

UCSF

UC San Francisco Previously Published Works

Title

Bihemispheric network dynamics coordinating vocal feedback control

Permalink

<https://escholarship.org/uc/item/5q04m46b>

Journal

Human Brain Mapping, 37(4)

ISSN

1065-9471

Authors

Kort, Naomi S

Cuesta, Pablo

Houde, John F

et al.

Publication Date

2016-04-01

DOI

10.1002/hbm.23114

Peer reviewed

Bihemispheric Network Dynamics Coordinating Vocal Feedback Control

Naomi S. Kort,^{1*} Pablo Cuesta,² John F. Houde,³ and
Srikantan S. Nagarajan^{3,4}

¹*Department of Psychiatry, SFVAMC 116-D, Building 8, 4150 Clement St, San Francisco, California*

²*Laboratory of Cognitive and Computational Neuroscience (UCM-UPM), Centre for Biomedical Technology (CTB), Technical University of Madrid (UPM), Campus Montegancedo, Pozuelo De Alarcón, Madrid 28223, Spain*

³*Department of Otolaryngology- Head and Neck Surgery, University of California, San Francisco, San Francisco, California*

⁴*Department of Radiology, University of California, San Francisco, San Francisco, California*



Abstract: Modulation of vocal pitch is a key speech feature that conveys important linguistic and affective information. Auditory feedback is used to monitor and maintain pitch. We examined induced neural high gamma power (HGP) (65–150 Hz) using magnetoencephalography during pitch feedback control. Participants phonated into a microphone while hearing their auditory feedback through headphones. During each phonation, a single real-time 400 ms pitch shift was applied to the auditory feedback. Participants compensated by rapidly changing their pitch to oppose the pitch shifts. This behavioral change required coordination of the neural speech motor control network, including integration of auditory and somatosensory feedback to initiate change in motor plans. We found increases in HGP across both hemispheres within 200 ms of pitch shifts, covering left sensory and right premotor, parietal, temporal, and frontal regions, involved in sensory detection and processing of the pitch shift. Later responses to pitch shifts (200–300 ms) were right dominant, in parietal, frontal, and temporal regions. Timing of activity in these regions indicates their role in coordinating motor change and detecting and processing of the sensory consequences of this change. Subtracting out cortical responses during passive listening to recordings of the phonations isolated HGP increases specific to speech production, highlighting right parietal and premotor cortex, and left posterior temporal cortex involvement in the motor response. Correlation of HGP with behavioral compensation demonstrated right

Contract grant sponsor: National Institutes of Health; Contract grant number: R01DC013979; R01DC010145; Contract grant sponsor: National Science Foundation; Contract grant number: BCS 1262297, BCS-0926196; Contract grant sponsor: Bay Area Affective Sciences Training Program predoctoral fellowship; Contract grant number: T32 MH20006-13; Contract grant sponsor: National Institute of Mental Health; Contract grant number: T32 MH089920

*Correspondence to: Srikantan S. Nagarajan or John F. Houde, Speech Neuroscience Laboratory, 513 Parnassus Avenue, S679,

San Francisco, CA 94143-0628. E-mail: sri@ucsf.edu, houde@phy.ucsf.edu

We declare no conflicts of interest.

Received for publication 30 November 2015; Revised 23 December 2015; Accepted 3 January 2016.

DOI: 10.1002/hbm.23114

Published online in Wiley Online Library (wileyonlinelibrary.com).

frontal region involvement in modulating participant's compensatory response. This study highlights the bihemispheric sensorimotor cortical network involvement in auditory feedback-based control of vocal pitch. *Hum Brain Mapp* 37:1474–1485, 2016. © 2016 Wiley Periodicals, Inc.

Key words: speech; magnetoencephalography; sensorimotor cortex; sensory-motor performance; functional neuroimaging; phonation

INTRODUCTION

Vocal control of pitch is essential for human speech. However, despite extensive study, the understanding of sensory-motor control of pitch production is incomplete. Traditional theories of speech production [Broca, 1861; Dronkers, 1996; Hickok et al., 2011] emphasize functional lateralization of the left hemisphere, neglecting the role of the right hemisphere in speech production despite neuroimaging studies consistently showing bilateral neural activity during speech perception and production [Price, 2010]. Recent work has challenged this notion by suggesting sensory-motor interactions occur bilaterally for word repetition [Cogan et al., 2014]. Yet sensory-motor transformations do not only occur in speech repetition but also continuously occur during ongoing speech as auditory feedback is monitored and errors in production are rapidly recognized and corrected. Vocal control of pitch employs sensory-motor transformations during auditory feedback for online control and long-term maintenance of pitch production: when auditory feedback is available, it is used to make rapid adjustments in pitch production [Burnett et al., 1998] and the control of pitch production is impaired when feedback is absent [Lane and Webster, 1991].

Current models of speech production describe the manner in which sensory feedback is monitored and errors are recognized and corrected as a special case of neural predictive coding [Guenther and Vladusich, 2012; Houde and Nagarajan, 2011]. In these current models of speech motor control [Guenther and Vladusich, 2012; Hickok et al., 2011; Houde and Nagarajan, 2011], premotor cortex sends a forward model encoding the predicted auditory feedback to auditory cortex, where the expected sound is compared to the actual sound during speaking. When the perceived auditory feedback matches the predicted auditory feedback, the prediction error is minimized and auditory cortical responses are suppressed [Chang et al., 2013; Flinker et al., 2010; Greenlee et al., 2011; Houde et al., 2002; Kort et al., 2014; Muller-Preuss and Ploog, 1981; Niziolek et al., 2013]. Yet, when there is a mismatch between the perceived and predicted auditory feedback, there is a prediction error corresponding to an enhanced neural response [Behroozmand and Larson, 2011; Behroozmand et al., 2009; Chang et al., 2013; Greenlee et al., 2013b; Kort et al., 2014; Niziolek et al., 2013]. This prediction error can then be passed to higher levels to update the motor map, make online changes to the motor plan, and refine future predictions [Guenther and Vladusich, 2012;

Hickok et al., 2011; Houde and Nagarajan, 2011; Niziolek et al., 2013]. The neural substrates of the corresponding transformations can be probed using time resolved neuroimaging. Specifically, the neural mechanisms that allow errors to be detected and responded to can be broken down into three stages: (1) detection of the error/neural representation of the prediction error, (2) coordination of the motor change, which involves integrating the somatosensory and auditory feedback, updating the motor plan and creating new sensory predictions, (3) detection of the sensory consequence of the new motor plan and using this information to update the state estimation. The process of monitoring and responding to sensory feedback is vocal feedback control. The goal of this study is to elucidate the neural substrates of these three stages of compensation for pitch feedback alterations.

Studies using fMRI lack the temporal resolution to distinguish the stages of sensory-motor transformations mentioned above [Parkinson et al., 2012; Zarate and Zatorre, 2008]. Conversely, EEG studies [Behroozmand et al., 2009, 2011] lack spatial resolution to make inferences about what brain regions are involved in these stages sensory-motor processing. ECoG studies are limited by grid coverage and have focused on the left hemisphere [Chang et al., 2013; Greenlee et al., 2013a]. A previous study using magnetoencephalography, focusing on low-frequency evoked responses, found bilateral sensory-motor responses during vocal control of pitch which did not correlate with the participants' motor behavior [Kort et al., 2014]. Space- and time-resolved imaging of induced neural activity are important to better determine neural computations related to vocal feedback control. The high time and spatial resolution of magnetoencephalographic imaging provides this resolution and offers an intermediate scale between invasive ECoG and fMRI. Here, we focus on the induced high gamma band (65–150 Hz) as the prominent frequency band modulated in previous ECoG studies [Chang et al., 2013; Greenlee et al., 2013a]. Induced measures of neural activity include responses that are not phase-locked to the stimulus. High gamma band activity has been shown to correlate with fMRI [Mukamel et al., 2005] and spiking activity of neurons [Crone et al., 2006].

In this study, we sought to investigate the spatial-temporal dynamics involved in feedback control of speech production. Using magnetoencephalography to measure the neural response to an unexpected shift in pitch allowed us to investigate neural responses to unexpected reafferent information. With this paradigm, we examined which cortical regions are involved in voice pitch control

removed from linguistic and emotional context, how these cortical responses to an error in pitch production evolve over time, how these cortical responses relate to behavioral responses, and the neural connectivity between nodes in the pitch production network.

MATERIALS AND METHODS

Participants

Fifteen right-handed English speaking volunteers with normal speech and hearing participated in the study. Only nonsingers were included in this study since unique behavioral and neural responses to shifted auditory feedback during singing for musicians [Zarate and Zatorre, 2008] have been reported. Three participants (2 male and 1 female) were eliminated from the study due to the failure of the pitch tracking algorithm to reliably track the pitch. Twelve volunteers (6 female) participants showed behavioral compensation to the auditory perturbation and were included in all analyses. After procedures had been fully explained, all participants gave their informed consent. The study was performed with the approval of the University of California, San Francisco Committee for Human Research.

MEG Recording

The task was completed during whole head MEG neural recording in awake participants lying in the supine position. The MEG system (CTF, Coquitlam, British Columbia, Canada) consists of 275 axial gradiometers whose data were recorded with a sampling rate of 1200 Hz. Three fiducial coils were placed on the nasion and left/right preauricular points to triangulate the position of the head relative to the MEG sensor array. In a separate session, high-resolution anatomical MRIs were obtained for each participant. The fiducial markers points were later coregistered with an anatomical MRI to generate head shape.

Experimental Design and Procedure

The experiment was administered in a block design with a Speaking Condition and Listening Condition. In both the Speaking Condition and the Listening Condition, participants watched a screen with a projected image in their line of sight. Prior to each block, participants were verbally cued to the instructions of the block, described below. For every trial, the background of the screen was black. A trial was initiated when three large white dots appeared in the center of the screen. Each dot disappeared one by one, simulating a countdown. When all three dots disappeared, participants either phonated or passively listened, corresponding to the block. A visual cue—the number of remaining trials preceding a break—appeared to signal the end of the trial.

In the Speaking Condition, participants were instructed to speak the vowel /a/ until the termination cue. The partici-

pants spoke into an MEG-compatible optical microphone (Optimic MEG, Optoacoustics, <http://www.optoacoustics.com/medical/optimic-meg/features>) and received auditory feedback through MEG-compatible earplug earphones (ER-3A insert headphone, Etymotic, <http://www.etymotic.com/auditory-research/accessories/er3a.html>). During the phonation, the participants heard one 100 cent pitch perturbation lasting 400 ms whose onset was jittered in time from speech onset, beginning between 500 and 1000 ms after the visual prompt. The 100 cent pitch perturbation was selected since it is sufficiently large to induce a robust neural and behavioral response, but sufficiently small that it is in line with naturally occurring speech errors and the speech is still identified as self-produced. An equal number of pitch shifts that either raised or lowered the perceived pitch were pseudorandomly distributed across the experiment. The jittered perturbation onset prevented the participant from anticipating the timing of the perturbation while the pseudorandom selection of raising or lowering the pitch prevented the participant from anticipating the direction of the perturbation. In the Listening Condition, participants received the same visual prompts but passively listened to the recording of their perturbed voice feedback obtained in the previous Speaking Condition block. The auditory input through earphones in both conditions was identical.

The experiment contained 4 blocks of 74 trials with brief, self-paced breaks every 15 trials. The blocks alternated between the two conditions: blocks 1 and 3 were the Speaking Condition; blocks 2 and 4 were the Listening Condition. Prior to the start of the experiment, the volume of auditory input through the earphones was adjusted, so that participants reported their auditory feedback was the same as or slightly louder than they normally would hear their voice during air conduction. This was to ensure that the participants perceived the auditory feedback through their headphones as natural.

Speech alteration system

The speech alteration system performed the pitch perturbation using the following methods. The system used an analysis–synthesis process to repeatedly digitize 3 ms frames of the participant's speech (32 time samples at an 11,025 Hz sampling rate) directly from the microphone. These 3 ms frames were analyzed, modified, and resynthesized into new frames that were the altered audio output. The system used a 400-sample (36 ms) buffer that was analyzed by computing a narrow-band magnitude frequency spectrum. The pitch was estimated from the harmonic spacing of the narrow-band spectrum, and was modified independently from other speech features (formants and total frame energy) before being recombined to make the new narrowband magnitude spectrum. The new spectrum was used to create the next frame of audio output using sinusoidal synthesis [McAulay and Quatieri, 1986]. This synthesis method does not require the phase spectrum of the original input speech; instead, each harmonic peak in the new

narrow-band magnitude spectrum specifies the frequency and amplitude of a sinusoid; these sinusoids simply were added together to create the next frame of output speech. The output audio of the speech was then converted back to an analog signal that was fed through the participant's MEG-compatible earphones.

Audio analysis

Pitch analysis was performed on each trial of each participant's audio data. The microphone and feedback signals were recorded and analyzed, sampled at 11025 Hz. The data was recorded in 32-sample frames. Waveform data was analyzed in successive 30 ms frames, with an advance of 3 ms between frames. Pitch was estimated for each of these frames using the standard autocorrelation method [Parsons, 1986]. Initial pitch bounds for pitch tracking were 30–300 Hz, which were then adjusted for each subject to minimize octave errors in the pitch estimation. The resulting frame-by-frame pitch contour was then smoothed with a 20 Hz, fifth-order, low-pass Butterworth filter. Using visual inspection, trials with erroneous pitch contours were removed. The visual inspection identified trials with large, nonphysiologic deviations that can be a consequence of the pitch tracking software. A pre-perturbation interval was analyzed to calculate the mean and standard deviation for each participant's pitch contour. The baseline was chosen as the largest possible interval that fit the constraint of the minimum time between voice onset and perturbation onset. The baseline window size varied across trials since the onset of the pitch shift was jittered in respect to speech onset. By using a baseline window, we avoid a confound that could arise if a participant has a tendency to either raise or lower the pitch of their voice during steady-state vocalization and allow for comparison across subjects. The magnitude and onset of the compensation were determined for each participant and then averaged to create the grand-average compensation magnitude and response onset. A participant's response onset was conservatively set to occur when the mean pitch time-course deviates from the baseline by two standard deviations. The absolute value of pitch contours for each perturbation type, +100 cent or –100 cent, was averaged together for each participant. The pitch contour in cents was calculated from absolute frequency (Hertz) by

$$\Delta\text{cents} = 100 \times \left[12 \times \log_2 \left(\frac{\text{pitch response peak (Hz)}}{\text{mean pitch frequency of pre-perturbation baseline (Hz)}} \right) \right]$$

The calculation for the mean percent compensation was

$$\% \text{compensation} = -100 \times \frac{\Delta\text{cents}}{\text{applied pitch shift (cents)}}$$

The negative sign ensures a positive-percent compensation.

MEG data preprocessing

The MEG sensor data were manually marked at the speech onset and at the perturbation onset. Third gradient noise correction filters were applied to the data and the data were corrected for a DC offset based on the whole trial. Artifact rejection of abnormally large signals due to EMG, head movement, eye blinks, or saccades was performed qualitatively through visual inspection and trials with artifacts were eliminated from the analysis. Sensor data was notch filtered around 120 Hz with a width of 4 Hz. Only trials with a minimum for 400 ms between voice onset and perturbation onset were included in subsequent analyses. On average, there were 136 trials for each condition per participant that fit these criteria.

MEG data analysis

A time–frequency optimized spatially adaptive filtering algorithm implemented in NUTMEG [Dalal et al., 2008] was used to localize the induced activity in the high gamma band (65–150 Hz) to the individual participant's spatially normalized MRIs. The NUTMEG toolbox has been previously described in detail in Dalal et al. [2011]. Noise-corrected pseudo-F ratios were computed between the active windows (following the perturbation onset) and the prestimulus control baseline (the window preceding the onset of the perturbation) [Dalal et al., 2008; Sekihara and Nagarajan, 2010]. The windows were 100 ms with an overlap of 25 ms, and since four overlapping windows are required to reconstruct one 25 ms window, we were able to interrogate the times 100 ms following the perturbation onset to 300 ms following the perturbation onset in 25 ms steps. Group statistics were computed using the NUTMEG time–frequency statistics toolbox [Dalal et al., 2008, 2011] with statistical nonparametric mapping (SnPM) [Singh et al., 2003].

A multisphere lead field (forward model) was calculated for every 5 mm voxel in the brain. The lead field describes the magnetic field strength at each MEG sensor that would arise from a single dipole source originating in each voxel. Source localization was then calculated for the high gamma band activity using both the lead field and sensor covariance. Noise-corrected pseudo-F ratios were computed between the active windows (following the perturbation onset) and the prestimulus control baseline (the window preceding the onset of the perturbation). Deep brain nuclei are a greater distance from the MEG sensors with low signal-to-noise ratio, especially for high-gamma signals, resulting in spatial blur and greater uncertainty in estimating their activity. Therefore, we are restricting our analysis to the cortical surface by removing voxels in deep brain structures.

The results from the time–frequency beam former for each participant, computed in their native space, were normalized to the MNI template brain using SPM2. For every time–frequency point, three-dimensional maps average and variance

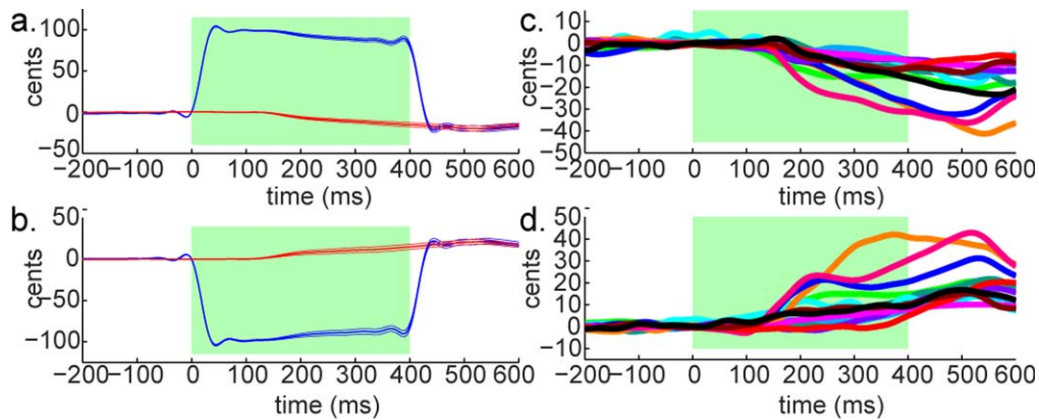


Figure 1.

Vocal responses to the shift in pitch of audio feedback. Grand-average across participants vocal responses (a) to +100 cent pitch shift and (b) to -100 cent pitch shift. In both (a) and (b), thick center blue line shows the mean time course of the audio input heard by participants (flanked by thin \pm standard error lines) and thick center red line shows participants' mean vocal

production (again, flanked by thin \pm standard error lines). Individual participant mean responses to the (c) +100 cent pitch shift and (d) -100 cent pitch shift. The shift onset occurs at 0 ms and is sustained for 400 ms, denoted by the green region. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

maps were calculated. The variance maps were then smoothed with a Gaussian kernel with a half-width of $20 \times 20 \times 20 \text{ mm}^3$. Using this, a pseudo- t statistic was obtained at each voxel and time window. Active and control labels across trials were permuted 2^{Number} of participants times to create nonparametric null distributions to derive p -values. The neurobehavioral correlations comparing individual participant's mean compensation with neural responses were calculated by computing Pearson's correlation coefficients for activations for all voxels again behavioral compensation. All images were cluster-corrected to 30 voxels, $P < 0.01$, so that only clusters remained that contained 30 contiguous voxels with $P < 0.01$ using the Nutmeg software toolbox.

RESULTS

Left Sensory Cortex Responses and Right Inferior Frontal Regions Guide Pitch Shift Detection

Participants responded to the transient pitch shifts by compensating: they rapidly changed their pitch production to oppose the shifts that either increased or decreased the pitch of their voice. The mean f_0 contours of each participant in response to the pitch shift are shown in Figure 1. The mean compensation for each participant was 21.79% (range: 10.4–41.65). The mean compensation onset latency was 187.22 ms (range: 124.69–300.29), whereas the mean peak compensation latency was 522.94 ms (range: 458.48–625.37). There were no correlations between mean compensation, compensation onset latency, and peak compensation latency across participants.

Within the first 125 ms after pitch shift onset, presumably the first stage of pitch feedback error detection, we observed widespread increases in induced high gamma power (HGP) over left primary and secondary sensory regions and right premotor, parietal, and frontal regions, after which HGP enhancement becomes right dominant (Figure 2 and Table I). In the left hemisphere, primary sites were in sensory regions across auditory cortices in the temporal lobe, and dorsally including somatosensory cortex (SSC). In the right hemisphere, activations included the inferior parietal lobe (IPL), the supramarginal gyrus (SMG), right premotor cortex (PMC), and anterior regions including the insula and inferior frontal gyrus (IFG).

Cortical activity prior to onset of behavioral compensation (187.22 ms) are presumably involved in processing the pitch feedback error and preparing for a behavioral change. The left hemisphere primary and secondary sensory regions, with the exception of left posterior middle temporal gyrus (MTG), had peak activity during this interval. In contrast, only two regions in the right hemisphere had activations peaking in this interval: right IFG-pars triangularis and anterior superior temporal gyrus (STG).

Right Premotor and Parietal Cortex, Left Posterior Temporal Cortex Initiate Compensation, and Prepare for Change in Feedback

Cortical regions that show peak activity concurrent with the compensation onset suggest involvement in both

inducing compensation and preparing for the new feedback that will result from the motor change. The onset of compensation occurs in the window 175–200 ms following the pitch shift. During this window the HGP continued to increase in the right hemisphere with peaks in right premotor cortex and right dorsal IPL, while the HGP response in the left hemisphere was restricted to left posterior MTG.

Right Frontal and Parietal Regions Monitor the Feedback Change Feedback

As participants change their pitch to compensate for the shift, their auditory feedback also changes. Cortical regions that show peak activity after the onset of compensation suggest involvement in continued compensation and monitoring of the new auditory feedback. In this segment of the pitch shift response, between 200 and 300 ms, right-hemisphere activity continued to increase, including peaks in right frontal (middle and inferior frontal gyrus), insula, temporal (MTG), and parietal areas (SMG). Throughout the entire window of the pitch shift, right SMG demonstrates the largest sustained power increase, indicating an involvement in coordinating the detection of and response to the pitch error. By 300 ms following the pitch shift, induced activity in the right hemisphere persists across SMG, PMC, and frontal regions.

Right Parietal and Premotor Cortex, Left Posterior Temporal Lobe Showed Speaking Specific Activity

The motor act of vocalization is necessarily accompanied by concurrent auditory input of the acoustic consequence of one's own vocalization. By comparing the cortical responses during the beginning of compensation to passively listening to the same auditory input, we can identify responses that are associated with initiating vocal compensation as opposed to passively perceiving a change in pitch. Despite the widespread bilateral early activity to the unexpected pitch shift in the speaking condition, only right IPL, SMG, and PMC and left posterior MTG and middle occipital gyrus (MOG) showed greater responses in the speaking condition than in the passive listening condition (Figure 3 and Table II). The enhancement during speaking as compared to during passive listening of the response to an unexpected pitch shift, termed Speaking Perturbation Response Enhancement (SPRE) [Chang et al., 2013; Kort et al., 2014], had peak activity in right dorsal IPL and right PMC immediately prior to (PMC) or concurrent with (IPL) the onset of compensation, indicating their involvement in coordinating the motor change driving compensation. Regions in the left hemisphere that show SPRE, left posterior MTG and MOG, and one region in the right

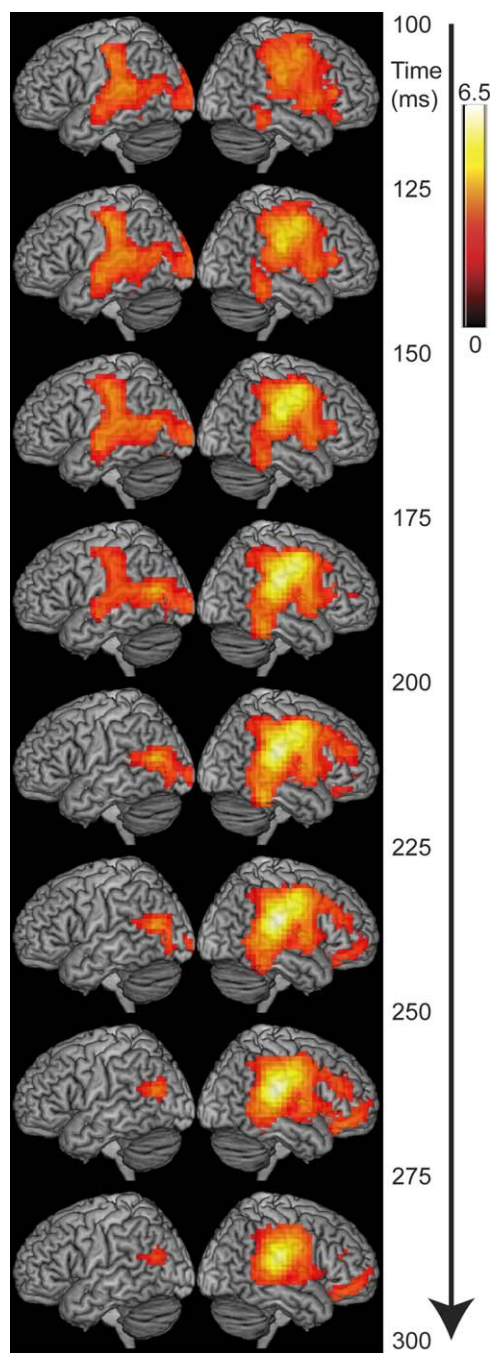


Figure 2.

Cortical responses to the pitch shift during the speaking condition. MEG cortical responses aligned to perturbation onset are enhanced in the speaking condition in response to the pitch shift compared to steady-state vocalization. Images are cluster corrected, 30 voxels, $P < 0.01$. Color scale represents t-value. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE I. Regions of significant enhancement in response to the pitch shift during speaking compared to steady-state vocalization

Region	Perturbed Speech–Speech MNI peak voxel	Perturbed Speech–Speech Time-course	<i>t</i> -value	<i>P</i> -value
Left transverse temporal gyrus	−45.0 −30.0 10.0	Peak: 100–125 ms Duration: 100–200 ms	3.96685	2.4×10^{-4}
Left somatosensory cortex	−55.0 −25.0 40.0	Peak: 125–150 ms Duration: 100–200 ms	4.02157	9.8×10^{-4}
Left posterior middle temporal gyrus	−60.0 −70.0 15.0	Peak: 175–200 ms Duration: 100–300 ms	4.54941	4.9×10^{-4}
Left middle occipital lobe	−10.0 −100.0 10.0	Peak: 125–150 ms Duration: 100–150 ms	3.34456	0.0081
Left middle temporal gyrus	−65.0 −10.0 −5.0	Peak: 150–175 ms Duration: 100–200 ms	3.83291	0.0012
Right premotor cortex	55.0 −10.0 40.0	Peak: 175–200 ms Duration: 100–300 ms	5.57166	2.4×10^{-4}
Right inferior parietal lobe	55.0 −30.0 40.0	Peak: 175–200 ms Duration: 100–300 ms	4.49757	4.9×10^{-4}
Right supramarginal gyrus	55.0 −25.0 20.0	Peak: 225–250 ms Duration: 100–300 ms	6.35952	2.4×10^{-4}
Right middle temporal gyrus	60.0 −40.0 −15.0	Peak: 200–225 ms Duration: 100–275 ms	3.63265	9.8×10^{-4}
Right insula	40.0 0.0 5.0	Peak: 250–275 ms Duration: 100–300 ms	4.56657	4.9×10^{-4}
Right inferior frontal gyrus-pars triangularis	60.0 30.0 10.0	Peak: 100–125 ms Duration: 100–200 ms	3.49622	0.002
Right anterior superior temporal gyrus	65.0 10.0 −5.0	Peak: 150–175 ms Duration: 100–275 ms	3.62678	0.0017
Right middle frontal gyrus	60.0 30.0 25.0	Peak: 200–225 ms Duration: 200–275 ms	3.27512	0.0068
Right inferior frontal gyrus-pars orbitalis	30.0 35.0 −10.0	Peak: 250–275 ms Duration: 200–300 ms	3.52064	0.0015

Table contains regions, peak voxel location, duration, peak latency, *t*-value and *P*-value of regions of significant enhancement in response to the pitch shift during speaking compared to steady-state vocalization. Regions included are cluster corrected, 30 voxels, $P < 0.01$.

hemisphere, SMG, showed significant enhancement when the sensory consequences of the compensation should be detected.

Right Frontal and Left Anterior and Posterior Temporal Regions Correlated with Individual Participant’s Mean Compensation

While all participants in this analysis compensated for the unexpected shift in pitch, the amount each participant opposed the shift varied. To address the neural underpinnings of this behavioral variability, we performed neurobehavioral correlations of HGP during speaking and individual participant’s mean compensation. Neurobehavioral correlations with mean compensation revealed a large cluster of right frontal regions, including IFG-pars triangularis, IFG-pars orbitalis, and middle frontal gyrus (MFG) that were significantly positively correlated with individual participant behavioral compensation to the

pitch shift (Figure 4 and Table III). In the left hemisphere, one large cluster extending from posterior temporal lobe to MOG was significantly positively correlated with behavioral compensation. On the other hand, left anterior MTG showed a strong negative correlation with behavioral compensation. The neurobehavioral correlations with mean compensation occur in the first 200 ms following the pitch shift. Given the timing of the correlations with mean compensation, we can infer that the neurobehavioral correlations relate to how observing the initial detection and processing of the pitch shift impacts the total amount of compensation.

DISCUSSION

Online control of pitch using auditory feedback is important for communication. In this study, we examined the cortical mechanisms involved in online control of pitch by studying the response to a shift in pitch feedback. We

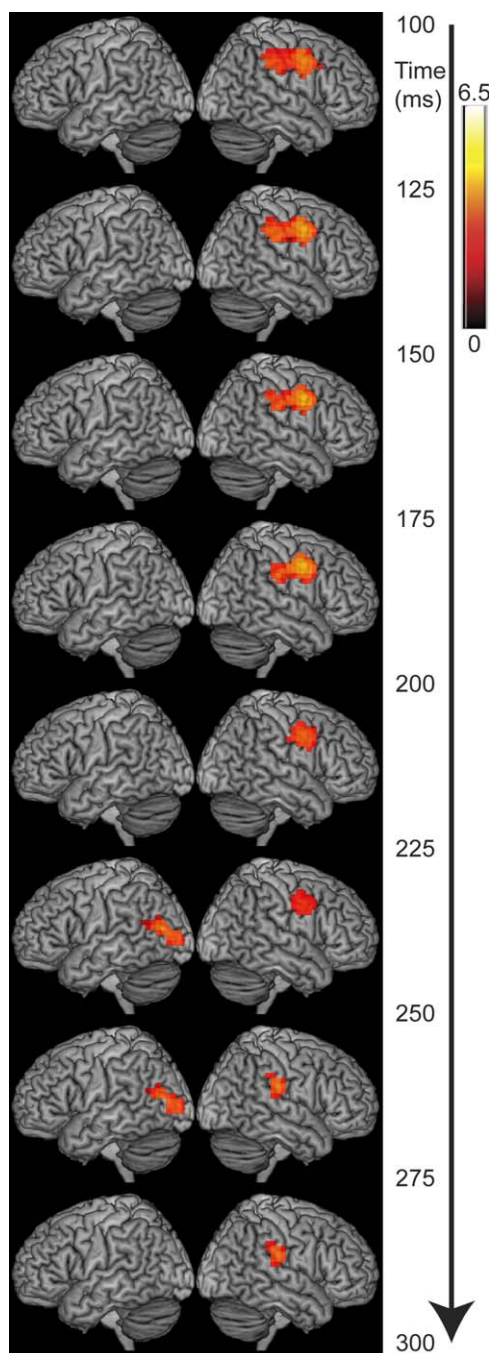


Figure 3.

Cortical responses to the pitch shift that are greater in the speaking condition than the listening condition (SPRE). MEG cortical responses that are greater in the speaking condition in response to the pitch shift than in the passive listening condition. Images are cluster corrected, 30 voxels, $P < 0.01$. Color scale represents t-value. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

demonstrated bihemispheric involvement in feedback control of speech. Importantly, we showed early left hemisphere sensory responses preceding right hemisphere frontal, parietal, and premotor responses. However, when controlling for passive auditory perception, we found right hemisphere parietal and premotor as well as left posterior temporal responses. In addition, we showed bihemispheric neurobehavioral correlations with mean compensation. These results highlight the importance of the left hemisphere in sensory feedback error processing of the pitch shift, and the importance of the later right hemisphere activations in higher level pitch feedback processing that drives the behavioral compensation.

The timing of the cortical activity gives insight to the mechanisms of error detection and compensation, and the role of each cortical region in the circuit. In the period prior to compensation onset, left-hemisphere primary and secondary sensory regions across the temporal lobe and into somatosensory cortex show enhanced responses to the pitch shift in the speaking condition, but the absence of SPRE in these regions implies a very similar pattern of responses to the pitch shift in the listening condition. The large response in left occipital regions was surprising as models of speech production do not involve a role for the occipital lobe. This response could be due to presence of the visual prompt. Similarly, in the right hemisphere, IFG-pars triangularis shows an early peak in the speaking condition that correlates with the amount of compensation, but does not show SPRE. Taken together, left sensory regions and right IFG appear to be involved in sensory detection and processing of the pitch shift. In terms of current models of speech production, these data provide evidence that feedback prediction errors (in terms of SFC model) [Houde and Nagarajan, 2011] and auditory and somatosensory error maps (in terms of DIVA model) [Guenther and Vladusich, 2012] exist in left sensory regions.

In the period during compensation onset, right dorsal IPL and right PMC showed their peak activity. The timing of these peaks indicates their involvement in coordinating the motor change. Interestingly, only left posterior temporal lobe showed peak activity concurrent with the compensation onset, but had a later SPRE peak. This could indicate the posterior temporal lobe's involvement in processing both the initial pitch shift and the sensory consequence to the motor compensation. Given the timing of the activations in right parietal and left posterior temporal lobe, these regions indicate involvement in integrating multisensory information and performing coordinate transformations between prediction errors and motor commands. In terms of current models of speech production, this would ascribe right parietal areas the computational role of the Kalman filter in the SFC model [Houde and Nagarajan, 2011] and the pseudoinverse of the Jacobian matrix (feedback mapping) in the DIVA model [Guenther and Vladusich, 2012]. The involvement of right PMC in

TABLE II. Regions of significant enhancement in response to the pitch shift during speaking compared to passive listening

Region	SPRE MNI peak voxel	SPRE Time-course	<i>t</i> -value	<i>P</i> -value
Left posterior middle temporal gyrus/occipital lobe	-55.0 -75.0 15.0	Peak: 225–250 ms Duration: 225–275 ms	3.77336	4.9×10^{-4}
Left middle occipital gyrus	-35.0 -80.0 10.0	Peak: 250–275 ms Duration: 225–275 ms	3.31917	0.0076
Right premotor cortex	55.0 -5.0 40.0	Peak: 175–200 ms Duration: 100–250 ms	4.08262	2.4×10^{-4}
Right inferior parietal lobe	60.0 -25.0 30.0	Peak: 175–200 ms Duration: 125–200 ms	3.40243	0.0017
Right supramarginal gyrus	65.0 -25.0 25.0	Peak: 250–275 ms Duration: 250–300 ms	3.53236	0.0012

Table contains regions, peak voxel location, duration, peak latency, *t*-value and *P*-value of regions of significant SPRE, enhancement in response to the pitch shift during speaking compared to passive listening. Regions included are cluster corrected, 30 voxels, $P < 0.01$.

initiating compensating corresponds with the DIVA model [Guenther and Vladusich, 2012], which assigns the feedback control map to right PMC. The SFC model [Houde and Nagarajan, 2011] similarly assigns premotor cortex the role of adding the auditory- and somatosensory-based state corrections to the previous state estimation to create the subsequent state estimate, but is agnostic to hemispheric lateralization.

Following compensation onset, the new motor activity continually increases compensation and the sensory consequence of the compensation is being detected and processed. During this time, we see right SMG, right MTG, and several right frontal regions show their peak activity. This window of continued monitoring, maintenance, and memory formation is not well described in either model of speech motor control, and is only alluded to in the SFC

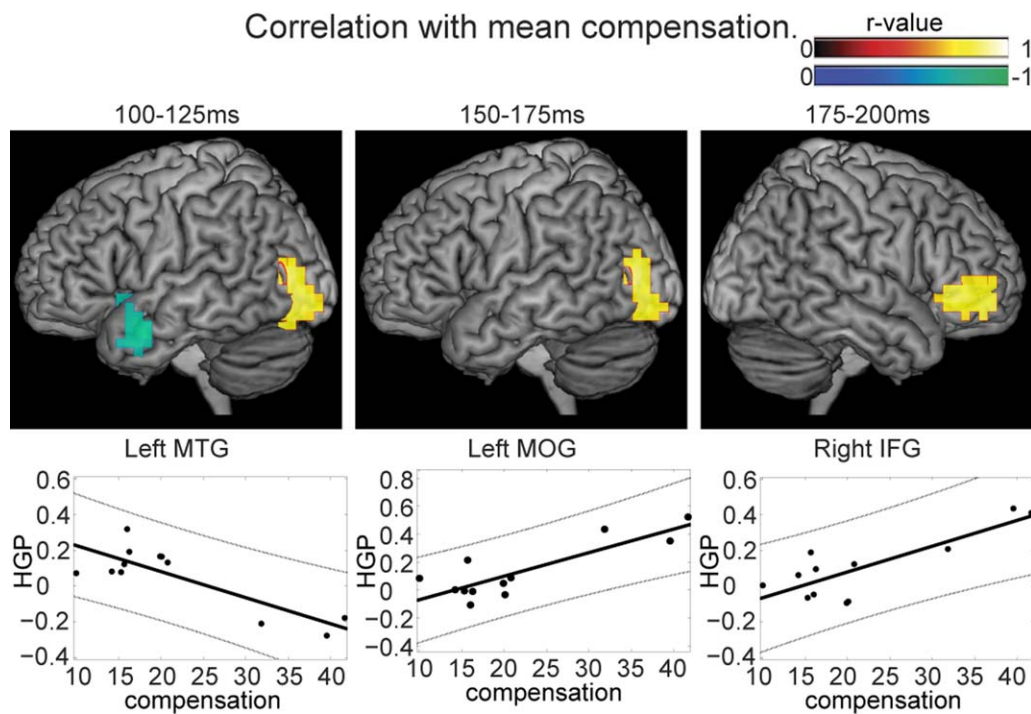


Figure 4. Neurobehavioral correlations across participants with mean compensation. Images are cluster corrected, 30 voxels, $P < 0.01$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE III. Significant neurobehavioral correlations with individual participant mean compensation

Region	MNI peak voxel	Time course	Robust R^2	P -value
Left middle temporal gyrus	-55.0 0.0 -25.0	Peak: 100–125 ms Duration: 100–150 ms	0.6134	5.1×10^{-4}
Left middle occipital gyrus/ posterior temporal lobe	-35.0 -80.0 5.0	Peak: 150–175 ms Duration: 100–225 ms	0.617	3.3×10^{-4}
Right inferior frontal gyrus	60.0 20.0 5.0	Peak: 125–150 ms Duration: 100–150 ms	0.4404	0.004
Right inferior frontal gyrus	35.0 35.0 0.0	Peak: 175–200 ms Duration: 100–225 ms	0.576	8.2×10^{-4}
Right middle frontal gyrus	45.0 55.0 -5.0	Peak: 175–200 ms Duration: 100–225 ms	0.5736	0.0022

Table contains regions, peak voxel location, duration, peak latency, robust r^2 , and P -value of regions with significant neurobehavioral correlations with individual participant mean compensation. Regions included are cluster corrected, 30 voxels, $P < 0.01$.

model as the task goals expressed in other regions of frontal cortex [Houde and Nagarajan, 2011].

Two previous studies have looked at HGP cortical responses to pitch-altered feedback using electrocorticography (ECoG) finding enhanced HGP in response to pitch-altered feedback in bilateral STG [Greenlee et al., 2013a] and in left posterior temporal lobe and premotor cortex [Chang et al., 2013]. Within the left hemisphere, Chang et al. [2013] reported the latency in enhanced responses proceeded from auditory to motor regions, and in one exemplary participant showed a similar trend in the right hemisphere. These results are in accord with our study, which found left-hemisphere sensory HGP enhancement preceded right-hemisphere PMC enhancement. But due to the limitation of the size of the ECoG grid, the ECoG studies could not address the role of the right hemisphere, or bilateral frontal regions.

In a previous ROI-based low-frequency evoked response study conducted by our group, we found cortical responses to an unexpected pitch shift enhanced in bilateral vSMG/pSTS, bilateral premotor cortex, right primary auditory cortex, and left higher order auditory cortex, but this enhancement was not correlated with vocal compensatory behavior. In this study, we expand upon the initial study by examining the induced high gamma band across the cortical mantle and their correlations with behavior. These findings extend our initial finding of cortical involvement in both hemispheres in response to an unexpected shift in auditory feedback by describing the cortical networks involved in recognizing and responding to a perceived error in pitch production. Importantly, here we examine not only the differential response to hearing a pitch shift during active speaking and passive listening, but also examine the high gamma power change from monitoring continuous of correct auditory feedback during speech.

Previous work studying the entire speech-motor network in fMRI with pitch [Parkinson et al., 2012], formant [Tourville et al., 2008], and somatosensory [Golfinopoulos et al., 2011] perturbations has provided insight into the

cortical network involved in responding to sensory feedback. Parkinson et al.’s [2012] study of pitch-altered feedback showed bilateral STG enhancement. In contrast, other fMRI studies of sensory feedback errors sustained across the whole trial have shown enhanced responses to altered auditory feedback in bilateral perisylvian, right ventral somatosensory, motor, and premotor cortices [Tourville et al., 2008], and enhanced responses to altered somatosensory feedback in bilateral ventral motor, right anterior SMG, right IFG-pars triangularis, and right ventral PMC [Golfinopoulos et al., 2011]. The restricted network reported in the Parkinson et al.’s study could potentially be attributed to task design, where they include responses to unshifted vocalizations, onset of a shift, and the offset of a shift, whereas the other fMRI studies [Golfinopoulos et al., 2011; Tourville et al., 2008] included sustained responses to feedback alterations. However, due to the temporal limitation of fMRI, all the aforementioned studies were not able to address the timing of these regions, and may miss cortical regions whose responses do not persist for the duration of the shift and cannot address an error mid-utterance. Furthermore, the timing of activity is essential in determining the possible functional roles of the region, for example, early responses indicate detection of the speech error while later responses are involved in generating the corrective motor response. Despite some similarities between the current findings and the cortical responses to formant-altered feedback [Tourville et al., 2008], there are also several noteworthy differences. For instance, the current findings include regions across the a greater extent of the left temporal lobe and into SSC while in the Tourville study, the left hemisphere enhancement did not extend beyond the posterior Perisylvian region. Similarly, in this study, right-hemisphere regions include IPL, SMG, and PMC that were not seen in the Tourville et al.’s study. Given the timing and power of right SMG and right dorsal IPL in our study, the role of these regions in auditory feedback control is an important addition for any model of speech production.

In this study, we saw a very large response in right parietal regions throughout the auditory error, particularly in right SMG. Yet right SMG was not correlated with the individual participants behavioral compensation. Taken together, right SMG is indicated as highly involved in coordinating the compensation response. But, importantly, this does not directly correspond with coordinating a larger motor change, but instead could be involved in resolving the conflict between the two sensory modalities. The exact role of SMG in the network, either facilitating or dampening the motor response, can be further interrogated with virtual or clinical lesions.

The work presented here has strong implications for our understanding of the neuroscience of speech. The results challenge conventional models of speech production that posit left lateralization of speech production [Dronkers, 1996; Hickok et al., 2011]. Instead, this work provides evidence that auditory error detection occurs in both hemispheres. These findings challenge models of speech production to address the feedback control of speech and the importance of interhemispheric communication to describe the neuroscience of speech production. Importantly, this work also impacts models of motor control by clarifying the role of reafferent information and sensory feedback during complex, ongoing movements.

ACKNOWLEDGMENTS

The authors thank Susanne Honma, Danielle Mizuiru, Anne Findlay, Caroline Niziolek, Leighton Hinkley, Zariyah Agnew, and members of the Biomagnetic Imaging Laboratory and Speech Neuroscience Lab for technical assistance and comments on the manuscript.

REFERENCES

- Behroozmand R, Karvelis L, Liu H, Larson CR (2009): Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin Neurophysiol* 120:1303–1312.
- Behroozmand R, Larson CR (2011): Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neurosci* 12:54.
- Behroozmand R, Liu H, Larson C (2011): Time-dependent neural processing of auditory feedback during voice pitch error detection. *J Cognit Neurosci* 23:1205–1217.
- Broca P (1861): Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche. *Bull Soc Anthropol* 6:235–238.
- Burnett TA, Freedland MB, Larson CR, Hain TC (1998): Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103:3153–3161.
- Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013): Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc Natl Acad Sci USA* 110:2653–2658.
- Cogan GB, Thesen T, Carlson C, Doyle W, Devinsky O, Pesaran B (2014): Sensory-motor transformations for speech occur bilaterally. *Nature* 507:1–7.
- Crone NE, Sinai A, Korzeniewska A (2006): High-frequency gamma oscillations and human brain mapping with electrocorticography. *Progr Brain Res* 159:275–295.
- Dalal SS, Guggisberg AG, Edwards E, Sekihara K, Findlay AM, Canolty RT, Berger MS, Knight RT, Barbaro NM, Kirsch HE, Nagarajan SS (2008): Five-dimensional neuroimaging: Localization of the time–frequency dynamics of cortical activity. *NeuroImage* 40:1686–1700.
- Dalal SS, Zumer JM, Guggisberg AG, Trumpis M, Wong DDE, Sekihara K, Nagarajan SS (2011): MEG/EEG source reconstruction, statistical evaluation, and visualization with NUTMEG. *Comput Intelli Neurosci* 2011:1–17.
- Dronkers NF (1996): A new brain region for coordinating speech articulation. *Nature* 384:159–161.
- Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010): Single-trial speech suppression of auditory cortex activity in humans. *J Neurosci* 30:16643–16650.
- Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Nieto-Castanon A, Guenther FH (2011): fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage* 55:1324–1338.
- Greenlee JDW, Behroozmand R, Larson CR, Jackson AW, Chen F, Hansen DR, Oya H, Kawasaki H, Howard MAIII (2013a): Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS ONE* 8:e60783
- Greenlee JDW, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H, Howard MA (2011): Human auditory cortical activation during self-vocalization. *PLoS One* 6:e14744.
- Greenlee J, Behroozmand R, Narayanan N, Kingyon JR, Larson C, Oya H, Kawasaki H, Howard MA (2013b): Sensorimotor integration during human self-vocalization: Insights from invasive electrophysiology. *J Acoust Soc Am* 133:3520–3520.
- Guenther FH, Vladusich T (2012): A neural theory of speech acquisition and production. *J Neurolinguistics* 25:408–422.
- Hickok G, Houde J, Rong F (2011): Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69:407–422.
- Houde JF, Nagarajan SS (2011): Speech production as state feedback control. *Front Hum Neurosci* 5:1–14.
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002): Modulation of the auditory cortex during speech: An MEG study. *J Cognit Neurosci* 14:1125–1138.
- Kort NS, Nagarajan SS, Houde JF (2014): A bilateral cortical network responds to pitch perturbations in speech feedback. *NeuroImage* 86:525–535.
- Lane H, Webster JW (1991): Speech deterioration in postlingually deafened adults. *J Acoust Soc Am* 89:859–866.
- McAulay RJ, Quatieri TF (1986): Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans Acoustics Speech Signal Process ASSP* 34:744–754.
- Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R (2005): Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309:951–954.
- Muller-Preuss P, Ploog D (1981): Inhibition of auditory cortical neurons during phonation. *Brain Research* 215:61–76.
- Niziolek CA, Nagarajan SS, Houde JF (2013): What does motor efference copy represent? Evidence from speech production. *J Neurosci* 33:16110–16116.
- Parkinson AL, Flagmeier SG, Manes JL, Larson CR, Rogers B, Robin DA (2012): Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage* 61:314–322.

- Parsons TW (1986): Voice and Speech Processing. New York, NY: McGraw-Hill Book Company.
- Price CJ (2010): The anatomy of language: A review of 100 fMRI studies published in 2009. *Ann New York Acad Sci* 1191:62–88.
- Sekihara K, Nagarajan SS (2010): 245 Adaptive Spatial Filters for Electromagnetic Brain Imaging.
- Singh KD, Barnes GR, Hillebrand A (2003): Group imaging of task-related changes in cortical synchronisation using nonparametric permutation testing. *NeuroImage* 19:1589–1601.
- Tourville JA, Reilly KJ, Guenther FH (2008): Neural mechanisms underlying auditory feedback control of speech. *NeuroImage* 39:1429–1443.
- Zarate JM, Zatorre RJ (2008): Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage* 40:1871–1887