Brain & Language 132 (2014) 7–13

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

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

Sabina G. Flagmeier [a], Kimberly L. Ray [a], Amy L. Parkinson [a], Karl Li [a], Robert Vargas [a], Larry R. Price [c], Angela R. Laird [a,b], Charles R. Larson [d], Donald A. Robin [a,e,f,g,h,*]

[a] Research Imaging Institute, University of Texas Health Science Center at San Antonio, United States
[b] Department of Physics, Florida International University, Miami, FL, United States
[c] Department of Mathematics and College of Education, Texas State University, San Marcos, TX, United States
[d] Department of Communication Sciences and Disorders, Northwestern University, Evanston, United States
[e] Neurology, University of Texas Health Science Center at San Antonio, United States
[f] Biomedical Engineering, University of Texas Health Science Center at San Antonio, United States
[g] Radiology, University of Texas Health Science Center at San Antonio, United States
[h] Honor's College, University of Texas San Antonio, San Antonio, United States

ARTICLE INFO

Article history:
Accepted 4 February 2014
Available online 27 March 2014

Keywords:
Voice control
Error detection and correction
Structural equation modeling
Connectivity
fMRI
Sparse sampling

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. 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, Mittal, Larson, & Hain, 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, Freedland, Larson, & Hain, 1998; Larson, 1998; Parkinson et al., 2012).

Vocal control utilizes the accurate perception and integration of the auditory signal and somatosensory information generated by the individual (Burnett, Senner, & Larson, 1997; Golfinopoulos et al., 2011; Hain et al., 2000; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; 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; (Behroozmand & Larson, 2011;

* Corresponding author. Address: Research Imaging Institute – MC 6240, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, United States.
E-mail address: robind@uthscsa.edu (D.A. Robin).

http://dx.doi.org/10.1016/j.bandl.2014.02.001
0093-934X/Published by Elsevier Inc.

Guenther, Ghosh, & Tourville, 2006; Parkinson et al., 2012; Tourville, Reilly, & Guenther, 2008; Zarate & 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, 2009; Parkinson et al., 2012; Tourville et al., 2008). 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 & Gonzalez-Lima, 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. 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 (5 s). During vocalization the subject's voice was shifted ±100 cents (200 ms; randomized direction; >250 ms 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 off-line. 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-min 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, 2009; Parkinson et al., 2012; Zarate & Zatorre, 2008). 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.