degree of imitation in the tasks of deliberate, unconscious, and inhibited imitation.

	Deliberate imitation			Unconscious imitation			Inhibited imitation		
	Slope	r	p	Slope	r	p	Slope	r	p
S1	0.90	0.96	<0.001	0.88	0.98	<0.001	0.15	0.52	=0.005
S2	0.98	0.98	<0.001	0.29	0.59	=0.001	0.16	0.48	=0.012
S3	1.00	0.98	<0.001	0.83	0.77	<0.001	0.10	0.22	=0.26
S4	0.87	0.96	<0.001	0.78	0.97	<0.001	−0.09	−0.21	=0.285
S5	0.87	0.95	<0.001	0.19	0.36	=0.005	−0.11	−036	=0.069
S6	0.62	0.93	<0.001	0.11	0.38	=0.005	0.03	0.19	=0.336
S7	0.65	0.89	<0.001	0.27	0.75	<0.001	0.06	0.32	=0.099
S8	0.93	0.96	<0.001	0.69	0.86	<0.001	0.51	0.70	<0.001
S9	0.89	0.99	<0.001	0.57	0.88	<0.001	0.22	0.31	=0.112
S10	0.98	0.99	<0.001	0.98	0.99	<0.001	0.65	0.85	<0.001
S11	0.78	0.93	<0.001	0.45	0.74	<0.001	0.01	0.02	=0.952
S12	0.94	0.96	<0.001	0.89	0.94	<0.001	−0.86	−0.40	=0.039
S13	0.87	0.99	<0.001	0.88	0.99	<0.001	0.29	0.61	<0.001
S14	0.78	0.90	<0.001	0.39	0.80	<0.001	0.10	0.35	=0.078
S15	1.00	0.96	<0.001	0.44	0.77	<0.001	−0.20	−0.50	=0.009
S16	0.75	0.93	<0.001	0.46	0.66	<0.001	0.18	0.35	=0.076

Results that are statistically significant (p < 0.05) are indicated in black and non significant results are indicated in grey.

the stimuli (slope coefficient of -0.20, $r = -0.500$, $p = 0.009$), indicating a strategy of overcompensation of the convergence effect.

NEURAL ACTIVATIONS
Systems of speech perception and production
The classical neural networks for speech perception and production were observed in the reference tasks of passive vowel perception and vowel production from visual instructions. Surface rendering of brain activity observed in these reference tasks is displayed in the top left panels of (**Figures 3**, **4**).

Vowel perception and production reference tasks. Passive vowel perception induced large bilateral activation of the superior temporal gyrus (STG), from its anterior part to the temporo-parietal junction. Maximum activity was displayed in the primary, secondary, and associative auditory cortices, extending to the middle temporal gyrus (MTG), the Insula, and the rolandic operculum. Bilateral activations were also observed in the inferior frontal gyrus (IFG), within the pars opercularis and triangularis, extending ventrally to the pars orbitalis in the left hemisphere, and rostrodorsally to BA46 in the right hemisphere. Additional frontal activations were observed bilaterally in the dorsolateral prefrontal cortex, the premotor cortex, and the supplementary motor area (SMA). Superior and inferior parietal activations were observed bilaterally in the supramarginal gyrus (SMG), the rolandic operculum, and in the left angular gyrus. Further activity was displayed in limbic structures, in particular in the thalamus and the cingulate cortex (anterior part in the left hemisphere and middle part in the right hemisphere), and in the basal ganglia (right caudate nucleus).

Vowel production from visual instructions induced bilateral activations of the premotor, primary motor, and sensorimotor cortices, and of the SMA. Bilateral activations were also observed in the IFG (pars opercularis and triangularis) and in the STG, extending to the rolandic operculum and the SMG. Additional activations were displayed bilaterally in superior and posterior parts of the parietal cortex, including the precuneus, the associative cortex, and the angular gyrus. Further activity was found in the left inferior temporal gyrus, and bilaterally in the cerebellum, the cingulate cortex (anterior and middle part in the left hemisphere, middle part only in the right hemisphere), and the visual cortex.

Speech perception and production with various types and degrees of imitation. The typical neural network for speech perception was found again in the three other tasks of active perception, in preparation of deliberate, unconscious, or inhibited imitations (NoGo trials). Surface rendering of the conjunction of the brain activity observed in all the perception tasks is displayed in the right panel of **Figure 3**, with a summary of the maximum activation peaks.

This shared perception network involves bilateral activation of the STG, extending to the rolandic operculum and to the left Insula. Frontal regions participate in this network in the left hemisphere only, in particular Broca's area (pars opercularis and triangularis of the IFG), and the frontal region BA8. It also involves inferior parietal regions in both hemispheres: the SMG, extending to the rolandic operculum on the left side, and the angular gyrus on the right side. Further shared activations were found in the limbic system (right thalamus and left posterior cingulate cortex). A significant activation of the dorsolateral prefrontal region BA46 was also observed during the perception step of deliberate and inhibited imitations (NoGo trials, see **Figure 1**). However, the activity of this region was not found to vary significantly between the 4 different
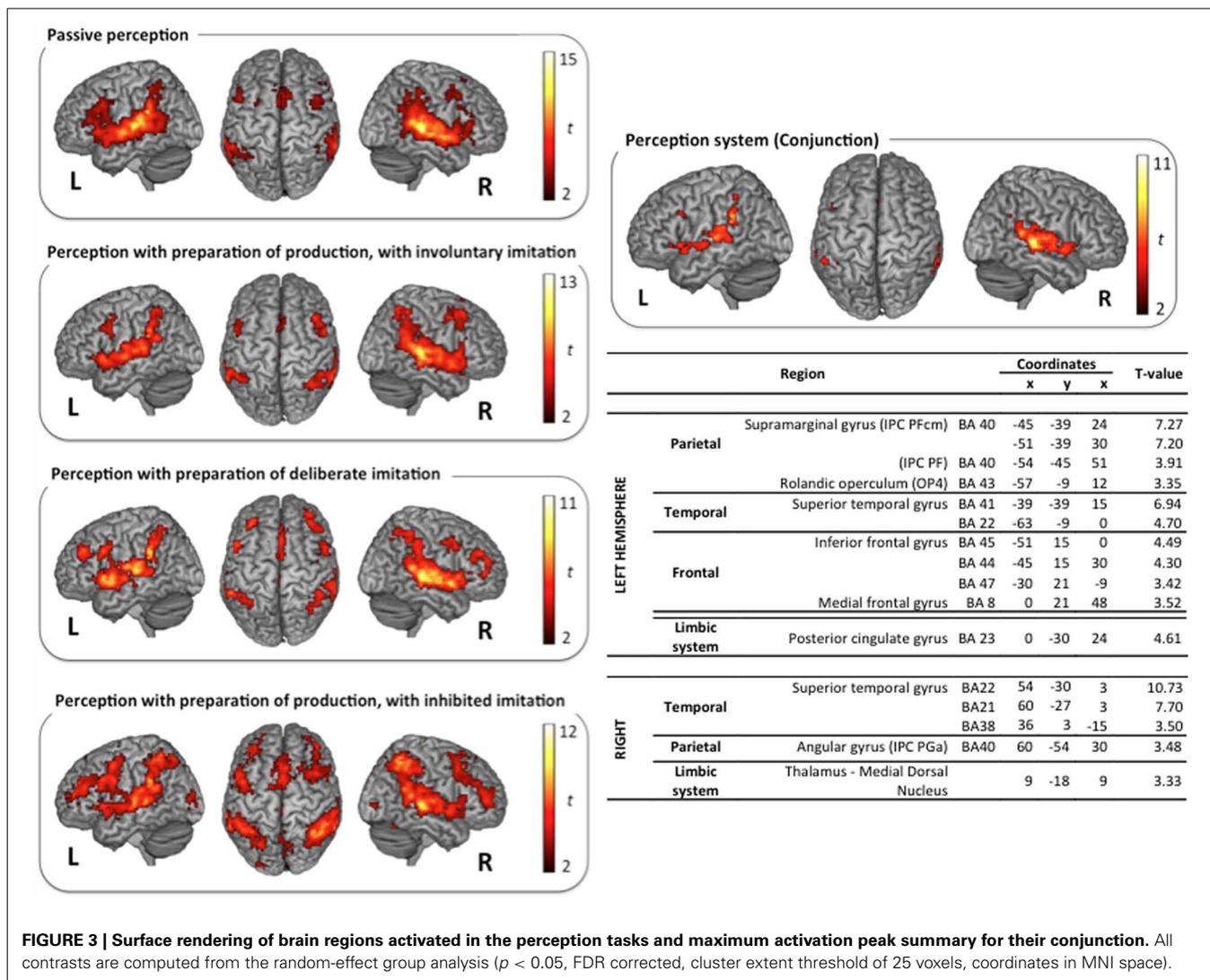
FIGURE 3 | Surface rendering of brain regions activated in the perception tasks and maximum activation peak summary for their conjunction. All contrasts are computed from the random-effect group analysis ($p < 0.05$, FDR corrected, cluster extent threshold of 25 voxels, coordinates in MNI space).

speech perception tasks (see paragraph Whole Brain Analysis below).

Similarly, the typical network for speech production was also observed in the three other speech production tasks with deliberate, unconscious, or inhibited imitations (t-contrast between the Go and the NoGo trials). Surface rendering of the conjunction of the brain activity observed in all the perception tasks is displayed in the right panel of **Figure 4**, with a summary of the maximum activation peaks.

This shared production network involves bilateral activations in the premotor, primary motor and sensorimotor cortices, extending to the IFG (pars triangularis) and to the SMA. It also involves the primary auditory cortex in the STG, extending to the rolandic operculum, and to the Insula. Further shared activations were found in posterior parietal regions, including the precuneus and the associative cortex, as well as in the limbic system (anterior cingulate gyrus, thalamus), the cerebellum, the putamen, the red nucleus, and the right basal ganglia (substantia nigra).

Comparison of deliberate, unconscious and inhibited imitations
Whole brain analysis. Using a corrected statistical analysis, no brain region was found to be significantly more or less activated between the four speech perception tasks.

No brain region was found to be significantly modulated in activation between the four speech production tasks either.

ROI analysis: differences between tasks. **Tables 4**, **5** summarize the results of further analysis and comparison of brain activity, more specifically in regions of interest of the dorsal stream.

The ROI analysis showed a significant modulation of brain activity in the auditory cortex and Wernicke's area, bilaterally, between the four vowel perception tasks with varying degrees and types of imitation. No tendency was observed toward increasing or decreasing activation with the degree of imitation.

For the production tasks, the ROI analysis again highlighted the right auditory cortex as a brain region of the dorsal stream whose activity is significantly modulated between the four vowel production tasks with varying degrees and types of imitation.
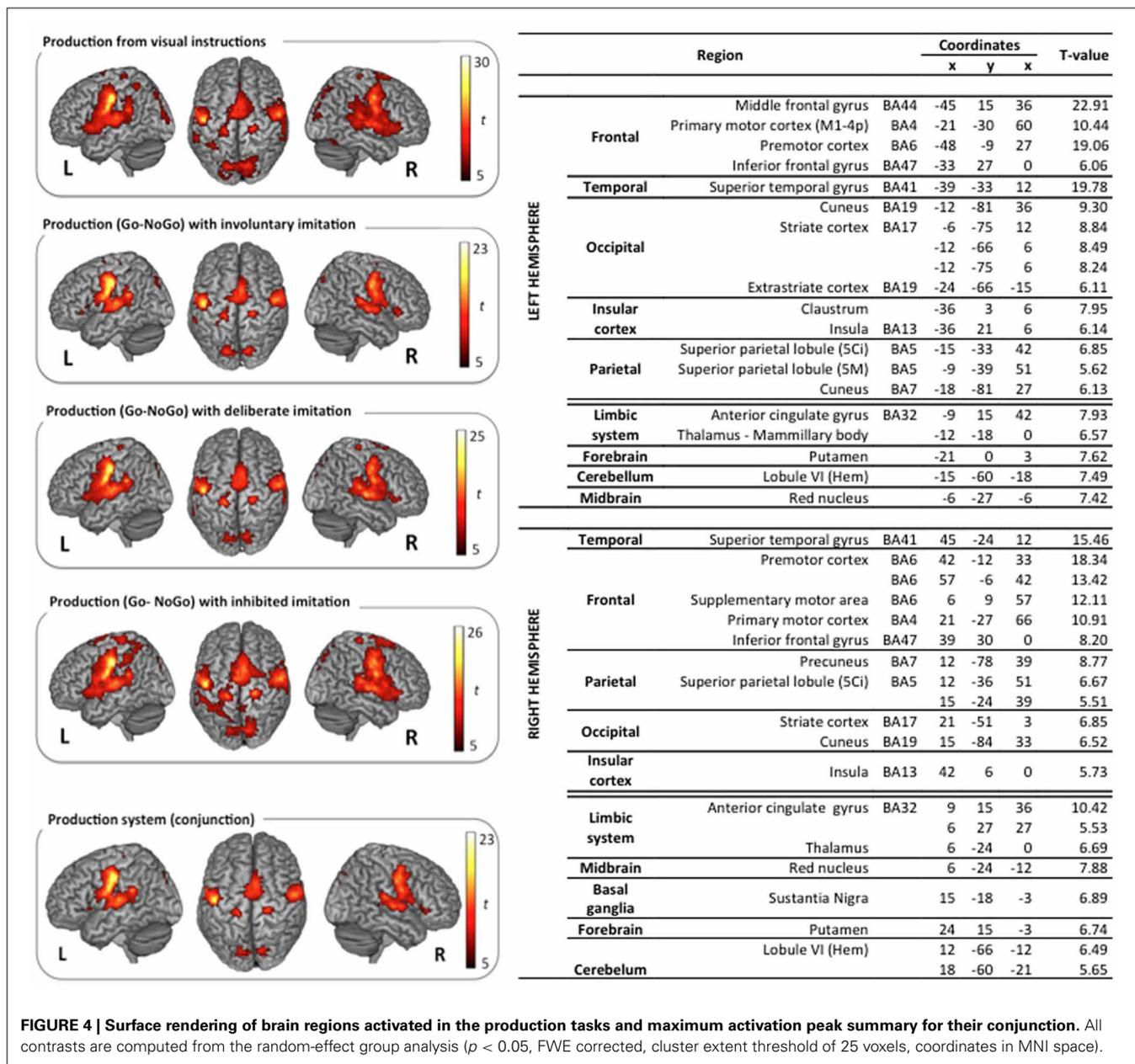
Production from visual instructions

Production (Go-NoGo) with involuntary imitation

Production (Go-NoGo) with deliberate imitation

Production (Go- NoGo) with inhibited imitation

Production system (conjunction)

	Region		Coordinates			T-value
			x	y	x	
LEFT HEMISPHERE						
Frontal	Middle frontal gyrus	BA44	-45	15	36	22.91
	Primary motor cortex (M1-4p)	BA4	-21	-30	60	10.44
	Premotor cortex	BA6	-48	-9	27	19.06
	Inferior frontal gyrus	BA47	-33	27	0	6.06
Temporal	Superior temporal gyrus	BA41	-39	-33	12	19.78
Occipital	Cuneus	BA19	-12	-81	36	9.30
	Striate cortex	BA17	-6	-75	12	8.84
			-12	-66	6	8.49
			-12	-75	6	8.24
	Extrastriate cortex	BA19	-24	-66	-15	6.11
Insular cortex	Claustrum		-36	3	6	7.95
	Insula	BA13	-36	21	6	6.14
Parietal	Superior parietal lobule (5Ci)	BA5	-15	-33	42	6.85
	Superior parietal lobule (5M)	BA5	-9	-39	51	5.62
	Cuneus	BA7	-18	-81	27	6.13
Limbic system	Anterior cingulate gyrus	BA32	-9	15	42	7.93
	Thalamus - Mammillary body		-12	-18	0	6.57
Forebrain	Putamen		-21	0	3	7.62
Cerebellum	Lobule VI (Hem)		-15	-60	-18	7.49
Midbrain	Red nucleus		-6	-27	-6	7.42
RIGHT HEMISPHERE						
Temporal	Superior temporal gyrus	BA41	45	-24	12	15.46
Frontal	Premotor cortex	BA6	42	-12	33	18.34
		BA6	57	-6	42	13.42
	Supplementary motor area	BA6	6	9	57	12.11
	Primary motor cortex	BA4	21	-27	66	10.91
	Inferior frontal gyrus	BA47	39	30	0	8.20
Parietal	Precuneus	BA7	12	-78	39	8.77
	Superior parietal lobule (5Ci)	BA5	12	-36	51	6.67
			15	-24	39	5.51
Occipital	Striate cortex	BA17	21	-51	3	6.85
	Cuneus	BA19	15	-84	33	6.52
Insular cortex	Insula	BA13	42	6	0	5.73
Limbic system	Anterior cingulate gyrus	BA32	9	15	36	10.42
			6	27	27	5.53
	Thalamus		6	-24	0	6.69
Midbrain	Red nucleus		6	-24	-12	7.88
Basal ganglia	Sustantia Nigra		15	-18	-3	6.89
Forebrain	Putamen		24	15	-3	6.74
Cerebelum	Lobule VI (Hem)		12	-66	-12	6.49
			18	-60	-21	5.65

FIGURE 4 | Surface rendering of brain regions activated in the production tasks and maximum activation peak summary for their conjunction. All contrasts are computed from the random-effect group analysis ($p < 0.05$, FWE corrected, cluster extent threshold of 25 voxels, coordinates in MNI space).

The left Insula and the right SMG were two additional regions of the dorsal stream that demonstrated a significant modulation of their neural activation. No tendency was found toward increasing or decreasing activation of these regions with the degree of imitation.

ROI analysis: correlations with behavioral data. **Table 6** summarizes the results of Pearson's correlation analysis that examined the correlation between brain activity in regions of interest of the dorsal stream and the degree of imitation in the deliberate and unconscious imitation tasks. The degree of imitation was evaluated from the slope of the variation in the produced f_0, as a function of the f_0 of the stimuli.

Again, in the two active perception tasks, preparing a deliberate or unconscious imitation, brain activity in both left and right auditory cortices was found to correlate significantly with the degree of following imitation. So did left Wernicke's area and the SMG, bilaterally.

On the other hand, no brain region was found to vary in activation with a significant correlation with the degree of imitation, for the production step of the deliberate and unconscious imitation only tasks (Go-NoGo) or for the entire process of these tasks (Go).

DISCUSSION

In line with previous studies on phonetic convergence, our results show that speakers follow and unconsciously imitate the phonetic features (here f_0) of voices they are exposed to, even when unaware of this imitation phenomenon and without being instructed to do so. Our results, however, differ from previous

Table 4 | Results of the ROI analysis, comparing brain activity in several regions of the dorsal stream, between the four speech perception tasks [passive perception, perception in preparation of vowel production (phonetic convergence), of deliberate imitation, or inhibited imitation].

		Region	Main effect		Contrast estimates			
			F	p	Passive	Conv.	Imit.	Inhib.
LEFT	Temporal	Auditory cortex (TE 1.0, TE 1.1, TE 1.2)	4.672	0.007	0.102	0.052	0.045	0.038
		Wernicke's area (TE3)	6.312	0.001	0.110	0.072	0.050	0.063
	Parietal	Supramarginal gyrus (IPC PF, PFm, Pfcm)	1.302	0.286	0.041	0.026	0.018	0.032
	Frontal	BA6 (Premotor cortex and SMA)	0.121	0.947	0.009	0.001	0.002	0.004
		Broca's area (BA44)	0.453	0.717	0.028	0.018	0.026	0.027
		Broca's area (BA45)	1.696	0.205	0.023	0.002	0.005	0.010
	Insular cortex	Insula	0.217	0.884	0.026	0.019	0.021	0.017
RIGHT	Temporal	Auditory cortex (TE 1.0, TE 1.1, TE 1.2)	4.643	0.007	0.095	0.063	0.051	0.038
		Wernicke's area (TE3)	3.164	0.034	0.103	0.075	0.081	0.059
	Parietal	Supramarginal gyrus (IPC PF, PFm, Pfcm)	0.874	0.462	0.039	0.029	0.022	0.031
	Frontal	BA6 (Premotor cortex and SMA)	0.204	0.893	0.009	−0.001	0.001	0
		Broca's area (BA44)	0.174	0.914	0.029	0.023	0.023	0.021
		Broca's area (BA45)	1.498	0.229	0.027	0.010	0.003	0.014
	Insular cortex	Insula	1.998	0.179	0.028	0.013	0.012	0.011

Results that are statistically significant (p < 0.05) are indicated in black and non significant results are indicated in grey.

Table 5 | Results of the ROI analysis, comparing brain activity in several regions of the dorsal stream, between the four speech production tasks (production from visual instructions, production with phonetic convergence, deliberate imitation, inhibited imitation).

		Region	Main effect		Contrast estimates			
			F	p	Prod.	Conv.	Imit.	Inhib.
LEFT	Temporal	Auditory cortex (TE 1.0, TE 1.1, TE 1.2)	0.971	0.415	0.197	0.175	0.205	0.203
		Wernicke's area (TE3)	0.846	0.477	0.102	0.082	0.105	0.097
	Parietal	Supramarginal gyrus (IPC PF, PFm, Pfcm)	0.925	0.437	0.076	0.051	0.066	0.068
	Frontal	BA6 (Premotor cortex and SMA)	1.745	0.173	0.084	0.064	0.072	0.096
		Broca's area (BA44)	1.792	0.163	0.086	0.057	0.077	0.095
		Broca's area (BA45)	1.113	0.355	0.018	0.009	0.028	0.033
	Insular cortex	Insula	3.678	0.019	0.100	0.046	0.080	0.091
RIGHT	Temporal	Auditory cortex (TE 1.0, TE 1.1, TE 1.2)	3.294	0.030	0.186	0.134	0.182	0.178
		Wernicke's area (TE3)	1.439	0.245	0.109	0.087	0.096	0.115
	Parietal	**Supramarginal gyrus (IPC PF, PFm, Pfcm)**	3.559	0.022	0.056	0.014	0.038	0.037
	Frontal	BA6 (Premotor cortex and SMA)	2.694	0.058	0.086	0.063	0.078	0.103
		Broca's area (BA44)	2.780	0.053	0.113	0.072	0.089	0.120
		Broca's area (BA45)	1.295	0.289	0.037	0.021	0.040	0.058
	Insular cortex	Insula	2.614	0.640	0.093	0.052	0.086	0.081

Results that are statistically significant (p < 0.05) are indicated in black and non significant results are indicated in grey. Corresponding ROIs are indicated in bold.

studies by a much greater order of magnitude in the degree of unconscious imitation. Indeed, slope coefficients in unconscious imitation of f_0 ranged from 0.11 to 0.99 in our study (0.57 on average), whereas previous studies rather reported slope coefficients of 0.08 (Sato et al., 2013). Participants were debriefed after the experiment and they all said that they had not inferred that the study dealt with imitation, before accomplishing the deliberate imitation task (T4), and that they had neither thought nor tried to imitate the stimuli during the second vowel production task (T2). This discrepancy between our results and those of previous studies cannot be interpreted in terms of interactive vs. non-interactive protocols, as Sato et al. (2013) also used non-interactive tasks of vowel perception and production, similar to this study. One possible explanation is that such a high degree

Table 6 | Results of Pearson's correlation analysis, examining the correlation between the degree of imitation during the tasks of deliberate and unconscious imitation, and the brain activity in several regions of the dorsal stream, during the whole imitative process (Go trials), during the perception step only (NoGo trials) and during the production step only (Go-Nogo contrast).

		Region	Perception step (NoGo)		Production step (Go-NoGo)		Whole imitation process (Go)	
			r	*p*	*r*	*p*	*r*	*p*
LEFT	Temporal	Auditory cortex (TE 1.0, TE 1.1, TE 1.2)	−0.45	0.013	−0.03	0.86	−0.29	0.13
		Wernicke's area (TE3)	−0.38	0.040	0.04	0.82	−0.20	0.28
	Parietal	Supramarginal gyrus (IPC PF, PFm, Pfcm)	−0.38	0.038	0.10	0.62	−0.22	0.24
	Frontal	BA6 (Premotor cortex and SMA)	−0.28	0.14	−0.12	0.52	−0.31	0.10
		Broca's area (BA44)	−0.01	0.98	−0.14	0.47	−0.13	0.49
		Broca's area (BA45)	−0.25	0.18	0.04	0.82	−0.12	0.53
	Insular cortex	Insula	−0.27	0.14	0.05	0.78	−0.12	0.52
RIGHT	Temporal	Auditory cortex (TE 1.0, TE 1.1, TE 1.2)	−0.37	0.045	0.10	0.59	−0.13	0.49
		Wernicke's area (TE3)	−0.20	0.30	−0.16	0.40	−0.25	0.19
	Parietal	Supramarginal gyrus (IPC PF, PFm, Pfcm)	−0.46	0.010	0.09	0.62	−0.21	0.26
	Frontal	BA6 (Premotor cortex and SMA)	−0.26	0.18	−0.09	0.64	−0.25	0.19
		Broca's area (BA44)	−0.32	0.08	−0.12	0.54	−0.27	0.16
		Broca's area (BA45)	−0.34	0.06	0.01	0.94	−0.21	0.27
	Insular cortex	Insula	−0.02	0.91	0.03	0.88	−0.15	0.45

Results that are statistically significant (p < 0.05) are indicated in black and non significant results are indicated in grey.

of unconscious imitation may come from the vowel production task, which may be closer to singing than to speech. Another explanation is that we used participant-specific stimuli in this experiment, varying in f_0 from −20% to +20% of each participant's average habitual pitch. The pitch of these stimuli was consequently closer to the own pitch of each participant than in other studies on phonetic convergence, where the same stimuli are used for all the participants. One may suggest that phonetic convergence is not a linear phenomenon and that speakers are more disposed to completely mimic a voice already similar to their own voice, because the target is reachable and does not require much more effort to produce than their usual intra-individual variations. On the contrary, speakers may demonstrate a lesser degree of unconscious imitation toward voices that are very different from their own, because the target is out of reach or would induce vocal discomfort. Supporting this idea, one participant of this experiment (S5) was found to follow completely the pitch of the stimuli for values below 5% of his average habitual pitch (slope of 1.0). Above that pitch height, he did not follow the stimuli further and "saturated" to a constant f_0 value (see **Figure 2**). Another argument comes from studies of intra-individual variations in habitual pitch, which is reported to vary as much as plus or minus three semitones, i.e., ~18% (Coleman and Markham, 1991). We can thus infer that the participants in our study have not made any particular effort to imitate the stimuli they were exposed to, which may explain why the degree of imitation is much greater than in the literature.

As in previous studies, we also observed a great inter-individual variability in the degree of deliberate imitation, with slope coefficients observed ranging from 0.62 to 1.00. Such a result is consistent with Pfordresher and Brown (2007), showing that about 15% of the population fails to imitate the pitch of a song by more than a semitone. Great inter-individual variability was also observed in the degree of unconscious imitation. In particular, a group of six participants demonstrated a similar degree of imitation for deliberate and unconscious imitation (slope coefficients from 0.78 to 1.00) whereas the other participants showed slope coefficients ranging from 0.11 to 0.69. Babel and Bulatov (2012) also reported a substantial inter-speaker variability in f_0 accommodation, with some participants even diverging from their interlocutor. Although it has been suggested that female talkers may be better imitators (Pardo, 2006), we could not relate the different imitation abilities to the participant's gender or to their level of musical training. It is more likely, as suggested by Postma and Nilsenova (2012) or Lewandowski (2009) that inter-individual differences in the ability to imitate f_0 or a foreign accent ("phonetic talent") may be related to the neurocognitive capacity to extract acoustic parameters (pitch, in particular) from the speech signal. Supporting this idea, in this study we found a significant correlation between the degree of imitation and brain activity in the auditory cortex, while the lateral Heschl gyrus is consensually considered as the "pitch processing center" (Bendor and Wang, 2006).

SENSORY-MOTOR INTERACTIONS IN SPEECH PERCEPTION, PRODUCTION, AND IMITATION

At the neural level, the typical networks for speech production and perception were observed, in agreement with previous studies on vowel production and perception (Özdemir et al., 2006; Sörös et al., 2006; Terumitsu et al., 2006; Brown et al., 2008; Ghosh

et al., 2008; Grabski et al., 2013). Our results also concord with previous studies on voluntary imitation of speech (Damasio and Damasio, 1980; Caramazza et al., 1981; Bartha and Benke, 2003; Molenberghs et al., 2009; Irwin et al., 2011; Reiterer et al., 2011) and fast overt repetition (Peschke et al., 2009), about the involvement of brain regions of the dorsal stream in deliberate imitation processes. Contrary to the "direct matching hypothesis," however, we did not observe the whole dorsal stream, and the IFG in particular, to be overall significantly more activated during deliberate imitation of speech (Irwin et al., 2011). In our results, only four ROIs of the dorsal stream: the auditory cortex and Wernicke's area, bilaterally, were found to vary significantly in activation from the passive perception task to the perception step of deliberate imitation. Unexpectedly, greater activation was observed for the passive perception task. Nevertheless, several ROIs of the dorsal stream, including the right SMG, showed an activity that correlated negatively with the degree of imitation during the perception step of the imitative process. Peschke and colleagues (2009) also identified such a region in the right inferior parietal area, though with a positive correlation.

The first questions addressed in this study were to determine whether phonetic convergence relied on the same mechanism and neural network as deliberate imitation, and to what extent brain regions related to sensori-motor integration were involved in that potentially shared network. Neither the whole brain analysis nor the ROI analysis showed any significant modulation in brain activation between the two tasks of deliberate and unconscious imitations (see **Tables 4**, **5**). Furthermore, the ROI analysis revealed that the activation of several regions in the dorsal stream—the auditory cortex and the SMG, bilaterally, as well as the left Wernicke area—negatively correlated with the degree of imitation, during the perception step of the imitation processes. All these observations support the idea that both deliberate and unconscious imitations are based on the same mechanism and neural network, involving regions of the dorsal stream. Unlike Leslie et al. (2004) who compared deliberate and unconscious face imitation, we did not observe a right lateralization in unconscious speech imitation, and more bilateral activations for deliberate speech imitation, which would support the idea of a "voice mirroring system" in the right hemisphere, as they suggested for face imitation.

Another question concerned the steps of the perception-production process at which the imitative process occurs: is imitation included in the perception process, in the production one, or in both? The ROI analysis revealed that the activation in several regions of the dorsal stream was significantly modulated between vowel production and perception reference tasks, and both the perception and production steps of unconscious imitation—in the auditory cortex and Wernicke's area, bilaterally, for the perception step; in the right auditory cortex, the supramarginal gyrus, and the left insula for the production step. In the case of deliberate imitation, however, significant changes in activation were also found in these ROIs, but only for the perception step, as compared to the vowel perception reference task. Finally, ROIs in the dorsal stream whose activation correlated with the degree of imitation were found for the perception step of imitative processes only. No such region was found for the production step,

or for the whole imitative processes. These observations support the idea that (1) the imitation process requires both perception and production steps of the sensori-motor loop, and that (2) the degree of imitation is determined by processes occurring during the perception step. The fact that the degree of imitation is determined by processes occurring during the perception step supports the hypothesis that perception intrinsically includes an automatic update of sensori-motor representations from the speech inputs.

A last question dealt with the degree of control and consciousness that one can have on phonetic convergence and its inhibition. The behavioral results of this study showed that phonetic convergence can be inhibited to some extent. A great inter-speaker variability was observed: some speakers were able to inhibit this unconscious imitation completely (or even with an overcompensation), others only partially, while some speakers could not inhibit it at all. At the neural level, no additional region or network, out of the typical networks of speech production and perception, appeared to be specifically involved in imitation inhibition. It is worthwhile noting that a significant activation was observed during the perception step of deliberate and inhibited imitation in the dorsolateral prefrontal region BA46, an area commonly associated with attention, resource allocation, and verbal self-monitoring (Indefrey and Levelt, 2004). That region is also known to have connections with temporal areas and to play a role in auditory processes (Romanski et al., 1999). However, the activity of that region was not found to be significantly greater in deliberate and inhibited imitation, as compared to passive perception or unconscious imitation, which does not enable us to speculate further on its role in imitation processes. On the contrary, modulated activation was observed in the left insular cortex, a brain region involved, amongst other functions, in self-awareness and inter-personal experience. This is consistent with previous studies showing that resisting motor mimicry involves cortical areas that are required to distinguish between self-generated and externally triggered motor representations (Brass et al., 2003, 2005; Spengler et al., 2010).

CONCLUSIONS AND PERSPECTIVES

The different behavioral and neural observations of this study support the hypothesis that phonetic convergence may not only be driven by social or communicative motivations, but that it may primarily be the consequence of an automatic process of sensorimotor recalibration. This has some important implications on speech production and perception, for the comprehension of how internal models and phonetic representations are learnt and updated. Indeed, many previous studies had shown how speakers modify their speech production to compensate for perturbations of their auditory or proprioceptive feedback (Abbs and Gracco, 1984; Houde and Jordan, 1998; Jones and Munhall, 2000; Villacorta et al., 2007; Shiller et al., 2009; Cai et al., 2010). After-effects of these compensations were observed on both speech production and perception (Nasir and Ostry, 2009; Shiller et al., 2009), reflecting an update of sensori-motor representations, in response to modifications of the sensory feedbacks. Complementory to these studies, our experiment brings new arguments supporting the idea that sensorimotor representations and internal models that map speech motor commands

onto their sensory consequences are continuously updated, not only from the comparison between sensory feedbacks and the predicted consequences of actions, but also from the comparison between our own production and external speech inputs provided by others. This idea of "comparison" was explored by several neurofunctional studies that suggested the existence of an "auditory-error module," supposed to be located in the posterior STG, more specifically in the planum temporale (Tourville et al., 2008). The same authors also proposed the existence of a "somatosensory-error module," assumed to be located in the SMG and the left anterior Insula, modulated when the somatosensory feedback is perturbed (Golfinopoulos et al., 2011). Interestingly, these regions are exactly those whose activation was modulated between the different imitative tasks of our study, and whose activation correlated significantly with the degree of imitation.

This possible involvement of sensorimotor recalibration processes also has implications on the communicative and social aspects of phonetic convergence. Imitation may facilitate communication not only by improving our likeability or our intelligibility for the interlocutor, but also by helping *us* to better understand our interlocutor (his/her feelings, attitudes and speech). Thus, it was shown that imitating someone else's accent improves after our appreciation of this accent (Adank et al., 2013), or that covert imitation facilitates the prediction of upcoming words in sentences in adverse listening conditions (Adank et al., 2010) and to some more limited extent, the recognition of single words (Nguyen et al., 2012).

From these findings, the involvement of sensori-motor recalibration processes in phonetic convergence, and its potential explanation of higher-level communicative and social effects (inter-individual differences and phonetic talent, i.e., the ability to learn a second language, empathy and likability, intelligibility enhancement, …) remain to be investigated in future studies.

ACKNOWLEDGMENTS

REFERENCES

Abbs, J. H., and Gracco, V. L. (1984). Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech. *J. Neurophysiol.* 51, 705–723.

Adank, P., Hagoort, P., and Bekkering, H. (2010). Imitation improves language comprehension. *Psychol. Sci.* 21, 1903–1909. doi: 10.1177/0956797610389192

Adank, P., Stewart, A. J., Connell, L., and Wood, J. (2013). Accent imitation positively affects language attitudes. *Front. Psychol.* 4:280. doi: 10.3389/fpsyg.2013.00280

Aubanel, V., and Nguyen, N. (2010). Automatic recognition of regional phonological variation in conversational interaction. *Speech Commun.* 52, 577–586. doi: 10.1016/j.specom.2010.02.008

Babel, M. (2009). *Phonetic and Social Selectivity in Speech Accommodation*. PhD dissertation, University of California.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phon.* 40, 177–189. doi: 10.1016/j.wocn.2011.09.001

Babel, M., and Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Lang. Speech* 55, 231–248. doi: 10.1177/0023830911417695

Bailly, G., and Lelong, A. (2010). "Speech dominoes and phonetic convergence," in *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association*, eds K. Takao, H. Keikichi, and N. Satoshi (Tokyo: INTERSPEECH, ISCA).

Bartha, L., and Benke, T. (2003). Acute conduction aphasia: an analysis of 20 cases. *Brain Lang.* 85, 93–108. doi: 10.1016/S0093-934X(02)00502-3

Bendor, D., and Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Curr. Opin. Neurobiol.* 16, 391–399. doi: 10.1016/j.conb.2006.07.001

Bosshardt, H.-G., Sappok, C., Knipschild, M., and Hölscher, C. (1997). Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. *J. Psycholinguist. Res.* 26, 425–448. doi: 10.1023/A:1025030120016

Bourhis, R. Y., and Giles, H. (1977). "The language of intergroup distinctiveness," in *Language, Ethnicity and Intergroup Relations*, ed H. Giles (London: Academic Press), 119–135.

Brass, M., Derrfuss, J., Cramon, G.-v., and von Cramon, D. Y. (2003). Imitative response tendencies in patients with frontal brain lesions. *Neuropsychology* 17, 265–271. doi: 10.1037/0894-4105.17.2.265

Brass, M., Derrfuss, J., and von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: a functional double dissociation. *Neuropsychologia* 43, 89–98. doi: 10.1016/j.neuropsychologia.2004.06.018

Brown, S., Ngan, E., and Liotti, M. (2008). A larynx area in the human motor cortex. *Cereb. Cortex* 18, 837–845. doi: 10.1093/cercor/bhm131

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., et al. (2004). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron* 42, 323–334. doi: 10.1016/S0896-6273(04)00181-3

Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *J. Soc. Pers. Relat.* 14, 417–431. doi: 10.1177/0265407597143008

Cai, S., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *J. Acoust. Soc. Am.* 128, 2033–2048. doi: 10.1121/1.3479539

Callan, D. E., Jones, J. A., Callan, A. M., and 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, 1182–1194. doi: 10.1016/j.neuroimage.2004.03.006

Caramazza, A., Basili, A. G., Koller, J. J., and Berndt, R. S. (1981). An investigation of repetition and language processing in a case of conduction aphasia. *Brain Lang.* 14, 235–271. doi: 10.1016/0093-934X(81)90078-X

Catani, M., and Jones, D. K. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16. doi: 10.1002/ana.20319

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910. doi: 10.1037/0022-3514.76.6.893

Chen, X., Striano, T., and Rakoczy, H. (2004). Auditory–oral matching behavior in newborns. *Dev. Sci.* 7, 42–47. doi: 10.1111/j.1467-7687.2004.00321.x

Coleman, R. F., and Markham, I. W. (1991). Normal variations in habitual pitch. *J. Voice* 5, 173–177. doi: 10.1016/S0892-1997(05)80181-X

Damasio, H., and Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain* 103, 337–350. doi: 10.1093/brain/103.2.337

Delvaux, V., and Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145–173. doi: 10.1159/000107914

Dick, A. S., Solodkin, A., and Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain Lang.* 114, 101–114. doi: 10.1016/j.bandl.2009.08.005

D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., and Fadiga, L. (2011). Vocal pitch discrimination in the motor system. *Brain Lang.* 118,

9–14. doi: 10.1016/j.bandl.2011.02.007

Estow, S., and Jamieson, J. P. (2007). Self-monitoring and mimicry of positive and negative social behaviors. *J. Res. Pers.* 41, 425–433. doi: 10.1016/j.jrp.2006.05.003

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402. doi: 10.1046/j.0953-816x.2001.01874.x

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *J. Phon.* 14, 3–28.

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377. doi: 10.3758/BF03193857

Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 517–528. doi: 10.1098/rstb.2002.1234

Garrod, S., and Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Lang. Cogn. Process.* 8, 101–126. doi: 10.1080/01690969308406950

Ghosh, S. S., Tourville, J. A., and Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *J. Speech Lang. Hear. Res.* 51, 1183. doi: 10.1044/1092-4388(2008/07-0119)

Giles, H. (1973). Accent mobility: a model and some data. *Anthropol. Linguist.* 15, 87–105.

Giles, H., Coupland, N., and Coupland, J. (1991). "1. Accommodation theory: communication, context, and consequence," in *Contexts of Accommodation: Developments in Applied Sociolinguistics*, eds H. Giles, N. Coupland, and J. Coupland (New York, NY: Cambridge University Press), 1–68.

Giles, H., Taylor, D. M., and Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Lang. Soc.* 2, 177–192. doi: 10.1017/S0047404500000701

Goldinger, S. D. (1997). "Words and voices: Perception and production in an episodic lexicon," in *Talker Variability in Speech Processing*, eds K. Johnson and J. W. Mullennix (San Diego, CA: Academic Press), 33–66.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical

access. *Psychol. Rev.* 105, 251. doi: 10.1037/0033-295X.105.2.251

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., and Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage* 55, 1324–1338. doi: 10.1016/j.neuroimage.2010.12.065

Grabski, K., Schwartz, J.-L., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., et al. (2013). Shared and distinct neural correlates of vowel perception and production. *J. Neurolinguist.* 26, 384–408. doi: 10.1016/j.jneuroling.2012.11.003

Gregory, S. W. Jr. (1986). A sociolinguistic indicator of group membership. *J. Psycholinguist. Res.* 15, 189–207.

Gregory, S. W. Jr., Green, B. E., Carrothers, R. M., Dagan, K. A., and Webster, S. (2000). Verifying the primacy of voice fundamental frequency in social status accommodation. *Lang. Commun.* 21, 37–60. doi: 10.1016/S0271-5309(00)00011-2

Gregory, S. W. Jr., and Hoyt, B. R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *J. Psycholinguist. Res.* 11, 35–46. doi: 10.1007/BF01067500

Gregory, S. W. Jr., and Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *J. Pers. Soc. Psychol.* 70, 1231. doi: 10.1037/0022-3514.70.6.1231

Gregory, S. W., Dagan, K., and Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *J. Nonverb. Behav.* 21, 23–43. doi: 10.1023/A:1024995717773

Gregory, S. W., Webster, S., and Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Lang. Commun.* 13, 195–217. doi: 10.1016/0271-5309(93)90026-J

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* 39, 350–365. doi: 10.1016/j.jcomdis.2006.06.013

Hawks, J. W. (1985). "Intelligibility of one's own speech relative to the speech of others. Independent Studies and Capstones," in *Program in Audiology and Communication Sciences* (Washington University School of Medicine). Available online

at: http://digitalcommons.wustl.edu/pacs_capstones/44

Heimann, M., Nelson, K. E., and Schaller, J. (1989). Neonatal imitation of tongue protrusion and mouth opening: methodological aspects and evidence of early individual differences. *Scand. J. Psychol.* 30, 90–101. doi: 10.1111/j.1467-9450.1989.tb01072.x

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4, 131–138. doi: 10.1016/S1364-6613(00)01463-7

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Houde, J. F., and Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science* 279, 1213–1216. doi: 10.1126/science.279.5354.1213

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science* 286, 2526–2528. doi: 10.1126/science.286.5449.2526

Indefrey, P., and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144. doi: 10.1016/j.cognition.2002.06.001

Irwin, J. R., Frost, S. J., Mencl, W. E., Chen, H., and Fowler, C. A. (2011). Functional activation for imitation of seen and heard speech. *J. Neurolinguist.* 24, 611–618. doi: 10.1016/j.jneuroling.2011.05.001

Jones, J. A., and Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *J. Acoust. Soc. Am.* 108, 1246–1251. doi: 10.1121/1.1288414

Kappes, J., Baumgaertner, A., Peschke, C., and Ziegler, W. (2009). Unintended imitation in nonword repetition. *Brain Lang.* 111, 140–151. doi: 10.1016/j.bandl.2009.08.008

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. California, CA: Pergamon Press Inc.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131.

Lelong, A., and Bailly, G. (2012). "Original objective and subjective

characterization of phonetic convergence," in *International Symposium on Imitation and Convergence in Speech*, (Aix en Provence). Available online at: http://hal.archives-ouvertes.fr/hal-00741686/

Leslie, K. R., Johnson-Frey, S. H., and Grafton, S. T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *Neuroimage* 21, 601–607. doi: 10.1016/j.neuroimage.2003.09.038

Lewandowski, N. (2009). "Sociolinguistic factors in language proficiency: Phonetic convergence as a signature of pronunciation talent," in *Language Talent and Brain Activity*, eds G. Dogil and S. M. Reiterer (Berlin: Mouton de Gruyter), 257–278.

Liberman, G., and Mattingly, A. M. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6

Meltzoff, A. N., and Moore, M. K. (1997). Explaining facial imitation: a theoretical model. *Early Dev. Parent.* 6, 179–192.

Miller, R. M., Sanchez, K., and Rosenblum, L. D. (2010). Alignment to visual speech information. *Atten. Percept. Psychophys.* 72, 1614–1625. doi: 10.3758/APP.72.6.1614

Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neurosci. Biobehav. Rev.* 33, 975–980. doi: 10.1016/j.neubiorev.2009.03.010

Molnar-Szakacs, I., Iacoboni, M., Koski, L., and Mazziotta, J. C. (2005). Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cereb. Cortex* 15, 986–994. doi: 10.1093/cercor/bhh199

Nagy, E. (2006). From imitation to conversation: the first dialogues with human neonates. *Infant Child Dev.* 15, 223–232. doi: 10.1002/icd.460

Namy, L. L., Nygaard, L. C., and Sauerteig, D. (2002). Gender differences in vocal accommodation: the role of perception. *J. Lang. Soc. Psychol.* 21, 422–432. doi: 10.1177/026192702237958

Nasir, S. M., and Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20470–20475. doi: 10.1073/pnas.0907032106

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social

desirability. *J. Pers. Soc. Psychol.* 32, 790–804. doi: 10.1037/0022-3514.32.5.790

Nguyen, N., Dufour, S., and Brunellière, A. (2012). Does imitation facilitate word recognition in a non-native regional accent? *Front Psychol.* 3:480. doi: 10.3389/fpsyg.2012.00480

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *J. Phonet.* 39, 132–142. doi: 10.1016/j.wocn.2010.12.007

Nyström, P. (2008). The infant mirror neuron system studied with high density EEG. *Soc. Neurosci.* 3, 334–347. doi: 10.1080/17470910701563665

Özdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635. doi: 10.1016/j.neuroimage.2006.07.013

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119, 2382–2393. doi: 10.1121/1.2178720

Pardo, J. S. (2010). "Expressing oneself in conversational interaction," in *Expressing Oneself/Expressing One's Self*, ed E. Morsella (Taylor & Francis), 183–196.

Peschke, C., Ziegler, W., Kappes, J., and Baumgaertner, A. (2009). Auditory–motor integration during fast repetition: the neuronal correlates of shadowing. *Neuroimage* 47, 392–402. doi: 10.1016/j.neuroimage.2009.03.061

Pfordresher, P. Q., and Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Percep.* 25, 95–115. doi: 10.1525/mp.2007.25.2.95

Poeppel, D., and Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition* 92, 1–12. doi: 10.1016/j.cognition.2003.11.001

Postma, E. O., and Nilsenova, M. (2012). "Individual differences in F0 imitation," in *Proceedings of Measuring Behavior 2012*, eds A. J. Spink, O. E. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, and P. H. Zimmerman (Utrecht), 154–157.

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., et al. (2011). Individual differences in audio-vocal speech imitation

aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Front. Psychol.* 2:271. doi: 10.3389/fpsyg.2011.00271

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194. doi: 10.1016/S0166-2236(98)01260-0

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (2002). "14 From mirror neurons to imitation: facts and speculations," in *The Imitative Mind: Development, Evolution, and Brain Bases*, Vol. 6, eds A. N. Meltzoff and W. Prinz (New York, NY: Cambridge University Press), 247–266.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141. doi: 10.1016/0926-6410(95)00038-0

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136. doi: 10.1038/16056

Sancier, M. L., and Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *J. Phonet.* 25, 421–436. doi: 10.1006/jpho.1997.0051

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J.-L., and Nguyen, N. (2013). Converging towards a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Front. Cogn. Sci.* 4:422. doi: 10.3389/fpsyg.2013.00422

Sato, W., and Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104, 1–18. doi: 10.1016/j.cognition.2006.05.001

Schwartz, J.-L., Basirat, A., Ménard, L., and Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): a perceptuo-motor theory of speech perception. *J. Neurolinguist.* 25, 336–354. doi: 10.1016/j.jneuroling.2009.12.004

Serkhane, J., Schwartz, J.-L., and Bessiere, P. (2005). Building a talking baby robot: A contribution

to the study of speech acquisition and evolution. *Interact. Stud.* 6, 253–286. doi: 10.1075/is.6.2.06ser

Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *J. Acoust. Soc. Am.* 125, 1103. doi: 10.1121/1.3058638

Shockley, K., Sabadini, L., and Fowler, C. A. (2004). Imitation in shadowing words. attention, *Percept. Psychophys.* 66, 422–429. doi: 10.3758/BF03194890

Shockley, K., Santana, M. V., and Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326. doi: 10.1037/0096-1523.29.2.326

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101, 260–277. doi: 10.1016/j.bandl.2007.02.008

Sommerville, J. A., Woodward, A. L., and Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition* 96, B1–B11. doi: 10.1016/j.cognition.2004.07.004

Spengler, S., von Cramon, D. Y., and Brass, M. (2010). Resisting motor mimicry: control of imitation involves processes central to social cognition in patients with frontal and temporo-parietal lesions. *Soc. Neurosci.* 5, 401–416. doi: 10.1080/17470911003687905

Street, R. L., and Giles, H. (1982). Speech accommodation theory: a social cognitive approach to language and speech behavior. *Soc. Cogn. Commun.* 193–226.

Sörös, P., Sokoloff, L. G., Bose, A., McIntosh, A. R., Graham, S. J., and Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *Neuroimage* 32, 376–387. doi: 10.1016/j.neuroimage.2006.02.046

Tajfel, H., and Turner, J. C. (1979). "An integrative theory of intergroup conflict," in *The Social Psychology of Intergroup Relations*, eds W. G. Austin and S. Worchel (Monterey, CA: Brooks-Cole), 33–47.

Tanaka, S., and Inui, T. (2002). Cortical involvement for action imitation of hand/arm postures versus finger configurations: an fMRI study. *Neuroreport* 13, 1599–1602.

doi: 10.1097/00001756-200209160-00005

Terumitsu, M., Fujii, Y., Suzuki, K., Kwee, I. L., and Nakada, T. (2006). Human primary motor cortex shows hemispheric specialization for speech. *Neuroreport* 17, 1091–1095. doi: 10.1097/01.wnr.0000224778.97399.c4

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054

Vallabha, G. K., and Tuller, B. (2004). Perceptuomotor bias in the imitation of steady-state vowels. *J. Acoust. Soc. Am.* 116, 1184. doi: 10.1121/1.1764832

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J. Acoust. Soc. Am.* 122, 2306–2319. doi: 10.1121/1.2773966

Wilson, S. M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702. doi: 10.1038/nn1263