Published in final edited form as:
Neuroimage. 2011 August 15; 57(4): 1561–1571. doi:10.1016/j.neuroimage.2011.05.067.

On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception

Pascale Tremblay[1] and **Steven L. Small**[1]

[1] The University of Chicago, Department of Neurology, 5841 S. Maryland Ave. MC-2030, 60637, Chicago, IL, USA

Abstract

What is the nature of the interface between speech perception and production, where auditory and motor representations converge? One set of explanations suggests that during perception, the motor circuits involved in producing a perceived action are in some way enacting the action without actually causing movement (covert simulation) or sending along the motor information to be used to predict its sensory consequences (i.e., efference copy). Other accounts either reject entirely the involvement of motor representations in perception, or explain their role as being more supportive than integral, and not employing the identical circuits used in production. Using fMRI, we investigated whether there are brain regions that are conjointly active for both speech perception and production, and whether these regions are sensitive to articulatory (syllabic) complexity during both processes, which is predicted by a covert simulation account. A group of healthy young adults (1) observed a female speaker produce a set of familiar words (perception), and (2) observed and then repeated the words (production). There were two types of words, varying in articulatory complexity, as measured by the presence or absence of consonant clusters. The simple words contained no consonant cluster (e.g. "**pa**lace"), while the complex words contained one to three consonant clusters (e.g. "**pla**net"). Results indicate that the left ventral premotor cortex (PMv) was significantly active during speech perception and speech production but that activation in this region was scaled to articulatory complexity only during speech production, revealing an incompletely specified efferent motor signal during speech perception. The right planum temporal (PT) was also active during speech perception and speech production, and activation in this region was scaled to articulatory complexity during both production and perception. These findings are discussed in the context of current theories theory of speech perception, with particular attention to accounts that include an explanatory role for mirror neurons.

1. Introduction

From the first months of life, speech perception and production are closely related. Indeed, learning to speak through babbling involves learning a mapping between articulation and the resulting acoustic signal, and thus requires close interaction between perceptual and motor systems. By the end of the first year of life, a child's perception and production begin to reflect characteristics of his or her native language, and a concomitant synchronization of

Corresponding author: Pascale Tremblay, PhD., Center for Mind & Brain Sciences (CIMeC), Università degli studi di Trento, Via delle Regole, 101, I - 38060, Mattarello (TN), Italy. pascale.tremblay@unitn.it, Phone: (+39)0461282752.

the two systems (Vihman & de Boysson-Bardies, 1994). As pointed out by Liberman and colleagues (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967): "[…] t*he perceiver is also the speaker and must be supposed, therefore, to possess all the mechanisms for putting language through the successive coding operations that result eventually in the acoustic signal*" (p.452). Thus, the idea that perception and production must overlap at the neural level is not surprising. Nevertheless, the nature and extent of these interactions remain unclear.

Proponents of the Motor Theory of Speech Perception (MTSP), developed in the 1950s (Liberman et al., 1967; Liberman & Mattingly, 1985), have contended that speech perception and production are intimately linked, and that conversion from acoustic signal to articulatory patterns occurs within a biologically specialized "speech" (phonetic) brain module. According to this view, each speech sound (phoneme) is associated with a specific combination of motor commands, such as "tongue retraction" and "jaw opening". The ability to categorize the speech sounds forming the incoming speech stream is accomplished by tracking the intended articulatory patterns, and thus, the intended articulatory patterns represent the ultimate objects of speech perception, and "perception tracks articulation" (see Galantucci, Fowler, & Turvey, 2006, for a contemporary view of this theoretical perspective).

More recently, the close relationship between the neural mechanisms for speech perception and production has received support with the discovery, in the 1990s, of a special class of neurons, the mirror neurons (MNs), first identified in area F5 of the macaque ventral premotor cortex (PMv). These MNs discharge both when a monkey produces an action and when he observes another monkey or a person performing the same or a similar action (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Kohler et al., 2002; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). It has been suggested that activity in MNs during action observation reflects the enactment (covert simulation) of the motor program(s) involved in producing the action being observed, a mechanism that would be essential to action recognition and understanding. As noted by Hickok (2009), the existence of human MNs has been inferred beginning with the earliest mirror neuron reports, yet their existence remains to be demonstrated (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti & Arbib, 1998; but see also Turella, Pierno, Tubaldi, & Castiello, 2009). By inference from homology and evidence from brain imaging, the localization of putative human MNs in the frontal lobe is typically assumed to lie in PMv and/or adjacent pars opercularis of the inferior frontal gyrus (pIFG), which would be the analogues of macaque area F5 (Rizzolatti & Arbib, 1998; Rizzolatti & Craighero, 2004, but see also Petrides, Cadoret, & Mackey, 2005). Despite some controversy, the discovery of MNs is important for it provides a neuroanatomical substrate for the link between perception and production. For some theorists, the presence of MNs also provides a neurophysiological mechanism (covert simulation) by which this link is instantiated.

Following the discovery of MNs, evidence for the binding of speech perception and production mechanisms has grown. A number of neuroimaging studies have shown activation in PMv during passive listening to syllables and phonemes (Pulvermuller, Shtyrov, Ilmoniemi, & Marslen-Wilson, 2006; Wilson & Iacoboni, 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). Moreover, passively watching videos of a person telling a story activates PMv/pIFG more strongly than listening to the same stories (without seeing the talker), suggesting a role for PMv/pIFG in tracking articulation visually (Skipper, Nusbaum, & Small, 2005). The left pIFG has also been shown to be sensitive to perceived phonetic categories independent of sensory properties; these supra-modal categorical representations may be used to process and produce speech, although this remains to be demonstrated empirically (Hasson, Skipper, Nusbaum, & Small, 2007). In addition to being activated

during passive auditory and audiovisual tasks, a number of phonological tasks also recruit the PMv/pIFG area, including phonetic discrimination (i.e. judging whether two auditory presented syllables start/end with the same consonant) (Zatorre, et al., 1992, Siok, Jin, Fletcher, & Tan, 2003, Burton et al., 2000, Burton & Small, 2006), same/different syllable judgments performed on visual-only videos depicting individuals producing meaningless syllables (Fridriksson et al., 2008), discrimination of non-speech and speech sounds (Joanisse & Gati, 2003), phonemic identification/discrimination (Siok et al., 2003, Callan et al., 2010) and sound-picture matching tasks performed in babbling noise (Wong, et al. 2008), to name a few.

Additional evidence for a role for PMv/pIFG during speech perception and production comes from transcranial magnetic stimulation (TMS) studies. TMS can be used to transiently disrupt the function of a given cortical region, thereby creating a reversible, focal "virtual" lesion. TMS studies have shown that when stimulating the face/mouth area of the left primary motor cortex (M1), motor-evoked potentials (MEP) recorded from the lip and tongue are enhanced during passive speech perception, suggesting access to motor representations during speech perception (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Sundara, Namasivayam, & Chen, 2001; Watkins & Paus, 2004; Watkins, Strafella, & Paus, 2003). Furthermore, when applied to PMv, TMS interferes with participants' ability to discriminate speech sounds in noise (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007). Interestingly, however, applying TMS to PMv in the absence of ambient noise has little or no effect on participants' ability to perceive or categorize speech sounds (Sato, Tremblay, & Gracco, 2009), suggesting that speech perception may not be based on, or constrained by, articulatory mechanisms. One possibility is that speech production mechanisms may be used in an auxiliary fashion, i.e., in situations in which there is perceptual ambiguity, such as in the presence of noise, in the context of hearing loss, during exposure to an unfamiliar accent, or, more generally speaking, when performing a difficult perception task. In such situations, motor representations could be used to further constrain speech sound identification. This position is consistent with the Perception-for-Action-Control Theory (PACT) (Schwartz, Basirat, & Sato, 2010), which posits that speech production mechanisms have two functions in speech perception: first, they are used to co-structure auditory categories with learned motor routines, which leads to the integration of articulatory information into perceptual categories, and, second, they mediate speech perception through a predictive mechanism (i.e., efference copy) similar to the one proposed in Skipper et al. (2006a;b), but which is only used when there is perceptual ambiguity, to recover the missing information. On the other hand, there is little reason to believe that brain mechanisms are ever turned off, so it is more likely that production systems are exploited more fully during perceptual ambiguity than when there is no such ambiguity.

While the PMv/pIFG area may be a site of importance for perceptual-motor convergence for speech, there are good reasons to believe that perception and production also interface in auditory regions. One view that emphasizes the role of auditory cortex in this interface suggests that speech production mechanisms depend on speech perception, and that the shared representations are auditory instead of motor (Guenther, Ghosh, & Tourville, 2006; Hickok & Poeppel, 2000, 2004, 2007).Consistent with this hypothesis, several brain imaging studies have shown that the left planum temporale (PT), which corresponds to the caudal end of the superior temporal plane, just anterior to the supramarginal gyrus (SMG), is activated during speech production (Bohland & Guenther, 2006; Dhanjal, Handunnetthi, Patel, & Wise, 2008; Karbe, Herholz, Weber-Luxenburger, Ghaemi, & Heiss, 1998; Peschke, Ziegler, Kappes, & Baumgaertner, 2009; Riecker, Brendel, Ziegler, Erb, & Ackermann, 2008; Saito et al., 2005; Schulz, Varga, Jeffires, Ludlow, & Braun, 2005; Tourville, Reilly, & Guenther, 2008; Wise et al., 2001; Zheng, Munhall, & Johnsrude, 2010), but also during (silent) speech rehearsal, which does not involve self-generated

auditory feedback (Buchsbaum, Hickok, & Humphries, 2001; Callan, Callan, Tajima, & Akahane-Yamada, 2006; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Hickok et al., 2000; Huang, Thangavelu, Bhatt, C, & Wang, 2002; Okada, Smith, Humphries, & Hickok, 2003; Pa & Hickok, 2008; Papathanassiou et al., 2000; Shergill et al., 2002; Wise et al., 2001). Importantly, it has also been shown that the caudal part of PT and the adjacent SMG[1], is more sensitive to sub-vocal rehearsal than to the perception of auditory stimuli (Buchsbaum et al., 2001; Hickok et al., 2003), further suggesting that this region is not exclusively tied to auditory feedback processing, but could also play a role in integration of information about speech perception and production.

In sum, evidence abounds that speech perception and production are linked at the level of the cerebral cortex. While this might signify that perceptual-motor representations for speech are overlapping, even perhaps inseparable, drawing such a conclusion at present would be premature, because too little is known about the general and specific roles of PMv, pIFG and PT in speech perception and production. In the present study, we sought to further characterize the contribution of these regions to the perception and production of speech using functional MRI (fMRI), in order to shed light on the nature of the link between these processes. In a first step, we identified regions that are active during both single word perception and single word production. Then, in a second step, we examined whether regions with overlapping activation for perception and production are sensitive to articulation. To do this, we manipulated the articulatory complexity of the words, using both simple words, containing no consonant cluster (e.g. “palace”), and complex words, containing one to three consonant clusters (e.g. “planet”). A consonant cluster (or blend) is a sequence of two or more consonants (C) occurring together in the same syllable without a vowel (V) between them (e.g. CCV). In most languages, syllables containing consonant clusters are less common than consonant-vowel syllable (CV) sequences. In fact, many languages of the world do not permit consonant clusters at all (e.g., Maori, Hawaiian, Fijian), or restrict their position in words (e.g., Arabic). In English, one third of monosyllables begin with a consonant cluster, and consonant clusters predominate in word-final position (Locke, 1983). Compared to CV sequences, consonant clusters represent an increased difficulty for the speaker because they necessitate the rapid production of two consonant gestures, which requires a quick reconfiguration of the articulators to change the manner in which the airflow is obstructed. The protracted development of consonant clusters in normally developing children illustrates the complexity associated with producing these types of sequences.

Consonant clusters present various degrees of articulatory complexity, which depends upon the degree of articulatory similarity of the adjacent consonants. The more dissimilar the consonants, the higher the difficulty, and the later these sequences are mastered by typically developing children (Smit et al., 1990; Templin, 1957; Higgs, 1968, see also McLeod et al., 2001 for a review of the literature on consonant cluster acquisition). Two-element consonant clusters are mastered before more complex three-element clusters (McLeod et al., 2001). There is also evidence to suggest that consonant clusters continue to present an increased difficulty in adulthood compared to simpler syllable structures. For instance, reading words beginning with consonant clusters is slower than reading words beginning with a single consonant (Santiago et al., 2000). Moreover consonant clusters in the word initial position have been shown to increase the probability of stuttering (Howell et al., 2000). There is also evidence that clusters of maximal articulatory difficulty for the speaker, i.e., those containing maximally different adjacent phonemes, which require greater articulatory travel

[1]“Based on functional brain imaging data, but without clear anatomy, some investigators have called this general brain area “Spt”.

(e.g. /gz/), are either avoided completely or produced rarely in both English and Spanish (Saporta, 1955).

The intrinsic difficulty in producing consonant clusters is further demonstrated by the robust finding that children or adults with apraxia of speech, a disorder of speech motor control, present high error rate for the production of consonant clusters, which are often reduced to a single consonant (e.g. Lewis et al., 2004, Jacks et al., 2006, Aichert and Ziegler, 2004). Finally, three previous fMRI studies have shown that words containing consonant clusters are associated with increased activity in a number of areas associated with speech production compared to words containing simpler syllabic structure (Bohland & Guenther, 2006; Riecker et al., 2008; McGettigan et al., 2010). Based on the literature, we hypothesized that if activity in PMv/pIFG reflects motor simulation during perception, then it follows that these regions should be modulated by articulatory complexity during both speech production and speech perception.

2. Methods

2.1.Participants

Twenty-one healthy right-handed (Oldfield, 1971) native speakers of American English (mean age 23.7 ± 5.5; range: 18-38 years; 11 females), with a mean of 15.1 ± 2 years of education (range: 12-18) participated in this experiment. The data from one participant could not be used because of a technical problem with the stimulus presentation, leaving twenty participants in the analysis. All participants had normal or corrected-to-normal vision and no self-reported history of speech, language or neurological disorder. All participants had normal hearing, as assessed using a standard audiometric testing procedure (pure-tone air conduction thresholds for the following frequencies: 250, 500, 1000, 2000, 3000, 4000, 6000, and 8000Hz). In addition, a standard speech discrimination testing procedure was used to evaluate participants' ability to identify speech sounds. Speech discrimination procedures measure a person's ability not only to hear words but also to identify them. We used the Northwestern University auditory test number six (form A). The procedure includes the presentation of 50 monosyllabic words at an easily detectable intensity level and the calculation the percentage of words correctly identified. The Institutional Review Board of the Biological Sciences Division of The University of Chicago approved the study.

2.2. Stimuli and Procedures

The experiment consisted of two tasks: (1) observation of a set of short video clips showing a female actor producing single words (perception), and (2) observation of a set of similar videos followed by repetition of the word produced by the speaker (production). A resting condition (crosshair fixation) was also included as the baseline condition. The tasks were performed within separate runs. Participants always completed the perception task first, in order to avoid covert rehearsal of the words during perception. Moreover, participants did not know that they would be required to produce words until the beginning of the production task: this was done to avoid covert rehearsal during perception. Within each run, the experimental trials (simple, complex) were interleaved with rest trials; the order of the conditions and the number and (jittered) duration of rest trials were optimized using OPTseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/). Trials were separated by a minimum of 3.5 sec and a maximum of 7 sec of rest. This data set was acquired as part of a larger project that also included two 5-min resting scans, a sentence listening task and a hand movement observation task, which will not be discussed here.

The stimuli were 120 short video clips of a native English-speaking female actor articulating a set of bisyllabic nouns matched for their stress pattern (all words had the stress on the first syllable). We chose to use audiovisual word stimuli because they approximate more closely

naturalistic face-to-face verbal communications. During such interactions, visual information from the face and lips can help disambiguate speech, particularly in challenging contexts, such as when speaking in a noisy environment. Because visual information processing was not a factor of interest, the stimuli were matched for the amount of visual speech information that they provided (see explanation below).

The words were divided into two classes based on articulatory complexity: simple and complex, resulting in a 2×2 experimental design (Task, Complexity). Articulatory complexity was measured in terms of the presence or absence of a consonant cluster.[2] The simple words contained no cluster; all had a CV-CV structure or a CV-CVC structure. There was never any adjacent consonant, either within or across syllables. The complex words, in contrast, contained on average 1.73 consonant clusters (range 1-3), and included words with different combinations of four syllable types: CV, VCV, CCV and VCVC. The complex words contained groups of adjacent consonants within and across syllables (see Supplementary Table 1 for a list of all words and a description of their syllabic structure). The simple words had an average familiarity[3] score of 537 (±55), an average concreteness score of 490 (±112), and an average number of visemes[4] of 2.86 (±.44). The complex words had an average familiarity of 520 (±49), an average concreteness of 468 (±109), and an average number of visemes of 2.81 (±.48). These differences were not statistically significant. All stimuli were presented using Presentation Software (Neurobehavioral System, CA, USA). Visual stimuli were delivered to a custom rear projection screen placed inside the bore of the magnet approximately 24” from the subject. The subject viewed the stimuli via a mirror attached to the head coil. Auditory stimuli were delivered via a high quality full frequency range auditory amplifier (Avotec Inc., FL, USA). We used a noise cancellation method to record subject's overt responses using an MRI compatible microphone without the scanner noise. The noise suppression method creates a template of the noise from the scanner during the dummy period. The template is subtracted from the output of the microphone located in the bore of the magnet near the subject's mouth. The algorithm reduces the scanner noise by 20 dB.

2.3 Image acquisition

The data were acquired on a whole-body Siemens 3.0 Tesla Tim Trio MRI scanner (Siemens Medical Solutions, Erlangen, Germany) at Northwestern University (Chicago, IL, USA). Subjects wore MR compatible headphones (Avotec Inc., FL, USA). Thirty-two axial slices (3*1.7*1.7 mm, no gap) were acquired in interleaved order using a multislice EPI sequence (TR = 2sec, TE = 20ms; FOV = 200*207*127mm; 128*128 matrix; Flip angle: 75). Two experimental runs (6.3 minutes each) resulted in the acquisition of 380 T2*-weighted BOLD images (120 experimental trials and 60 baseline trials). High-resolution T1-weighted volumes were acquired for anatomical localization (176 sagittal slices, 1*1*1 mm resolution, TR = 23ms, TE = 2.91, FOV = 256*256*176 mm). Throughout the procedure, each subject's head was immobilized by means of a set of cushions and pads.

[2]It is important to distinguish between consonant clusters and digraphs. A digraph is a group of two or more orthographic symbols that stand for one sound (usually a consonant). Fort instance, in the English word /chat/, the sequence /ch/ represents a single sound and contains no consonant cluster.

[3]The familiarity and concreteness indices were extracted from the MRC psycholinguistic database available at : http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

[4]Visemes are a subset of visual speech movements that are sufficient for phonetic classification, that is, they provide access to phonetic information in the absence of accompanying auditory speech signal (e.g. Jackson, 1988; Preminger, Lin, Payen, & Levitt, 1998). In this study, we used various measures of viseme content (Bement, et al., 1988).

2.4 Image analysis

All time series were spatially registered, motion-corrected, time-shifted, de-spiked and mean-normalized using AFNI (Cox, 1996). In addition, we censored time points occurring during excessive motion, defined as >1 mm (Johnstone et al., 2006). For each subject we created separate regressors for each of our four experimental conditions (perception simple, perception complex, production simple, production complex); additional regressors were the mean, linear, and quadratic trend components, and the 6 motion parameters (x, y, z and roll, pitch and yaw). To remove additional sources of spurious variance unlikely to represent signal of interest, we also regressed signal from the lateral ventricles (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Fox et al., 2005). A linear least squares model was used to establish a fit to each time point of the hemodynamic response function for each condition. We modeled the entire trial duration (i.e. TR = 2sec), which was the same across perception and production runs. Hence, the modeled intervals for speech production contained both stimulus and response related effects. Event-related signals were calculated by linear interpolation, beginning at stimulus onset and continuing at 2-sec intervals for 12 sec, using AFNI's tent function (i.e. a piecewise linear spline model). The fit was examined at these different time lags to identify the time points showing the strongest hemodynamic response in our regions of interest. All subsequent analyses focused on the beta values averaged across the 4-6 sec post-stimulus onset time lag and the 6-8 sec post-stimulus onset time lag.

We used FreeSurfer (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999) to create surface representations of each participant's anatomy by inflating each hemisphere of the anatomical volumes to a surface representation and aligning it to a template of average curvature. SUMA was used to import the surface representations into the AFNI 3D space and to project the functional data from the 3-dimensional volumes onto the 2-dimensional surfaces. Data were smoothed on the surface to achieve a target smoothing value of 6mm using a Gaussian full width half maximum (FWHM) filter. Smoothing on the surface as opposed to the volume ensures that white matter values are not included, and that functional data situated in anatomically distant locations on the cortical surface are not averaged across sulci (Argall, Saad, & Beauchamp, 2006; Desai, Liebenthal, Possing, Waldron, & Binder, 2005). Whole-brain, group analyses were performed using SUMA on the subjects' smoothed beta values resulting from the first level analysis. This analysis focused on the main effect of the perception and production conditions, as well as any differences between them.

The surface-based group analyses were corrected for multiple comparisons using a Monte Carlo simulation procedure on surface data, which implements the cluster-size threshold procedure as a protection against Type I error. Based on the simulation, we determined that a family-wise error (FWE) rate of $p < 0.05$ is achieved with a minimum cluster size of 196 contiguous surface nodes, each significant at $p < 0.005$. In addition to the whole-brain analyses, we also profiled brain areas involved in both perception and production by examining the “conjunction” (Nichols, Brett, Andersson, Wager, & Poline, 2005) of brain activity from the whole-brain contrasts (corrected for multiple comparisons). This analysis overlapped activity in the perception simple, perception complex, production simple and production complex, yielding an intersection map of *perception* ∩ *production*.

Anatomical ROI analysis—In addition to the whole-brain and conjunction analyses, we also conducted an analysis of anatomical regions of interest (ROI) on a set of *a priori* selected sensory and motor regions. ROIs were anatomically defined on each individual's cortical surface representation using an automated parcellation scheme (Fischl et al., 2004; Desikan et al., 2006). This procedure uses a probabilistic labeling algorithm that incorporates the anatomical conventions of Duvernoy (1991) and has a high accuracy

approaching that of manual parcellation (Fischl et al., 2002, 2004; Desikan et al., 2006). We augmented the parcellation manually with further subdivisions.

The ROIs resulting from the automated Freesurfer parcellation were divided into two anatomical classes: (i) frontal-parietal regions, which included pIFG, PMv, the ventral primary motor area (M1v), and the ventral somatosensory area (S1v), and (ii) temporal-parietal regions, which included PT, the transverse temporal gyrus (TTG), the transverse temporal sulcus (TTS), and the supramarginal gyrus (SMG). These ROIs were defined as follows (see also Figure 1). (1) p**IFG**: Unedited FreeSurfer ROI, defined as the gyrus immediately anterior to the precentral gyrus. PIFG is bounded caudally by the precentral sulcus, and rostrally by pars triangularis. (2) **PMv:** For PM, we edited the FreeSurfer precentral sulcus and gyrus regions, by subdividing them into ventral (PMv) and dorsal (PMd) segments at the level of the junction of the inferior frontal sulcus and the precentral sulcus. The resulting PMv is bounded rostrally by the IFG pars opercularis, caudally by the central sulcus, and dorsally by PMd, and includes the precentral sulcus. PMv was then further subdivided into two halves along the dorsal/ventral axis. All ROI analyses of this region focused on the dorsal part of this region. (3) **M1v:** For M1, we edited the FreeSurfer central sulcus region by subdividing it into a ventral (M1v) and a dorsal (M1d) segment at the level of the junction of the inferior frontal sulcus and the precentral sulcus. M1 is bounded rostrally by the precentral gyrus and caudally by the postcentral gyrus. (4) **S1v:** For S1, we edited the FreeSurfer postcentral gyrus region by subdividing it into ventral (S1v) and dorsal (S1d) segments at the level of the junction of the inferior frontal sulcus and the precentral sulcus. S1 is bounded rostrally by the central and caudally by the postcentral sulcus. (5) **PT:** Unedited FreeSurfer ROI, defined as the part of the superior temporal plane immediately posterior to the transverse temporal sulcus, bounded medially by the Sylvian fissure, and posteriorly by the supramarginal gyrus. (6) **TTG**: Unedited FreeSurfer ROI, bounded rostrally by the rostral extent of the transverse temporal sulcus, caudally by the caudal portion of the insular cortex, medially by the superior temporal gyrus and laterally by the lateral fissure. (7) **TTS**: Unedited FreeSurfer ROI, located immediately anterior to PT and posterior to the TTG. (8) **SMG**: Unedited FreeSurfer ROI, bounded rostrally by the caudal extent of the superior temporal gyrus, caudally by the rostral extent of the superior parietal gyrus, medially by the lateral banks of the intraparietal sulcus, and laterally by the medial banks of the lateral fissure and/or the superior temporal gyrus, respectively (Desikan et al., 2006).

For each ROI and each subject, we first extracted the mean percentage of BOLD signal change. Next we examined a set of four FDR-corrected t-tests ($q^* = .05$) (Benjamini and Hochberg, 1995; Genovese et al., 2002) (Rosenthal, Rosnow, & Rubin, 2000), focused on a set of specific hypotheses: (1) *perception > zero* ($n = 16$ one-sample t-test), *(2) production > zero* ($n = 16$ one-sample t-tests). For regions that were significantly active in both of these contrasts (N = 7 ROIs), we examined the effect of task (perception < production), and we also tested the following two additional hypotheses: (3) perception of complex words > perception of simple words, and (4) production of complex words > production of simple words, using a set of FDR-corrected paired-sample t-tests. Finally, we also tested for a Complexity × Task interaction contrast using difference scores for the complexity effect in perception and production [(perception complex – perception simple) – (production complex – production complex)].

3. Results

3.1 Behavioral data

Overall participants' performance was very high, with only 15 total errors in over 960 trials (representing fewer than 2% of all trials). Eight of these errors occurred during the simple trials, and seven during the complex trials.

3.2 Imaging data

3.2.1 Whole brain analysis—As illustrated in Figure 2, for perception (in blue), we found several clusters of positively activated nodes in sensory areas bilaterally, including TTG, the posterior superior temporal sulcus (STS), and PT, as well as the occipital lobe, including the caudal end of the calcarine sulcus and the occipital pole. There was also significant activation in premotor areas, including PMv (along the precentral sulcus and gyrus) and pre-SMA, bilaterally. A list of all clusters is presented in Table 1A. As shown in Figure 2 (in red), for production, we also found several clusters of positively activated nodes in sensory areas bilaterally, including TTG, posterior STS, PT, the caudal end of the calcarine sulcus and the occipital pole. There was also extensive bilateral activation in frontal motor and premotor areas, including PMv, pIFG, M1v and S1v, and pre-SMA. A list of all clusters is presented in Table 1B. The direct comparison (t-test) of production and perception revealed a cluster of activation located around PMv, bilaterally, as well as clusters of activation in the left inferior frontal sulcus, most of the STG including TTG, in the pre-SMA bilaterally, and in the occipital lobe. These results are detailed in Table 1C.

3.2.2 Conjunctions—Figure 2 (purple) shows the positive activation common to perception and production; this included activation in the left PMv, right inferior frontal sulcus, and bilateral superior temporal gyri, including TTG, TTS, PT and posterior STS. These results are presented in Table 2.

3.2.3 ROI analysis—First we examined whether activation magnitude in each ROI differed from zero (all p values followed by an asterisk survive an FDR correction, $q^* = 0.05$, $i = 32$). This analysis revealed that activations in the left PMv (production: $p = .0000002$*, perception: $p = .001$*), right PMv (production: $p = .00001$*, perception: $p = .003$*), left TTG (production: $p = .000003$*, perception: $p = .000001$*), right TTG (production: $p = .000001$*, perception: $p = .000002$*), left TTS (production: $p = .000001$*, perception: $p = .000002$*), right TTS (production: $p = .000001$*, perception: $p = .000001$*), left PT (production: $p = .000004$*, perception: $p = . .0000004$*) and right PT (production: $p = .0000004$*, perception: $p = .00000002$*) were significantly different from zero for both perception and production, consistent with the results of the whole-brain conjunction analysis; for the other ROIs (IFGp, M1v, S1v, SMG), activation was significantly different from zero only for production, with the exception of the SMG, which was not significantly active for either perception or production. Next, we examined the effect of task (perception, production) in the seven ROIs that showed significant activation magnitude for both perception and production (left PMv, right PMv, bilateral TTG, bilateral TTS, and bilateral PT). This analysis revealed that all ROIs, with the exception of PT bilaterally, were significantly more active for perception than production. In PT, activation magnitude was identical for perception and production.

Finally, we examined the effect of articulatory complexity in the same set of seven ROIs that showed significant activation magnitude for both perception and production. In the left PMv, the results revealed an effect of articulatory complexity for production (simple < complex, $p = .026$*), but not for perception ($p = .95$), and a highly significant Task × Complexity interaction ($t_{(19)} = 4.82$, $p = .0001$*). In the right PMv, there was no effect of

complexity for production (p = .32), or perception (p = .47), and the Task × Complexity interaction was not significant ($t_{(19)}$ = -.14, p = .89). These results are illustrated in Figure 3. In the left PT, there was no effect of complexity for production (p = .39), or perception (p = .18), and the Task × Complexity interaction was not significant ($t_{(19)}$ = .56, p = .59). In the right PT, we found a significant effect of Complexity for production (simple < complex, p = .019*), but not for perception, despite a trend in that direction (p = .06). The interaction was not significant ($t_{(19)}$ = -.27, p = .79). These results are illustrated in Figure 4. For the other temporal ROIs (bilateral TTS and bilateral TTG), there was no main effect of complexity for either perception or production, and no interaction.

4. Discussion

The discovery of mirror neurons (MN), active during both execution and observation of actions, has motivated the development and/or resurrection of various hypotheses about the links between speech perception and production, primarily by providing a potential neurophysiological mechanism to help explain the interaction between these processes. According to Rizzolatti and Craighero (2004): "*Each time an individual sees an action done by another individual, neurons that represent that action are activated in the observer's premotor cortex. This automatically induced, motor representation of the observed action corresponds to that which is spontaneously generated during active action and whose outcome is known to the acting individual*." With regard to spoken language, this account could be taken to suggest that perception of a syllable, e.g., /ba/, activates the same neural circuits involved in the production of /ba/. Consistent with this hypothesis, several imaging studies have shown overlap in the regions involved in speech perception and speech production. Of particular interest with respect to the MN account is the finding of overlap in PMv and pIFG, which are usually considered to have homology with macaque area F5, known to contain MNs. However, the question that naturally arises from these observations is the extent to which such activation during perception represents action simulation, or whether these representations are used in some other way (or not at all) to improve perception.

In fact, despite the existing accounts, very little is known about the specific role that PMv and pIFG play in speech perception and speech production. To address this issue, we used fMRI to examine whether there are brain regions involved in both perceiving and producing single words, and if so, whether these overlap regions show similar sensitivity to articulatory complexity (i.e., syllabic structure) during production and perception. Equivalent sensitivity to articulatory complexity in observation and execution would provide evidence for an action simulation mechanism in speech perception. In the present study, while we found that several regions were modulated by articulatory complexity during speech production, we found no region modulated by complexity during perception. Based on prior studies, two groups of cortical regions were examined (bilaterally): four frontal-parietal regions (pIFG, PMv, M1v and S1v) and four temporal-parietal regions (TTG, TTS, PT and SMG). In the following paragraph, we discuss our results separately for these two groups of regions.

4.1 Frontal-Parietal activation in speech perception and production

Unlike the other ROIs in this group (M1v, S1v or IFGp), the left PMv was significantly active during both speech perception and speech production, consistent with previous reports (e.g., Wilson et al., 2004, Pulvermuller et al., 2006, Skipper et al. 2007, Callan et al., 2010), though with a much stronger activation magnitude during production compared with perception. This finding confirms PMv as a site where perception and production are likely to interact, and suggests that speech perception should not be conceptualized as a set of strictly auditory processes.

A key finding of the current study is that, during speech production, the left PMv is modulated by articulatory complexity, operationalized here as syllable complexity, consistent with recent finding of syllable-level processing in this region (Peeva et al., 2010), while, during speech perception, no such modulation was found. The robustness of this finding is shown by the highly significant Task × Complexity interaction that was found in this region. Taken together, these findings appear at odds with the results of Wilson & Iacoboni (2006), who showed that the contrast of native and non-native phonemes modulates activation magnitude in the PMv during a passive perception task (native < non-native), suggesting that PMv is sensitive to whether or not a region is part of a speaker's phonological inventory, and consequently, to a speaker's motor repertoire, which emphasizes a link between perceptual and motor processes. However, the authors also report that PMv is not sensitive to the producibility of the non-native phonemes, consistent with the present finding (syllables containing adjacent consonants are indeed less easily articulated than syllables with a simpler syllabic structure). While Wilson & Iacoboni (2006) did not specifically address that question, it is likely that PMv would have been modulated by producibility during speech production. Indeed, a number of studies, including the present one, have shown that words containing consonant clusters are associated with increased activity in a number of areas associated with speech production compared to words containing simpler syllabic structure (Bohland & Guenther, 2006; Riecker et al., 2008), hence that producibility modulates the motor system during speech production.

Two potential interpretations of the present findings are (1) that action simulation is not solely driven by the motor characteristics of the observed action, but also by the manner in which the perceptual information is acted upon (*i.e.*, the task) or (2) that activation in PMv does not reflect a motor simulation process at all. Indeed, if activation in PMv during perception reflects the enactment of the same neural circuits involved in speech production, then one would predict that activation pattern in PMv would be identical in perception and production, and hence reflect syllable structure during both perception and production. Such a pattern of activation was not found. However, action simulation may be context- or task-dependent. Indeed, in the absence of a specific task (for example, during a passive speech perception task like the one that we used in the present study), access to a detailed phonological representation of the incoming auditory signal may not be necessary. Consequently, in that context, the signal coming from premotor centers may be limited to only necessary information, that is, it may not contain a fully specified set of motor commands, a suggestion that is consistent with the principle of parsimony. In this case, the speech motor system would be tuned to task-specific requirements and only generate as much information as needed. Spontaneous changes in the environment and task-specific requirements would trigger a recalibration of the system to generate increased or decreased amount of information.

The reliance of speech perception upon more detailed phonological representations, presumably associated with stronger PMv activation, can be triggered in various naturalistic situations. These situation include, but are not limited to, perception in a noisy environment resulting in a degraded auditory signal, processing a foreign accent or trying to parse an unknown language, or performing a phonological task (e.g. phoneme identification). Additional circumstances include the auditory degradation that comes from alterations in the perceiver, such as hearing loss or neurological injury.

The idea of a context-dependent contribution of PMv to speech perception is supported by a series of recent fMRI and rTMS studies. For example, using rTMS, Sato, Tremblay and Gracco (2009) demonstrated that, in the absence of ambient noise, stimulation of the left PMv has little or no effect on participants' ability to perceive or categorize speech sounds. In contrast, when the auditory signal is degraded, repetitive TMS to PMv has a stronger impact

on speech perception (Meister et al., 2007). Consistent with this finding, Callan et al. (2010) used fMRI to show that correct identification of phonemes in noise was associated with increased activation in PMv relative to trials in which the phonemes were incorrectly identified. The authors interpreted this as reflecting the use of an internal model (i.e. motor simulation) to facilitate speech perception. In summary, the present findings, together with previous studies, suggest a context-dependent contribution of PMv to speech perception. Future work is needed to characterize the type of information that is contributed by PMv in different contexts, as well as the origin (top-down vs. bottom-up) of this information, and the respective contribution of general and context-dependent motor information to speech perception.

4.2 Audiovisual speech perception and PMv

In the present study, we used audiovisual stimuli to characterize the interface between speech perception and speech production. In day-to-day situations, verbal communication consists primarily of face-to-face interactions during which both auditory and visual (face and lips) speech information can help speech recognition. Indeed, it has been shown that adding visual information enhances speech recognition (e.g., Sumby and Pollack, 1954, Reisberg et al., 1987), which could be associated with decreased reliance upon motor information during speech perception. This finding could explain the relatively low activation level found in PMv during speech perception in the present study. However, increased activation in IFGp/PMv has been shown in the context of audiovisual speech perception compared to auditory speech perception (Skipper et al., 2005; 2007). Because the objective of this study was not to examine the specific contribution of visual information to speech processing, all stimuli were matched (across tasks and complexity levels) for the amount of visual speech information that they provided. Hence, the activation level in PMv in the present study reflects the processing of audiovisual speech stimuli in the context of a passive perception task, and an overt speech production task. Indeed, visual information from the face and lips was equally available during speech production and perception yet activation in PMv was scaled to syllable structure only in speech production, a finding that does not appear to bear a relationship on the audiovisual nature of the stimuli that were used.

4.3 Temporal-Parietal activation in speech perception and production

As discussed in the Introduction, speech perception and production appear to interact not only in premotor regions, but also in cortical auditory association regions. One view presents these auditory association cortices as the primary sites for this interaction, *i.e.*, speech production depends on speech perception mechanism and not the reverse. Based on fMRI data and studies of aphasic patients, Hickok and Poeppel (2000, 2004, 2007) have suggested that auditory association areas play an important part in speech production. In particular, they emphasized the contribution of a region of the caudal temporal plane, incorporating the planum temporale (PT) and the adjacent supramarginal gyrus (SMG), to this process. PT is a large cortical area (most probably containing multiple functional fields) located immediately caudal to Heschl's sulcus (Von Economo and Horn, 1930, Pfeifer, 1936, Galaburda and Sanides, 1980). The results of the present study support a role for PT (but not SMG) in processing speech that goes beyond perception, with significant activation bilaterally during both perception and production. In this respect, PT behaves like PMv, and appears to be a site for perceptual-motor interaction and/or integration. However, unlike PMv, PT, bilaterally, was equally activated during perception and production, revealing a similar contribution to both processes (recall that the contribution of PMv to perception was weaker than its contribution to production). This suggests that PT activation was not driven by the presence or absence of self-generated auditory or proprioceptive feedback (which are absent in perception). Additional evidence that PT is not simply involved in processing self-generated feedback comes from the finding that the activation magnitude in the right PT was

scaled to articulatory complexity during production, and there was a similar trend during speech perception. This pattern was not found in the left PT, the TTG or the TTS, which were not modulated by articulatory complexity, either during perception or production. These results suggest that a fully specified phonological-motor representation of the perceived audiovisual speech signal is available at the level of the right PT during a passive speech perception task as well as during an auditory-triggered speech repetition task. Taken together, these findings suggest that the right PT is not involved in processing self-generated feedback, but instead, may be involved in converting external auditory input into a phonological representation for categorizing speech sounds in the incoming auditory stream, or for generating the corresponding motor commands during speech production. This interpretation is consistent with the hypothesis of a role for PT in sensorimotor transformation occurring within a larger dorsal route network (Hickok and Poeppel, 2000; 2004, 2007).

While in the present study we did not examine functional connectivity, we nevertheless report similar activation patterns in PT and PMv during speech production, providing indirect evidence of a link between motor regions and PT, and support for a role for this region in sensorimotor transformation for speech production, consistent with previous studies (Hickok et al, 2000, Bushbaum et al., 2001 Hickok *et al.*, 2003, Okada et al., 2003, Pa and Hickok, 2008). As discussed in section 4.1, PMv was not sensitive to articulatory complexity during speech perception, which may reflect the lack of a need to access detailed phonological information during speech perception, supporting our hypothesis of a context-dependent contribution of motor/premotor areas to speech perception. When speech production is contingent upon the processing of an auditory speech signal, such as in word repetition, such sensorimotor transformation is necessary to establish parity between perceptual and motor representations. Our results suggest that this parity may be established though a connection between the right PT and the left PMv.

In contrast to the right PT, the left PT was similarly involved in speech perception and production, suggesting a role in converting audiovisual speech into a phonological representation. However, the lack of an effect of articulatory (syllabic) complexity on either production or perception at the level of the left PT may indicate the generation of a cruder phonological representation, which could be used for a different purpose than the representation generated in the right PT. Laterality effects in PT are abundant in the literature, and are generally supported by a known leftward asymmetry (Geschwind and Levitsky, 1968; Galaburda et al., 1978; Steinmetz and Galaburda, 1991), which has been interpreted as reflecting a foundation of the left cerebral hemisphere for language (e.g. Tzourio et al., 1998; Marshall, 2000), though findings are not entirely consistent (e.g. Binder et al., 1996; Dorsaint-Pierre, et al., 2006; Eckert et al., 2006). For example, a recent large-scale (N = 100) study focusing on the relationship of PT asymmetry and language dominance (as determined by comparing a single word comprehension task with a tone perception task), revealed no relationship between these two variables (Eckert et al., 2006). These and other studies suggest that PT is involved in processing a wide range of auditory stimuli; the functional correlates of the leftward asymmetry, as well as the specific roles of the left and right PT remain poorly understood. In summary, in the present study, the left and right PT manifested the same general pattern of activation, suggesting similar function, but only the right PT was sensitive to the syllabic structure of the words that were perceived and produced. Further work is needed to better understand potential distinct contribution of the left and right PT to speech sound processing.

4.4 Summary and conclusion

Our results support the idea that the mechanisms for perception and production overlap at the level of the cerebral cortex, in both premotor and auditory areas of the frontal and

temporal lobes, suggesting bidirectional influence of sensory and motor systems on perception and production. However, our findings also reveal important context-related differences in the contribution of motor and sensory areas to the perception and production of speech.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Margaret Flynn, Blythe Buchholz, Anthony S. Dick and Michael Andric for their help collecting the data, and Matthew Schiel and Priya Santhanam for their help pre-processing the fMRI data. Thanks also to all participants. This study was supported by the National Institutes of Health under NIDCD grants R33 DC008638 and R01 DC003378 to S.L. Small, and by a postdoctoral fellowship from the Canadian Institute for Health research (CIHR) to P. Tremblay. Their support is gratefully acknowledged.

References

Argall BD, Saad ZS, Beauchamp MS. Simplified intersubject averaging on the cortical surface using SUMA. Human Brain Mapping. 2006; 27(1):14–27. [PubMed: 16035046]

Bement L, Wallber J, DeFilippo C, Bochner J, Garrison W. A new protocol for assessing viseme perception in sentence context: the lipreading discrimination test. Ear Hear. 1988; 9:33–40. [PubMed: 3342942]

Boë L, Ménard L, S J, Birkholz P, Badin P, Canault M. La croissance de l'instrument vocal: contrôle, modélisation, potentialités acoustiques et conséquences perceptives. Revue française de linguistique appliquée. 2008; XIII(2):59–80.

Bohland JW, Guenther FH. An fMRI investigation of syllable sequence production. Neuroimage. 2006; 32(2):821–841. [PubMed: 16730195]

Buchsbaum B, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for perception and production. Cognitive Science. 2001; 25:663–678.

Burton MW, Small SL, Blumstein SE. The role of segmentation in phonological processing: an fMRI investigation. J Cogn Neurosci. 2000; 12:679–690. [PubMed: 10936919]

Burton MW, Small SL. Functional neuroanatomy of segmenting speech and nonspeech. Cortex. 2006; 42(4):644–651. [PubMed: 16881272]

Callan AM, Callan DE, Tajima K, Akahane-Yamada R. Neural processes involved with perception of non-native durational contrasts. Neuroreport. 2006; 17(12):1353–1357. [PubMed: 16951584]

Callan DE, Jones JA, Callan AM, Akahane-Yamada R. 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. 2004; 22:1182–1194. [PubMed: 15219590]

Callan D, Callan A, Gamez M, Sato MA, Kawato M. Premotor cortex mediates perceptual performance. Neuroimage. 2010; 51:844–858. [PubMed: 20184959]

Dale AM, Fischl B, Sereno MI. Cortical surface-based Analysis: I. Segmentation and surface reconstruction. NeuroImage. 1999; 9(2):179–194. [PubMed: 9931268]

Desai R, Liebenthal E, Possing ET, Waldron E, Binder JR. Volumetric vs. surface-based alignment for localization of auditory cortex activation. NeuroImage. 2005; 26(4):1019–1029. [PubMed: 15893476]

Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage. 2006; 31(3):968–980. [PubMed: 16530430]

Dhanjal NS, Handunnetthi L, Patel MC, Wise RJ. Perceptual systems controlling speech production. J Neurosci. 2008; 28(40):9969–9975. [PubMed: 18829954]

di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. Understanding motor events: a neurophysiological study. Experimental Brain Research. 1992; 91(1):176–180.

Dick AS, Goldin-Meadow S, Hasson U, Skipper JI, Small SL. Co-speech gestures influence neural activity in brain regions associated with processing semantic information. Human Brain Mapping. 2009

Duvernoy, HM. The human brain: Structure, three-dimensional sectional anatomy and MRI. New York: Springer-Verlag; 1991.

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. Science. 1971; 171:303–306. [PubMed: 5538846]

Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. European Journal of Neuroscience. 2002; 15(2):399–402. [PubMed: 11849307]

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33(3):341–355. [PubMed: 11832223]

Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. NeuroImage. 1999; 9(2):195–207. [PubMed: 9931269]

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, et al. Automatically parcellating the human cerebral cortex. Cerebral Cortex. 2004; 14(1):11–22. [PubMed: 14654453]

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences. 2005; 102(27):9673–9678.

Fridriksson J, Moss J, Davis B, Baylis GC, Bonilha L, Rorden C. Motor speech perception modulates the cortical language areas. Neuroimage. 2008; 41(2):605–613. [PubMed: 18396063]

Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. Psychon Bull Rev. 2006; 13(3):361–377. [PubMed: 17048719]

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. Brain. 1996; 119(2):593–609. [PubMed: 8800951]

Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 2006; 96(3):280–301. [PubMed: 16040108]

Hasson U, Skipper JI, Nusbaum HC, Small SL. Abstract coding of audiovisual speech: beyond sensory representation. Neuron. 2007; 56(6):1116–1126. [PubMed: 18093531]

Hickok G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J Cogn Neurosci. 2009; 21(7):1229–1243. [PubMed: 19199415]

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci. 2003; 15(5):673–682. [PubMed: 12965041]

Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, et al. A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. Neurosci Lett. 2000; 287(2): 156–160. [PubMed: 10854735]

Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 2000; 4(4):131–138. [PubMed: 10740277]

Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition. 2004; 92(1-2):67–99. [PubMed: 15037127]

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8(5): 393–402. [PubMed: 17431404]

Huang B, Thangavelu M, Bhatt S, JS C, Wang S. Prenatal diagnosis of 45,X and 45,X mosaicism: the need for thorough cytogenetic and clinical evaluations. Prenatal Diagnosis. 2002; 22(2):105–110. [PubMed: 11857613]

Jackson PL. The theoretical minimal unit for visual speech perception: visemes and coarticulation. Volta Rev. 1988; 90(5):9–115.

Joanisse MF, Gati JS. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage. 2003; 19(1):64–79. [PubMed: 12781727]

Johnstone T, Ores Walsh KS, Greischar LL, Alexander AL, Fox AS, Davidson RJ, et al. Motion correction and the use of motion covariates in multiple-subject fMRI analysis. Human Brain Mapping. 2006; 27:779–788. [PubMed: 16456818]

Karbe H, Herholz K, Weber-Luxenburger G, Ghaemi M, Heiss WD. Cerebral networks and functional brain asymmetry: evidence from regional metabolic changes during word repetition. Brain and Language. 1998; 63(1):108–121. [PubMed: 9642023]

Kent RD. Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. J Speech Hear Res. 1976; 19(3):421–447. [PubMed: 979206]

Kent RD, Murray AD. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. J Acoust Soc Am. 1982; 72(2):353–365. [PubMed: 7119278]

Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G. Hearing sounds, understanding actions: action representation in mirror neurons. Science. 2002; 297(5582):846–848. [PubMed: 12161656]

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychological Review. 1967; 74(6):431–461. [PubMed: 4170865]

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21(1): 1–36. [PubMed: 4075760]

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Current Biology. 2007; 17:1692–1696. [PubMed: 17900904]

Moffitt AR. Consonant cue perception by twenty- to twenty-four-week-old infants. Child Dev. 1971; 42:717–731. [PubMed: 5121099]

Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the minimum statistic. Neuroimage. 2005; 25(3):653–660. [PubMed: 15808966]

Okada K, Smith KR, Humphries C, Hickok G. Word length modulates neural activity in auditory cortex during covert object naming. Neuroreport. 2003; 14(18):2323–2326. [PubMed: 14663184]

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971; 9(1):97–113. [PubMed: 5146491]

Pa J, Hickok G. A parietal-temporal sensory-motor integration area for the human vocal tract: evidence from an fMRI study of skilled musicians. Neuropsychologia. 2008; 46(1):362–368. [PubMed: 17709121]

Papathanassiou D, Etard O, Mellet E, Zago L, Mazoyer B, Tzourio-Mazoyer N. A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. Neuroimage. 2000; 11(4):347–357. [PubMed: 10725191]

Peeva MG, Guenther FH, Tourville JA, Nieto-Castanon A, Anton JL, Nazarian B, Alario FX. Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. Neuroimage. 2010; 50:626–638. [PubMed: 20035884]

Peschke C, Ziegler W, Kappes J, Baumgaertner A. Auditory-motor integration during fast repetition: the neuronal correlates of shadowing. Neuroimage. 2009; 47(1):392–402. [PubMed: 19345269]

Petrides M, Cadoret G, Mackey S. Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. Nature. 2005; 435(7046):1235–1238. [PubMed: 15988526]

Preminger JE, Lin HB, Payen M, Levitt H. Selective visual masking in speechreading. J Speech Lang Hear Res. 1998; 41(3):564–575. [PubMed: 9638922]

Pulvermuller F, Shtyrov Y, Ilmoniemi RJ, Marslen-Wilson WD. Tracking speech comprehension in space and time. Neuroimage. 2006; 31(3):1297–1305. [PubMed: 16556504]

Reisberg, D.; McLean, J.; Goldfield, A. Easy to hear but hard to understand: a lipreading advantage with intact auditory stimuli. In: D, B.; C, R., editors. Hearing by eye: the psychology of lipreading. Erlbaum; Hillsdale: 1987. p. 97-114.

Riecker A, Brendel B, Ziegler W, Erb M, Ackermann H. The influence of syllable onset complexity and syllable frequency on speech motor control. Brain Lang. 2008; 107(2):102–113. [PubMed: 18294683]

Rizzolatti G, Arbib MA. Language within our grasp. Trends in Neurosciences. 1998; 21:188–194. [PubMed: 9610880]

Rizzolatti G, Craighero L. The mirror-neuron system. Annual Review of Neuroscience. 2004; 27:169–192.

Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. Brain Research. Cognitive Brain Research. 1996; 3(2):131–141. [PubMed: 8713554]

Saito DN, Yoshimura K, Kochiyama T, Okada T, Honda M, Sadato N. Cross-modal binding and activated attentional networks during audiovisual speech integration: A function MRI study. Cerebral Cortex. 2005; 15:1750–1760. [PubMed: 15716468]

Sato M, Tremblay P, Gracco VL. A mediating role of the premotor cortex in phoneme segmentation. Brain and Language. 2009; 111(1):1–7. [PubMed: 19362734]

Schulz GM, Varga M, Jeffires K, Ludlow CL, Braun AR. Functional Neuroanatomy of Human Vocalization: An H215O PET Study. Cerebral Cortex. 2005 Epub ahead of print.

Schwartz J, Basirat A, M L, Sato M. The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. Journal of Neurolinguistics. 2010

Shergill SS, Brammer MJ, Fukuda R, Bullmore E, Amaro E Jr, Murray RM, et al. Modulation of activity in temporal cortex during generation of inner speech. Hum Brain Mapp. 2002; 16(4):219–227. [PubMed: 12112764]

Siok WT, Jin Z, Fletcher P, Tan LH. Distinct brain regions associated with syllable and phoneme. Human Brain Mapping. 2003; 18(3):201–207. [PubMed: 12599278]

Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. Neuroimage. 2005; 25(1):76–89. [PubMed: 15734345]

Skipper JI, van Wassenhove V, Nusbaum HC, Small SL. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cerebral Cortex. 2006a; 17:2387–2399. [PubMed: 17218482]

Skipper, JI.; Nusbaum, HC.; Small, SL. Lending a helping hand to hearing: Another motor theory of speech perception. In: Arbib, MA., editor. Action to language via the mirror neuron system. Cambridge, UK: Cambridge University Press; 2006b. p. 250-285.Vol

Sumby WH, Pollack I. Visual Contribution to Speech Intelligibility in Noise. J Acoust Soc Am. 1954; 26:212–215.

Sundara M, Namasivayam AK, Chen R. Observation-execution matching system for speech: a magnetic stimulation study. Neuroreport. 2001; 12(7):1341–1344. [PubMed: 11388407]

Trehub SE. Infants' sensitivity to vowel and tonal contrasts. Developmental Psychology. 1973; 9:91–96.

Trehub SE, Rabinovitch MS. Auditory-linguistic sensitivity in early infancy. Developmental Psychology. 1972; 6:74–77.

Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. Neuroimage. 2008; 39(3):1429–1443. [PubMed: 18035557]

Tzourio N, Nkanga-Ngila B, Mazoyer B. Left planum temporale surface correlates with functional dominance during story listening. Neuroreport. 1998; 9:829–833. [PubMed: 9579674]

Turella L, Pierno AC, Tubaldi F, Castiello U. Mirror neurons in humans: consisting or confounding evidence? Brain Lang. 2009; 108(1):10–21. [PubMed: 18082250]

Vihman MM, de Boysson-Bardies B. The nature and origins of ambient language influence on infant vocal production and early words. Phonetica. 1994; 51(1-3):159–169. [PubMed: 8052670]

Watkins K, Paus T. Modulation of motor excitability during speech perception: the role of Broca's area. Journal of Cognitive Neuroscience. 2004; 16(6):978–987. [PubMed: 15298785]

Watkins K, Strafella A, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 2003; 41(8):989–994. [PubMed: 12667534]

Wilson S, Iacoboni M. Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. Neuroimage. 2006; 33(1):316–325. [PubMed: 16919478]

Wilson SM, Iacoboni M. Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. Neuroimage. 2006; 33(1):316–325. [PubMed: 16919478]

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. Nature Neuroscience. 2004; 7(7):701–702.

Wise R, Chollet F, Hadar U, Friston K, Hoffner E, Frackowiak R. Distribution of cortical neural networks involved in word comprehension and word retrieval. Brain. 1991; 114(Pt 4):1803–1817. [PubMed: 1884179]

Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. Separate neural subsystems within ‘Wernicke's area’. Brain. 2001; 124(Pt 1):83–95. [PubMed: 11133789]

Wong PC, Uppunda AK, Parrish TB, Dhar S. Cortical mechanisms of speech perception in noise. J Speech Lang Hear Res. 2008; 51(4):1026–1041. [PubMed: 18658069]

Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. Science. 1992; 256(5058):846–849. [PubMed: 1589767]

Zheng ZZ, Munhall KG, Johnsrude IS. Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. J Cogn Neurosci. 2010; 22(8):1770–1781. [PubMed: 19642886]

Research highlights

- The mechanisms for speech perception and production overlap in the cortex.
- Overlap occurs in premotor and auditory areas of the frontal and temporal lobes.
- No motor region was sensitive to motor complexity during speech perception.
- The contribution of motor regions to speech perception and production is context-dependent.

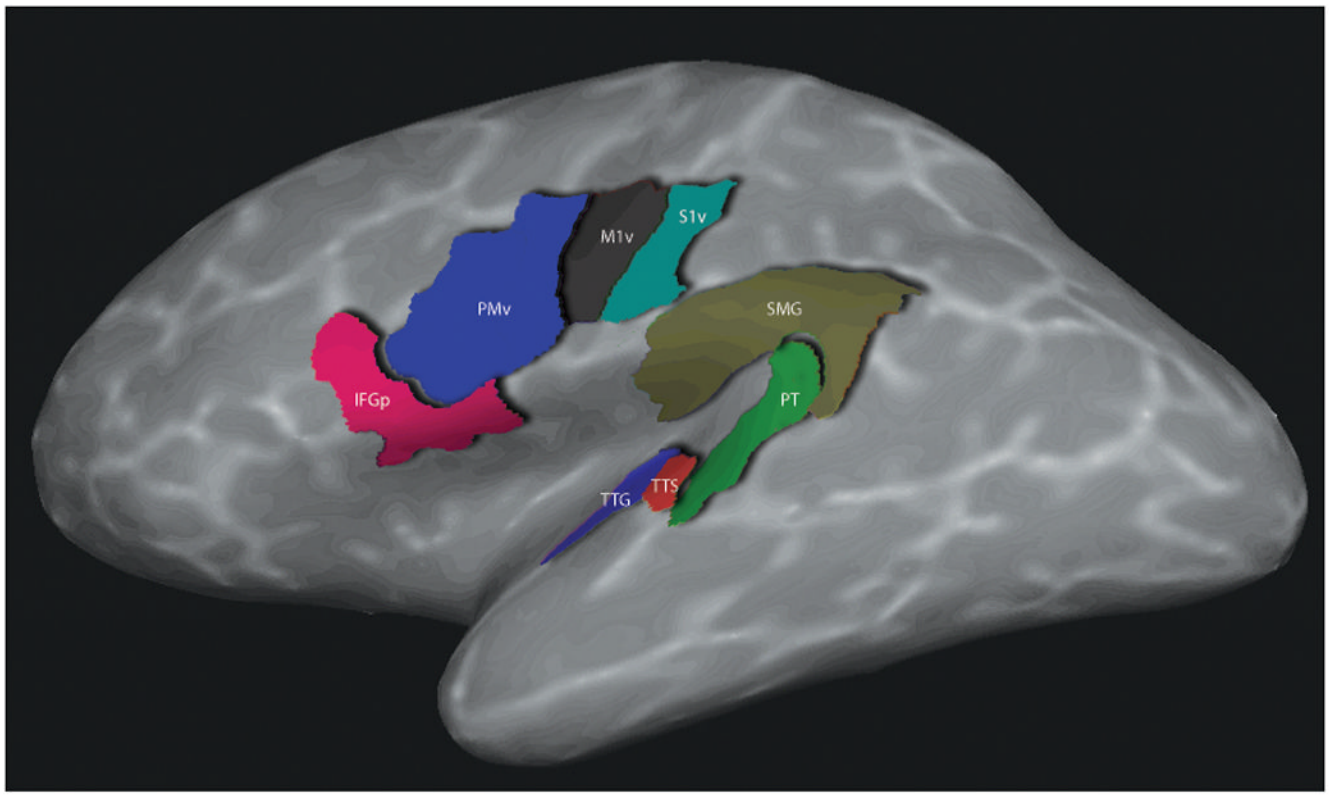

Figure 1.
Anatomical regions of interest (ROIs) in the current study, shown on the left surface hemisphere of one subject. IFGp = pars opercularis of the inferior frontal gyrus; PMv = ventral premotor cortex; M1v = ventral primary motor cortex; S1v = ventral primary somatosensory cortex; PT = planum temporale; TTG = transverse temporal gyrus; TTS = transverse temporal sulcus; SMG = supramarginal gyrus. Only the left ROIs are shown in the image, but all ROIs were defined bilaterally.

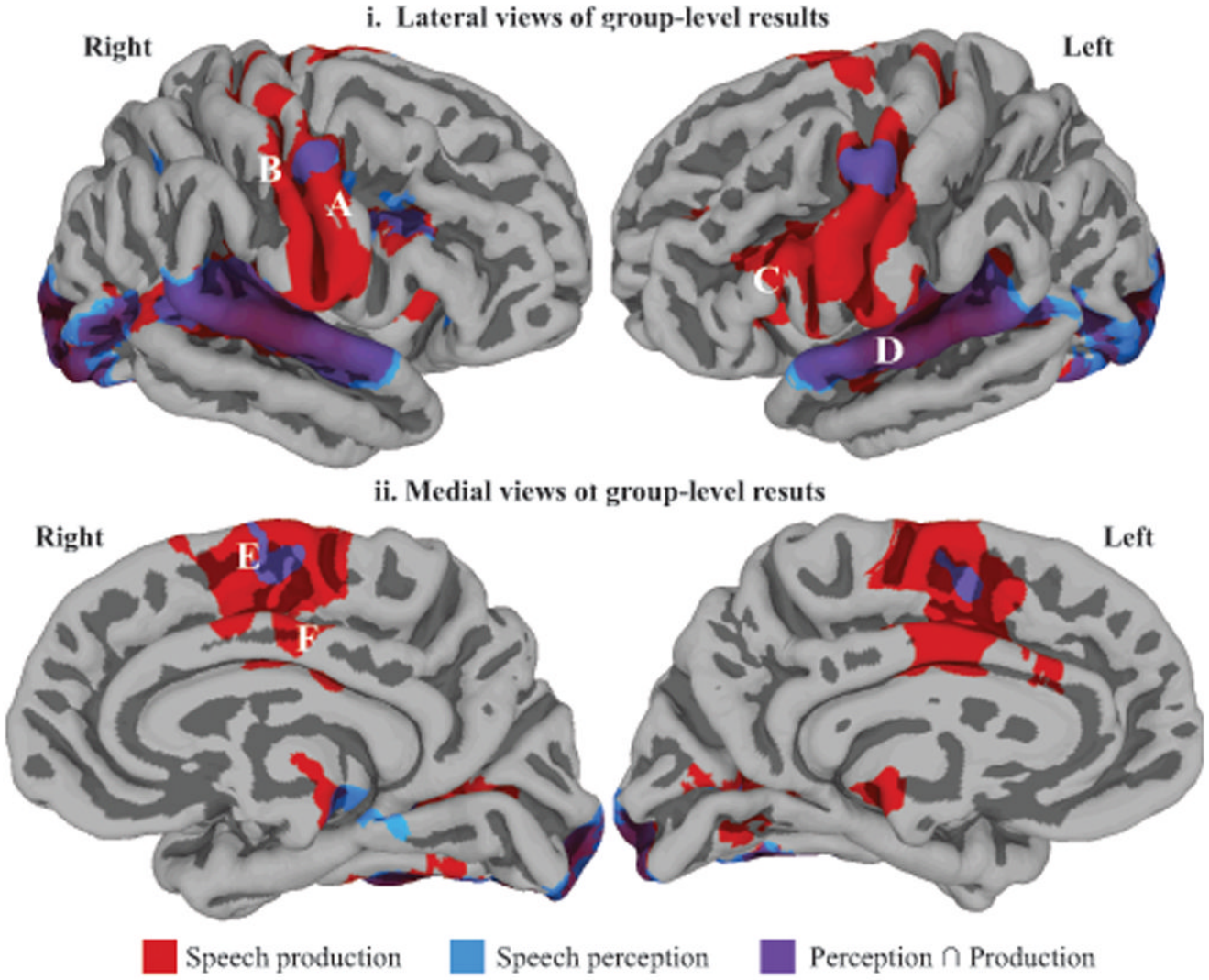

Figure 2.
Regions significantly active, at the group-level, for perception (in blue), production (in red) and for the conjunction of perception and production (in purple), shown on lateral (i) and medial views of the cerebral hemispheres (ii). Activation is shown on the group average smoothed white matter folded surface. Some relevant anatomical landmarks are identified on the brains: A = ventral precentral gyrus, B = ventral postcentral gyrus, C = pOFG, D = STG, E = medial frontal gyrus (pre-SMA), F = cingulate gyrus.

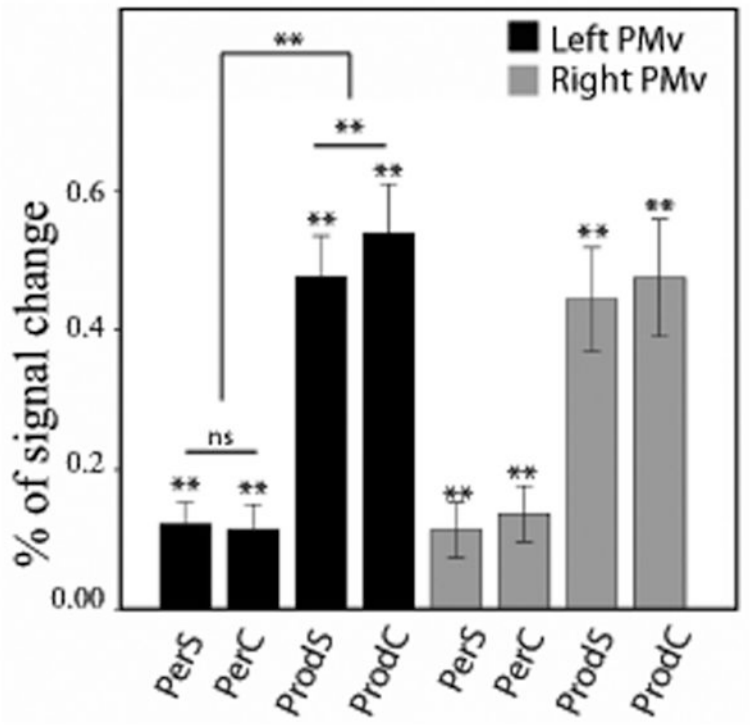

Figure 3.

Brain activity (expressed as a percentage of signal change) for speech perception and speech production, in the left (black bars) and right sPMv (gray bars). Double asterisks indicate a significant (FDR corrected, q = 0.05) difference (paired sample t-test), either between two conditions, or against zero (one-sample t-test). PerS = perception of Simple words; PerC = perception of complex words; ProdS = production of simple words; ProdC = production of complex words. The error bars represent the standard error (SE) of the mean.

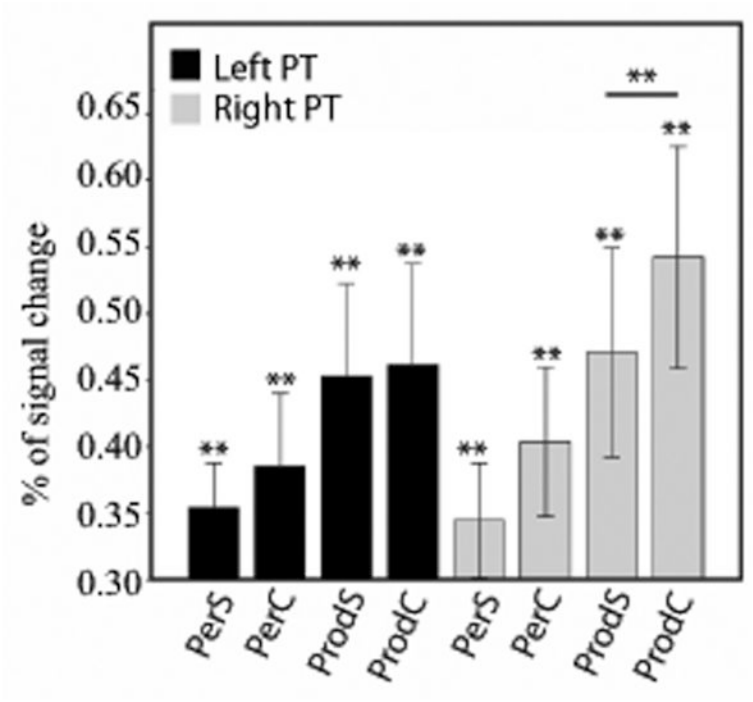

Figure 4.

Brain activity (expressed as a percentage of signal change) for speech perception and speech production, in the left (black bars) and right PT (gray bars). Double asterisks indicate a significant (FDR corrected, q = 0.05) difference, either between two conditions (paired sample t-test), or against zero (one-sample t-test). A single asterisk indicates uncorrected significance. PerS = perception of Simple words; PerC = perception of complex words; ProdS = production of simple words; ProdC = production of complex words. The error bars represent the standard error (SE) of the mean.

Table 1

FWE-corrected group-level (N = 20), whole brain results, for the contrast of Perception against rest (A), Production against rest (B), and the contrast of Production against Perception (C). Coordinates are in Talairach space and represent the peak surface node for each of the cluster (minimum cluster size: 196 contiguous surface nodes, each significant at $p < 0.005$).

Description	Hemi.	x	y	z	t	p	Nodes
A. Perception							
Transverse temporal gyrus, extending anteriorly and posteriorly into the superior temporal gyrus, and in the posterior temporal sulcus.		-43	-22	4	7.81	0.00000	9004
Occipital pole, extending into the posterior end of the calcarine sulcus medially, and ventrally into the fusiform gyrus		-18	-95	3	10.48	0.00000	8030
Precentral gyrus and sulcus (PMv)		-45	-7	45	5.64	0.00002	605
Anterior insula.	Left	-26	25	11	3.96	0.00084	522
Gyrus subcentral.		-46	-10	14	4.26	0.00042	601
Medial frontal gyrus (pre-SMA).		-8	8	53	4.42	0.00029	279
Body of the calcarine sulcus.		-15	-71	7	4.19	0.00050	249
Transverse temporal gyrus, extending anteriorly and posteriorly into the superior temporal gyrus, and in the posterior temporal sulcus.		49	-16	4	8.46	0.00000	10102
Occipital pole, extending into the posterior end of the calcarine sulcus medially, and ventrally into the fusiform gyrus		16	-91	-10	6.74	0.00000	9172
Inferior frontal sulcus, extending into the precentral sulcus.		34	19	27	3.54	0.00219	1126
Anterior insula.		28	28	8	4.57	0.00021	714
Anterior end of the calcarine sulcus.		21	-48	1	3.30	0.00377	694
Medial frontal gyrus (pre-SMA).	Right	8	3	57	4.37	0.00033	429
Ventral precentral sulcus (PMv)		33	2	30	3.45	0.00268	404
Gyrus subcentral.		54	-9	13	3.80	0.00121	309
Precentral gyrus (PMv)		47	-5	44	4.89	0.00010	258
Orbital gysus/anterior insula.		38	31	0	4.05	0.00068	226
Intraparietal sulcus.		24	-52	43	4.36	0.00034	205
B. Production							
Ventral central sulcus (M1), precentral and postcentral gyri, extending rostrally into the ventral precentral sulcus, and into the inferior frontal gyrus, pars opercularis.		-38	-11	30	7.96	0.00000	11098
Transverse temporal gyrus, extending anteriorly and posteriorly into the superior temporal gyrus, and in the temporal sulcus.		-43	-22	4	5.67	0.00002	10969
Medial frontal gyrus (pre-SMA, SMA-proper), extending ventrally into the cingulate sulcus and gyrus.	Left	-6	1	60	10.00	0.00000	6472
Occipital pole, extending into the posterior end of the calcarine sulcus medially.		-20	-89	2	6.68	0.00000	3964
Anterior insula.		-25	29	5	7.16	0.00000	2986

B. Production							
Fusiform gyrus.		-38	-62	-19	4.50	0.00025	1648
Dorsal central sulcus.		-16	-28	55	5.84	0.00001	1459
Anterior calcarine sulcus.		-23	-61	6	4.79	0.00013	1396
Posterior insula.		-30	-11	17	4.50	0.00025	404
Middle frontal sulcus.		-24	37	24	4.06	0.00067	317
Lingual gyrus.		-12	-70	-9	3.83	0.00113	244
Fusiform gyrus, occipital pole, extending into the posterior end of the calcarine sulcus medially, and rostrally ad dorsally into the transverse temporal gyrus, extending anteriorly and posteriorly into the superior temporal gyrus, and in the temporal sulcus.		35	-65	-15	4.63	0.00018	19536
Ventral central sulcus (M1), precentral and postcentral gyri, extending rostrally into the ventral precentral sulcus.		35	-11	34	7.07	0.00000	7618
Medial frontal gyrus (pre-SMA, SMA-proper), extending ventrally into the cingulate sulcus and gyrus.	Right	6	0	63	8.92	0.00000	6695
Anterior and body of the calcarine sulcus.		22	-50	1	3.61	0.00187	2135
Inferior frontal sulcus, extending into the inferior frontal gyrus pasr opercularis, and into the precentral sulcus (PMv).		33	12	25	4.05	0.00068	1341
Anterior insula		27	23	6	4.76	0.00014	1200
C. Production > Perception							
Ventral central sulcus (M1), precentral and postcentral gyri, extending rostrally into the ventral precentral sulcus, and into the inferior frontal gyrus, pars opercularis.		-37	-11	31	8.24	0.00000	6388
Medial frontal gyrus (pre-SMA, SMA-proper), extending ventrally into the cingulate sulcus and gyrus.		-6	1	61	8.23	0.00000	4192
Anterior insula.		-25	30	4	4.26	0.00042	1604
Dorsal central sulcus.		-15	-27	57	5.17	0.00005	1572
Medial transvers sulcus.		-32	-35	12	4.39	0.00031	678
Lingual gyrus.		-12	-66	-4	4.38	0.00032	542
Inferior frontal gyrus, pars opercularis.	Left	-45	21	19	3.38	0.00314	535
Superior frontal sulcus.		-12	37	49	4.44	0.00028	359
Precentral sulcus, inferior end.		-43	10	9	3.22	0.00451	305
Dorsal precentral sulcus.		-22	-8	44	3.18	0.00493	273
Posterior insula.		-31	-12	18	4.12	0.00058	270
Posterior cingulate sulcus.		-9	-17	37	3.84	0.00110	264
Intraparietal sulcus.		-20	-61	47	3.57	0.00204	246
Ventral central sulcus (M1), precentral and postcentral gyri.		35	-12	34	6.50	0.00000	5199
Dorsal central sulcus.		18	-29	5	6.37	0.00000	1819
Medial frontal gyrus (pre-SMA, SMA-proper), extending ventrally into the cingulate sulcus.	Right	6	-1	63	7.37	0.00000	1723
Cingulate sulcus, extending drsally into the medial frontal gyrus (pre-SMA)		10	12	35	4.92	0.00010	900

C. Production > Perception						
Body of the calcarine sulcus, extending into the parieto-occipital sulcus.	22	-63	7	3.21	0.00461	477
Transverse temporal gyrus.	42	-23	7	3.61	0.00187	317
Intraparietal sulcus.	21	-39	52	4.09	0.00062	298
Dorsal precentral sulcus.	16	-17	63	4.41	0.00030	230
Anterior insula.	31	12	2	4.10	0.00061	217

Table 2

FWE-corrected group-level (N = 20) conjunction analyses overlapping activity in Perception Simple, Perception Complex, Production Simple and Production Complex, yielding an intersection map of Perception ∩ Production (minimum cluster size: 196 contiguous surface nodes, each significant at $p < 0.005$).

Description	Hemi.	x	y	z	Nodes
Transverse temporal gyrus, extending anteriorly and posteriorly into the superior temporal gyrus, and in the posterior temporal sulcus.	Left	-47	-32	9	7652
Occipital pole, extending into the posterior end of the calcarine sulcus medially.		-23	-96	-8	1924
Fusiform gyrus.		-37	-59	-18	930
Occipital sulcus, anterior part.		-39	-66	4	707
Precentral gyrus and sulcus (PMv)		-45	-4	47	419
Occipital gyrus.		-37	-79	-12	223
Transverse temporal gyrus, extending anteriorly and posteriorly into the superior temporal gyrus, and in the posterior temporal sulcus.	Right	49	-26	3	8853
Occipital pole, extending into the posterior end of the calcarine sulcus medially.		31	-82	-1	4345
Fusiform gyrus.		36	-56	-16	1339
Inferior frontal sulcus.		40	19	22	333