The neural changes in connectivity of the voice network during voice pitch perturbation

Sabina G Flagmeier[1], **Kimberly L Ray**[1], **Amy L Parkinson**[1], **Karl Li**[1], **Robert Vargas**[1], **Larry R. Price**[3], **Angela R Laird**[1,2], **Charles R Larson**[4], and **Donald A Robin**[1]

[1]Research Imaging Institute, University of Texas Health Science Center San Antonio

[2]Department of Physics, Florida International University, Miami, FL

[3]Department of Mathematics and College of Education, Texas State University, San Marcos, TX

[4]Department of Communication Sciences and Disorders, Northwestern University, Evanston

Abstract

Voice control is critical to communication. To date, studies have used behavioral, electrophysiological and functional data to investigate the neural correlates of voice control using perturbation tasks, but have yet to examine the interactions of these neural regions. The goal of this study was to use structural equation modeling of functional neuroimaging data to examine network properties of voice with and without perturbation. Results showed that the presence of a pitch shift, which was processed as an error in vocalization, altered connections between right STG and left STG. Other regions that revealed differences in connectivity during error detection and correction included bilateral inferior frontal gyrus, and the primary and pre motor cortices. Results indicated that STG plays a critical role in voice control, specifically, during error detection and correction. Additionally, pitch perturbation elicits changes in the voice network that suggest the right hemisphere is critical to pitch modulation.

1.0 Introduction

The use of the human voice is essential for oral communication and is controlled by complex neural processing that drives feedforward and feedback mechanisms. Given the primacy of auditory feedback in voice control a neurobiological model of phonation based on sensory feedback is essential. Peripheral mechanisms of voice control, including respiratory, laryngeal and articulatory systems, have been heavily studied and are well understood; however, information related to neural mechanisms of voice control remains elusive (Bauer et al., 2006). The study of the underlying properties associated with systems-level neural network of vocalization can provide insight into the relations between vocal output and sensory feedback. Recent developments in neuroimaging not only allow for the

identification of regions involved in this complex system but also allow for the development of effective connectivity models. Here, we developed models of neural causal linkage using data from a pitch shift auditory feedback paradigm where the pitch of self voice feedback was unexpectedly changed during vocalization (Burnett et al., 1998, Parkinson et al., 2012, Larson, 1998).

Vocal control utilizes the accurate perception and integration of the auditory signal and somatosensory information generated by the individual (Burnett et al., 1997, Hain et al., 2000, Heinks-Maldonado et al., 2005, Golfinopoulos et al., 2011, Parkinson et al., 2012). During vocalization a shift is perceived as an error in production and triggers corrective mechanisms whereby subjects respond to the pitch-shift by changing their own voice fundamental frequency (F0) in the opposite direction to the shift. In speech and voice systems the presence of error signals are generated as a result of a mismatch between a predicted outcome and sensory feedback. Both functional imaging and ERP analyses using perturbation paradigms have previously indicated that the superior temporal gyrus is a key brain region involved in coding mismatches between expected and actual auditory signals and that the right hemisphere is especially involved in pitch processing; (Parkinson et al., 2012, Behroozmand and Larson, 2011, Tourville et al., 2008, Guenther et al., 2006, Zarate and Zatorre, 2008) however, it is well known that the brain operates as a network rather than as isolated modules. As a result, this study aims to extend previous reports on the voice *network* and identify how that network changes as a response to a detected error in pitch. Consequently, we developed two independent data-driven models of best fit for a shift and a no shift condition.

Brain imaging can uncover much about the neural control of the voice. Effective connectivity analyses allow for study of interactive processes and causal relations in the underlying neural network associated with vocalization and other motor activities. Structural equation modeling (SEM) utilizes knowledge gained from imaging modalities and provides a model of the effective connectivity in a given neural system (Laird et al., 2008). For example, using a stacked modeling approach, Tourville et al. used SEM to model network connectivity involved in speech with and without first formant frequency (F1) shifts to examine connectivity as it relates to a computational speech model (DIVA). This analysis showed that an unexpected F1 shift of participants' speech resulted in significant influence from bilateral auditory regions to frontal regions indicating that corrective mechanisms from auditory error cells are sent to regions of motor control in response to errors during speech (Tourville et al., 2008). While this analysis gives important insight into perceived error in speech it differs from our analysis in two keys ways. Firstly, unlike F0, F1 shifts are typically used during normal speech to change phonemic categories. As a result, F1 shifts are likely different from shifts in F0. Secondly, the stacked model approach tested a fully constrained model. The approach employed by this study is minimally constrained; consequently, this approach removes bias that could result from a priori constraint and uncovers pathways that best fit the model from an unbiased standpoint. Therefore, further investigation of the neural network responsible for voice control is warranted.

Here, we examined the effective connectivity of voice control using a data-driven approach to SEM. We utilized data from a previously published fMRI dataset (Parkinson et al., 2012)

that employed the pitch shift paradigm during vocalization. We created two models (shift/no shift) examining bilateral cortical brain regions previously identified as being involved in vocalization, including the superior temporal gyrus (STG), premotor cortex (PMC), primary motor cortex (M1), and inferior frontal gyrus (IFG) (Brown S, 2008, Tourville et al., 2008, Parkinson et al., 2012). We hypothesized that our models would confirm differences in connectivity between models for regions involved in audio-vocal integration. Differences between models were identified through the absence or presence of pathways as well as connection strengths. The path coefficients represents the direct proportional functional influence one region has on another (McIntosh 1994). Furthermore, due to previous work that showed differences in processing during perturbation in bilateral STG, we hypothesized that bilateral STG would show changes in modulation between the two models (Parkinson et al. 2012). We expected that this would result in a greater degree of involvement in error processing (shift condition) than in typical vocalization (no shift) between regions, which would be indicated by a larger path coefficient.

2.0 Methods

2.1 Participants

Subject data was obtained from a previous functional imaging study (Parkinson et al., 2012). This sample included ten right-handed English-speaking subjects. Two of these subjects were omitted from the current analysis due to lack of activations in the no shift vs. rest condition in two or more seed regions and two additional subjects scanned since publication of the above study were included. This provided ten subjects (4 males, 6 females, mean age 30) with no history of neurological disorder. Prior to functional imaging, subjects underwent pre-screening to ensure that all subjects showed a vocal response to the pitch-shift paradigm (Change in baseline of pitch magnitude in the upward or downward direction following a pitch shift). This has been standard practice for over a decade of testing and less than five percent of subjects do not show a response. No subjects were eliminated due to this criterion for our experiment. Inclusion criteria also required that subjects were safe for MRI scanning, had normal hearing, reported no neurological deficits, no speech or voice disorders and no formal musical experience in the past 10 years. The institutional review board of the University of Texas Health Science Center at San Antonio approved all study procedures.

2.2 Experimental Procedure

A detailed description of MRI scanning procedures and imaging acquisition can be found in Parkinson et al. 2012. In summary, subjects lay in the scanner with electrostatic headphones (Koss KSP 950) and viewed a monitor screen displaying a visual cue, "ahhh". Each trial began with the presentation of a speech or rest visual cue. Subjects vocalized until the cue disappeared from the screen (5s). During vocalization the subject's voice was shifted ±100 cents (200ms; randomized direction; > 250ms post onset) during shift trials, and had no shift during vocalization only conditions. When presented with a rest cue, subjects remained silent. Data were stored to a PC workstation and analyzed offline. An experimental block consisted of 64 trials, 48 vocalization trials (16 shift-up, 16 shift-down, 16 no- shift) and 16 rest trials. The trials were presented in a random order. Each subject performed 3 experimental blocks within the session and there was a 2-minute rest period between each

block. All structural and fMRI data were acquired on a Siemens Trio 3T scanner. Three full-resolution structural images were acquired using a T1-weighted, 3D TurboFlash sequence with an adiabatic inversion contrast pulse with a resolution of 0.8 mm isotropic. The scan parameters were TE = 3.04, TR=2100, TI=78 ms, flip angle=13, 256 slices, FOV=256 mm, 160 transversal slices. The three structural images were combined to create an average, which was then used to register the brain of each subject to their functional data. The functional images were acquired using a sparse sampling technique. T2* weighted BOLD images were acquired using the following parameters; FOV 220 mm, slice acquisition voxel size = $2 \times 2 \times 3$mm, 43 slices, matrix size = 96×96, flip angle=90, TA=3000 ms, TR=11,250 ms and TE=30 ms. Slices were acquired in an interleaved order with a 10% slice distance factor. Each experimental run of the task consisted of 64 volumes. Functional data were obtained using a sparse sampling technique triggered by a digital pulse sent from the stimulus computer for each event.

2.3 Region of Interest (ROI) selection

Prior studies have found that primary motor cortex, superior temporal gyrus, anterior cingulate cortex, supplementary motor area, premotor cortex, insula, thalamus, putamen, and cerebellum are all part of the vocalization network (Brown S, 2008, Zarate and Zatorre, 2008, Parkinson et al., 2012). While all regions found in the cited works are contributors to vocalization and are important, we were unable to include all regions in our model as this would cause a loss in statistical power. As a result, we chose 8 regions consistent with the above reports that showed robust activation in the Parkinson et al. 2012 paper. The regions selected were examined bilaterally due to differential processing between hemispheres. Regions in our models included bilateral superior temporal gyrus (STG), bilateral inferior frontal gyrus (IFG), bilateral premotor cortex (PMC), and bilateral primary motor cortex (M1). In Parkinson 2012, superior temporal gyrus demonstrated increased activation during shift conditions when compared to no shift vocalization. Furthermore, it is involved in auditory-vocal integration and processing of predicted and actual vocal output (Zarate and Zatorre, 2005). Additionally, we investigated IFG, which was shown as an imperative part of the speech/vocalization network and has been identified as a site for additional sensory processing for motor planning and control of vocalization (Tourville et al., 2008, Zarate et al., 2010, Parkinson 2013). Premotor cortex has been identified as a location for selecting alternatives to already programed learned responses as well as generating motor commands for speech and vocalization (Zarate et al., 2010, Tourville et al, 2008). Primary motor cortex was selected for its involvement in sending motor commands to be executed. Primary motor cortex is functionally connected with IFG giving rise to speech and vocalization making it an optimal candidate for this analysis (Greenlee et al., 2004). Given the limited number of data points made available by sparse sampling, subcortical regions were not included in the bilateral model. Instead, we focused on cortical contributions to vocalization with and without shifted feedback. Separate models were created for the shift and no shift conditions. Specific coordinates for regions of interest were identified from the unshifted vocalization versus rest contrast in a group analysis (Table 1). Individual ROIs were created (125mm³ cubic volume centered around the specified MNI coordinate) for each of the above listed regions using the multi-image analysis GUI (Mango) image processing software (http://ric.uthscsa.edu/mango/) (Lancaster et al., 2012). Individual ROIs were converted from the

normalized MNI space back to native subject space allowing for the extraction of raw data from each individual subject while ensuring that data were extracted from the identical sites across subjects.

2.4 Data Analysis

Preprocessing was performed using the FSL 4.1.4 (FMRIB Software Library) software package. Head motion was corrected using MCFLIRT and non-brain was removed from the structural image using the BET brain extraction tool. The functional EPI images were smoothed using a FWHM of 5mm and transformed to MNI space using FSLs registration tools. The FMRI BOLD signal was extracted from each ROI for each subject's data set and experimental condition. Data points for the shift condition and no shift conditions were averaged and separated into respective conditions from the data set using FSL MEANTS (FSL tool used to average times series data) and concatenated across the three scanning sessions for each subject resulting in 480 data points per condition per ROI. Extracted data for each ROI was then normalized to a mean of zero and standard deviation of one.

Effective connectivity of regions activated during shift and no-shift paradigms was assessed using path analysis within a structural equation modeling framework (AMOS version 19.0, SPSS, IBM). While the typical strategy for SEM is to implement *a priori* hypotheses to fully constrain the SEM models as seen in the Tourville 2008 study, this can be misleading. Instead, we chose to employ an approach with minimal a priori constraint which allowed for the production of data driven models for vocalization (Laird et al., 2008, (Laird et al., 2008; Hastie, Tibshirani & Friedman, 2009). While the results from Tourville's stacked model are important, our goal differed from the Tourville study. Our goal was to provide a data driven model that reduced bias introduced by a priori models. Bias is the result of a fully constrained model requiring assumptions to be made which can potentially limit the identification of vital connections within a system. Due to our data driven approach, we were able to examine key pathways that may not have been identified a priori. Furthermore, our model started with a full comprehensive model that included all possible paths from our point of origin. To establish a starting connection for each structural equation model, we imposed a prior assumption identifying superior temporal gyrus as the initial region receiving auditory input. The use of STG as the initial region of input is supported by research indicating that information from an auditory stimulus reaches STG approximately 12 – 17ms from the stimulus onset (Steinschneider et al., 1999, Inui et al., 2006). Thus, it was hypothesized that STG interacts with one or more of the remaining variables/regions. Paths connecting the STG to all other regions were established and a specification search was employed to determine the best combination of connected regions following the guidelines of Burnham & Anderson (2002). Specification search allows for multiple candidate models to be tested using optional unidirectional path loadings. The Browne-Cudeck criterion value (BCC) is an information-theoretic index that represents the predictive fit index and is used to select among competing models fit to the same data (Schumacker & Lomax, 2010, p. 230). In this analysis, the model with the lowest BCC value was selected as the model that best represented the data (Laird et al. 2008). The next sets of candidate pathways were identified in an exploratory manner through the use of modification indices (MI). Paths with the highest MI were chosen as the next likely paths. The new paths were

added to the model, and an additional specification search was conducted. This search procedure continued in an iterative manner until a root mean square error of approximation (RMSEA) value of less than .08 was achieved. An RMSEA of less than .08 is indicative of a close fit of the sample (empirical) covariance matrix to the population matrix (Browne and Cudeck, 1993). Goodness of fit of the overall model was determined using descriptive statistics such as the likelihood ratio chi-square statistic, ^{2}, (models with a ^{2} of zero indicate a theoretical model that fits the data perfectly), p-value (high p-values indicate a model is unlikely to be refuted in other independent samples), and a root mean square error of approximation (RMSEA) index of less than .08 indicating minimal discrepancy between the empirical or sample covariance matrix and the population. The class of models evaluated in this study was nonrecursive. In nonrecursive SEMs the presence of bidirectional feedback loops creates the possibility of a non-stable system resulting in biased parameter estimates. In our models, stability of the nonrecursive system was evaluated using the stability index based on the work of Bentler and Freeman (1983). In all models the stability index was between −1.0 and 1.0 verifying that the nonrecursive models were stable. Separate nonrecursive models were created for the shift and no shift conditions.

3.0 Results

3.1 No Shift condition

The no shift condition revealed connectivity associated with vocalization without error with a chi square fit index of 31.411, RMSEA =.071. Not surprisingly, we found that there are many connections between and within hemispheres. Connections presented in the left hemisphere include left M1 to left PMC, left STG, and left IFG which emphasizes the extent of connectivity necessary with the motor cortex to execute speech accurately. Left IFG showed coupling with left PMC regions commonly associated with the voice and speech network and contributors to speech articulation and retrieval of speech sounds. Left STG showed a relationship with left IFG and likely contributes to voice perception and processing. Right hemisphere connections include right M1 to right IFG and right PMC. A negative connection from right IFG to right M1 was also observed. The connections in the right hemisphere contribute to pitch processing. Cross hemisphere connections include, left STG to right M1, left IFG to right M1, left STG to right STG, and left IFG to right PMC. Lastly, a negative connection is visible from right PMC to left IFG. These cross-hemisphere connections indicate that vocalization requires crosstalk from both hemispheres to ensure accurate vocalization. No shift connectivity is shown in black (Figure 1).

3.2 Shift condition

The shift condition consisted of rapid 200ms shifts presented to the subject. These quick deviations from the subjects' intended vocal output were likely processed as errors. Therefore, changes in connectivity between the no shift and shift conditions are likely due to this detected error and the processes associated with error correction. Here we present the resulting shift model which yielded a chi square fit index of 32.302, and RMSEA =.072. Connections within the left hemisphere included left PMC to left M1, left STG to left IFG, left IFG to left PMC and a negative connection between left PMC and left IFG. Right intra-hemispheric connections include right M1 to right IFG, right PMC to right M1 and right

STG to right IFG. A negative coupling is seen from right IFG to right STG as well. Interestingly, negative pathways are generated during the shift condition that are not present in the no shift condition. This change of circuitry indicates differential processing necessary during the detection and correction of perceived vocal error. Cross-hemispheric connections include right primary motor cortex to left primary motor cortex, and left STG. Left IFG is coupled with right PMC. Importantly, a connection between left STG to right STG is observed. Additionally, a negatively correlated connection is present from right STG to Left STG (Figure 2).

4.0 Discussion

The focus of this study was to use effective connectivity modeling of fMRI data to determine neural networks involved in vocal control and identify pathways that are key to detecting and correcting vocal errors. Vocalization is a highly complex motor skill that requires coordination amongst multiple effector systems (e.g., respiratory and vocal) at a rapid pace. In order to execute voluntary actions with precision, both feedforward and feedback systems are integrated. Feedforward models compare anticipated changes to be imposed with the actual output (Jeannerod et al., 1979). Therefore, it is the difference between the actual and predicted sensory feedback that results in a sensory error, which is used to correct the current state estimate (Chang et al., 2013, Wolpert et al., 1995). Given that we delivered perturbation to the subjects during mid vocalization, these perturbations are processed as errors in self-vocalization (Behroozmand et al., 2011, Liu et al., 2010). As a result, we predicted that STG would serve as a vital region in error detection; therefore, STG would show differences in connectivity when an error was present compared to unperturbed vocalization. Consistent with our hypothesis, we found differences in neural connectivity of the voice network associated with vocal perturbations. Data support the idea that STG plays a crucial role in vocalization and shift processing as evidenced by our model. Our analysis also revealed the emergence of negative pathways that we interpret as feedback loops for during shifted vocalization that are not present with unperturbed productions.

4.1 The critical role of STG

Coupling between right STG and left STG in the no shift condition indicated that this path is critical to vocalization. Using a simple effect size computation (r^2), one can see that approximately 5% of the variance in the direct relationship between left STG to right STG is accounted for in the no shift model; however, in the shift condition 50% of the variance is accounted for by this pathway. Interestingly, when an error was present, coupling between right STG and left STG changed further by uncovering a pathway with a negative path coefficient from right STG to left STG creating a negative feedback loop. This feedback loop to and from bilateral STG regions is likely used for the rapid fine-tuning of motor commands. In SEM, feedback loops represent reciprocal connections between neural regions. The presence of these feedback loops is a result of functional differences between shift and no shift conditions; however, these differences are discussed with caution due to the inability to interpret connectivity relative to the sign of the path (positive/negative) (McIntosh, 1994). These differences are discussed below.

Studies have indicated that STG acts as a location for efference copy mechanisms which involve comparison of afferent vocal feedback and efferent motor and sensory predictions (Parkinson et al., 2012, Chang et al., 2013, Heinks-Maldonado et al., 2005). Parkinson et al. (2012) used fMRI to uncover neural regions involved in vocalization and error detection. A subtraction analysis revealed increased activity in STG during shift when contrasted with the no shift condition and revealed increased neural activity related to error detection and correction during vocalization (Parkinson et al., 2012). Studies using event related potentials (ERPs) show that responses to predicted vocal output are suppressed compared to listening to a playback of one's own voice; however, when the predicted output does not match the resulting output, there is an enhancement in the ERP response to self vocalization (Heinks-Maldonado et al., 2005, Behroozmand and Larson, 2011). ERP literature supports the idea that increased computation and fine-tuning of the neural signal is required for error detection and correction. High-resolution invasive intracranial recordings have confirmed this phenomenon, revealing a suppressed response to vocalization specifically in the superior temporal gyrus in response to self-vocalization (Greenlee et al., 2011). ERP and ECoG findings in conjunction with findings from our study, support forward models of voice control and suggest that efference copies of motor commands modulate the activity in bilateral STG.

The feedback loop generated in the shift condition may be the result of the need for fine-tuning from specialized regions to correct for the detected error. It has been suggested by previous studies that right and left hemispheres are specialized and respond to the auditory feedback differently with the right hemisphere showing specialization for spectral information (frequency) and the left showing sensitivity to temporal information (Hickok et al., 2011, Johnsrude et al., 2000, Zatorre and Belin, 2001, Zatorre et al., 1992, Behroozmand et al., 2012, Robin et al., 1990). For example, Robin et al. (1990) examined patients with left temporoparietal lesions, right temporoparietal lesions and healthy controls during temporal and spectral tone discrimination tasks. Results demonstrated that lesions to the left hemisphere resulted in impaired ability to perceive temporal information but did not impair ability to perceive spectral information. The right hemisphere lesion group displayed an ability to process temporal information but not spectral. Behroozmand and colleagues (2012) produced data that further supported this idea when examining +200 cent shifts during and auditory feedback task of self-vocalization, complex tones and pure tones with missing fundamental. Zatorre 1988 showed that patients with right surgical excisions of the right auditory cortex (left intact) are impaired at perceiving pitch in complex tones with missing fundamental. Furthermore, in a pitch discrimination task, patients with right but not left temporal lobe excisions showed significantly elevated thresholds for directional changes of pitch (Johnsrude et al., 2000). Increased communication between these two regions during a shift could be the result of fine-tuning necessary during error detection that is not needed for vocalization without error.

4.2 The role of IFG

Our analysis indicated that the detection of an error resulted in the presence of a feedback loop between right IFG and right STG. This change in coupling properties indicates the need for these regions in the right hemisphere in error detection during voice production and

further fine-tuning of the actual execution of the motor command. Studies have shown that connections between IFG and STG specifically, are important to pitch processing and are therefore necessary in the detection and correction of errors in vocal performance. The neural network for pitch processing, which includes the pars triangularis of Broca's area and the right superior temporal gyrus (STG), plays a vital role in melodic and lexical pitch processing (Nan and Friederici, 2012). Evidence that pitch processing is similar for both tonal speech and music supports the idea that IFG plays a large role in pitch processing regardless of modality and could be consistent with the link between right STG and right IFG (Nan and Friederici, 2012). Additionally, support for increased activity between these regions stems from work examining song where a predominance of right IFG contribution to melody is thought to be due to elongated vowels (Merrill et al., 2012). Finally, Tourville et al. observed increased activation of IFG during shift vs. no shift of the F1. Authors concluded that IFG was responsible for additional processing of sensorimotor information in response to error detection (STG). Our findings support this conclusion.

In our model, the connection left STG to left IFG as well as left IFG to left PMC is present in both shift and no shift conditions. Similar to the right hemisphere, the presence of an unexpected pitch shift resulted in a feedback loop from left PMC to left IFG. This finding suggests a change in circuitry during vocalization in the presence of an error and is in accordance with models of speech control which indicate that left IFG and left premotor areas are crucial to audio-vocal integration which is necessary in error detection and correction (Hickok et al., 2009, Guenther et al, 2006). Guenther and colleagues (Guenther et al., 2006) posed that the left STG is the site responsible for *sound error maps* while left IFG contains *speech sound maps* and plays a role in motor programming in the DIVA model (Guenther et al., 2006, Golfinopoulos et al., 2011). This aligns nicely with our model, which implies increased influence between these regions during error processing. Additionally, Papoutsi et al. (2009) supports the existence of a "dorsal stream" proposed by Hickok for speech processing, which suggests that inferior frontal gyrus, premotor area and sPT are a core network in speech production (Papoutsi et al., 2009). Given this, it is possible that the similarities between the shift and no shift condition are indicative of the necessity of coupling between left IFG and left premotor cortex in vocalization. Furthermore, the development of the feedback loop in our analysis is likely due to the increased need for processing corrective motor commands to be sent to M1 thus contributing to this change in circuitry.

Results showed coupling of inferior frontal gyri and the primary motor cortices regardless of the presence of a shift. This is likely a result of IFG's critical involvement in speech production and functional connections with the primary motor cortex. The coupling observed between IFG and the primary motor cortices is supported by invasive surface recording data. Using this technique, Greenlee et al. determined that stimulation in IFG resulted in recorded evoked potentials in orofacial motor cortex and stimulation in orofacial motor cortex resulted in evoked potentials in IFG (Greenlee et al., 2004). These data provided evidence of a functional connection between these two regions and supports our findings.

The role of motor cortices in vocalization—Our analysis also showed several connections with the primary motor cortices. This is not a surprising finding given the need for motor commands to be sent from these regions for vocalization. Activation from bilateral motor cortex is likely a result of the vocal folds being bilaterally innervated. The shift condition did result in a cross-hemispheric excitatory connection from right M1 to left M1 that is not seen in the no shift condition. While bilateral motor cortex does play a role in vocalization regardless of the presence of a shift, the coupling induced by the shift is likely due to increased demand for error correction that is not necessary during the no shift condition.

5.0 Possible limitations

While the findings in this study provide insights into feedback control of the human voice, there are limitations that must be noted. First, we acknowledge that more optimal networks may exist or the inclusion of additional regions may enhance the analysis, For instance, the cerebellum is thought to play a role in error detection, possibly acting as a site of comparison between predicted and actual feedback or as the location responsible for sending corrective motor control signals to the motor cortices (Ito, 2013, Blakemore et al., 1998, Knolle et al. 2012a, Knolle et al. 2012b, Knolle et al. 2013). However, we selected regions we found important to vocal control and error detection given our previous study and existing literature that allow for a reliable SEM analysis that is not lacking in statistical power and cerebellar activations did not survive our analysis. Secondly, the method of data collection (ie, sparse sampling) necessary for our experimental design limited the number of data points used in this analysis. While this is a drawback, SEM is an ideal method of analysis for sparse sampling as it does not require a time series when calculating the path coefficients. Other modeling methods such as dynamic causal modeling, however, do have a requirement for an accurate time series. Lastly, the differences observed between the shift and no shift networks are qualitative in nature however we still obtain valuable information regarding changes in connectivity elicited from error detection and correction and have identified models that best represent the data set.

6.0 Conclusions

In conclusion, we used structural equation modeling to examine differences in connectivity during no shift and shifted vocalization. Our analysis indicated coupling between left STG to right STG in both the shift and no shift conditions; however, the shift condition introduced a negative path from right STG to left STG. These results in conjunction with previous literature, confirms our hypothesis that STG plays a vital role in error detection and correction. Furthermore, the presence of a shift alters the network circuitry between many of the regions in our model specifically introducing feedback loops between right IFG and right STG, and left IFG and left premotor when an error is detected. Previous literature suggests that the right hemisphere, is specialized for pitch processing and may play a key role in the development of these loops as an attempt to complete high-level processing required for error detection and correction of vocalization. Understanding how these networks are connected during vocalization and how they change as a result of detected errors is critical to understanding voice regulation.

8.0 References

Bauer JJ, Mittal J, Larson CR, Hain TC. Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. J Acoust Soc Am. 2006; 119:2363–2371. [PubMed: 16642849]

Behroozmand R, Korzyukov O, Larson CR. ERP correlates of pitch error detection in complex tone and voice auditory feedback with missing fundamental. Brain Research. 2012; 1448:89–100. [PubMed: 22386045]

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 CR. Time-dependent neural processing of auditory feedback during voice pitch error detection. J Cogn Neurosci. 2011; 23:1205–1217. [PubMed: 20146608]

Blakemore SJ, Wolpert DM, Frith CD. Central cancellation of self-produced tickle sensation. Nat Neurosci. 1998; 1:635–640. [PubMed: 10196573]

Bentler PM, Freeman EH. Tests for stability in linear structural equation systems. Psychometrika. 1983; 48:289–308.

Brown S, NELM. A larynx area in the human motor cortex. Cereb Cortex. 2008; 18:837–845. [PubMed: 17652461]

Browne, MW.; Cudeck, R. Alternative Ways of Assessing Model Fit. In: Bollen, K.; Long, J., editors. Testing Structural Equation Models. Sage; Newbury Park, CA: 1993. p. 136-162.

Burnham, KP.; Anderson, DR. Model selection and multimodal inference: a practical information-theoretic approach. 2. New York: Springer-Verlag; 2002.

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. J 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. Proc Natl Acad Sci U S A. 2013; 110:2653–2658. [PubMed: 23345447]

Golfinopoulos E, Tourville JA, Bohland JW, Ghosh SS, Nieto-Castanon A, Guenther FH. fMRI investigation of unexpected somatosensory feedback perturbation during speech. Neuroimage. 2011; 55:1324–1338. [PubMed: 21195191]

Greenlee JD, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H, Howard MA 3. Human auditory cortical activation during self-vocalization. PLoS ONE. 2011; 6:e14744. [PubMed: 21390228]

Greenlee JDW, Oya H, Kawasaki H, Volkov IO, Kaufman OP, Kovach C, Howard MA, Brugge JF. A functional connection between inferior frontal gyrus and orofacial motor cortex in human. J Neurophysiol. 2004; 92:1153–1164. [PubMed: 15056683]

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

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]

Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning. 2. New York: Springer; 2009.

Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. Psychophysiology. 2005; 42:180–190. [PubMed: 15787855]

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. J Neurophysiol. 2009; 101:2725–2732. [PubMed: 19225172]

Inui K, Okamoto H, Miki K, Gunji A, Kakigi R. Serial and parallel processing in the human auditory cortex: a magnetoencephalographic study. Cereb Cortex. 2006; 16:18–30. [PubMed: 15800024]

Ito M. Error detection and representation in the olivo-cerebellar system. Front Neural Circuits. 2013; 7:1. [PubMed: 23440175]

Jeannerod M, Kennedy H, Magnin M. ScienceDirect.com - Neuropsychologia - Marc Jeannerod, Henry Kennedy, Michel Magnin. Neuropsychologia. 1979

Johnsrude IS, Penhune VB, Zatorre RJ. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain. 2000; 123 (Pt 1):155–163. [PubMed: 10611129]

Knolle F, Schroger E, Baess P, Kotz SA. The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. J Cogn Neurosci. 2012a; 24:698–706. [PubMed: 22098261]

Knolle F, Schroger E, Kotz SA. Cerebellar contribution to the prediction of self-initiated sounds. Cortex. 2012b

Knolle F, Schroger E, Kotz SA. Prediction errors in self- and externally-generated deviants. Biol Psychol. 2013; 92:410–6. [PubMed: 23246535]

Laird AR, Robbins JM, Li K, Price LR, Cykowski MD, Narayana S, Laird RW, Franklin C, Fox PT. Modeling motor connectivity using TMS/PET and structural equation modeling. Neuroimage. 2008; 41:424–436. [PubMed: 18387823]

Lancaster JL, Laird AR, Eickhoff SB, Martinez MJ, Fox PM, Fox PT. Automated regional behavioral analysis for human brain images. Front Neuroinform. 2012; 6:23. [PubMed: 22973224]

Larson CR. Cross-modality influences in speech motor control: the use of pitch shifting for the study of F0 control. J Commun Disord. 1998; 31:489–502. quiz 502–3– 553. [PubMed: 9836138]

Liu H, Behroozmand R, Larson CR. Enhanced neural responses to self-triggered voice pitch feedback perturbations. NeuroReport. 2010; 21:527–531. [PubMed: 20386347]

McIntosh AR, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping. 1994; 2:2–22.

Merrill J, Sammler D, Bangert M, Goldhahn D, Lohmann G, Turner R, Friederici AD. Perception of words and pitch patterns in song and speech. Front Psychol. 2012; 3:76. [PubMed: 22457659]

Nan Y, Friederici AD. Differential roles of right temporal cortex and Broca's area in pitch processing: Evidence from music and Mandarin. Hum Brain Mapp. 2012

Papoutsi M, de Zwart JA, Jansma JM, Pickering MJ, Bednar JA, Horwitz B. From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. Cereb Cortex. 2009; 19:2156–2165. [PubMed: 19181696]

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]

Parkinson AL, Korzyukov O, Larson CL, Litvack V, Robin DA. Modulation of effective connectivity during vocalization with perturbed auditory feedback. Neuropsychologia. 2013; 51:1471–1480. [PubMed: 23665378]

Robin DA, Tranel D, Damasio H. Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. Brain Lang. 1990; 39:539–555. [PubMed: 2076495]

Schumacker, RE.; Lomax, RG. A beginner's guide to structural equation modeling. 3. New York: Taylor & Francis/Routledge; 2010.

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J Neurophysiol. 1999; 82:2346–2357. [PubMed: 10561410]

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

Wolpert DM, Ghahramani Z, Jordan MI. An Internal Model for Sensorimotor Integration. Science. 1995; 269:1880–1882. [PubMed: 7569931]

Zarate JM, Wood S, Zatorre RJ. Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. Neuropsychologia. 2010; 48:607–618. [PubMed: 19896958]

Zarate JM, Zatorre RJ. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. N.Y. Acad Sci. 2005; 1060:404–408.

Zarate JM, Zatorre RJ. Experience-dependent neural substrates involved in vocal pitch regulation during singing. Neuroimage. 2008; 40:1871–1887. [PubMed: 18343163]

Zatorre RJ. Pitch perception of complex tones and human temporal-lobe function. J Acoust Soc Am. 1988; 84:566–572. [PubMed: 3170948]

Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. Cereb Cortex. 2001; 11:946–953. [PubMed: 11549617]

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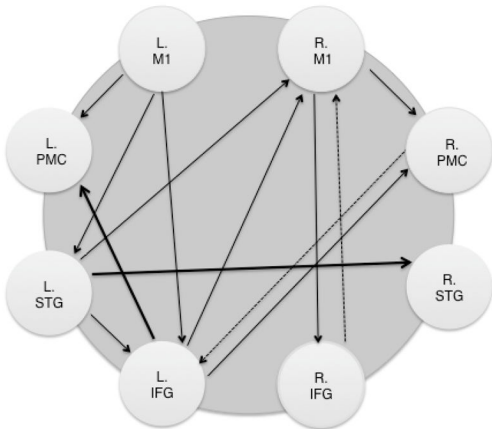
Highlights

- We modeled voice networks with and without auditory shifts using structural equation modeling.

- We examined differences in connectivity between the two models.

- We determined that STG is critical to vocalization and error detection/correction.
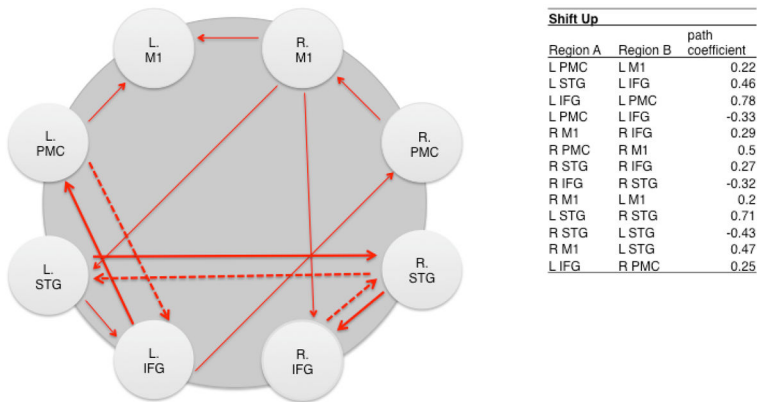
SEM: No Shift

No Shift		
Region A	Region B	path coefficient
L M1	L PMC	0.2
L STG	L IFG	0.41
L IFG	L PMC	0.54
L M1	L STG	0.29
L M1	L IFG	0.16
R M1	R IFG	0.55
R IFG	R M1	-0.3
R M1	R PMC	0.46
L STG	R STG	0.24
L STG	R M1	0.36
L IFG	R M1	0.33
L IFG	R PMC	0.29
R PMC	L IFG	-0.19

FIGURE 1.

No Shift connectivity model – chi sq = 31.411, df = 15, P = .008, RMSEA = .071, BIC = 161, BCC = 74.

SEM Shift

Shift Up		
		path
Region A	Region B	coefficient
L PMC	L M1	0.22
L STG	L IFG	0.46
L IFG	L PMC	0.78
L PMC	L IFG	-0.33
R M1	R IFG	0.29
R PMC	R M1	0.5
R STG	R IFG	0.27
R IFG	R STG	-0.32
R M1	L M1	0.2
L STG	R STG	0.71
R STG	L STG	-0.43
R M1	L STG	0.47
L IFG	R PMC	0.25

FIGURE 2.

Shift connectivity model – chi sq = 32.302, df = 15, P = .006, RMSEA = .072, BIC = 161, BCC = 75. Negative path coefficients are represented with a dashed line.

TABLE 1

ROIs for SEM analysis were 5mm^3 in volume. Coordinates are presented in MNI space.

Region	BA	x	y	z
Right superior temporal gyrus	42	60	−33	10
Left superior temporal gyrus	42	−61	−13	1
Right inferior frontal gyrus	44	61	6	14
Left inferior frontal gyrus	44	−57	7	16
Right premotor cortex	6	54	−4	45
Left premotor cortex	6	−57	1	25
Right primary motor cortex	4	44	−12	36
Left primary motor cortex	4	−47	−8	42