Published in final edited form as:

Neuropsychologia. 2013 July ; 51(8): 1471–1480. doi:10.1016/j.neuropsychologia.2013.05.002.

Modulation of effective connectivity during vocalization with perturbed auditory feedback

Amy L. Parkinson^{a,b,*}, Oleg Korzyukov^c, Charles R. Larson^c, Vladimir Litvak^d, and Donald A. Robin^{a,b}

^aResearch Imaging Institute, University of Texas Health Science Center San Antonio, San Antonio, TX, USA

^bDepartment of Neurology, University of Texas Health Science Center San Antonio, San Antonio, TX, USA

^cDepartment of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA

^dWellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, UK

Abstract

The integration of auditory feedback with vocal motor output is important for the control of voice fundamental frequency (F0). We used a pitch-shift paradigm where subjects respond to an alteration, or shift, of voice pitch auditory feedback with a reflexive change in F0. We presented varying magnitudes of pitch shifted auditory feedback to subjects during vocalization and passive listening and measured event related potentials (ERP's) to the feedback shifts. Shifts were delivered at +100 and +400 cents (200 ms duration). The ERP data were modeled with Dynamic Causal Modeling (DCM) techniques where the effective connectivity between the superior temporal gyrus (STG), inferior frontal gyrus and premotor areas were tested. We compared three main factors; the effect of intrinsic STG connectivity, STG modulation across hemispheres and the specific effect of hemisphere. A Bayesian model selection procedure was used to make inference about model families. Results suggest that both intrinsic STG and left to right STG connections are important in the identification of self-voice error and sensory motor integration. We identified differences in left to right STG connections between 100 cent and 400 cent shift conditions suggesting that self and non-self voice error are processed differently in the left and right hemisphere. These results also highlight the potential of DCM modeling of ERP responses to characterize specific network properties of forward models of voice control.

Keywords

Vocalization; Auditory feedback; ERP; DCM; Audio-vocal integration; Pitch shift

© 2013 Elsevier Ltd. All rights reserved.

*Corresponding author: Parkinson@uthscsa.edu, Tel: +1 210-567-8136 ; Fax: +1 210-567-8152.

Present address: Research Imaging Institute, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Vocalization is used to communicate in most air-breathing vertebrate animals and humans. In humans the importance of vocalization cannot be overstated. For instance, personality and emotional state are reflected in prosodic features of the voice (Williams & Stevens, 1972; Tompkins & Mateer, 1985; Banziger & Scherer, 2005). Moreover, people with voice disorders suffer significant difficulties with communication and interaction with others. While much is known about the human voice, an understanding of the neural basis of voice production is still needed. Such information will allow for greater insight into the diagnosis and treatment of voice disorders. One needed critical issue is the understanding of how auditory feedback (the primary sensory system involved in vocalization) is processed and used to control vocal pitch or fundamental frequency (F0) and amplitude. There are clear differences in the neural processing of speaking and listening, and it is well described that attenuation of the auditory cortex is modulated by how closely the auditory feedback from self-vocalization matches expected output (Houde, et al., 2002; Heinks-Maldonado et al., 2006; Behroozmand and Larson, 2011). Several studies have shown that perturbations in voice auditory feedback are important for both feedforward and feedback control of vocalization. (e.g. Larson et al., 2000; Jones and Munhall, 2002; Houde and Jordan 2002). The current study used a pitch-shifted auditory feedback paradigm (Larson, 1998) where the pitch of self-voice feedback is changed in real time during vocal production. When presented with a brief shift in vocal pitch of auditory feedback during vocalization, subjects will briefly respond to the perturbation by changing the F0 of their own voice in the opposite direction to the shift.

Several studies have examined event-related potentials (ERP's) in response to varying magnitude of pitch-shifted stimuli during vocalization (Behroozmand and Larson, 2011; Liu et al., 2011; Korzyukov et al. 2012). Several ERP components are elicited by unexpected shifts in the pitch of self-voice feedback during vocalization. The N1 (negative component with peak latency around 100ms post stimulus onset) and P2 (positive component with a peak latency within 200–300ms post stimulus onset) components both show prominent amplitudes to the pitch shift stimulus when the perturbations occur after vocalization onset (Behroozmand et al., 2009). When the perturbations are presented at voice onset, there is suppression of the N1 or M1 (MEG) component (Houde et al., 2002, Heinks-Maldonado et al., 2005) relative to the potential elicited by passive listening to the same pitch-shifted vocalization. That is to say, vocal onset triggers an ERP that is smaller in magnitude than the ERP triggered by passive listening to the same self-vocalization. However, if the vocalization is altered in pitch or other characteristics, the ERP it generates is not suppressed relative to the playback and passive listening to the signal. These observations have been interpreted to mean that at vocal onset, one of the first steps in neural processing of voice auditory feedback is identification of self or non-self vocalization (see Fig. 5, Hain et al., 2001). Subsequent studies have shown that the suppression is reduced with increasing shift magnitude and is almost completely eliminated in response to a 400 cent stimulus (Behroozmand and Larson, 2011). These findings suggest that the N1 component is modulated for different levels of error between actual and predicted voice feedback, and thus the 400 cent shift may be processed differently than the smaller shifts (100 cent) as a result of increasing error levels between the actual and predicted feedback.

It is likely that these neural responses to auditory feedback perturbations rely upon information provided by an internal forward model system similar to forward models suggested in speech motor control (Aliu et al., 2009; Golfinopuluos et al., 2011; Hickok et al., 2011). The idea of the forward model is that forward predictions about the state of the motor articulators and the sensory consequences of the predicted action are made and used to update the motor system. Forward predictions of the action are thought to be generated by

an internal model that receives an efference copy signal (prediction) of the motor commands and integrates this with the current state of the system (Hickok 2011). This model uses auditory feedback and an efference copy signal in the online error detection and correction processes to adjust the motor drive for vocal output. The cortical regions involved in controlling vocal responses to auditory feedback have been identified using functional neuroimaging techniques (Tourville et al., 2008; Zarate and Zatorre, 2008; Zarate et al., 2010; Parkinson et al., 2012). Results of these studies identified brain regions involved in audio-vocal integration including the superior temporal gyrus (STG), pre-motor cortex (PM), inferior frontal gyrus (IFG) anterior insula and primary motor regions (M1). Moreover, an increase in bilateral STG activity reported in conditions where subjects received altered feedback of their own voice identified this as a key region involved in sensory motor integration.

Various forward models of speech and vocal production contain brain regions that are involved in sensory motor integration. The DIVA (Directions Into Velocities of Articulators) model of speech production (Guenther et al., 1998; Guenther et al., 2006; Tourville & Guenther, 2011) incorporates auditory feedback and feedforward commands for voice and speech control (Guenther, et al., 2006). The DIVA model (Guenther et al., 2006) predicts bilateral ventral precentral gyrus involvement in receiving corrective error commands from the auditory error cells in the STG. Hickok (2011) has also proposed a dual stream model of speech processing involving a ventral and dorsal stream of connected brain regions. This model suggests that lexical information for speech processing occurs in a ventral stream that encompasses bilateral dorsal STG, and a strongly left dominant dorsal stream of connected brain regions, supporting sensory-motor integration. Yet little is known about how these specific brain areas interact during these neural responses to auditory feedback perturbations.

In efforts to further understand how these brain regions are connected for vocal control responses within various models of voice production, Tourville et al (2008) performed structural equation modeling (SEM) of fMRI responses to auditory feedback. In this study, the first formant frequency (F1) of participants' speech was shifted unexpectedly. Results of this modeling identified increases in connectivity from left STG to right premotor cortex, left STG to right STG and from right STG to right IFG, indicating increased influence of bilateral auditory cortical areas on right frontal areas. These results suggest that auditory feedback control of speech is mediated by projections from auditory error cells in the STG to motor correction cells in the right frontal cortex, however left hemisphere frontal regions were not included in the model and so cannot be ruled out as also being involved. However, the paradigm of the Tourville study and ours are different. Our goal is to understanding sensory motor integration during vocalization, not speech motor control, which may involve more complex models. In Tourville, the formant shift resulted in a phonemic shift whereas our perturbation does not. Thus, the comparison between speech and vocalization models requires empirical study which is a purpose of the current paper.

An important methodological limitation to date is that the interactive processes and causal relations within the underlying neuronal network involved in these sensory feedback mechanisms are difficult to identify by conventional imaging or EEG analysis alone. Dynamic causal modeling (DCM) has recently been extended to allow for effective connectivity of electroencephalography (EEG) data (David et al., 2006; Kiebel et al., 2006, 2009). DCM for EEG allows for causal inferences about how connections within a network of brain regions are modulated by an experimental condition. A connection (or a set of connections) is said to be modulated if changing its strength allows it to fit the condition-specific evoked responses while keeping other connections the same for all conditions. DCM using event related potentials (ERP's) takes advantage of the exceptional temporal

resolution of electrophysiological signals for precise modeling of temporal dynamics within a network of specified brain regions. DCM fits these dynamics with a neurophysiologically realistic neural mass model. The model fitting is framed as a Bayesian model evidence optimization problem. This allows model comparisons taking into account both model accuracy and model complexity (Penny et al. 2004, 2010).

In the present study we used DCM to model effective connectivity of brain regions involved in sensory control of the voice. These regions were identified from bold fMRI activation of pitch-shifted vocalization in a recent study (Parkinson et al., 2012). Based on the work discussed above, our *a priori* hypothesis was that DCM modeling would identify different connectivity patterns based on vocalization with pitch-shifted feedback at different magnitudes (100 cents and 400 cents). The underlying hypothesis is that voice auditory feedback that is different from the actual voice being produced may not be recognized as self-voice, and hence different neural processing mechanisms may be activated by vocalizations that are shifted by 400 cents.

From earlier studies it is known that STG, IFG and motor regions of the brain are involved with sensory-motor integration and error detection/correction mechanisms (Parkinson et al., 2012). We have previously identified the importance of bilateral STG regions in sensory control of the voice (Parkinson et al., 2012) and wish to model the coupling between these regions. We compared families of models examining various combinations of STG involvement in the error detection/correction process. In the first comparison we examined differences between models both with and without intrinsic STG connections, the second comparison was concerned with STG connections across hemispheres (e.g. left STG to right STG) and the final comparison examined the effect of hemisphere by examining bilateral versus left versus right connections of other cortical regions (e.g. left STG to PM, left STG to left IFG and left IFG to STG being modulated with no right hemisphere connections modulated in the left hemisphere models). We specified a model involving the right hemisphere and including STG as a driving area in a feed-forward system to control vocal output.

2. Methods

2.1 Participants

Ten right-handed speakers of American English (8 females and 2 males, mean age: 20.6 years $\pm$ 1.8 years) with no history of neurological disorder participated in the study. All potential subjects underwent a pre-screening session to check eligibility for the study. This consisted of monitoring audio-vocal responses to the pitch shift paradigm and a bilateral pure-tone hearing-screening test at 20dB SPL (octave frequencies between 250–8000 Hz). The Northwestern University institutional review board approved all study procedures including recruitment, data acquisition and informed consent, and subjects were monetarily compensated for their participation. Written informed consent was received from all participants.

2.2 Experimental design

During the test, subjects were seated in a sound-attenuated room and were instructed to sustain the vowel sound /a/ for approximately 3–4 seconds at their conversational pitch and loudness levels whenever they felt comfortable, i.e., without a cue. They were informed that their voice would be played back to them during their vocalizations, and they were asked to ignore pitch-shifts in the feedback of their voice. While they were vocalizing, subjects watched a variety of color pictures of landscapes, flowers, animals, and humans changing randomly every 30–35 sec. Subjects typically paused for 2–3 sec between vocalizations to take a breath. During each vocalization a pitch-shift stimulus (200 ms duration) was

presented (Figure 1). The stimulus in each vocalization occurred 500–1000 ms after voice onset. All pitch-shift stimuli were in the upward direction and were randomly varied from trial to trial in magnitude (100 or 400 cents). The pitch-shift unit, “cents”, is a logarithmic value related to the 12-tone musical scale, where 100 cents equals one semitone. The rise time of the pitch shift was 10–15 ms. There were approximately 200 active vocalization and 200 playback trials (100 with pitch shifted vocalization and 100 with unshifted vocalization) per subject at each of the 100 cent and 400 cent magnitudes. Subjects were instructed to keep their eyes open throughout the recording session.

Subjects’ voices were recorded with an AKG boomset microphone (model C420), amplified with a Mackie mixer (model 1202-VLZ3), and pitch-shifted through an Eventide Eclipse Harmonizer. The time delay from vocal onset of the duration, direction, and magnitude of pitch shifts were controlled by MIDI software (Max/MSP v.5.0 Cycling 74). Voice and auditory feedback were sampled at 10 kHz using PowerLab A/D Converter (Model ML880, AD Instruments) and recorded onto a laboratory computer utilizing Chart software (AD Instruments). Subjects maintained their conversational F0 levels and voice loudness of about 70–75 dB, and the feedback signal (i.e., the subject’s pitch-shifted voice) was delivered through Etymotic earphones (model ER1-14A) at about 80–85 dB. The 10 dB gain between voice and feedback channels (controlled by a Crown amplifier D75) was used to partially mask air-borne and bone-conducted voice feedback. The subject’s voice that was recorded during the first block of trials was played back to the subjects during the passive listening experimental condition.

2.3 EEG acquisition

The electroencephalogram (EEG) signals were recorded from 64 sites on the subject’s scalp using an Ag-AgCl electrode cap (EasyCap GmbH, Germany) in accordance with the extended international 10–20 system (Oostenveld and Praamstra, 2001) including left and right mastoids. Recordings were made using the average reference montage in which outputs of all of the amplifier channels were averaged, and this averaged signal was used as the common reference for each channel. Scalp-recorded brain potentials were low-pass filtered with a 400-Hz cut-off frequency (anti-aliasing filter), digitized at 2 kHz, and recorded using a BrainVision QuickAmp amplifier (Brain Products GmbH, Germany). Electrode impedances were kept below 5 k Ω for all channels. The electro-oculogram (EOG) signals were recorded using two pairs of bipolar electrodes placed above and below the right eye to monitor vertical eye movements and at the canthus of each eye to monitor horizontal eye movements.

2.4 Behavioural Data Analysis

The voice and feedback signals were processed in Pratt (Boersma, 2001), which generated pulses corresponding to each cycle of the glottal waveform, and from this, the F0 of the voice signal (F0 contour) was calculated. These signals along with TTL pulses corresponding to the stimulus onset were then processed in IGOR PRO (Wavemetrics, Lake Oswego, OR).

2.5 ERP analysis

ERP pre-processing and data analysis steps were also performed offline using the SPM8 software package (update number 4667) (<http://www.fil.ion.ucl.ac.uk/spm>, Litvak et al. Comput Intell Neurosci. 2011). The data were epoched to single trials, with a peri-stimulus window of –100 to 500 ms, downsampled to 128 Hz and band-pass filtered (Butterworth) between 0.5 and 30Hz. Artifact removal was implemented with robust averaging. Separate analyses were done for 100 cents and 400 cents stimulus directions, active vocalization, and passive listening experimental conditions.

2.6 DCM analysis

Dynamic causal modeling (DCM) (David et al., 2006) was used within SPM8 to examine the connections between neural regions involved in the proposed model of processing auditory feedback during vocalization. DCM was originally created for connectivity analysis of fMRI data, and subsequently extended to model ERP's (David et al., 2006; Kiebel et al., 2006). The DCM method uses neural mass models to describe neural activity and estimate effective connectivity. Source time courses generated by a neurobiologically realistic model of the network of interest are projected on the scalp using a spatial forward model (Boundary Element Model in our case). The parameters of both the source model and the neural model are optimized using a variational Bayesian approach to match as closely as possible the observed EEG data. Data were modeled between 1 and 200ms following the pitch-shift stimulus with an onset of 60ms. The onset parameter determines when the stimulus, presented at time 0ms, is assumed to activate the cortical area. Hanning window was applied to the data and a detrend parameter of 1 was used with 8 modes. The Hanning window will reduce the influence of the beginning and end of the time series and a detrend parameter of 1 models the mean of the data at the sensor level. The evoked responses were modeled using the IMG (imaging) option, which models each source as a patch on the cortical surface (Daunizeau et al., 2009). The data with different pitch shift magnitudes (100 cents and 400 cents) were averaged separately and for each pitch-shift magnitude, vocalization and playback conditions were modeled together allowing particular connections in the model to vary to explain the difference between the two.

2.7 Model Identification and Selection

Our model network architecture for the first experiment was motivated by results from a previous fMRI study of pitch-shifted vocalization (Parkinson et al., 2012) and results from Tourville's (2008) structural equation modeling. The present study identified brain regions in both hemispheres involved in vocalization and processing pitch-shifted feedback. The coordinates chosen as source regions for the models were based on the peak MNI coordinates reported in the literature for vocalization (Parkinson et al., 2012). Regions showing significant activation in the group contrast of active vocalization compared to rest were identified. We also wanted to select similar regions to those modeled previously (Tourville et al., 2008). The regions selected were superior temporal gyrus (STG), inferior frontal gyrus (IFG) and premotor (PM) cortex in both left and right hemispheres (Table 1). We specified a basic model including modulated connections from STG to PM, PM to STG and STG to IFG in both hemispheres, and included variations in modulations across hemisphere from STG to STG and within intrinsic STG connections. We specified a bilateral input to STG. There were a total of eighteen variations of the model (Figure 3).

The first model characteristic we chose to examine (factor 1) was the effect of intrinsic coupling within STG regions. We grouped models into one of two families. The first family (Ext) consisted of models with only extrinsic connections specified to be modulated by the condition. The second family (Int) consisted of models that were identical to those in family 1 with the addition of intrinsic STG connections being modulated (Table 2, Figure 3).

Based on previous evidence of the involvement of bilateral STG regions in sensory voice control (Parkinson et al., 2012), we proposed to examine differences in connectivity across hemispheres between the left and right STG regions in a second separate analysis (factor 2). For this analysis we split the eighteen models into three different families consisting of either left to right STG (LtoR), right to left STG (RtoL) or bilateral (Bilat) connections being modulated by the condition. All other model parameters were identical (Table 2, Figure 3)

Finally, in a third analysis (Factor 3), we chose to examine the effect of hemisphere. Reasoning behind this was based on previous literature suggesting that the right hemisphere is involved in pitch processing (Divenyi and Robinson, 1989; Binder et al., 1997; Johnsrude et al., 2000; Zatorre and Belin, 2001). The right hemisphere ventral premotor region is also suggested to contain a *feedback control map* where auditory and somatosensory error signals are transformed into corrective motor commands if actual and predicted feedback do not match (Tourville et al., 2011). Based on this literature we specified connections between STG to PM, PM to STG and STG to IFG as being modulated by the experimental effect. We examined three different families of models where we specified left hemisphere, right hemisphere or bilateral connections as being modulated by the experimental effect (Vocalization versus playback) (Table 2, Figure 3).

Model comparison was performed for each pitch shift magnitude separately with a Bayesian model selection (BMS) family level inference procedure (Penny et al., 2010). Family level inference identifies the “best family of models” which is the one with the highest log-evidence for a given family over the other families across subjects. We used BMS for random effects (Stephan et al. 2009) to compare families across each of our 3 factors examined for each shift magnitude. Family model inference removes uncertainty about certain aspects of model structure other than the specific factor of interest. The family of models with the highest exceedance probability, i.e. the highest relative probability compared to any other model tested, was identified. We then used these models to make inference about model structure

3. Results

3.1 Behavioral results

Figure 1.b. displays a typical vocal response to both the 100 and 400 cent shifts which are well established in the literature (e.g. Burnett et al., 1998; Hain et al., 2000). Vocal pitch responses compensated for the upward shift in pitch by decreasing. The magnitudes of the compensatory vocal pitch responses were greater for the 400 cent shift than the 100 cent shift and response onsets were typically seen around 200ms post stimulus onset.

3.2 ERP results

ERP responses to pitch shifted stimuli are already well established in the literature (Liu et al., 2011; Behroozmand and Larson, 2011; Korzyukov et al., 2012). Grand averaged ERP's from all 10 subjects at the Cz electrode for 100 cent shifts (Figure 1.c) and 400 cent shifts (Figure 1.d.) are shown. N1 and P2 responses to 400 cent shifts were larger than responses to 100 cent shifts during both active vocalization and passive playback, and responses to 400 cents were larger than those to 100 cents for both active vocalization and passive listening. Figure 2 identifies scalp potential distribution of responses for each condition over the timeframe selected for DCM analysis. This figure shows spatial variation across the scalp 1–200ms post stimulus onset and provides justification for including separate nodes for the DCM analysis.

3.3 DCM Model comparison

3.3.1 Factor 1 – Effect of intrinsic connections—BMS identified the family of models with intrinsic STG connections as the best family to represent both the 100 cent data (0.97 random effects model exceedance probability, i.e. the probability of one model being more likely than any other model examined) and the 400 cent data (0.89 random effects model exceedance probability). The Int family differed from the Ext family in that intrinsic STG connections were modulated by the experimental effect (Figure 3).

3.3.2 Factor 2 – Effect of STG modulation across hemispheres—BMS of the three families identified to examine factor 2 selected the family with left to right STG connections (LtoR) as the best family of models to represent the 400 cent shift data (0.92 random effects model exceedance probability). In this family the connections from left STG to right STG were modulated by the experimental effect but connections from right STG to left STG across hemispheres were not modulated (Figure 4). There was no clear “winning” family chosen to represent the data for the 100 cent condition.

3.3.3 Factor 3 – Effect of bilateral, left or right connections—There was no significant “winning” family for either 100 cent or 400 cent conditions when comparing the left, right and bilateral families (Figure 4).

4. Discussion

The present study was designed to examine the neural networks and associated connectivity between brain regions involved in processing voice auditory feedback. Our findings indicated that a larger pitch feedback perturbation magnitude elicited larger neural responses, complementary to similar ERP findings (Behroozmand et al., 2009; Liu et al., 2011). Given the robust effect of basic ERP findings, the understanding of network coupling properties and their causal links is critical. To that end family level inference of models identified two main findings. Firstly that intrinsic STG connectivity is significantly modulated during vocalization with pitch shifted feedback. The second main finding was that connections between STG regions across hemispheres are modulated differently based on the magnitude of the pitch shift. We identified connections between STG regions across hemispheres, specifically from left to right STG, as being modulated for the larger magnitude (400 cent) shifts but not for the smaller (100 cent) shifts. This finding identifies the importance of the STG region and cross hemisphere connections in error detection and correction during vocalization. We did not identify any differences between models involving connections in just the left hemisphere, right hemisphere or bilaterally, as being modulated by pitch shifted vocalization. However, the finding from examining cross hemisphere STG connections, where left to right STG regions are significantly modulated for larger magnitude shifts, suggests increased involvement of the right hemisphere in error detection and correction during vocalization. These connectivity differences identified between STG regions in each hemisphere relative to shift magnitude may reflect differences in the processing of self and non-self vocalizations.

This first study using DCM of ERP data to examine auditory feedback during self-vocalization identified a causal model with intrinsic STG connections across both left and right hemispheres for both the 100 and 400 cent magnitude shifts. This finding supports the idea that STG is one of the primary regions involved with identification of self-voice error. A recent study using Graph theory to examine functional networks in people with absolute pitch has shown increased functional activations and clustering within the STG region when compared to a control group providing further evidence for the role of the STG in pitch processing and how this is enhanced in a subject group who may show increased engagement of multisensory integration (Loui et al., 2012). Using SEM of fMRI data, Tourville et al (2008) also identified a model including the STG region in response to auditory error via a shift in the first formant (F1). These authors suggest a right-lateralized rather than bilateral, motor mechanisms for auditory error correction. One important difference between the Tourville et al (2008) study and ours, that limits comparison across models, is that they were concerned with speech production and not vocalization per se. Thus, shifting F1 results in a phonemic change in which the vowel sound shifts (e.g. *bet* → *bat* for an upward F1 shift and *bet* → *bit* for a downward shift).

To understand the networks involved in the pitch shift responses associated with self and non-self voices it is necessary to interpret the neural systems underlying efference copy (EC). The idea of an EC system that differentiates auditory feedback of self-voice from external voices has recently been suggested based on data derived from MEG (Houde & Jordan, 2002), ERP (Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005) and fMRI (Parkinson et al., 2012) studies. The suppression of activation (motor induced suppression) in the auditory cortex is graded such that when expected and actual output match (i.e., during self-vocalization) and are identical, there is maximal suppression, which may reflect a mechanism for identification of self-vocalization. As the difference in pitch grows (100 to 400), there is a greater mismatch (e.g. pitch shifted feedback) and suppression is reduced (Aliu et al., 2009; Heinks-Maldonado & Houde, 2005; Houde & Jordan, 2002; Liu et al., 2010), which may suggest an ambiguity about self versus other voice.

As noted above, several ERP studies have demonstrated the greatest suppression of the N1 component during vocalization with unaltered feedback and reduced suppression in conditions where the voice feedback is shifted (Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005). Our current results from modeling responses to auditory feedback shed light on the interpretation of the degree of suppression linked to the amount of pitch shift and the role of the identified network of brain regions in error correction and detection. A second aim of this study was to identify how the left and right hemisphere brain networks interact to process auditory feedback during vocalization. We examined families of models containing only left, only right or bilateral modulated connections. We have previously identified the importance of bilateral STG regions in the vocal responses to pitch shifted auditory feedback (Parkinson et al., 2012) and were interested to see how these areas interact with other regions of the network based on the magnitude of the pitch shift. The rationale for examining left versus right hemisphere differences in our model was based on previous reports of involvement of the left hemisphere premotor region in speech perception studies (e.g. Wilson et al., 2008; Hickock et al., 2009) and it has been identified using DCM of fMRI data (Osnes et al., 2011). We did not see any clear winning family when comparing the individual hemispheres. This finding suggests involvement of a similar network of regions in both hemispheres. Differences in modulation of connections in different hemispheres related to the pitch shift task may be minor in relation to the modulation of connections involved in the process of vocalization and this may explain why we do not see any specific differences in coupling parameters when comparing across conditions.

The primary differences we found between the left and right hemispheres were seen across hemisphere rather than within hemisphere in the model selected to represent the data for the 400cent shift compared to the 100 cent shift. This difference in modulated connections of the STG across hemispheres identified for 100 and 400 cent shifts provides converging evidence that there may be suppression of auditory cortex when the error between actual and predicted feedback is less. In this case for 100 cents the connection from left to right STG was not significantly modulated as it was for the 400 cent condition. At 400 cent shifts, additional network cross hemisphere connections are present whereas at 100 cent shifts they are not evoked, possibly due to the vocalization being perceived as self voice. This unique finding from the DCM model increases the specificity of the “EC” driven suppression because at a 100 cent shift, cross hemisphere coupling of STG regions were no longer causally driven by the shift.

Interpretation of our differences in the model selected to represent pitch shift magnitude when examining differences between STG connections across hemispheres may also relate to the fact that the left and right hemispheres are differentially specialized in relation to the processing of auditory temporal and spectral information (Divenyi and Robinson, 1989; Binder et al., 1997; Johnsruide et al., 2000; Zatorre and Belin, 2001). While it is clear that

both the left and right hemispheres are involved in the vocal pitch error detection and correction process as identified in the current study, different processing demands likely affect the causal network coupling properties across hemispheres. A considerable amount of evidence supports the idea that the right hemisphere auditory areas are responsible for the processing of pitch. Examination of the specialization and lateralization of the auditory cortex to spectral and temporal information has shown that damage to the right hemisphere superior temporal cortex affects a variety of spectral and tonal processing tasks (Zatorre, 1985; Divenyi and Robinson, 1989; Robin et al., 1990). More specifically lesions to the right but not left primary auditory cortical areas impaired processing of the direction of pitch change (Johnsrude et al., 2000). In direct contrast, patients with damage to the left hemisphere temporoparietal region were impaired in their ability to perceive temporal information, but the perception of spectral information was normal (Robin et al., 1990). The increased coupling from the left to right hemisphere for a 400 cent shift may be a result of identification of the increased magnitude of the error and therefore increased engagement of the right hemisphere. A smaller magnitude error may be related to more internal error correction mechanisms whereas a larger magnitude error may be attributed to non-self-voice error. One hypothesis is that the relative contribution of the right hemisphere through interhemispheric connectivity may be generalized to the actual external error identification and correction rather than specific to the pitch identification, and interaction between the two hemispheres is induced as a result of the pitch identification.

Finally, we note limitations in the current study. We recognize that additional brain regions or even more optimal networks may exist in regard to error detection and correction mechanism. For instance, we selected the model fit for the peristimulus time of 1 to 200ms. It is possible that a different time window of analysis may add to our explanation of sensory motor control of the voice. Also, we selected regions that have already been identified as being involved in auditory feedback control of the voice and to ensure our results could be compared to previous modeling of similar data (Tourville et al., 2008). We based the models in the current study on *a priori* hypotheses of the regions involved and we did not test every possible combination of connections. We also note that quantitative analysis of differences between additional regions in the left and right hemisphere (e.g. supplementary motor regions, M1 and parietal regions) are not possible due to a limitation to the number of nodes in a DCM model, thus precluding direct comparison of all these areas across hemispheres. Also, when examining coupling strengths of connections within the models, our data did not show any specific differences between shift magnitudes. This may be due to the coupling properties of regions within the model being similar for the vocal error correction process. Although we hypothesize different network processes exist for the different magnitude of shifts, the vocal response is still adjusted irrespective of shift magnitude.

In summary, in using DCM of ERP data we have shown differing connectivity between left and right hemisphere STG regions related to the magnitude of pitch shift. This suggests a different network of brain regions is involved in identifying and correcting vocal output based on the magnitude of error in the auditory feedback. These results confirm our initial hypothesis that self and non-self voice error are processed differently. We show that connections between STG across hemisphere and modulation of intrinsic STG connections are important in the identification of self-voice error. We anticipate that these insights will help to guide future models and further our understanding of sensory control of the voice.

Acknowledgments

This work was supported by National Institute of Health grant 1R01DC006243.

References

- Aliu SO, Houde JF, Nagarajan SS. Motor-induced suppression of the auditory cortex. *J Cogn Neurosci*. 2009; (4):791–802. In Press. [PubMed: 18593265]
- Bänziger T, Scherer KR. The role of intonation in emotional expressions. *Speech Communication*. 2005; 46:252–267.
- 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, Karvelis L, Liu H, Larson CR. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin Neurophysiol*. 2009; 120(7): 1303–1312. [PubMed: 19520602]
- Binder, JR.; Frost, JA.; Hammeke, TA.; Cox, RW.; Rao, SM.; Prieto, T. Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. 1997. p. 1-10.
- Boersma P. Praat, a system for doing phonetics by computer. *Glott International*. 2001
- Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am*. 1998; 103(6):3153–3161. [PubMed: 9637026]
- Daunizeau J, Kiebel SJ, Friston KJ. Dynamic causal modelling of distributed electromagnetic responses. *Neuroimage*. 2009; 47(2):590–601. [PubMed: 19398015]
- David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ. Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*. 2006; 30(4):1255–1272.10.1016/j.neuroimage.2005.10.045 [PubMed: 16473023]
- Divenyi PL, Robinson AJ. Nonlinguistic auditory capabilities in aphasia. *Brain Lang*. 1989; 37(2): 290–326. [PubMed: 2475202]
- Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Nieto-Castanon A, Guenther FH. fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*. 2011; 55(3):1324–1338. [PubMed: 21195191]
- Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang*. 2006; 96(3):280–301. [PubMed: 16040108]
- Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*. 1998; 105(4):611–633. [PubMed: 9830375]
- 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. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*. 2000; 130(2):133–141. [PubMed: 10672466]
- Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. *Psychophysiology*. 2005; 42(2):180–190. [PubMed: 15787855]
- Heinks-Maldonado TH, Nagarajan SS, Houde JF. Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*. 2006; 17(13):1375–1379. [PubMed: 16932142]
- Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*. 2011; 69(3):407–422. [PubMed: 21315253]
- Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol*. 2009; 101(5):2725–2732. [PubMed: 19225172]
- Houde JF, Jordan MI. Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of speech, language, and hearing research : JSLHR*. 2002; 45(2):295–310.
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci*. 2002; 14(8):1125–1138. [PubMed: 12495520]
- Johnsrude IS, Penhune VB, Zatorre RJ. Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain : a journal of neurology*. 2000; 123(1):155–163. [PubMed: 10611129]
- Jones JA, Munhall KG. The role of auditory feedback during phonation: studies of Mandarin tone production. *Journal of Phonetics*. 2002; 30:303–320.
- Kiebel SJ, David O, Friston KJ. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *NeuroImage*. 2006; 30(4):1273–1284. [PubMed: 16490364]

- Kiebel, S.J.; Garrido, M.I.; Moran, R.; Chen, C.C.; Friston, K.J. Dynamic causal modeling for EEG and MEG. In: Salmelin, R.; Baillet, S., editors. *Human brain mapping*. Vol. 30. 2009. p. 1866-1876.
- 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(1):71–78. [PubMed: 22044551]
- Larson CR. Cross-modality influences in speech motor control: the use of pitch shifting for the study of F0 control. *J Commun Disord*. 1998; 31(6):489–502. quiz 502–3–553. [PubMed: 9836138]
- Larson CR, Burnett TA, Kiran S, Hain TC. Effects of pitch-shift velocity on voice Fo responses. *J Acoust Soc Am*. 2000; 107(1):559–564. [PubMed: 10641664]
- Liu H, Meshman M, Behroozmand R, Larson CR. Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology*. 2011; 122(5):951–957. [PubMed: 20869305]
- Loui P, Zamm A, Schlaug G. Enhanced functional networks in absolute pitch. *Neuroimage*. 2012; 63(2):632–640. [PubMed: 22836173]
- Oostenveld, R.; Praamstra, P. The five percent electrode system for high-resolution EEG and ERP measurements. In: Praamstra, P., editor. *Clinical Neurophysiology*. Vol. 112. 2001. p. 713-719.
- Osnes B, Hugdahl K, Specht K. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage*. 2011; 54(3):2437–2445. [PubMed: 20932914]
- 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(1):314–322. [PubMed: 22406500]
- Penny WD, Stephan KE, Mechelli A, Friston KJ. Modelling functional integration: a comparison of structural equation and dynamic causal models. *NeuroImage*. 2004; 23(Suppl 1):S264–74. [PubMed: 15501096]
- Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP. Comparing Families of Dynamic Causal Models. *PLoS Comput Biol*. 2010; 6(3):e1000709. [PubMed: 20300649]
- Robin DA, Tranel D, Damasio H. Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language*. 1990; 39(4):539–555. [PubMed: 2076495]
- Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. *Neuroimage*. 2009; 46(4):1004–1017. [PubMed: 19306932]
- Tompkins CA, Mateer CA. Right hemisphere appreciation of prosodic and linguistic indications of implicit attitude. *Brain Lang*. 1985; 24(2):185–203. [PubMed: 3978402]
- Tourville JA, Guenther FH. The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*. 2011; 26(7):952–981. [PubMed: 23667281]
- Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*. 2008; 39(3):1429–1443. [PubMed: 18035557]
- Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*. 1972; 52(4):1238–1250. [PubMed: 4638039]
- Zarate JM, Wood S, Zatorre RJ. Neural Networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*. 2010; 48(2):607–618. [PubMed: 19896958]
- Zarate JM, Zatorre RJ. Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*. 2008; 40(4):1871–1887. [PubMed: 18343163]
- Zatorre RJ. Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia*. 1985; 23(1):31–41. [PubMed: 3974853]
- Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*. 2001; 11(10):946–953. [PubMed: 11549617]

Highlights

- We use a pitch shift paradigm to examine use of auditory feedback in voice control
- We use dynamic causal modeling to identify superior temporal gyrus connectivity
- We identified increased right hemisphere involvement with larger shifts
- Processing of self and non-self voice is different in left and right hemispheres

Figure 1.

a) Schematic illustration of the voice perturbation paradigm. b) Vocal responses to the pitch shift stimulus in one representative subject. Voice F0 responses to the 100cent (dashed line) and 400cent (solid line) are displayed with stimulus onset identified by the dashed vertical line. c) Grand average ERP responses to pitch-shifted voice feedback at 100cents and 400 cents stimulus. Responses from central (Cz) EEG channels are overlaid for playback with pitch shift stimulus (dashed line) and vocalization with pitch shift stimulus (solid line). The vertical dashed line marks the stimulus onset.

Figure 2.

Scalp distribution map of grand average activity for all four conditions. Time points between 0 and 200ms are displayed to identify the distribution of scalp potentials during the time period selected for DCM analysis.

Figure 3.

Eighteen versions of the model were combined into families and analyzed for both 100 and 400 cent conditions. The input to the model was set to the STG regions. Model families varied based on one of three factors; i) intrinsic connections of STG being modulated ii) differences in cross hemisphere STG connections being modulated or iii) bilateral, left or right connections between STG, PM and IFG being modulated. Blue lines identify modulated connections.

Figure 4.

BMS results for families of models examining (a) effect of models with and without intrinsic STG connections (b) effect of STG connections across hemispheres and (c) effects of laterality. The family of models with intrinsic connections was identified as the best family to fit the data for both the 100 and 400 cent shift conditions. When comparing families of models differing by the modulation of STG connections across hemispheres (factor 2), the family of models with left to right STG connections being modulated was selected over the other families for the 400 cent shift condition. For 100 cents there was no clear winning family for this comparison. In examining the effect of laterality on modulated connections there was no clear winning family of models for either condition.

Table 1

Source location coordinates in MNI space

Sources	MNI-space (x,y,z)		
Left STG	-59	-16	6
Right STG	63	-11	6
Left PM	-57	2	30
Right PM	60	14	34
Left IFG	-32	31	3
Right IFG	56	32	24

Table 2

Three separate analyses of model families were performed

Analysis	Family name and description	Models included in family
Factor 1 Effect of Intrinsic Connections	Ext – Models without Intrinsic STG connections	1,2,3,7,8,9,13,14,15
	Int – Models with Intrinsic STG connections	4,5,6,10,11,12,16,17,18
Factor 2 Effect of STG modulation across hemispheres	LtoR – Models with Left to Right STG modulated	1,4,7,10,13,16
	RtoL – Models with Right to Left STG modulated	2,5,8,11,14,17
	Both – Models with Left to right and right to Left STG modulated	3,6,9,12,15,18
Factor 3 Effect of bilateral, left or right connections	Bilat – Bilateral connections between STG, PM and IFG modulated	1,2,3,4,5,6
	Left – Only Left hemisphere connections between STG, PM and IFG modulated	7,8,9,10,11,12
	Right – Only Right connections between STG, PM and IFG modulated	13,14,15,16,17,18