Cerebral Response to Different Voice Production: A Functional Magnetic Resonance Imaging Study

K. K. Peck[1], J. Galgano[2], R. Branski[3], M. Ho[3], A. Holodny[4], and D. Kraus[5]

[1]Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, New York, United States, [2]Teacher's College, Columbia University, [3]Surgery, Memorial Sloan-Kettering Cancer Center, [4]Radiology, Memorial Sloan-Kettering Cancer Center, [5]Surgery, Memorial Sloan-Kettering Cancer Center, New York, United States

Introduction

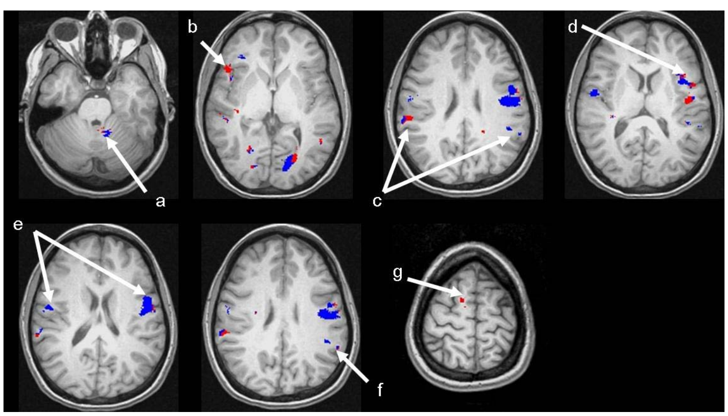
Pitch or frequency adjustments are necessary to meet the demands of conversation/speaking or singing. Moreover, intricate coordination of the subsystems of respiration and phonation during voice production or phonation is required in order to produce voice at a comfortable pitch as well as while regulating voice at different frequencies. The neural correlates of laryngeal movement for voice production at comfortable as well as other pitch levels remain poorly characterized. This is due, in part, to difficulties controlling for experimental confounds between the sensorimotor and cognitive-linguistic aspects of speech phonation/voicing. Thus, investigations aiming to isolate the neural mechanisms of voice production, and in particular, pitch modulation, can be a particular challenge. The current study seeks to describe the central mechanisms responsible for voice production in healthy subjects using functional magnetic resonance imaging (fMRI).

Subjects and Functional Tasks

12 right-handed healthy controls, ranging in age from 20 to 30 years, without a previous history of neurological illness or voice disorder were recruited. Three fMRI paradigms consisting of voice production tasks were used. Specifically, the subject generated the “uh” vowel sound for a maximum duration of 4 sec in an active state (repetition=10) with minimal mouth movement, and the task followed a rest state of 16 sec. The “uh” voice production task was performed using three different intonations consisting of high, comfortable, and low pitches. E-Prime was used to control and present experimental stimuli and to coordinate timing of stimulus presentation with scanner operations. The subject was instructed to perform the tasks, which were displayed on screens that were visible with specially made glasses. In addition, all subjects underwent a standardized voice evaluation serving as a secondary endpoint. The evaluation included a battery of acoustic and aerodynamic tests as part of a standard voice assessment protocol.

Method and Data Analysis

26 T1-weighted axial slices were acquired for the anatomical reference. Functional images were acquired with a 3T GE scanner and a gradient echo EPI sequence (TR=2000 ms; TE=30 ms; 128×128 matrix; 240 mm FOV; 4.5 mm in thickness). 3D T1-weighted images were acquired with a spoiled gradient-recalled acquisition in a steady state sequence. Image analysis was performed with AFNI (1). The reconstructed fMRI data was aligned using a 3D rigid-body registration method. Spatial smoothing, using a gaussian filter of 4mm, and temporal filtering were applied. A deconvolution analysis was applied to the signal time series to generate statistical maps and to derive the hemodynamic impulse response function on a voxel-wise basis. The inherent difference in the time scales of BOLD and voice induced signal changes has been used to minimize voice-correlated motion artifacts by discarding the first few images after voice production (2). Functional and structural images were normalized to Talairach space. For group analysis, the area under the curve was calculated based on each individual subject’s data using the IRF obtained from the deconvolution analysis. Group statistical maps were produced as t scores of relative signal change between active and baseline tasks compared to a null hypothesis of no change with the use of a t-test. Functional activation maps for the t statistic were averaged across all subjects. In order to address the multiple comparison correction, AlphaSim (in AFNI), based on Monte-Carlo simulations was used. Statistical significance was thresholded at $p<0.005$ ($t>5.0$) and a minimum cluster size of 100µl. The volume of activation within the activated areas was measured. A conjunction analysis was applied to localize common activated areas during the pitch tasks. A contrast analysis using a paired t-test was applied to image high versus comfortable pitch and low versus comfortable pitch.



Results

All aerodynamic and acoustic measurements were found to be within normal limits at the time of the assessment. All subjects were able to produce voice at comfortable, high, and low pitches at the time of the assessment. *Conjunction Analysis (figure)*: Areas of activation common to all tasks includes the following: (a) cerebellum, (b) inferior frontal gyrus, (c) inferior parietal lobe, (d) insula, e) precentral gyrus, (f) supramaginal gyrus, and (g) medial frontal gyrus (blue represents combined comfortable, high and low pitch; red represents combined high and low pitch only). In addition, superior temporal gyrus, postcentral gyrus, and posterior cingulate gyrus were commonly activated in all three tasks. Thalamus, basal ganglia and medial frontal gyrus activities were shown only with the low pitch task. *Contrast Analysis:* Two contrasts (high-comfortable pitch, H-C; low-comfortable pitch, L-C) were obtained. Areas obtained from the analysis include the following: H-C: left cerebellum, left cingulate gyrus, left STG, medial FG. L-C: left inferior frontal gyrus (IFG), Superior parietal lobe (SPL), medial frontal gyrus, cingulate gyrus.

Discussion

Activated regions common to all tasks reveal a large, connected network for motor preparation and execution as well as sensory feedback for voice production. Interestingly, Inferior frontal gyrus activity was dominant in the right hemisphere for all three tasks. Contrast analyses revealed increased activity in cerebellum and SPL for high and low pitch productions, respectively, pointing towards the need to recruit centers reported to be involved in controlled coordination and association of sensory information (3). Additionally, further contrast analyses revealed activation in the cingulate gyrus during high and low pitches production. This area, reported to be involved in emotion formation, indicates the need for involvement of the limbic system to produce vocal utterances with emotional overtones.

References

1 Cox R, *Comput Biomed Res.* (1996) 29:162-173; 2 Barch DM *et al.*, *NeuroImage* (1999) 10:642-657; 3 Janata P et al., *Science* (2002) v298, 2167-2170