Check for updates

# *T'ain't What You Say, It's the Way That You Say It*—Left Insula and Inferior Frontal Cortex Work in Interaction with Superior Temporal Regions to Control the Performance of Vocal Impersonations

Carolyn McGettigan[1,2], Frank Eisner[3], Zarinah K. Agnew[1], Tom Manly[4], Duncan Wisbey[1], and Sophie K. Scott[1]

## Abstract

■ Historically, the study of human identity perception has focused on faces, but the voice is also central to our expressions and experiences of identity [Belin, P., Fecteau, S., & Bedard, C. Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129–135, 2004]. Our voices are highly flexible and dynamic; talkers speak differently, depending on their health, emotional state, and the social setting, as well as extrinsic factors such as background noise. However, to date, there have been no studies of the neural correlates of identity modulation in speech production. In the current fMRI experiment, we measured the neural activity supporting controlled voice change in adult participants performing spoken impressions. We reveal that deliberate modulation of vocal identity recruits the left anterior insula and inferior frontal gyrus, supporting the planning of novel articulations. Bilateral sites in posterior superior temporal/inferior parietal cortex and a region in right middle/anterior STS showed greater responses during the emulation of specific vocal identities than for impressions of generic accents. Using functional connectivity analyses, we describe roles for these three sites in their interactions with the brain regions supporting speech planning and production. Our findings mark a significant step toward understanding the neural control of vocal identity, with wider implications for the cognitive control of voluntary motor acts. ■

## INTRODUCTION

Voices, like faces, express many aspects of our identity (Belin, Fecteau, & Bedard, 2004). From hearing only a few words of an utterance, we can estimate the speaker's gender and age, their country or even specific town of birth, as well as more subtle evaluations on current mood or state of health (Karpf, 2007). Some of the indexical cues to speaker identity are clearly expressed in the voice. The pitch (or fundamental frequency, F0) of the voice of an adult male speaker tends to be lower than that of adult women or children, because of the thickening and lengthening of the vocal folds during puberty in human men. The secondary descent of the larynx in adult men also increases the spectral range in the voice, reflecting an increase in vocal tract length.

However, the human voice is also highly flexible, and we continually modulate the way we speak. The Lombard effect (Lombard, 1911) describes the way that talkers automatically raise the volume of their voice when the auditory environment is perceived as noisy. In the social context of conversations, interlocutors start to align their behaviors, from body movements and breathing patterns to pronunciations and selection of syntactic structures (Pardo, 2006; Garrod & Pickering, 2004; McFarland, 2001; Chartrand & Bargh, 1999; Condon & Ogston, 1967). Laboratory tests of speech shadowing, where participants repeat speech immediately as they hear it, have shown evidence for unconscious imitation of linguistic and paralinguistic properties of speech (Kappes, Baumgaertner, Peschke, & Ziegler, 2009; Shockley, Sabadini, & Fowler, 2004; Bailly, 2003). Giles and colleagues (Giles, Coupland, & Coupland, 1991; Giles, 1973) put forward the Communication Accommodation Theory to account for processes of convergence and divergence in spoken language pronunciation—namely, they suggest that talkers change their speaking style to modulate the social distance between them and their interlocutors, with convergence promoting greater closeness. It has been argued by others that covert speech imitation is central to facilitating comprehension in conversation (Pickering & Garrod, 2007). Aside from these short-term modulations in speech, changes in vocal behavior can also be observed over much longer periods—the speech of Queen Elizabeth II has shown a gradual progression toward standard southern

[1]Institute of Cognitive Neuroscience, University College London, [2]Department of Psychology, Royal Holloway, University of London, [3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, [4]MRC Cognition and Brain Sciences Unit, Cambridge, UK

© 2013 Massachusetts Institute of Technology Published under a Creative Commons Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0) license.

*Journal of Cognitive Neuroscience 25:11, pp. 1875–1886*
*doi:10.1162/jocn_a_00427*

Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

British pronunciation (Harrington, Palethorpe, & Watson, 2000).

Although modulations of the voice often occur outside conscious awareness, they can also be deliberate. A recent study showed that student participants could change their speech to sound more masculine or feminine, by making controlled alterations that simulated target-appropriate changes in vocal tract length and voice pitch (Cartei, Cowles, & Reby, 2012). Indeed, speakers can readily disguise their vocal identity (Sullivan & Schlichting, 1998), which makes forensic voice identification notoriously difficult (Eriksson et al., 2010; Ladefoged, 2003). Notably, when control of vocal identity is compromised, for example, in Foreign Accent Syndrome (e.g., Scott, Clegg, Rudge, & Burgess, 2006), the change in the patient's vocal expression of identity can be frustrating and debilitating. Interrogating the neural systems supporting vocal modulation is an important step in understanding human vocal expression, yet this dynamic aspect of the voice is a missing element in existing models of speech production (Hickok, 2012; Tourville & Guenther, 2011).

Speaking aloud is an example of a very well practised voluntary motor act (Jurgens, 2002). Voluntary actions need to be controlled in a flexible manner to adjust to changes in environment and the goals of the actor. The main purpose of speech is to perform the transfer of a linguistic/conceptual message. However, we control our voices to achieve intended goals on a variety of levels, from acoustic–phonetic accommodation to the auditory environment (Cooke & Lu, 2010; Lu & Cooke, 2009) to socially motivated vocal behaviors reflecting how we wish to be perceived by others (Pardo, Gibbons, Suppes, & Krauss, 2012; Pardo & Jay, 2010). Investigations of the cortical control of vocalization have identified two neurological systems supporting the voluntary initiation of innate and learned vocal behaviors, where expressions such as emotional vocalizations are controlled by a medial frontal system involving the ACC and SMA, whereas speech and song are under the control of lateral motor cortices (Jurgens, 2002). Thus, patients with speech production deficits following strokes to lateral inferior motor structures still exhibit spontaneous vocal behaviors such as laughter, crying, and swearing, despite their severe deficits in voluntary speech production (Groswasser, Korn, Groswasser-Reider, & Solzi, 1988). Electrical stimulation studies show that vocalizations can be elicited by direct stimulation of the anterior cingulate (e.g., laughter; described by Sem-Jacobsen & Torkildsen, 1960) and lesion evidence shows that bilateral damage to anterior cingulate prevents the expression of emotional inflection in speech (Jurgens & von Cramon, 1982).

In healthy participants, a detailed investigation of the lateral motor areas involved in voluntary speech production directly compared voluntary inhalation/exhalation with syllable repetition. The study found that the functional networks associated with laryngeal motor cortex were strongly left-lateralized for syllable repetition but bilaterally organized for controlled breathing (Simonyan, Ostuni, Ludlow, & Horwitz, 2009). However, that design did not permit further exploration of the modulation of voluntary control within either speech or breathing. This aspect has been addressed in a study of speech prosody, which reported activations in left inferior frontal gyrus (IFG) and dorsal premotor cortex for the voluntary modulation of both linguistic and emotional prosody, that overlapped with regions sensitive to the perception of these modulations (Aziz-Zadeh, Sheng, & Gheytanchi, 2010).

Some studies have addressed the neural correlates of overt and unintended imitation of heard speech (Reiterer et al., 2011; Peschke, Ziegler, Kappes, & Baumgaertner, 2009). Peschke and colleagues found evidence for unconscious imitation of speech duration and F0 in a shadowing task in fMRI, in which activation in right inferior parietal cortex correlated with stronger imitation of duration across participants. Reiterer and colleagues (2011) found that participants with poor ability to imitate non-native speech showed greater activation (and lower gray matter density) in left premotor, inferior frontal, and inferior parietal cortical regions during a speech imitation task, compared with participants who were highly rated mimics. The authors interpret this as a possible index of greater effort in the phonological loop for less skilled imitators. However, in general, the reported functional imaging investigations of voluntary speech control systems have typically involved comparisons of speech outputs with varying linguistic content, for example, connected speech of different linguistic complexities (Dhanjal, Handunnetthi, Patel, & Wise, 2008; Blank, Scott, Murphy, Warburton, & Wise, 2002) or pseudowords of varying length and phonetic complexity (Papoutsi et al., 2009; Bohland & Guenther, 2006).

To address the ubiquitous behavior of voluntary modulation of vocal expression in speech, while holding the linguistic content of the utterance constant, we carried out an fMRI experiment in which we studied the neural correlates of controlled voice change in adult speakers of English performing spoken impressions. The participants, who were not professional voice artists or impressionists, repeatedly recited the opening lines of a familiar nursery rhyme under three different speaking conditions: normal voice (N), impersonating individuals (I), and impersonating regional and foreign accents of English (A). The nature of the task is similar to the kinds of vocal imitation used in everyday conversation, for example, in reporting the speech of others during storytelling. We aimed to uncover the neural systems supporting changes in the way speech is articulated, in the presence of unvarying linguistic content. We predicted that left-dominant orofacial motor control centers, including the left IFG, insula, and motor cortex, as well as auditory processing sites in superior temporal cortex, would be important in effecting change to speaking style and monitoring the auditory consequences. Beyond this, we aimed to measure whether the goal of



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

the vocal modulation—to imitate a generic speaking style/accent versus a specific vocal identity—would modulate the activation of the speech production network and/or its connectivity with brain regions processing information relevant to individual identities.

## METHODS

### Participants

Twenty-three adult speakers of English (seven women; mean age = 33 years 11 months) were recruited who were willing to attempt spoken impersonations. All had healthy hearing and no history of neurological incidents nor any problems with speech or language (self-reported). Although some had formal training in acting and music, none had worked professionally as an impressionist or voice artist. The study was approved by the University College London Department of Psychology Ethics Committee.

### Design and Procedure

Participants were asked to compile in advance lists of 40 individuals and 40 accents they could feasibly attempt to impersonate. These could include any voice/accent with which they were personally familiar, from celebrities to family members (e.g., "Sean Connery," "Carly's Mum"). Likewise, the selected accents could be general or specific (e.g., "French" vs. "Blackburn").

Functional imaging data were acquired on a Siemens Avanto 1.5-T scanner (Siemens AG, Erlangen, Germany) in a single run of 163 echo-planar whole-brain volumes (repetition time = 8 sec, acquisition time = 3 sec, echo time = 50 msec, flip angle = 90°, 35 axial slices, 3 mm × 3 mm × 3 mm in-plane resolution). A sparse-sampling routine (Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999) was employed, with the task performed during a 5-sec silence between volumes.

There were 40 trials of each condition: normal voice (N), impersonating individuals (I), impressions of regional and foreign accents of English (A), and a rest baseline (B). The mean list lengths across participants were 36.1 ($SD$ = 5.6) for condition I and 35.0 ($SD$ = 6.9) for A (a nonsignificant difference; $t(1, 22) = .795$, $p = .435$). When submitted lists were shorter than 40, some names/accents were repeated to fill the 40 trials. Condition order was pseudorandomized, with each condition occurring once in every four trials. Participants wore electrodynamic headphones fitted with an optical microphone (MR Confon GmbH, Magdeburg, Germany). Using MATLAB (Mathworks, Inc., Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997) and a video projector (Eiki International, Inc., Rancho Santa Margarita, CA), visual prompts ("Normal Voice," "Break" or the name of a voice/accent, as well as a "Start speaking" instruction) were delivered onto a front screen, viewed via a mirror on the head coil. Each trial began with a condition prompt triggered by the onset of a whole-brain acquisition. At 0.2 sec after the start of the silent period, the participant was prompted to start speaking and to cease when the prompt disappeared (3.8 sec later). In each speech production trial, participants recited the opening line from a familiar nursery rhyme, such as "Jack and Jill went up the hill," and were reminded that they should not include person-specific catchphrases or catchwords. This controlled for the linguistic content of the speech across the conditions. Spoken responses were recorded using Audacity (audacity.sourceforge.net). After the functional run, a high-resolution T1-weighted anatomical image was acquired (HIRes MP-RAGE, 160 sagittal slices, voxel size = 1 $mm^3$). The total time in the scanner was around 35 min.

### Acoustic Analysis of Spoken Impressions

Because of technical problems, auditory recordings were only available for 13 participants. The 40 tokens from the three speech conditions—Normal Voice, Impersonations, and Accents—was entered into a repeated-measures ANOVA with Condition as a within-subject factor for each of the following acoustic parameters: (i) duration (sec), (ii) intensity (dB), (iii) mean F0 (Hz), (iv) minimum F0 (Hz), (v) maximum F0 (Hz), standard deviation of F0 (Hz), (vi) spectral center of gravity (Hz), and (vii) spectral standard deviation (Hz). Three Bonferroni-corrected post hoc paired $t$ tests compared the individual conditions. Table 1 illustrates the results of these analyses, and Figure 1 illustrates the acoustic properties of example trials from each speech condition (taken from the same participant).

### fMRI Analysis

Data were preprocessed and analyzed using SPM5 (Wellcome Trust Centre for Neuroimaging, London, UK). Functional images were realigned and unwarped, coregistered with the anatomical image, normalized using parameters obtained from unified segmentation of the anatomical image, and smoothed using a Gaussian kernel of 8 mm FWHM. At the first level, the condition onsets were modeled as instantaneous events coincident with the prompt to speak, using a canonical hemodynamic response function. Contrast images were calculated to describe each of the four conditions (N, I, A and B), each speech condition compared with rest (N > B, I > B, A > B), each impression condition compared with normal speech (I > N, A > N), and the comparison of impression conditions (I > A). These images were entered into second-level, one-sample $t$ tests for the group analyses.

The results of the conjunction analyses are reported at a voxel height threshold of $p < .05$ (corrected for family-wise error). All other results are reported at an uncorrected voxel height threshold of $p < .001$, with a cluster extent correction of 20 voxels applied for a whole-brain $\alpha$ of $p < .001$ using a Monte Carlo simulation (with 10,000



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

**Table 1.** Acoustic Correlates of Voice Change during Spoken Impressions

| Acoustic Parameter | Mean Normal | Mean Voices | Mean Accents | ANOVA | | t Test N vs. V | | t Test N vs. A | | t Test V vs. A | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | *F* | *Sig.* | *t* | *Sig.* | *t* | *Sig.* | *t* | *Sig.* |
| Duration (sec) | 2.75 | 3.10 | 2.98 | 9.96 | **.006** | 3.25 | **.021** | 3.18 | **.024** | 2.51 | .081 |
| Intensity (dB) | 47.4 | 51.3 | 51.3 | 49.25 | **.000** | 10.15 | **.000** | 7.62 | **.000** | 0.88 | 1.00 |
| Mean F0 (Hz) | 155.9 | 207.2 | 186.3 | 24.11 | **.000** | 5.19 | **.001** | 4.87 | **.001** | 3.89 | **.006** |
| Min F0 (Hz) | 94.4 | 104.9 | 102.1 | 3.71 | **.039** | 2.20 | .144 | 2.18 | .149 | 0.77 | 1.00 |
| Max F0 (Hz) | 625.0 | 667.6 | 628.5 | 1.28 | .295 | 1.31 | .646 | 0.10 | 1.00 | 2.15 | .158 |
| *SD* F0 (Hz) | 117.3 | 129.9 | 114.7 | 1.62 | .227 | 1.26 | .694 | .240 | 1.00 | 3.30 | **.019** |
| Spec CoG (Hz) | 2100 | 2140 | 2061 | 0.38 | .617 | 0.37 | 1.00 | 0.39 | 1.00 | 1.49 | .485 |
| Spec *SD* (Hz) | 1647 | 1579 | 1553 | 2.24 | .128 | 1.17 | .789 | 2.05 | .188 | 0.89 | 1.00 |

F0 = fundamental frequency, *SD* = standard deviation, Spec = spectral, CoG = center of gravity. Significance levels are Bonferroni-corrected (see Methods), with significant effects shown in **bold**.

iterations) implemented in MATLAB (Slotnick, Moo, Segal, & Hart, 2003).

Conjunction analyses of second-level contrast images were performed using the null conjunction approach (Nichols, Brett, Andersson, Wager, & Poline, 2005). Using the MarsBaR toolbox (Brett, Anton, Valabregue, & Poline, 2002), spherical ROIs (4 mm radius) were built around the peak voxels—parameter estimates were extracted from these ROIs to construct plots of activation.

A psychophysiological interaction (PPI) analysis was used to investigate changes in connectivity between the conditions I and A. In each participant, the time course of activation was extracted from spherical volumes of interest (4 mm radius) built around the superior temporal peaks in the group contrast I > A (right middle/anterior STS: [54 −3 −15], right posterior STS: [57 −36 12], left posterior STS: [−45 −60 15]). A PPI regressor described the interaction between each volume of interest and a psychological regressor for the contrast of interest (I > A)—this modeled a change in the correlation between activity in these STS seed regions and the rest of the brain across the two conditions. The PPIs from each seed region were evaluated in a first-level model that included the individual physiological and psychological time courses as covariates of no interest. A random-effects, one-sample *t* test assessed the significance of each PPI in the group (voxelwise threshold: $p < .001$, corrected cluster threshold: $p < .001$).

Post hoc pairwise *t* tests using SPSS (version 18.0; IBM, Armonk, NY) compared condition-specific parameter estimates (N vs. B and I vs. A) within the peak voxels in the voice change conjunction ((I > N) ∩ (A > N)). To maintain independence and avoid statistical "double-dipping," an iterative, hold-one-out approach was used in which the peak voxels for each participant were defined from a group statistical map of the conjunction ((I > N) ∩ (A > N)) using the other 22 participants. These subject-specific peak locations were used to extract condition-specific parameter estimates from 4-mm spherical ROIs built around the peak voxel (using MarsBaR). Paired *t* tests were run using a corrected α level of .025 (to correct for two tests in each ROI).

The anatomical locations of peak and subpeak voxels (at least 8 mm apart) were labeled using the SPM Anatomy Toolbox (version 18; Eickhoff et al., 2005).

## RESULTS AND DISCUSSION

### Brain Regions Supporting Voice Change

Areas of activation common to the three speech output conditions compared with a rest baseline (B; (N > B) ∩ (I > B) ∩ (A > B)) comprised a speech production network of bilateral motor and somatosensory cortex, SMA,

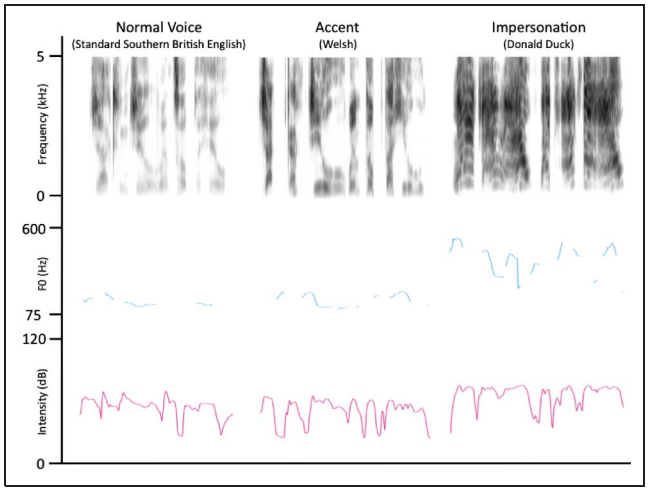



**Figure 1.** Examples of the phrase "Jack and Jill went up the hill" spoken by a single participant in the conditions Normal Voice, Accents, and Impersonations. Top: A spectrogram of frequency against time (where darker shading indicates greater intensity). Middle: The fundamental frequency (F0) profile across each utterance. Bottom: The intensity contour in decibels.



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

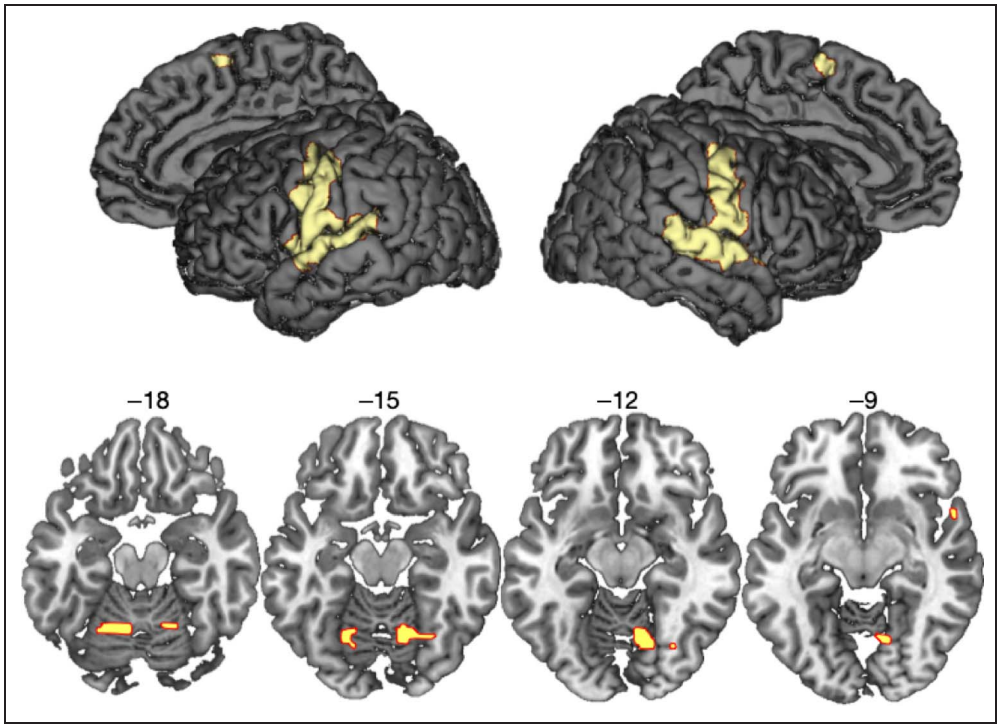



**Figure 2.** Activations common to the three speech conditions (Normal, Impersonations, and Accents) compared with a rest baseline (voxel height threshold $p < .05$, FWE-corrected). Numbers indicate the $z$ coordinate in Montreal Neurological Institute (MNI) stereotactic space.

superior temporal gyrus (STG), and cerebellum (Figure 2 and Table 2; Simmonds, Wise, Dhanjal, & Leech, 2011; Tourville & Guenther, 2011; Tourville, Reilly, & Guenther, 2008; Bohland & Guenther, 2006; Riecker et al., 2005; Blank et al., 2002; Wise, Greene, Buchel, & Scott, 1999). Activation common to the voice change conditions (I and A) compared with normal speech ((I > N) ∩ (A > N)) was found in left anterior insula, extending laterally onto the IFG (orbital and opercular parts) and on the right STG (Figure 3 and Table 3). Planned post hoc comparisons showed that responses in the left frontal sites were equivalent for impersonations and accents (two-tailed, paired $t$ test; $t(22) = -0.068$, corrected $p = 1.00$) and during normal speech and rest ($t(22) = 0.278$, corrected

**Table 2.** Activation Common to the Three Speech Output Conditions

| *Contrast* | *No. of Voxels* | *Region* | *Coordinate* | | | *t* | *z* |
|---|---|---|---|---|---|---|---|
| | | | *x* | *y* | *z* | | |
| All Speech > Rest ((N > B) ∩ (I > B) ∩ (A > B)) | 963 | Left postcentral gyrus/STG/precentral gyrus | −48 | −15 | 39 | 14.15 | 7.07 |
| | 852 | Right STG/precentral gyrus/postcentral gyrus | 63 | −15 | 3 | 13.60 | 6.96 |
| | 21 | Left cerebellum (lobule VI) | −24 | −60 | −18 | 7.88 | 5.38 |
| | 20 | Left SMA | −3 | −3 | 63 | 7.77 | 5.34 |
| | 34 | Right cerebellum (lobule VI), right fusiform gyrus | 12 | −60 | −15 | 7.44 | 5.21 |
| | 35 | Right/left calcarine gyrus | 3 | −93 | 6 | 7.41 | 5.19 |
| | 5 | Left calcarine gyrus | −15 | −93 | −3 | 6.98 | 5.02 |
| | 7 | Right lingual gyrus | 15 | −84 | −3 | 6.73 | 4.91 |
| | 1 | Right area V4 | 30 | −69 | −12 | 6.58 | 4.84 |
| | 3 | Left calcarine gyrus | −9 | −81 | 0 | 6.17 | 4.65 |
| | 2 | Left thalamus | −12 | −24 | −3 | 6.15 | 4.64 |
| | 2 | Right calcarine gyrus | 15 | −69 | 12 | 6.13 | 4.63 |

Conjunction null analysis of all speech conditions (Normal, Impersonations, and Accents) compared with rest. Voxel height threshold $p < .05$ (FWE-corrected). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space.



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

$p = 1.00$). The right STG, in contrast, was significantly more active during impersonations than accents (two-tailed, paired $t$ test; $t(22) = 2.69$, Bonferroni-corrected $p = .027$) and during normal speech compared with rest ($t(22) = 6.64$, corrected $p < .0001$). Thus, we demonstrate a partial dissociation of the inferior frontal/insular and sensory cortices, where both respond more during impressions than in normal speech, but where the STG shows an additional sensitivity to the nature of the voice change task—that is, whether the voice target is associated with a unique identity.

Acoustic analyses of the impressions from a subset of participants ($n = 13$) indicated that the conditions involving voice change resulted in acoustic speech signals that were significantly longer, more intense, and higher in fundamental frequency (roughly equivalent to pitch) than normal speech. This may relate to the right-lateralized temporal response during voice change, as previous work has shown that the right STG is engaged during judgments of sound intensity (Belin et al., 1998). The right temporal lobe has also been associated with processing nonlinguistic information in the voice, such as speaker identity (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005; Kriegstein & Giraud, 2004; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) and emotion (Schirmer & Kotz, 2006; Meyer, Zysset, von Cramon, & Alter, 2005; Wildgruber et al., 2005), although these results tend to implicate higher-order regions such as the STS.

The neuropsychology literature has described the importance of the left IFG and anterior insula in voluntary speech production (Kurth, Zilles, Fox, Laird, & Eickhoff, 2010; Dronkers, 1996; Broca, 1861). Studies of speech production have identified that the left posterior IFG and insula are sensitive to increasing articulatory complexity of spoken syllables (Riecker, Brendel, Ziegler, Erb, & Ackermann, 2008; Bohland & Guenther, 2006), but not to the frequency with which those syllables occur in everyday language (Riecker et al., 2008), suggesting involvement in the phonetic aspects of speech output rather than higher-order linguistic representations. Ackermann and Riecker (2010) suggest that insula cortex may actually be associated with more generalized control of breathing, which could be voluntarily modulated to maintain the sustained and finely controlled hyperventilation required to produce connected speech. In finding that the left IFG and insula can influence the way we speak, as well as what we say, we have also shown that they are not just coding abstract linguistic

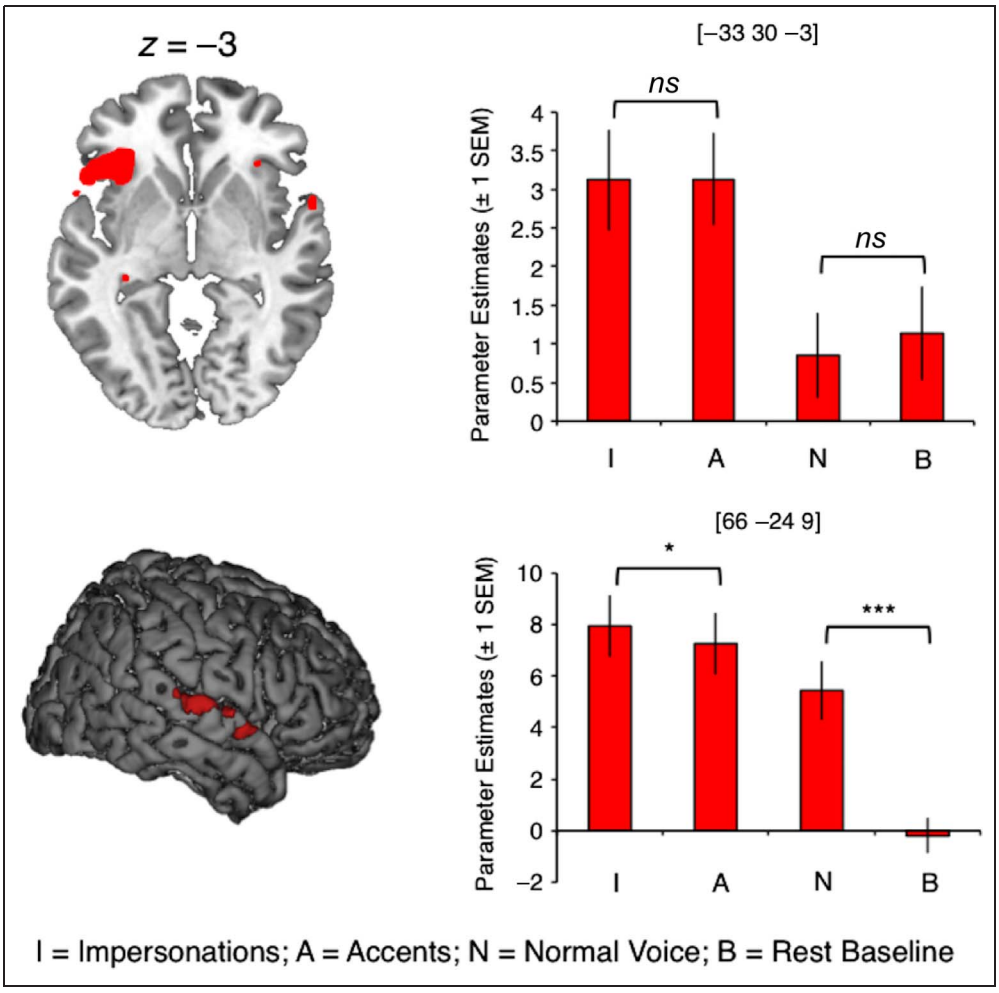



**Figure 3.** Brain regions supporting voice change. Bar plots show parameter estimates extracted from spherical ROIs centered on peak voxels. Annotations show the results of planned paired-sample $t$ tests (two-tailed, with Bonferroni correction; $*p < .05$, $***p < .0001$, $ns$ = nonsignificant). Coordinates are in MNI space.



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

**Table 3.** Neural Regions Recruited during Voice Change (Null Conjunction of Impersonations > Normal and Accents > Normal)

| Contrast | No. of Voxels | Region | Coordinate | | | t | z |
|---|---|---|---|---|---|---|---|
| | | | x | y | z | | |
| Impressions > Normal Speech ((I > N) ∩ (A > N)) | 180 | LIFG (pars orb., pars operc.)/insula | −33 | 30 | −3 | 8.39 | 5.56 |
| | 1 | Left temporal pole | −54 | 15 | −9 | 7.48 | 5.22 |
| | 19 | Right thalamus | 3 | −6 | 9 | 7.44 | 5.21 |
| | 17 | Right STG | 66 | −24 | 9 | 7.30 | 5.15 |
| | 16 | Right hippocampus | 33 | −45 | 3 | 7.17 | 5.10 |
| | 4 | Left thalamus | −12 | −6 | 12 | 7.11 | 5.07 |
| | 9 | Left thalamus | −27 | −21 | −9 | 6.80 | 4.94 |
| | 3 | Left hippocampus | −15 | −21 | −15 | 6.65 | 4.87 |
| | 6 | Right insula | 33 | 27 | 0 | 6.59 | 4.85 |
| | 1 | Right STG | 63 | −3 | 3 | 6.45 | 4.78 |
| | 1 | Left hippocampus | −24 | −39 | 9 | 6.44 | 4.78 |
| | 2 | Right STG | 66 | −9 | 6 | 6.44 | 4.78 |
| | 4 | Right temporal pole | 60 | 6 | −6 | 6.42 | 4.77 |
| | 1 | Left hippocampus | −15 | −42 | 12 | 6.30 | 4.71 |
| | 4 | Right caudate nucleus | 21 | 12 | 18 | 6.20 | 4.66 |
| | 2 | Left cerebellum (lobule VI) | −24 | −60 | −18 | 6.10 | 4.62 |

Voxel height threshold $p < .05$ (FWE-error corrected). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space. LIFG = left IFG; pars orb. = pars orbitalis; pars operc. = pars opercularis.

elements of the speech act. In agreement with Ackermann and Riecker (2010), we suggest that these regions may also play a role in more general aspects of voluntary vocal control during speech, such as breathing and modulation of pitch. In line with this, our acoustic analysis shows that both accents and impressions were produced with longer durations, higher pitches, and greater intensity, all of which are strongly dependent on the way that breathing is controlled (MacLarnon & Hewitt, 1999, 2004).

### Effects of Target Specificity: Impersonations versus Accents

A direct comparison of the two voice change conditions (I > A) showed increased activation for specific impersonations in right middle/anterior STS, bilateral posterior STS extending to angular gyrus (AG) on the left, and posterior midline sites on cingulate cortex and precuneus (Figure 4 and Table 4; the contrast A > I gave no significant activations). Whole-brain analyses of functional connectivity revealed areas that correlated more positively with the three sites on STS during impersonations than during accents (Figure 5 and Table 5). Strikingly, all three temporal seed regions showed significant interactions with areas typically active during speech perception and

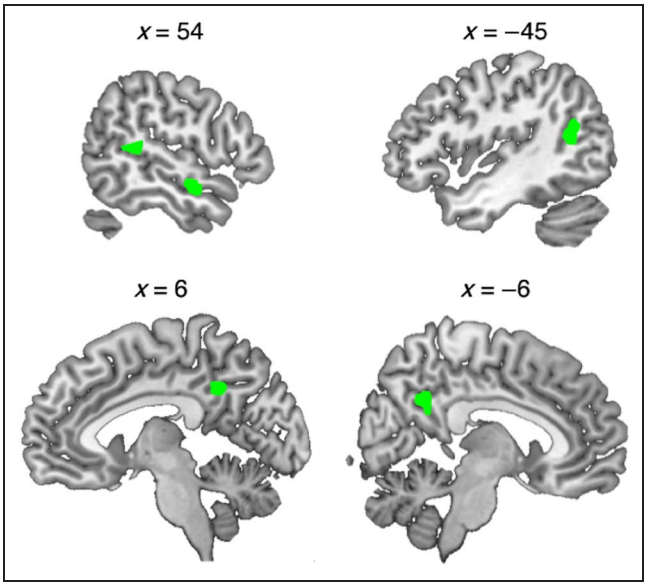



**Figure 4.** Greater activation for the production of specific impersonations (I) than for accents (A). Coordinates are in MNI space. Voxel height threshold $p < .001$, cluster threshold $p < .001$ (corrected).



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

**Table 4.** Brain Regions Showing Greater Activation during Specific Impersonations

| *Contrast* | *No. of Voxels* | *Region* | *Coordinate* | | | | |
|---|---|---|---|---|---|---|---|
| | | | *x* | *y* | *z* | *t* | *z* |
| Impersonation > Accents | 29 | Right STS | 54 | −3 | −15 | 5.79 | 4.46 |
| | 24 | Left STS | −45 | −60 | 15 | 4.62 | 3.82 |
| | 66 | Left middle cingulate cortex | −6 | −48 | 36 | 4.48 | 3.73 |
| | 32 | Right STG | 57 | −36 | 12 | 4.35 | 3.66 |

Voxel height threshold $p < .001$ (uncorrected), cluster threshold $p < .001$ (corrected). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space.

production, with notable sites of overlap in sensorimotor lobules V and VI of the cerebellum and left STG. However, there were also indications of differentiation of the three connectivity profiles. The left posterior STS seed region interacted with a speech production network including bilateral pre/postcentral gyrus, bilateral STG, and cerebellum (Price, 2010; Bohland & Guenther, 2006; Blank et al., 2002), as well as left-lateralized areas of anterior insula and posterior medial planum temporale. In contrast, the right anterior STS seed interacted with the left opercular part of the IFG and left SMA, and the right posterior STS showed a positive interaction with the left inferior frontal gyrus/sulcus, extending to the left frontal pole. Figure 5 illustrates the more anterior distribution of activations from the right-lateralized seed regions and the region of overlap from all seed regions in cerebellar targets.

Our results suggest that different emphases can be distinguished between the roles performed by these superior temporal and inferior parietal areas in spoken impressions. In a meta-analysis of the semantic system, Binder, Desai, Graves, and Conant (2009) identified the AG as a high-order processing site performing the retrieval and integration of concepts (Binder et al., 2009). The posterior left STS/AG activation has been implicated in the production of complex narrative speech and writing (Brownsett & Wise, 2010; Awad, Warren, Scott, Turkheimer, & Wise, 2007; Spitsyna, Warren, Scott, Turkheimer, & Wise, 2006) and, along with the precuneus, in the perceptual processing of familiar names, faces, and voices (von Kriegstein et al., 2005; Gorno-Tempini et al., 1998) and person-related semantic information (Tsukiura, Mochizuki-Kawai, & Fujii, 2006). We propose a role for the left STS/AG in accessing and integrating conceptual information related to target voices, in close communication with the regions planning and executing articulations. The increased performance demands encountered during the emulation of specific voice identities, which requires accessing the semantic knowledge of individuals, results in greater engagement of this left posterior temporo-parietal region and its enhanced involvement with the speech production network.

The interaction of right-lateralized sites on STS with left, middle, and inferior frontal gyrus and pre-SMA suggests higher-order roles in planning specific impersonations. Blank et al. (2002) found that the left pars opercularis of the IFG and left pre-SMA exhibited increased activation during production of speech of greater phonetic and linguistic complexity and variability and linked the pre-SMA to the selection and planning of articulations. In studies of voice perception, the typically right-dominant temporal voice areas in STS show stronger activation in response to vocal sounds of human men, women, and children compared with nonvocal sounds (Belin & Grosbras, 2010; Belin et al., 2000, 2002; Giraud et al., 2004), and right-hemisphere lesions are clinically associated with specific impairments in familiar voice recognition (Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010; Lang, Kneidl,

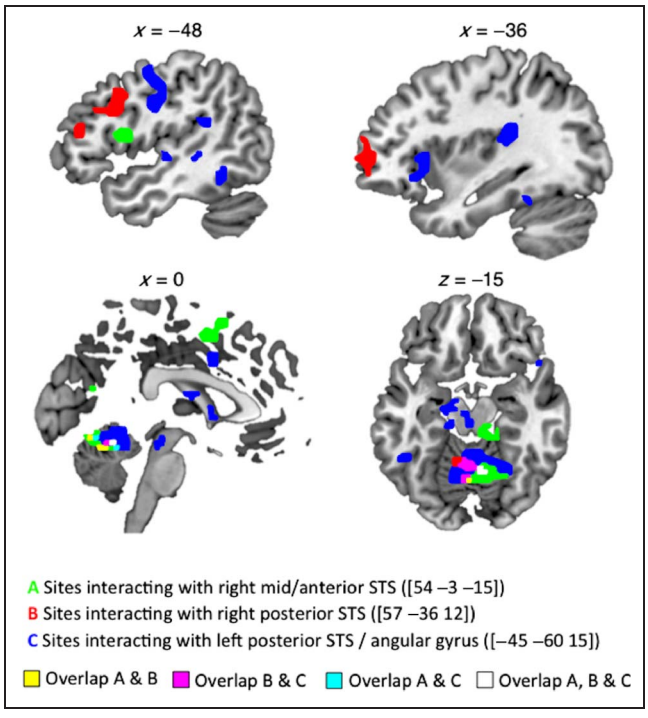



**Figure 5.** Regions of the brain showing stronger positive correlations during the production of impersonations compared with accents. The seed regions were defined using from the contrast Impersonations > Accents. Voxel height threshold $p < .001$ (uncorrected), cluster threshold $p < .001$ (corrected).



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

**Table 5.** Brain Regions Showing an Enhanced Positive Correlation with Temporo-parietal Cortex during Impersonations, Compared with Accents

| *Seed Region* | *No. of Voxels* | *Target Region* | *Coordinate* | | | *t* | *z* |
|---|---|---|---|---|---|---|---|
| | | | *x* | *y* | *z* | | |
| Right anterior STS | 66 | Left STG | −60 | −12 | 6 | 6.16 | 4.65 |
| | 98 | Right/left cerebellum | 9 | −63 | −12 | 5.86 | 4.50 |
| | 77 | Right cerebellum | 15 | −36 | −18 | 5.84 | 4.49 |
| | 21 | Left IFG (pars operc.) | −48 | 9 | 12 | 5.23 | 4.17 |
| | 65 | Right calcarine gyrus | 15 | −72 | 18 | 5.03 | 4.06 |
| | 48 | Left/right pre-SMA | −3 | 3 | 51 | 4.84 | 3.95 |
| | 37 | Right STG | 63 | −33 | 9 | 4.73 | 3.88 |
| Left posterior STS | 346 | Left rolandic operculum/left STG/STS | −33 | −30 | 18 | 6.23 | 4.68 |
| | 287 | Left/right cerebellum | 0 | −48 | −15 | 6.15 | 4.64 |
| | 306 | Right STG/IFG | 66 | −6 | −3 | 5.88 | 4.51 |
| | 163 | Right/left caudate nucleus and right thalamus | 15 | 21 | 3 | 5.72 | 4.43 |
| | 35 | Left thalamus/hippocampus | −12 | −27 | −6 | 5.22 | 4.17 |
| | 33 | Left hippocampus | −15 | −15 | −21 | 4.97 | 4.03 |
| | 138 | Left pre/postcentral gyrus | −51 | −6 | 30 | 4.79 | 3.92 |
| | 26 | Left/right mid cingulate cortex | −9 | 9 | 39 | 4.37 | 3.67 |
| | 21 | Left IFG/STG | −57 | 12 | 3 | 4.27 | 3.61 |
| | 23 | Right postcentral gyrus | 54 | −12 | 36 | 4.23 | 3.58 |
| | 37 | Left insula/IFG | −36 | 21 | 3 | 4.14 | 3.52 |
| Right posterior STS | 225 | Left middle/IFG | −39 | 54 | 0 | 5.90 | 4.52 |
| | 40 | Left STS | −66 | −36 | 6 | 5.63 | 4.38 |
| | 41 | Right postcentral gyrus/precuneus | 27 | −45 | 57 | 5.05 | 4.07 |
| | 20 | Right IFG | 42 | 18 | 27 | 4.79 | 3.92 |
| | 57 | Left/right cerebellum | −24 | −48 | −24 | 4.73 | 3.89 |
| | 29 | Left lingual gyrus | −18 | −69 | 3 | 4.64 | 3.83 |
| | 31 | Left STG | −63 | −6 | 0 | 4.35 | 3.66 |

Voxel height threshold $p < .001$ (uncorrected), cluster threshold $p < .001$ (corrected). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space. pars operc. = pars opercularis.

Hielscher-Fastabend, & Heckmann, 2009; Neuner & Schweinberger, 2000). Investigations of familiarity and identity in voice perception have implicated both posterior and anterior portions of the right superior temporal lobe, including the temporal pole, in humans and macaques (von Kriegstein et al., 2005; Kriegstein & Giraud, 2004; Belin & Zatorre, 2003; Nakamura et al., 2001). We propose that the right STS performs acoustic imagery of target voice identities in the Impersonations condition, and that these representations are used on-line to guide the modified articulatory plans necessary to effect voice change via left-lateralized sites on the inferior and middle frontal gyri. Although there were some acoustic differences between the speech produced under these two conditions—the Impersonations had a higher mean and standard deviation of pitch than the Accents (see Table 1)—we would expect to see sensitivity to these physical properties in earlier parts of the auditory processing stream, that is, STG rather than STS. Therefore, the current results offer the first demonstration that right temporal regions previously implicated in the perceptual processing and recognition of voices may play a direct role in modulating vocal identity in speech.

The flexible control of the voice is a crucial element of the expression of identity. Here, we show that changing the characteristics of vocal expression, without changing the linguistic content of speech, primarily recruits left anterior insula and inferior frontal cortex. We propose that therapeutic



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

approaches targeting metalinguistic aspects of speech production, such as melodic intonation therapy (Belin et al., 1996) and respiratory training, could be beneficial in cases of speech production deficits after injury to left frontal sites.

Our finding that superior temporal regions previously identified with the perception of voices showed increased activation and greater positive connectivity with frontal speech planning sites during the emulation of specific vocal identities offers a novel demonstration of a selective role for these voice-processing sites in modulating the expression of vocal identity. Existing models of speech production focus on the execution of linguistic output and monitoring for errors in this process (Hickok, 2012; Price, Crinion, & Macsweeney, 2011; Tourville & Guenther, 2011). We suggest that noncanonical speech output need not always form an error—for example, the convergence on pronunciations observed in conversation facilitates comprehension, interaction, and social cohesion (Garrod & Pickering, 2004; Chartrand & Bargh, 1999). However, there likely exists some form of task-related error monitoring and correction when speakers attempt to modulate how they sound, possibly along a predictive coding mechanism that attempts to reduce the disparity between predicted and actual behavior (Price et al., 2011; Friston, 2010; Friston & Price, 2001)—this could take place in the right superior temporal cortex (although we note that previous studies directly investigating the detection of and compensation for pitch/time-shifted speech have located this to bilateral posterior STG; Takaso, Eisner, Wise, & Scott, 2010; Tourville et al., 2008). We propose to repeat the current experiment with professional voice artists who are expert at producing convincing impressions and presumably also skilled in self-report on, for example, performance difficulty and accuracy. These trial-by-trial ratings could be used to interrogate the brain regions engaged when the task is more challenging to potentially uncover a more detailed mechanistic explanation for the networks identified for the first time in the current experiment.

We offer the first delineation of how speech production and voice perception systems interact to effect controlled changes of identity expression during voluntary speech. This provides an essential step in understanding the neural bases for the ubiquitous behavioral phenomenon of vocal modulation in spoken communication.

## Acknowledgments

This work was supported by a Wellcome Trust Senior Research Fellowship (WT090961MA) awarded to S.K.S.

Reprint requests should be sent to Carolyn McGettigan, Royal Holloway University of London, Egham Hill, Egham, Surrey TW20 0EX, United Kingdom, or via e-mail: Carolyn.McGettigan@rhul.ac.uk.

## REFERENCES

Ackermann, H., & Riecker, A. (2010). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language, 89,* 320–328.

Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2007). A common system for the comprehension and production of narrative speech. *Journal of Neuroscience, 27,* 11455–11464.

Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *Plos One, 5,* e8759.

Bailly, G. (2003). Close shadowing natural versus synthetic speech. *International Journal of Speech Technology, 6,* 11–19.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129–135.

Belin, P., & Grosbras, M.-H. (2010). Before speech: Cerebral voice processing in infants. *Neuron, 65,* 733–735.

Belin, P., McAdams, S., Smith, B., Savel, S., Thivard, L., Samson, S., et al. (1998). The functional anatomy of sound intensity discrimination. *Journal of Neuroscience, 18,* 6388–6394.

Belin, P., VanEeckhout, P., Zilbovicius, M., Remy, P., Francois, C., Guillaume, S., et al. (1996). Recovery from nonfluent aphasia after melodic intonation therapy: A PET study. *Neurology, 47,* 1504–1511.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport, 14,* 2105–2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research. Cognitive Brain Research, 13,* 17–26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403,* 309–312.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19,* 2767–2796.

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain, 125,* 1829–1838.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage, 32,* 821–841.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). *Region of interest analysis using an SPM toolbox*. Paper presented at the International Conference on Functional Mapping of the Human Brain, Sendai, Japan.

Broca, P. (1861). Perte de la parole, ramollissement chronique, et destruction partielle du lobe antérieur gauche du cerveau. *Bulletin de la Société Anthropologique, 2,* 235–238.

Brownsett, S. L. E., & Wise, R. J. S. (2010). The contribution of the parietal lobes to speaking and writing. *Cerebral Cortex, 20,* 517–523.

Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *Plos One, 7,* e31353.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76,* 893–910.

Condon, W. S., & Ogston, W. D. (1967). A segmentation of behavior. *Journal of Psychiatric Research, 5,* 221–235.

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *Journal of the Acoustical Society of America, 128,* 2059–2069.

Dhanjal, N. S., Handunnetthi, L., Patel, M. C., & Wise, R. J. (2008). Perceptual systems controlling speech production. *Journal of Neuroscience, 28,* 9969–9975.



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature, 384,* 159–161.
Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping, 7,* 89–97.
Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage, 25,* 1325–1335.
Eriksson, E., Sullivan, K., Zetterholm, E., Czigler, P., Green, J., Skagerstrand, A., et al. (2010). Detection of imitated voices: Who are reliable earwitnesses? *International Journal of Speech Language and the Law, 17,* 25–44.
Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11,* 127–138.
Friston, K. J., & Price, C. J. (2001). Dynamic representations and generative models of brain function. *Brain Research Bulletin, 54,* 275–285.
Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8,* 8–11.
Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics, 15,* 87–105.
Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *The contexts of accommodation* (pp. 1–68). New York: Cambridge University Press.
Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., et al. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex, 14,* 247–255.
Gorno-Tempini, M. L., Price, C. J., Josephs, O., Vandenberghe, R., Cappa, S. F., Kapur, N., et al. (1998). The neural systems sustaining face and proper-name processing. *Brain: A Journal of Neurology, 121,* 2103–2118.
Groswasser, Z., Korn, C., Groswasser-Reider, I., & Solzi, P. (1988). Mutism associated with buccofacial apraxia and bihemispheric lesions. *Brain and Language, 34,* 157–168.
Hailstone, J. C., Crutch, S. J., Vestergaard, M. D., Patterson, R. D., & Warren, J. D. (2010). Progressive associative phonagnosia: A neuropsychological analysis. *Neuropsychologia, 48,* 1104–1114.
Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping, 7,* 213–223.
Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? Elizabeth II's traditional pronunciation has been influenced by modern trends. *Nature, 408,* 927–928.
Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13,* 135–145.
Jurgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews, 26,* 235–258.
Jurgens, U., & von Cramon, D. (1982). On the role of the anterior cingulate cortex in phonation: A case report. *Brain and Language, 15,* 234–248.
Kappes, J., Baumgaertner, A., Peschke, C., & Ziegler, W. (2009). Unintended imitation in nonword repetition. *Brain and Language, 111,* 140–151.
Karpf, A. (2007). *The human voice: The story of a remarkable talent*. London: Bloomsbury Publishing PLC.
Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22,* 948–955.
Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: Functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure & Function, 214,* 519–534.
Ladefoged, P. (2003). Validity of voice identification. *Journal of the Acoustical Society of America, 114,* 2403.
Lang, C. J. G., Kneidl, O., Hielscher-Fastabend, M., & Heckmann, J. G. (2009). Voice recognition in aphasic and non-aphasic stroke patients. *Journal of Neurology, 256,* 1303–1306.
Lombard, É. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de L'Oreille et du Larynx, XXXVII,* 101–109.
Lu, Y., & Cooke, M. (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *Journal of the Acoustical Society of America, 126,* 1495–1499.
MacLarnon, A. M., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology, 109,* 341–363.
MacLarnon, A. M., & Hewitt, G. P. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology, 13,* 181–197.
McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech Language and Hearing Research, 44,* 128–143.
Meyer, M., Zysset, S., von Cramon, D. Y., & Alter, K. (2005). Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Brain Research. Cognitive Brain Research, 24,* 291–306.
Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., et al. (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia, 39,* 1047–1054.
Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition, 44,* 342–366.
Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25,* 653–660.
Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Broca's area in speech production. *Cerebral Cortex, 19,* 2156–2165.
Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119,* 2382–2393.
Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40,* 190–197.
Pardo, J. S., & Jay, I. C. (2010). Conversational role influences speech imitation. *Attention Perception & Psychophysics, 72,* 2254–2264.
Peschke, C., Ziegler, W., Kappes, J., & Baumgaertner, A. (2009). Auditory-motor integration during fast repetition: The neuronal correlates of shadowing. *Neuroimage, 47,* 392–402.
Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11,* 105–110.
Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191,* 62–88.
Price, C. J., Crinion, J. T., & Macsweeney, M. (2011). A generative model of speech production in Broca's and Wernicke's areas. *Front Psychol, 2,* 237.
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. *Frontiers in Psychology, 2,* 271.
Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021

syllable frequency on speech motor control. *Brain and Language, 107,* 102–113.
Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., et al. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology, 64,* 700–706.
Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24–30.
Scott, S. K., Clegg, F., Rudge, P., & Burgess, P. (2006). Foreign accent syndrome, speech rhythm and the functional neuronatomy of speech production. *Journal of Neurolinguistics, 19,* 370–384.
Sem-Jacobsen, C., & Torkildsen, A. (1960). Depth recording and electrical stimulation in the human brain. In Estelle R. Ramey and Desmond O'Doherty (Eds.), *Electrical studies on the unanesthetized brain*. New York.
Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66,* 422–429.
Simmonds, A. J., Wise, R. J. S., Dhanjal, N. S., & Leech, R. (2011). A comparison of sensory-motor activity during speech in first and second languages. *Journal of Neurophysiology, 106,* 470–478.
Simonyan, K., Ostuni, J., Ludlow, C. L., & Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *Journal of Neuroscience, 29,* 14912–14923.
Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research, 17,* 75–82.
Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2006). Converging language streams in the human temporal lobe. *Journal of Neuroscience, 26,* 7328–7336.
Sullivan, K. P. H., & Schlichting, F. (1998). A perceptual and acoustic study of the imitated voice. *Journal of the Acoustical Society of America, 103,* 2894.
Takaso, H., Eisner, F., Wise, R. J., & Scott, S. K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: A positron emission tomography study. *Journal of Speech Language and Hearing Research, 53,* 226–236.
Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26,* 952–981.
Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage, 39,* 1429–1443.
Tsukiura, T., Mochizuki-Kawai, H., & Fujii, T. (2006). Dissociable roles of the bilateral anterior temporal lobe in face-name associations: An event-related fMRI study. *Neuroimage, 30,* 617–626.
von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17,* 48–55.
von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience, 17,* 367–376.
Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage, 24,* 1233–1241.
Wise, R. J., Greene, J., Buchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet, 353,* 1057–1061.

Downloaded from http://mitprc.silverchair.com/jocn/article-pdf/25/11/1875/1779685/jocn_a_00427.pdf by MIT Libraries user on 17 May 2021



Downloaded from http://www.mitpressjournals.org/doi/pdf/10.1162/jocn_a_00427 by guest on 28 March 2024