A bilateral cortical network responds to pitch perturbations in speech feedback

Naomi S. Kort[a,b], **Srikantan S. Nagarajan**[a], and **John F. Houde**[c]

[a] Department of Radiology, University of California, San Francisco, USA

[b] Joint Graduate Group in Bioengineering, University of California, San Francisco and Berkeley, USA

[c] Department of Otolaryngology, University of California, San Francisco, USA

Abstract

Auditory feedback is used to monitor and correct for errors in speech production, and one of the clearest demonstrations of this is the pitch perturbation reflex. During ongoing phonation, speakers respond rapidly to shifts of the pitch of their auditory feedback, altering their pitch production to oppose the direction of the applied pitch shift. In this study, we examine the timing of activity within a network of brain regions thought to be involved in mediating this behavior. To isolate auditory feedback processing relevant for motor control of speech, we used magnetoencephalography (MEG) to compare neural responses to speech onset and to transient (400ms) pitch feedback perturbations during speaking with responses to identical acoustic stimuli during passive listening. We found overlapping, but distinct bilateral cortical networks involved in monitoring speech onset and feedback alterations in ongoing speech. Responses to speech onset during speaking were suppressed in bilateral auditory and left ventral supramarginal gyrus/ posterior superior temporal sulcus (vSMG/pSTS). In contrast, during pitch perturbations, activity was enhanced in bilateral vSMG/pSTS, bilateral premotor cortex, right primary auditory cortex, and left higher order auditory cortex. We also found speaking-induced delays in responses to both unaltered and altered speech in bilateral primary and secondary auditory regions, the left vSMG/ pSTS and right premotor cortex. The network dynamics reveal the cortical processing involved in both detecting the speech error and updating the motor plan to create the new pitch output. These results implicate vSMG/pSTS as critical in both monitoring auditory feedback and initiating rapid compensation to feedback errors.

Keywords

speech motor control; sensorimotor; auditory feedback; magnetoencephalography

1. Introduction

Speaking is a complex motor process where the goal is to produce sounds that convey an intended message; if the right sounds are produced, the listener will correctly comprehend the speaker's message. It is not surprising, therefore, that speakers monitor their sound output, and that this auditory feedback exerts a powerful influence on their speech. Indeed, the motor skill of speaking is very difficult to acquire without auditory feedback, and, once acquired, the skill is gradually lost in the absence of auditory feedback (Cowie et al., 1982). The control of the fundamental frequency of speech (f0), perceived as pitch, is rapidly lost in the absence of auditory feedback (Lane and Webster, 1991), demonstrating that pitch, along with other suprasegmental features, requires aural monitoring. However, the control of pitch during speech, given that the auditory feedback is both noisy and delayed, is still poorly understood. Precise control of pitch is essential for the prosodic content of speech- providing the speaker with information on emphasis, emotional content, and form of the utterance (eg. question or statement). The failure to properly modulate pitch is an impediment in communication, and the result of several neurological and psychiatric disorders, including schizophrenia and Parkinson's disease.

When auditory feedback is present, its alteration can have immediate effects on ongoing production. It has long been known, for example, that delaying auditory feedback can immediately render a speaker disfluent (Lee, 1950; Yates, 1963). More recently, experiments have altered specific features of auditory feedback, and the responses of speakers have been particularly revealing. In response to brief perturbations of the pitch, loudness, and formant frequencies of their auditory feedback, speakers will make quick adjustments to their speech that reduce the perceived effect of the perturbations on their auditory feedback (Chang-Yit et al., 1975; Houde and Jordan, 2002; 1998; Lane and Tranel, 1971; Lombard, 1911). These experiments, in which a feedback perturbation elicits a quick compensatory response, demonstrate the existence of speech sensorimotor pathways in the CNS that convey corrective information from auditory areas to speech motor areas during ongoing speaking. Behavioral experiments have further shown that auditory feedback is important for online control of pitch both in words and sentences. Altered pitch feedback on the first syllable of a nonsense word impacts the pitch of the second syllable, even when the first syllable is short and unstressed (Donath et al., 2002; Natke and Kalveram, 2001). Compensation to pitch-altered feedback influencing either the stress in a sentence (Patel et al., 2011) or the form of the sentence (Chen et al., 2007) has been observed. These studies have shown the importance of auditory feedback in controlling pitch both within a syllable, and on a suprasegmental level. This behavior may be used to compensate for disturbances in output pitch known to arise from a number of natural sources, including an error in the complex coordination of vocal fold tension (Lane and Webster, 1991), aerodynamic instability, and even heartbeat (Orlikoff and Baken, 1989). The rapid compensation to altered pitch feedback occurs both in continuous speech (sentence production) and during

single vowel phonation. This is not surprising given that phonation is an important part of speech. While the pitch perturbation response has been well characterized in behavioral studies, very little is known about the neural substrate of these sensorimotor pathways.

Until recently, the study of the neural circuitry monitoring self-produced speech has primarily focused on auditory cortex during correct (unaltered) vocalization. Work in non-human primates found that the majority of call-responsive neurons were inhibited during phonation (Eliades and Wang, 2008; 2005; 2002; Muller-Preuss and Ploog, 1981). Extensive work has been done to study this suppression effect in humans (Chang et al., 2013; Curio et al., 2000; Greenlee et al., 2011; Houde et al., 2002; Ventura et al., 2009). Studies using magnetoencephalography (MEG) in humans similarly found suppressed neural activity in auditory areas during self-produced speech compared to the neural activity while listening to the playback of recorded speech (Curio et al., 2000; Houde et al., 2002). This effect has been termed speaking-induced suppression (SIS). SIS is a specific example of the broader phenomenon of motor-induced suppression (MIS) (Aliu et al., 2009), where sensory responses to stimuli triggered by self-initiated motor act are suppressed. However, in these studies, it was difficult to localize the SIS effect to specific areas of auditory cortex. Better localization of SIS has been seen in studies based on intracranial recording (ECoG) in neurosurgery patients. One study found that the SIS response only occurred in circumscribed areas of auditory cortical areas, and in fact some areas show an anti-SIS effect (Greenlee et al., 2011). Another ECoG study focusing on responses in the left hemisphere found SIS primarily in electrodes clustered in posterior superior temporal cortex (Chang et al., 2013). However, the spatial coverage of ECoG is limited, and, as yet, no studies to date have examined more completely the spatial distribution of SIS along the speech sensorimotor pathways.

Monitoring feedback to confirm that speech motor acts give rise to the expected auditory outputs (resulting in SIS) is only one important role of the speech sensorimotor pathways. When feedback is altered and mismatches expectations, these pathways take on the additional role of conveying the mismatch to motor areas and generating a compensatory production change. What are the neural correlates of this process? Several SIS studies have showed that altering feedback at speech onset reduces SIS (Heinks-Maldonado et al., 2006; Houde et al., 2002). A few recent studies have looked at responses to feedback alterations during ongoing speech to pitch perturbations of auditory feedback using EEG (Behroozmand et al., 2009; Behroozmand and Larson, 2011; Behroozmand et al., 2011). These studies found that perturbations of ongoing vocal feedback evoked larger responses than did perturbations passively heard during the subsequent playback of feedback. We term this effect speech perturbation response enhancement (SPRE). Although EEG studies to date have not been able to localize SPRE to particular brain areas, in two recent ECoG studies the spatial distribution of SPRE was mapped (Chang et al., 2013; Greenlee et al., 2013). One study from our group looked at high gamma responses to pitch-altered feedback in the left hemisphere, finding SPRE responses clustered in ventral premotor cortex and posterior superior temporal cortex including the parietal-temporal junction (Chang et al., 2013). A second study, using ECoG, found enhanced evoked and high gamma responses in both left and right mid-to-anterior superior temporal gyrus (Greenlee et al., 2013). Coverage limitations of ECoG restrict the analysis to the individual subject's placement of the grid

electrodes. Furthermore, since each patient's grid is uniquely placed ECoG studies cannot easily compare results across subjects, or across hemispheres within a subject.

The relationship between SIS and SPRE is largely unexplored. A recent study in marmosets found auditory neurons that show suppression at the onset of vocalization are more likely to have an enhanced response to a perturbation, suggesting a direct link between the mechanisms suppressing self-produced speech to those recognizing errors in self-produced speech (Eliades and Wang, 2008). In contradiction with these findings, an ECoG study of SPRE found only a small number of electrodes displaying both SIS and SPRE, in contrast to a larger number of electrodes which preferentially display one or the other (Chang et al., 2013).

To more completely define the speech sensorimotor network, other studies have used whole-head functional imaging to look at how the speech motor system responds to feedback alterations. The spatial resolution of fMRI has allowed several studies to identify specific areas in the auditory and motor cortices that respond when auditory feedback is altered during speaking (Parkinson et al., 2012; Tourville et al., 2008; Toyomura et al., 2007). When the results from the altered auditory feedback fMRI studies are combined with other pertinent speech studies, several cortical regions emerge as likely computational nodes in processing and responding to an error in auditory feedback (Andersen et al., 1997; Buchsbaum et al., 2011; Fu et al., 2006; Gelfand and Bookheimer, 2003; Grefkes and Fink, 2005; Hickok et al., 2011; 2009; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Shadmehr and Krakauer, 2008; Tourville et al., 2008; Toyomura et al., 2007): primary auditory cortex, the superior temporal gyrus/middle temporal gyrus (STG/MTG), ventral supramarginal gyrus/posterior superior temporal sulcus (vSMG/pSTS) and premotor cortex, summarized in Table 1.

Unfortunately, the lack of temporal resolution in the fMRI studies and the limitations of ERP studies leave many questions unanswered. fMRI studies have helped to define the speech sensorimotor pathways, but due to the lack of temporal resolution, these studies have not revealed the dynamics of the network's behavior. In particular, do the areas identified in the fMRI studies also exhibit SPRE, and, if so, what is the time course of SPRE over these areas? Do the areas that exhibit SPRE also exhibit SIS? Do the areas that exhibit SPRE show correlations across subjects with behavior? The advent of new source localization algorithms for magnetoencephalography (MEG) gives us a unique, unexploited opportunity to answer these questions using the millisecond time resolution of MEG.

The present study used MEG to investigate the cortical neural responses at speech onset (to examine SIS) and during brief, unexpected shifts in the pitch of subjects' audio feedback (to examine SPRE) during the phonation of a single vowel. By using a single vowel utterance we were able to isolate phonation from additional (linguistic) aspects of speech to specifically study pitch production. In this study, we tested several hypotheses. First, given that SIS has been shown to be involved in auditory self-monitoring, we hypothesized that there would be spatial overlap between the monitoring role of SIS and the error recognition part of the SPRE network. Second, we hypothesized that SPRE would be seen to propagate through the speech sensorimotor network as the error is recognized and processed, and

ultimately induce a compensatory response. Third, we hypothesized that cortical responses to the perturbation during speaking would be correlated with compensation across subjects.

2. Materials and Methods

2.1 Subjects

Eleven healthy right-handed English-speaking (4 female) volunteers participated in this study. All participants gave their informed consent after procedures had been fully explained. The study was performed with the approval of the University of California, San Francisco Committee for Human Research.

2.2 MEG Recording

Neural data were recorded in awake subjects laying in the supine position with a whole head MEG system (MISL, Coquitlam, British Columbia, Canada) consisting of 275 axial gradiometers (sampling rate = 1200Hz). Three fiducial coils (nasion, left/right preauricular points) were placed to localize the position of the head relative to the sensor array. These points were later co-registered with an anatomical MRI to generate head shape.

2.3 Experimental Design and Procedure

The subjects spoke into an MEG-compatible optical microphone and received auditory feedback through earplug earphones. They observed a projection monitor directly in their line of sight. The screen background was black. A trial began when three white dots appeared in the center of the screen. Each dot disappeared one by one to simulate a count-down (3-2-1). When all three dots disappeared and the screen was completely black, subjects were instructed to follow the instructions corresponding to the block- either speaking or listening. The trial was terminated with the visual cue indicating the remaining trials before a break. A schematic of the experimental setup is shown in figure 1.

The experiment consisted of 4 blocks of 74 trials each, with brief, self-paced breaks every 15 trials. The blocks were arranged in an alternation of conditions: blocks 1 and 3 were the Speaking Condition; blocks 2 and 4 were the Listening Condition. In the Speaking Condition, subjects were instructed to produce a sustained utterance of the vowel /a/ until the termination cue. During the phonation the subjects heard one 100-cent pitch perturbation lasting 400ms whose onset was jittered in time from speech onset. In each trial, the direction of the pitch shift was either positive (i.e., raising the perceived pitch) or negative (lowering the perceived pitch). Equal numbers of positive and negative perturbation trials were pseudorandomly distributed across the experiment. The jittered perturbation onset prevented the subject from anticipating the timing of the perturbation while the pseudorandom selection of a negative and positive pitch shift prevented the subject from anticipating the direction of the perturbation. In the Listening Condition, subjects saw the same visual prompts but only 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, providing a method of comparison to extract speaking-specific activity. The auditory input through the earphones was adjusted prior to

commencement of the experiment to a level that subjects reported their auditory feedback was the same as or slightly louder than expected.

2.4 Audio analysis

Speech recordings were analyzed for pitch throughout the utterance. The data from ten subjects were included in this analysis. One subject was eliminated from the behavioral data analysis due to a corrupted file. Timing and magnitude of compensation was determined as follows. For each subject, the pitch contour of each perturbation type, +100 cent or −100 cent, were averaged together. Absolute frequency (hertz) was changed to cents peak response change from pre-perturbation baseline by: cents change = $100 \times [12 \times \log_2$(pitch response peak (Hz)/mean pitch frequency of pre-perturbation baseline (Hz))]. Mean percent compensation was calculated as −100*(cents change)/(cents of applied pitch shift). The negative sign makes the compensation a positive value. The pitch analysis was performed on each subject's single trial audio data. The audio data were sampled at 11025 Hz and both the microphone and feedback were recorded and analyzed. Each trial of the data was recorded in 32-sample frames and pitch was estimated for each of these frames using the standard autocorrelation method (Parsons, 1986). The resulting frame-by-frame pitch contour was then smoothed with a 20Hz, 5th order, low pass Butterworth filter. Trials with erroneous pitch contours were removed. The largest possible pre-perturbation interval was used to establish a baseline, constrained by the minimum time between voice onset and perturbation onset, and the mean and standard deviation of this baseline interval was calculated. A subject's response onset was conservatively set to occur when the mean pitch time course deviates from the baseline by two standard deviations. The magnitude and onset of the compensation were determined for each subject and then averaged to create the grand-average compensation magnitude and onset.

2.5 MEG data preprocessing

The MEG sensor data were 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.

2.6 Reconstructing source time courses

The time courses of source intensities at selected ROIs were computed using an adaptive spatial filtering technique using a signal bandwidth of 0-300 Hz (Bardouille et al., 2006; Oshino et al., 2007; Robinson and Vrba, 1999; Sekihara et al., 2004; Vrba and Robinson, 2001) with CTF software tools. A virtual sensor was created for each of the 4 defined locations for each task (speak or listen) and for each hemisphere (left or right) for every subject. The virtual channel data, i.e. source time course, were then filtered from 2 to 40 Hz. Each subject's data were normalized using the mean and standard deviation of the baseline activity of all sensors to compute a z-score. The z-scored results were used for averages and statistics. The latency and the magnitude of the peaks for each individual subject were calculated.

T-tests were computed over the eleven subjects comparing the speaking condition to the listening condition at each corresponding peak for both magnitude and latency differences. Multiple comparisons were corrected for using the 5% False Discovery Rate using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995), which set significance at $p < 0.0096$. Since only a few planned comparisons were performed on each virtual sensor (only peak values were tested), all results $p<0.05$ are reported, and all uncorrected p-values given.

2.7 ROI selection

When the results from the altered auditory feedback fMRI studies are combined with other pertinent speech studies and theoretical models, four well-defined neural regions emerge as likely computational nodes in processing and responding to an error in auditory feedback. These four cortical regions became the regions of interest used for the virtual sensor analysis. These four ROI's are primary auditory cortex, the superior temporal gyrus/middle temporal gyrus (STG/MTG), ventral supramarginal gyrus/posterior superior temporal sulcus (vSMG/pSTS) and premotor cortex, and the supporting evidence for their role as computational nodes are summarized in Table 1 (Andersen et al., 1997; Buchsbaum et al., 2011; Fu et al., 2006; Gelfand and Bookheimer, 2003; Grefkes and Fink, 2005; Hickok et al., 2011; 2009; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Shadmehr and Krakauer, 2008; Tourville et al., 2008; Toyomura et al., 2007). The virtual sensor seeds were anatomically determined as the center of mass of the Brodmann areas (BA) corresponding with the aforementioned sensorimotor nodes: 41/42, 21/22, 40 and 6. The time course represents the activity in the region of interest approximately 1 centimeter from the center seed voxel due to spatial blur of the MEG source localization.

3. Results

3.1 Auditory pitch perturbations induce compensation

The subjects responded to the pitch-shifted auditory feedback by adjusting the pitch of their voice to oppose the direction of the perturbation. The grand average across subjects is shown in figure 2. The average compensation to the perturbation across all subjects was 18.98% (range: 8.81% to 41.09%). The compensation for all subjects had a mean onset of 210.9ms (standard error = 19.87) after the perturbation onset and mean peak latency of 531.8ms (standard error = 14.07). The latency of the compensation onset was determined for each subject as the point in which the pitch contour crossed the two standard deviation line determined by the baseline period, and then averaged across subjects. Although the latency of the compensation onset in the averaged F0 contour seen in figure 2 appears to occur around 140ms, this is an inaccurate appearance of averaging across subjects. Compensation in response to the positive pitch shift and to the negative pitch shift were not statistically different either in magnitude (p=0.744) or in onset time (p = 0.187). Since the behavioral response to the positive and negative pitch shift trials was indistinguishable, all shifted trials were subsequently analyzed together.

3.2 Cortical neural activity in response to speech onset

Figure 3 shows the results for the four bilateral source time course reconstructions centered with zero at the time of speech onset in both the speaking and the listening conditions. In primary auditory cortex, we saw variations of the standard three-peak response to auditory input (M50, M100, M200). In the left primary auditory cortex this three peak pattern was clear in both the listening condition, (peaks at 30.9ms, 91.7ms, 188.6ms after speech onset), and in the speaking condition (peaks at 30.5ms, 92.9ms, 192.58ms after speech onset). In the right hemisphere, it was less clear that the pattern of response peaks fitted the standard three-peak auditory response pattern. In the listen condition, there were response peaks at 57.9ms, 84.8ms, 133.3ms, but there were also peaks at 31.7ms and 237.5ms. In the speaking condition, there were three clear peaks, but their correspondence to the standard auditory response pattern is complicated by the pattern of latencies (35.6ms, 118.0ms, 235.3ms) where the M50 peak is early, and the M200 peak is late, and by the unusually large amplitude of the M50 peak.

3.3 Speaking induced suppression (SIS)

The M100 response in left primary auditory cortex occurred at 92ms and showed SIS ($p = 0.0448$), with a mean SIS magnitude (z-scored), calculated by subtracting the speaking peak value from the listening peak value, equal to 9.264. In right primary auditory cortex, SIS was seen in both the M100 response occurring at 84.8ms (mean magnitude 6.3417, $p = 0.0312$), and in the M200 response occurring at 192.3ms (mean magnitude 2.1731, $p = 0.0184$). Interestingly, the earliest peak (35.6ms) in right primary auditory cortex had anti-SIS ($p = 0.0001$) with a mean magnitude of 6.1625.

Left STG/MTG did not show the SIS effect, but instead showed speaking-induced delays of the M100 response latency, with the M100 occurring 14.7ms later in the speaking condition at 114.5ms than in the listening condition at 99.8ms ($p = 0.0175$). In contrast, in right STG/MTG the m50 showed a listening-induced delay, with the first peak in the speaking condition (37.5ms) preceding the first peak in the listening condition (47.6ms) by 10.1ms ($p = 0.0273$). Right STG/MTG also showed SIS in the m100 response occurring at 87.7ms ($p = 0.0401$), with a mean magnitude of 5.493.

The most prominent and early SIS effects were seen in left vSMG/pSTS, peaking 84.8ms after the speech onset with an average magnitude equal to 5.0239 ($p = 0.0010$). In contrast, right vSMG/pSTS did not have a magnitude difference between the speaking and listening conditions, with the significance values for SIS at the three prominent peaks being $p=0.94$, $p=0.64$, and $p=0.80$. Neither left nor right vSMG/pSTS showed significant latency differences in the peaks between the speaking and listening conditions.

In premotor cortex, one might expect bilateral enhanced activity (i.e., anti-SIS) in the speaking condition compared to the listening condition, especially before speech onset. Yet the only significant enhancement in the speaking condition (magnitude 4.0518, $p = 0.0007$) was 38.3ms after speech onset in the right hemisphere. On the other hand, SIS was seen bilaterally in the premotor cortices. Left premotor cortex showed SIS at 141.2ms (magnitude 2.4004, $p = 0.0357$), while right premotor cortex showed SIS at 74.7ms (magnitude 4.5169,

p = 0.0269), at 218.9ms (magnitude 1.8163, p = 0.0459), and at 292.3ms (magnitude 2.023, p = 0.0331).

3.4 Cortical neural activity in response to auditory pitch perturbation

3.4.1 Primary auditory cortex responses to auditory pitch perturbation in speaking and listening—Source time courses for the four bilateral ROI's time-locked to the perturbation onset are shown in figure 4. Bilateral primary auditory cortex showed a pronounced three-peak response pattern to the auditory pitch perturbation in both the speaking and listening conditions. This three-peak response, while following the typical MEG auditory response pattern, was delayed as compared to the response pattern to the onset of speech in the first two peaks and in a late response peak. In the listening condition, left primary auditory cortex had peaks at 76.8ms, 140.3ms, 235.2ms following the perturbation onset while in the right hemisphere had peaks at 69.3ms, 148.6ms, 217.6ms following the perturbation onset. Similarly in the speaking condition, left primary auditory cortex had peaks at 96.0ms, 164.8ms, 239.1ms and right primary auditory cortex had peaks at 89.2ms, 146.6ms, and 226.3ms.

3.4.2 Speaking-specific responses prior to and during the onset of compensation—Responses in bilateral primary auditory cortex showed significant delays in the speaking condition when compared to the listening condition. In the left primary auditory cortex the first peak in the speaking condition (96.0ms) was delayed (p = 0.0005) by 19.2ms from the first peak in the listening condition (76.8ms), while in the right primary auditory cortex the first prominent peak in the speaking condition (89.1ms) was delayed (p = 0.00001) by 19.8ms from the peak in the listening condition (69.3ms). In the second peak, left primary auditory cortex showed a 24.5ms delay in the speaking condition (164.8ms) to the listening condition (140.3ms), p = 0.0002.

Bilateral STG/MTG had speaking induced delays, but at different latencies. The speaking condition in left STG/MTG had a peak at 178.2ms, 13.3ms delayed to the corresponding peak in the listening condition at 164.9ms (p = 0.0027). In contrast, right STG/MTG showed speaking induced delays early, with the first peak in the speaking condition occurring at 85.3ms, 21.2ms delayed to the peak in the listening condition at 64.1ms (p = 0.0002).

Left vSMG/pSTS had three distinct peaks with significant SPRE. The first, and largest SPRE in left vSMG/pSTS occurred at 149.6ms with a magnitude of 2.2757 (p = 0.0033). This SPRE had a corresponding speaking induced delay, with the peak in the speaking condition (149.6ms) 22.9ms after the peak in the listening condition (126.7ms), p < 0.0001. The following peak, at 186.7ms, also showed significant SPRE with a magnitude of 2.2304 (p = 0.005). The third SPRE peak in left vSMG/pSTS occurred following the compensation. Right vSMG/pSTS showed one wide SPRE peak at 167.3ms, with a magnitude of 2.3852 (p = 0.0306).

Bilateral premotor cortices (PMC) showed increasing activity in response to the pitch shifted feedback in both the speaking and the listening condition after the perturbation onset. In left premotor cortex the response in the speaking condition had a greater magnitude and duration than the response in the listening condition. Left premotor cortex showed a SPRE response

at 190.1ms with a magnitude of 2.9751 (p = 0.0399). Right premotor cortex had several peaks with significant speaking-induced delays and a late peak shows SPRE. The first peak with a speaking induced delay occurred at 66.8ms in the speaking condition, 8.4ms after the peak in the speaking condition at 58.4ms (p = 0.004). Later, the peak in the speaking condition at 125.1ms followed the corresponding peak in the listening condition (109.9ms) by 15.2ms (p = 0.002). Finally, the peak in the speaking condition at 181.9ms had a corresponding peak 21.2ms earlier in the listening condition at 160.7ms (p <0.0001).

3.4.3 Speaking-specific responses following the onset of compensation—

Following the compensation, left primary auditory cortex showed one speaking specific response, a 16.6ms speaking-induced delay (p = 0.0001), with the peak in the speaking condition occurring at 288.1ms and the peak in the listening condition occurring at 271.5ms. Right primary auditory cortex showed a 8.7ms speaking-induced delay between the peak in the speaking condition at 226.3ms and the peak in the listening condition at 217.6ms (p = 0.0120). SPRE was observed in right primary auditory cortex at 226.3ms after the perturbation with a magnitude of 1.583 (p = 0.0062), and at 326.2 ms (magnitude= 2.4026, p = 0.0064).

The only speaking specific response in bilateral STG/MTG following the compensation occurred in left STG/MTG. Left STG/MTG showed SPRE in a peak 300.4ms after the perturbation onset with a magnitude of 1.9203. Similarly, the only SPRE peak in vSMG/pSTS following the compensation was in left vSMG/pSTS. A late peak, occurring at 291.6ms, showed SPRE with a magnitude of 2.1761 (p = 0.0019). Conversely, right vSMG/pSTS had greater magnitude in the listening condition as compared to the speaking condition at 220.5ms with a magnitude of 1.9119 (p = 0.0456).

Left premotor cortex also showed SPRE in a late peak at 276.7ms, with a magnitude of 1.9772 (p = 0.0229). Similarly, right premotor cortex showed SPRE in the peak at 288.9ms showed SPRE (magnitude = 1.7373, p = 0.023). A late peak in right premotor cortex during the speaking condition (258.2ms) preceded the corresponding peak in the listening condition (243.5ms) by 14.7ms (p< 0.0001).

3.5 Correlation of cortical activity with behavioral compensation

Correlations between an individual subject's behavioral compensation and cortical activity during speaking for each peak showing SPRE were calculated. One SPRE peak showed significant correlation across subject compensation, the peak in right premotor cortex with the mean latency of 288.9ms. This response in this cortical area showed a significant, yet small, negative linear regression with behavioral compensation (p=0.049, r-squared = 0.401). The peak at 186.7ms in left vSMG/pSTS conversely showed a trend towards a significant positive correlation with compensation (p=0.098, r-squared = 0.305).

4. Discussion

In this study, we used MEG to examine auditory feedback processing within a subset of the speech sensorimotor network. This study addresses an integral part of speech, phonation. Phonation conveys important prosodic and affective information that are an invaluable part

of speech. Therefore understanding the neural circuitry that controls phonation is important and relevant to the understanding of speech. In this study, we looked at responses to speech onset and to brief, unexpected perturbations of pitch feedback, and compared responses seen during speaking with those seen during subsequent listening to playback of the feedback. In our analysis of speech onset responses, we found that speaking-induced suppression (SIS) occurs well beyond auditory cortices. While we replicate SIS in bilateral primary auditory cortex, we found that the most significant SIS effect occurs in left vSMG/pSTS, and interestingly, we document for the first time speaking induced delays in left STG/MTG. In our analysis of perturbation responses, we found speech perturbation response enhancement (SPRE) in a large bilateral network of regions, with maximal SPRE seen in the left vSMG/pSTS but also including right primary auditory, left STG/MTG, right vSMG/pSTS, and bilateral premotor cortex. We also observed significant speaking induced delays in perturbation responses within bilateral primary and secondary auditory cortices, left vSMG/pSTS, and right premotor cortex. We found little evidence of a correlation between SPRE and compensation across subjects. We discuss these findings in the context of the growing literature on responses to unaltered and altered speech feedback.

4.1 The effect of the onset of speech on MEG magnitude response

While previous work studying the neural response to unaltered self-produced speech has demonstrated SIS in auditory areas, these previous studies did not examine the spatial distribution of the SIS response beyond auditory cortices (Curio et al., 2000; Houde et al., 2002; Ventura et al., 2009). In this study we explored four nodes of the speech sensorimotor pathways and found SIS responses at a number of locations. Although SIS responses were seen in bilateral primary auditory cortex and right STG/MTG, the most prominent SIS responses were localized primarily to left vSMG/pSTS. Previous reports averaged sensor-space data from multiple sources across temporal cortex (Curio et al., 2000; Houde et al., 2002), and so include responses from sources in both primary auditory cortex and vSMG/pSTS. These studies also found weaker SIS effects in the right hemisphere. In this study, the SIS in the responses across the right temporal cortex did not survive conservative correction for multiple comparisons. Yet, if these responses from right primary and secondary auditory cortices were summed, significant SIS could occur. Of interest is the large early response (35.6ms) in right primary auditory cortex in the speaking condition. This temporally and spatially specific anti-SIS response combines with the surrounding SIS to create a pattern consistent with the distributed representation of SIS reported by Greenlee and colleagues using electrocorticography (ECoG) (Greenlee et al., 2011).

4.2 The effect of speaking on MEG response latency

Speaking specific responses to the onset of speech were not limited to SIS. Significant speaking induced delays were also found in this study. Even though latency differences have not received much attention in the speech production literature, their existence here suggests the possibility of additional speaking-specific processing, where information is conveyed by MEG peak delays in the speaking condition. Indeed, prior work in speech perception has already shown that peak latency can carry important information. For example, latency of the m100 response during passive listening varies significantly across vowels or across consonants (Gage et al., 1998; Poeppel et al., 1997). In this study, a speaking-induced delay

was found in secondary auditory areas (STG/MTG) in response to the onset of unaltered feedback. Previous studies have reported conflicting information on a latency difference between responses in active speaking and passive listening conditions. Curio et al. report that the m100 in the speaking condition was significantly delayed compared to the m100 in the listening condition in bilateral auditory areas (Curio et al., 2000). In another study, Houde et al. did not report a latency difference between the speaking and the listening conditions with unaltered auditory feedback, but when feedback was altered to produce noisy auditory feedback, the speaking condition was delayed (Houde et al., 2002).

Speaking induced delays could also play an important role in the preliminary auditory processing of the pitch-altered feedback. The results in this study show discrete delayed responses to the altered feedback in the speaking condition when compared to the peak latencies in the listening condition. Given that a change in fundamental frequency is a weaker acoustic event than the onset of speech, it is not surprising that the three-peak auditory response to the perturbation in both the speaking and listening conditions were delayed as compared to the three-peak response to the onset of speech. Yet, of particular interest are the peaks that show an additional selective delay in the speaking condition in response to the altered auditory feedback. These significant delays demonstrate processing that differs both from passive listening to the pitch shift and from speaking with unaltered feedback. These speaking induced delays in response to the altered auditory feedback were seen in the early responses in bilateral primary auditory cortex, bilateral STG/MTG, left vSMG/pSTS and in several peaks in right premotor cortex. The first speaking specific responses to the perturbation are the speaking induced delays in right primary auditory cortex and right STG/MTG followed shortly by left primary auditory cortex. The right temporal cortex has previously been demonstrated to be important in pitch perception (Samson and Zatorre, 1988; Sidtis and Volpe, 1988; Zatorre and Halpern, 1993). Importantly, the speaking induced delays cannot be attributed simply to a change in pitch, since previous studies have shown that varying the fundamental frequency of auditory stimulus is not sufficient to alter the latency of the m100 (Poeppel et al., 1997). One intriguing possible consequence of the observed speaking-induced delay in right primary auditory cortex and STG/MTG is additional processing in the speaking condition involving the recognition and initial processing of the pitch feedback error.

The speaking induced delays we observed might also help to explain why some results, such as the 2009 study of Behroozmand et al. differ from our results. In their previous EEG study, Behroozmand and colleagues found enhanced evoked responses in bilateral scalp electrodes to pitch-shifted auditory feedback during active vocalization compared to passive listening in both the P1 (73.51ms) and P2 components (199.55ms) (Behroozmand et al., 2009). While Behroozmand et al. do not report behavioral compensation; both SPRE peaks precede the onset of compensation found in our study. In the current study we do not see the early enhancement (73.51ms) previously reported, but we do see significant speaking induced delays in the early response components. These speaking induced delays, when summed over cortical regions and recorded on the scalp, could appear as a magnitude difference.

4.3 Altered auditory feedback induces behavioral compensation

The behavioral effect of the unexpected pitch shift during auditory feedback induces a rapid compensatory response that, on average, opposes the direction of the external shift for each subject. The mean compensation varied across subjects, replicating findings in previous studies (Burnett et al., 1998; 1997; Greenlee et al., 2013). The response latency measured in our study, with a mean of 210.9ms, is the same as one previous study, and slightly later than some previously reported mean latencies, but within the range of latencies. Previous studies have reported mean latencies of 210.91ms, standard error 31.84 (Jones and Munhall, 2002), 159ms, range 104-223ms (Burnett et al., 1997), 192ms (Burnett et al., 1998), and 184ms, range 115ms-285ms (Greenlee et al., 2013) and as early as around 100ms (Korzyukov et al., 2012; Larson et al., 2007). The point where the F0 contour crosses the two standard deviation line is a conservative measure of when the compensation has definitively begun, but is likely later than the actual initiation of the new vocalization. Thus, in order to account for this discrepancy, neural responses immediately preceding the compensation onset will be considered here as occurring during the onset of compensation.

Some previous work has shown that long duration pitch shifts can induce a multipeak response (Burnett et al., 1998; Hain et al., 2000). The two phase response was most clearly reported in a previous study in which subjects were specifically instructed to make a voluntary response to a pitch perturbation (Hain et al., 2000). We did not see a two-component response, which may be due to the fundamental task difference, since our subjects were neither instructed to respond to nor ignore the pitch alteration, and due to the fact that we looked at all trials averaged for each subject.

4.4 The effect of altered auditory feedback in speech on MEG magnitude response

In addition to the speaking induced delays, speaking specific neural enhancement occurred in later cortical responses. As opposed to other studies that found an unlocalized SPRE response in two early peaks (Behroozmand et al., 2011; 2009; Behroozmand and Larson, 2011; Korzyukov et al., 2011), we found maximal SPRE responses localized to left vSMG/pSTS, and additional SPRE distributed throughout the bilateral speech motor network. The sequence of SPRE can be divided into two intervals, the SPRE occurring before and during the behavioral compensation at 210.9ms and the SPRE occurring after onset of the behavioral compensation. We discuss these intervals separately because models of speech motor control postulate different processes are occurring during these different intervals (Hickok et al., 2011; Houde and Nagarajan, 2011; Ventura et al., 2009). According to such models, the SPRE occurring before and during the compensation is likely to be involved in the recognition of the speech error, the processing of that error and the motor plan change that causes the subsequent behavioral compensation. The SPRE occurring after the onset of the compensation is likely to be involved in the continuously changing motor command as the compensation increases, preparing for the new auditory input of the compensated pitch, and monitoring the new auditory feedback due to the updated motor plan.

4.4.1 Role of SPRE Responses: Prior to and during compensation initiation—

Two significant SPRE peaks occur prior to and during the initiation of compensation in left vSMG/pSTS and one SPRE peak occurs in right vSMG/pSTS, a region encompassing areas

of parietal cortex along the SMG and area Spt (Sylvian fissure at the parietal-temporal boundary). These areas have been implicated as crucial in sensorimotor translation (Andersen et al., 1997; Hickok et al., 2011; 2009; Shum et al., 2011). Given the previously established role of vSMG/pSTS in auditory motor translation, we interpret this enhancement as encompassing processing related not only to recognizing the pitch error, but also crucial computation to both induce and prepare for the resulting motor response. Further support for the role of vSMG/pSTS in changing the motor command is the SPRE immediately following in left premotor cortex. The SPRE in premotor cortex would be most naturally implicated in driving the compensatory motor response, but it is important to ensure that this assumption is causally reasonable given the timing of the SPRE compared to the compensation. Rödel et al found activation of the vocalis muscle 10.7-11.8 ms following cortical stimulation and activation of the cricothyroid muscle 10.1-10.8ms following cortical stimulation (Rödel et al., 2004). Given that work, the timing of SPRE in the premotor cortex reasonably allows for SPRE to be involved in compensation.

4.4.2 Role of SPRE Responses: Following compensation onset—The first SPRE response following compensation occurs in right primary auditory cortex. It is probable that this speaking specific activity is once again demonstrating the well-established role of the right hemisphere in pitch detection. Interestingly, following this pitch detection we observe a series of SPRE originating in left premotor cortex followed by right premotor cortex, and following the speech motor control network via left vSMG/pSTS through left STG/MTG and back to right primary auditory cortex. While a causal link cannot be ascertained with these data, the temporal sequence suggests efference copy of a corrected motor plan being fed back to auditory cortex. Consistent with this interpretation, we note that while the compensation onset occurs at 210.9ms, the compensation continues to ramp up for the duration of the pitch shift until the peak at 531.8ms.

4.4.3 Correlation of MEG response during speaking with behavioral compensation—Interestingly, our results did not show a strong positive correlation between cortical response to the perturbation in the speaking condition and individual subject compensation at the peaks that exhibit SPRE. While we did see a trend towards a positive correlation in left vSMG/pSTS, we did expect a stronger association within this region. Furthermore, we expected to see a positive correlation in premotor cortex with compensation. Instead, activity in left premotor cortex did not show any relationship with compensation, while right premotor cortex showed a weak negative correlation with compensation. These findings suggest that information influencing the degree of compensation was not captured in the low-frequency evoked response data analyzed in this study. In contrast, Chang et al. did find cortical areas whose activity correlated with compensation (Chang et al., 2013). What accounts for this difference in findings? First, Chang et al. looked for correlations with activation in the high gamma (50-150Hz) frequency band, instead of the lower frequency responses examined here. Chang et al. were also analyzing trial-to-trial correlations within individual subjects, whereas here we correlated mean responses with compensation across subjects. Finally, in the Chang et al. study, the correlation analysis was response locked, with the activity time course of each trial being aligned to the time of peak compensation for that trial, whereas here trials were

aligned to perturbation onset before being averaged. Overall, the Chang et al. study implicates high gamma responses are important in driving single trial variability in compensation within subjects. Whether mean compensation is also correlated to mean high gamma responses across subjects is yet to be addressed.

4.5 Relationship between SIS and SPRE

In addition to looking at the temporal sequence of SPRE, the design of this study also permitted study of the relationship of SIS and SPRE. The SIS response showed dissociation from the SPRE response in several cortical areas. Left STG/MTG and right vSMG/pSTS showed SPRE, but did not have a SIS response. Left primary auditory cortex and right STG/MTG showed SIS, but did not show SPRE. Right primary auditory cortex and bilateral premotor cortex showed both SIS and SPRE, but these peaks did not survive corrections for multiple comparisons. Only one region showed co-localization of a significant SIS and SPRE response, left vSMG/pSTS, the sensorimotor translation region.

The first attempt to relate cortical responses to unaltered vocalizations with responses to a feedback error was conducted in work done in non-human primates. Eliades and Wang found that the majority of neurons in the auditory cortex of marmoset monkeys (approximately three-quarters) were suppressed during vocalization, compared with their pre-vocalization activity (Eliades and Wang, 2008). Overall, the neurons that showed this suppression during vocalization showed increased activity during frequency-shifted feedback but maintained suppressed activity as compared to the period preceding the vocalization. These findings were interpreted to say that the vocalization induced suppression increases the neurons sensitivity to vocal feedback. As stated above, we did not see this direct relationship between SIS and SPRE in human left auditory cortex but we did see both SIS and SPRE in right primary auditory cortex. Instead, SIS and SPRE co-localized in left vSMG/pSTS. Fundamental differences in the task and definition of suppression and enhancement can account for the differences in results between these studies. While in the current study suppression and enhancement in speaking were defined in comparison to the neural response to a listening condition with identical auditory input, in the Eliades and Wang study suppression and enhancement were defined as the comparison between the active window and the pre-vocalization activity. The current study finds the population response in left vSMG/pSTS to show both SIS and SPRE, but further studies are necessary to test if the suppression at the onset of speech increases the neurons' sensitivity to a feedback error. One previous study has looked at the spatial relationship with SIS and SPRE in ECoG in the left hemisphere (Chang et al., 2013). This study found the majority of electrodes preferentially demonstrate SIS or SPRE, with a few overlapping electrodes scattered across left temporal cortex, left inferior parietal lobe, and one electrode in left ventral premotor cortex. The majority of the electrodes showing overlapping SIS and SPRE are clustered near the temporal parietal junction, a region encompassed in the vSMG/pSTS region in the current study, where SIS and SPRE were colocalized. This overlapping, yet dissociated pattern of responses to the unaltered speech onset and the pitch perturbation shows that the networks involved in recognizing self-produced speech are similar to, but not identical to, those involved in recognizing and responding to an error in auditory feedback. For several models of speech motor control, these results are somewhat problematic, since

such models assume the processes responsible for SPRE get their input from processes responsible for SIS (Chang et al., 2013). Thus, these models predict all cortical areas showing SPRE should also show SIS, which the results here show is not always the case.

4.6 Limitations

In order to understand speaking-specific cortical activity, we use passive listening to the subject's recorded voice from the previous block as a reference signal with which to discuss responses during speech. The auditory input during the passive listening matches the auditory input through the headphones during speech, but does not include the additional sound perceived during speaking by bone conduction. The minimal role of bone conduction in SIS has been indirectly addressed by a study demonstrating that motor induced suppression occurs in response to auditory input created by a volitional motor act (Aliu et al., 2009). In that study, where the auditory input was perfectly matched in the active and passive listening conditions, subjects experienced motor induced suppression when hearing a tone in response to a button press. It is thus unlikely that the suppression seen during speech is an artifact of bone conduction instead of speech motor induced suppression. The impact of bone conduction on the pitch perturbation response has also been well addressed by Burnett and colleagues in their 1998 study (Burnett et al., 1998). That study found the addition of pink masking noise to block the effect of bone conduction did not change the pitch perturbation response, importantly showing that bone conduction does not play a role in the perturbation response. The current study was designed to examine the role of auditory feedback in speech production, so a closely matched auditory task works as an effective reference signal. Yet during ongoing speech there are also complex and dynamically changing somatosensory cues influencing speech. The interplay between the role of somatosensory feedback and auditory feedback during speech production, which has only recently been addressed in a behavioral study (Lametti et al., 2012), is an exciting question for future studies.

5. Conclusion

Results from the present study suggest that there are distinct, yet overlapping networks involved in monitoring the onset of speech and in detecting and correcting for an error during ongoing speech. SIS was found in bilateral auditory cortex and left ventral supramarginal gyrus/posterior superior temporal sulcus (vSMG/pSTS). In contrast, during pitch perturbations, activity was enhanced in bilateral vSMG/pSTS, bilateral premotor cortex, right primary auditory cortex, and left higher order auditory cortex. The results in the study also suggest that the latency of cortical responses may also be important in understand cortical processing during speech. Importantly, vSMG/pSTS is involved in both monitoring the onset of speech and recognizing and responding to auditory errors during speech.

Acknowledgments

References

Aliu SO, Houde JF, Nagarajan SS. Motor-induced Suppression of the Auditory Cortex. Journal of Cognitive Neuroscience. 2009; 21:791–802. [PubMed: 18593265]

Andersen RA, Snyder LH, Bradley DC, Xing J. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual Review of Neuroscience. 1997; 20:303–330.

Bardouille T, Picton T, Ross B. Correlates of eye blinking as determined by synthetic aperture magnetometry. Clinical Neurophysiology. 2006; 117:952–958. [PubMed: 16564205]

Behroozmand R, Karvelis L, Liu H, Larson CR. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. Clinical Neurophysiology. 2009; 120:1303–1312. [PubMed: 19520602]

Behroozmand R, Larson CR. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci. 2011; 12:54. [PubMed: 21645406]

Behroozmand R, Liu H, Larson C. Time-dependent Neural Processing of Auditory Feedback during Voice Pitch Error Detection. Journal of Cognitive Neuroscience. 2011; 23:1205–1217. [PubMed: 20146608]

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995; 57:289–300.

Buchsbaum BR, Baldo J, Okada K, Berman KF, Dronkers N, D'Esposito M, Hickok G. Conduction aphasia, sensory-motor integration, and phonological short-term memory â€' An aggregate analysis of lesion and fMRI data. Brain and Language. 2011; 119:119–128. [PubMed: 21256582]

Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. 1998; 103:3153–3161. [PubMed: 9637026]

Burnett TA, Senner JE, Larson CR. Voice F0 responses to pitch-shifted auditory feedback: a preliminary study. Journal of Voice. 1997; 11:202–211. [PubMed: 9181544]

Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF. Human cortical sensorimotor network underlying feedback control of vocal pitch. Proceedings of the National Academy of Sciences. 2013

Chang-Yit R, Pick HL Jr. Siegel GM. Reliability of sidetone amplification effect in vocal intensity. Journal of Communication Disorders. 1975; 8:317–324. [PubMed: 802981]

Chen SH, Liu H, Xu Y, Larson CR. Voice F[sub 0] responses to pitch-shifted voice feedback during English speech. J. Acoust. Soc. Am. 2007; 121:1157–1163. [PubMed: 17348536]

Cowie R, Douglas-Cowie E, Kerr AG. A study of speech deterioration in post-lingually deafened adults. J. Laryngol. Otol. 1982; 96:101–112. [PubMed: 7057081]

Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. Hum. Brain Mapp. 2000; 9:183–191. [PubMed: 10770228]

Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on voice F[sub 0] contours in syllables. J. Acoust. Soc. Am. 2002; 111:357–366. [PubMed: 11831808]

Eliades SJ, Wang X. Sensory-Motor Interaction in the Primate Auditory Cortex During Self-Initiated Vocalizations. Journal of Neurophysiology. 2002; 89:2194–2207. [PubMed: 12612021]

Eliades SJ, Wang X. Dynamics of Auditory-Vocal Interaction in Monkey Auditory Cortex. Cerebral Cortex. 2005; 15:1510–1523. [PubMed: 15689521]

Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature. 2008; 453:1102–1106. [PubMed: 18454135]

Fu CHY, Vythelingum GN, Brammer MJ, Williams SCR, Amaro E, Andrew CM, Yaguez L, van Haren NEM, Matsumoto K, McGuire PK. An fMRI Study of Verbal Self-monitoring: Neural Correlates of Auditory Verbal Feedback. Cerebral Cortex. 2006; 16:969–977. [PubMed: 16195470]

Gage N, Poeppel D, Roberts TPL, Hickok G. Auditory evoked M100 reflects onset acoustics of speech sounds. Brain Research. 1998; 814:236–239. [PubMed: 9838140]

Gelfand JR, Bookheimer SY. Dissociating Neural Mechanisms of Temporal Sequencing and Processing Phonemes. Neuron. 2003; 38:831–842. [PubMed: 12797966]

Greenlee JDW, Behroozmand R, Larson CR, Jackson AW, Chen F, Hansen DR, Oya H, Kawasaki H, Howard MA III. Sensory-Motor Interactions for Vocal Pitch Monitoring in Non-Primary Human Auditory Cortex. PLoS ONE. 2013; 8:e60783. [PubMed: 23577157]

Greenlee JDW, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H, Howard MA. Human Auditory Cortical Activation during Self-Vocalization. PLoS ONE. 2011; 6:e14744. [PubMed: 21390228]

Grefkes C, Fink GR. REVIEW: The functional organization of the intraparietal sulcus in humans and monkeys. Journal of anatomy. 2005; 207:3–17. [PubMed: 16011542]

Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. Exp Brain Res. 2000; 130:133–141. [PubMed: 10672466]

Heinks-Maldonado TH, Nagarajan SS, Houde JF. Magnetoencephalographic evidence for a precise forward model in speech production. NeuroReport. 2006:1–5. [PubMed: 16361939]

Hickok G, Houde J, Rong F. Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. Neuron. 2011; 69:407–422. [PubMed: 21315253]

Hickok G, Okada K, Serences JT. Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing. Journal of Neurophysiology. 2009; 101:2725–2732. [PubMed: 19225172]

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

Houde JF, Jordan MI. Sensorimotor Adaptation in Speech Production. Science. 1998; 279:1213–1216. [PubMed: 9469813]

Houde JF, Jordan MI. Sensorimotor Adaptation of Speech I: Compensation and Adaptation. Journal of Speech, Language, and Hearing Research. 2002; 45:295–310.

Houde JF, Nagarajan SS. Speech production as state feedback control. Frontiers in Human Neuroscience. 2011; 5:1–14. [PubMed: 21283556]

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the Auditory Cortex during Speech: An MEG Study. Journal of Cognitive Neuroscience. 2002; 14:1125–1138. [PubMed: 12495520]

Jones JA, Munhall KG. The role of auditory feedback during phonation: studies of Mandarin tone production. Journal of Phonetics. 2002; 30:303–320.

Korzyukov O, Karvelis L, Behroozmand R, Larson CR. ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. International Journal of Psychophysiology. 2011; 83:71–78. [PubMed: 22044551]

Korzyukov O, Karvelis L, Behroozmand R, Larson CR. ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. International Journal of Psychophysiology. 2012; 83:71–78. [PubMed: 22044551]

Lametti DR, Nasir SM, Ostry DJ. Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. Journal of Neuroscience. 2012; 32:9351–9358. [PubMed: 22764242]

Lane H, Tranel B. The Lombard Sign and the Role of Hearing in Speech. Journal of Speech and Hearing Research. 1971; 14:677–709.

Lane H, Webster JW. Speech deterioration in postlingually deafened adults. J. Acoust. Soc. Am. 1991; 89:859–866. [PubMed: 2016436]

Larson CR, Sun J, Hain TC. Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F[sub 0] and amplitude control. J. Acoust. Soc. Am. 2007; 121:2862–2872. [PubMed: 17550185]

Lee BS. Some Effects of Side-Tone Delay. J. Acoust. Soc. Am. 1950; 22:639–640.

Lombard E. Le signe de l'elevation de la voix. Ann. maladies oreille larynx nez pharynx. 1911; 37:101–119.

Muller-Preuss P, Ploog D. Inhibition of auditory cortical neurons during phonation. Brain Research. 1981; 215:61–76. [PubMed: 7260601]

Natke U, Kalveram KT. Effects of Frequency-Shifted Auditory Feedback on Fundamental Frequency of Long Stressed and Unstressed Syllables. J Speech Lang Hear Res. 2001; 44:577–584. [PubMed: 11407562]

Orlikoff RF, Baken RJ. Fundamental frequency modulation of the human voice by the heartbeat: Preliminary results and possible mechanisms. J. Acoust. Soc. Am. 1989; 85:888–893. [PubMed: 2926003]

Oshino S, Kato A, Wakayama A, Taniguchi M, Hirata M, Yoshimine T. Magnetoencephalographic analysis of cortical oscillatory activity in patients with brain tumors: Synthetic aperture magnetometry (SAM) functional imaging of delta band activity. NeuroImage. 2007; 34:957–964. [PubMed: 17175174]

Parkinson AL, Flagmeier SG, Manes JL, Larson CR, Rogers B, Robin DA. Understanding the neural mechanisms involved in sensory control of voice production. NeuroImage. 2012; 61:314–322. [PubMed: 22406500]

Parsons, TW. Voice and Speech Processing. Mcgraw-Hill Book Company; New York, NY: 1986.

Patel R, Niziolek C, Reilly K, Guenther FH. Prosodic Adaptations to Pitch Perturbation in Running Speech. J Speech Lang Hear Res. 2011; 54:1051–1059. [PubMed: 21173388]

Poeppel D, Phillips C, Yellin E, Rowley HA, Roberts TPL, Marantz A. Processing of vowels in supratemporal auditory cortex. Neuroscience Letters. 1997; 221:145–148. [PubMed: 9121685]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 2009; 12:718–724. [PubMed: 19471271]

Robinson, S.; Vrba, J. Recent Advances in Biomagnetism. Tokyo University Press; Sendai: 1999. Functional neuroimaging by synthetic aperture magnetometry (SAM).; p. 302-305.

Rödel RMW, Olthoff A, Tergau F, Simonyan K, Kraemer D, Markus H, Kruse E. Human Cortical Motor Representation of the Larynx as Assessed by Transcranial Magnetic Stimulation (TMS). The Laryngoscope. 2004; 114:918–922. [PubMed: 15126757]

Samson S, Zatorre RJ. Melodic and harmonic discrimination following unilateral cerebral excision. Brain and Cognition. 1988; 7:348–360. [PubMed: 3401387]

Sekihara K, Nagarajan SS, Poeppel D, Marantz A. Performance of an MEG Adaptive-Beamformer Source Reconstruction Technique in the Presence of Additive Low-Rank Interference. IEEE Trans. Biomed. Eng. 2004; 51:90–99. [PubMed: 14723498]

Shadmehr R, Krakauer JW. A computational neuroanatomy for motor control. Exp Brain Res. 2008; 185:359–381. [PubMed: 18251019]

Shum M, Shiller DM, Baum SR, Gracco VL. Sensorimotor integration for speech motor learning involves the inferior parietal cortex. European Journal of Neuroscience. 2011; 34:1817–1822. [PubMed: 22098364]

Sidtis JJ, Volpe BT. Selective loss of complex-pitch or speech discrimination after unilateral lesion. Brain and Language. 1988; 34:235–245. [PubMed: 3401692]

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

Toyomura A, Koyama S, Miyamoto T, Terao A, Omori T, Murohashi H, Kuriki S. Neural correlates of auditory feedback control in human. Neuroscience. 2007; 146:499–503. [PubMed: 17395381]

Ventura M, Nagarajan S, Houde J. Speech target modulates speaking induced suppression in auditory cortex. BMC Neurosci. 2009; 10:58. [PubMed: 19523234]

Vrba J, Robinson SE. Signal Processing in Magnetoencephalography. Methods. 2001; 25:249–271. [PubMed: 11812209]

Yates AJ. Delayed auditory feedback. Psychological Bulletin. 1963; 60:213–232. [PubMed: 14002534]

Zatorre RJ, Halpern AR. Effect of unilateral temporal-lobe excision on perception and imagery of songs. Neuropsychologia. 1993; 31:221–232. [PubMed: 8492875]

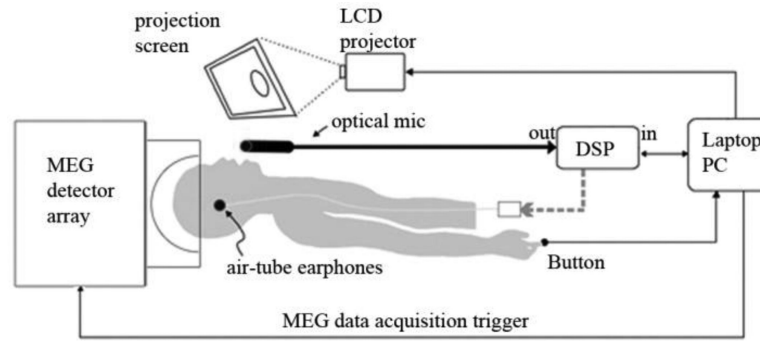**Figure 1.**
Schematic of experimental setup. Subjects are in the supine position in the MEG detector array. They speak into the optical microphone and receive auditory feedback through headphones. Their speech is passed through the Digital Signal Processor (DSP) both during altered and unaltered feedback. The laptop computer triggers the pitch-shifted stimulus at a jittered delay after speech onset.
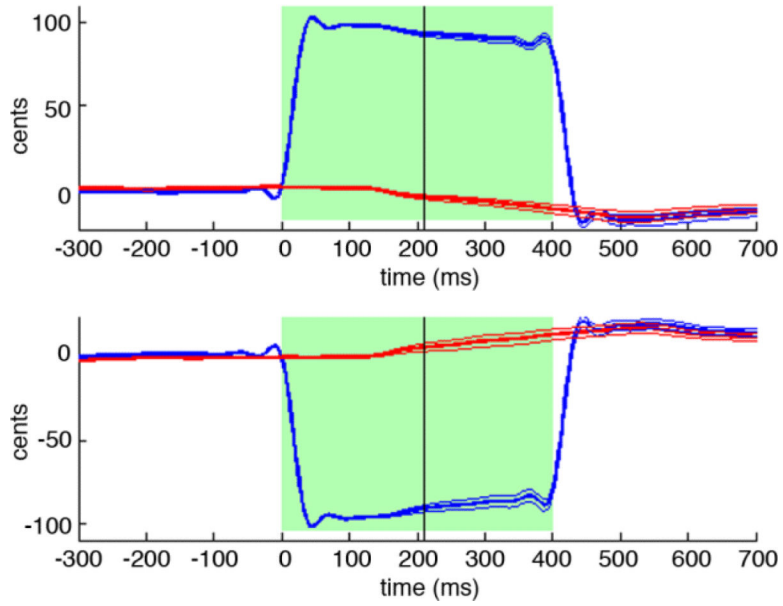
Figure 2.
Vocal responses to the pitch shifts of audio feedback. Upper and lower plots show time courses of the average subject response to the +100 cent and -100 cent pitch perturbations, respectively. In each plot, time is relative to perturbation onset at 0ms, and pitch is expressed in cents relative to the mean pitch of the pre-perturbation period. Green regions indicate the duration of the pitch perturbation. Blue traces show the mean time course of feedback heard by subjects, with thick blue line is the grand-average over subjects and flanking thin blue lines are +/− standard errors. In a similar fashion, red traces show the mean time course of the pitch produced by subjects. The black vertical line denotes the mean time when the subjects' mean pitch time course deviates from the baseline by two standard deviations.
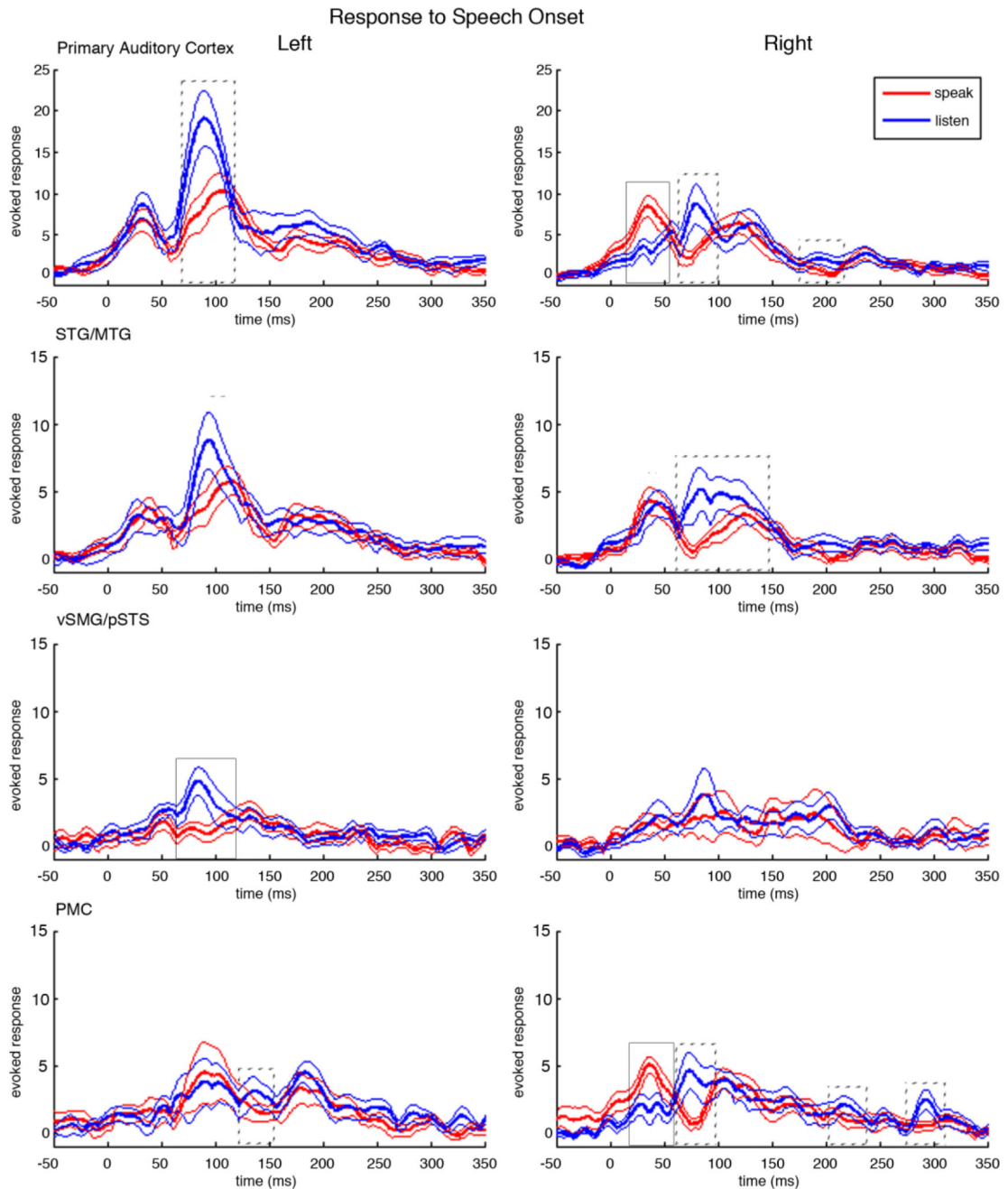
Figure 3.
Neural responses to the onset of speech in both the speaking and the listening condition. In each plot, time is relative to the onset of speech at 0ms. Red traces show the mean cortical responses at the corresponding anatomical location across all subjects in the speaking condition. The thick red line is the grand-average over subjects and the flanking thin red lines are the standard error bars. Blue traces show the mean cortical responses in the listening condition. The thick blue line is the grand-average over subjects and the thin blue lines are the standard error bars. Solid boxes surround peaks with a significant magnitude difference between speaking and listening, p< 0.0096, dotted boxes surround peaks with

p<0.05. Similarly, solid lines mark significantly different latencies of peaks between speaking and listening, p< 0.0096, and dotted lines mark peaks with p<0.05.
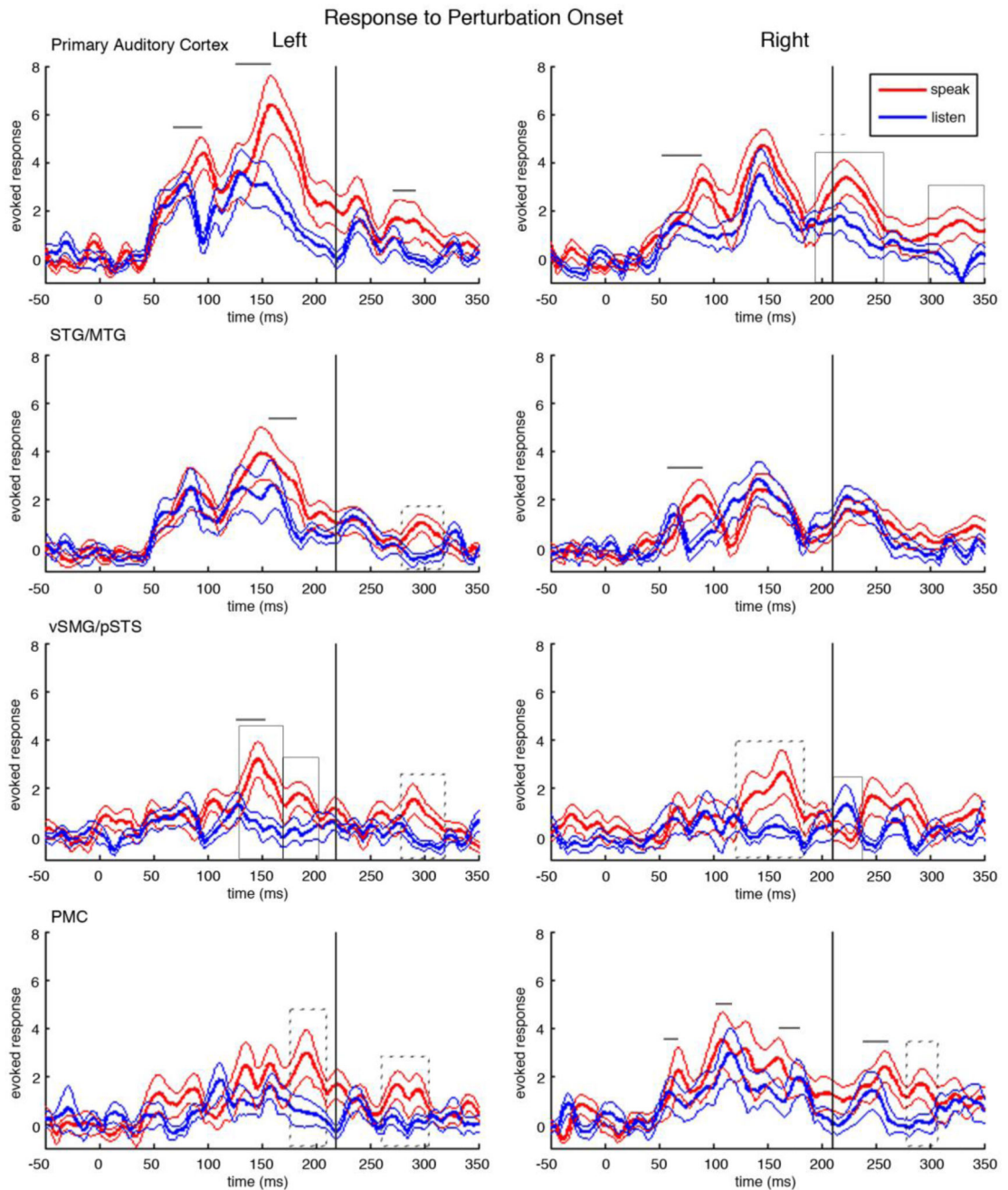
Figure 4.
Neural responses to the pitch-shifted feedback in both the speaking and the listening condition. In each plot, time is relative to the onset of the pitch-shifted feedback at 0ms. Red traces show the mean cortical responses at the corresponding anatomical location across all subjects in the speaking condition. The thick red line is the grand-average over subjects and the flanking thin red lines are the standard error bars. Blue traces show the mean cortical responses in the listening condition. The thick blue line is the grand-average over subjects and the thin blue lines are the standard error bars. Solid boxes surround peaks with a significant magnitude difference between speaking and listening, p< 0.0096, dotted boxes

surround peaks with p<0.05. Similarly, solid lines mark significantly different latencies of peaks between speaking and listening, p< 0.0096, and dotted lines mark peaks with p<0.05. The black vertical line denotes the mean time when the subjects' mean pitch time course deviates from the baseline by two standard deviations.

Table 1

The location of each virtual sensor is reported in both MNI and Talairach coordinates. The Brodmann area for which the voxel is the center of mass is listed and the corresponding functional area is given. Experimental and theoretical references are given that informed the choice of central voxel for the ROI analysis.

MNI Voxel Location	Talairach Voxel	Brodmann Area	Functional Area	References
[−54.3, −26.5, 11.6] [54.4, −26.7, 11.7]	[−51, −27, 13] [52, −27, 13]	41/42	Primary Auditory Cortex	pSTg: [−64,−30,14][11] PT: [−52,−34,16] [11] PT: [58,−28,12] [11] pSTG:[−62,−30,14][11] pdSTs: [58,−28,6] [11] pdSTS:[−60,−30,10][11] PT: [−62,−24,10][11] STG/42: T: [53, −20, 9][10] [57, −20, 15][10] [−57, −20, 20][10] Auditory Cortex[6],[8] BA 41/42[6],[12] STG/STS[8],[12] Sensory System[7]
[−58, −19.8, −6] [58.9, −20.1, −6.1]	[−54, −21, −1] [56, −21, 0]	21/22	Superior Temporal Gyrus/ Middle Temporal Gyrus	STA: [52 −10 −2][2] STS: [−60 −30 3][1] STS: [63 −24 0][1] MTG/BA 21: T: [57, −10, −7][10] [57, −30, −2][10] [−53, −13, −7][10] [−53, −26, −7][10] STG/BA 22: T: [57, −17, 9][10] [−57, −17, 4][10] adSTs: [56,−10,−4][11] Auditory Cortex[6],[8] BA 21/22[6],[12] STG/STS[8],[12] Sensory System[7]
[−51.2, −43.3, 40.0] [51.8, −43.3, 40.4]	[−50, −41, 38] [52, −40, 38]	40	Supermarginal Gyrus/posterior Superior Temporal Sulcus/ Inferior Parietal Lobule	IPS: [45, −36, 45][1] IPS: [32, −42, 44][2] Spt: [−51, −42, 21][1] Supramarginal Gyrus: T: [−44, −40, 40][3] Spt: T: [−52, −43, 28][4] Spt: T: [−40, −32, 26][4] PO: [−44,−34,24][11] IPS[5] IPL, BA 39&40[6] Parietal Cortex[7],[9] Spt[8]

MNI Voxel Location	Talairach Voxel	Brodmann Area	Functional Area	References
[−46, 0, 35] [46.5, 0, 35]	[−45, 0, 33] [45, 0, 32]	6	Premotor Cortex	Premotor: [−52, 8, 38][2] Precentral Gyrus: T: [48 −4 36][3] [−34 2 38][3] Inferior Frontal Gyrus: T:[−50 10 36][3] Precentral Sulcus: [−51 −9 42][1] vPMC: [−48,0,30][11] Premotor Cortex[6,7,8,12] BA 6[6]

[1] Buchsbaum, 2011

[2] Toyomura, 2007

[3] Gelfand, 2003

[4] Hickok, 2009

[5] Grefkes, 2005

[6] Rauschecker, 2009

[7] Shadmehr, 2008

[8] Hickok, 2011

[9] Andersen, 1997

[10] Fu, 2006

[11] Tourville, 2008

[12] Hickok, 2007