Neuroscience Letters 484 (2010) 51–55

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Task-related differences in temporo-parietal cortical activation during human phonatory behaviors

Richard D. Andreatta [a,*], Joseph C. Stemple [a], Ashwini Joshi [a], Yang Jiang [b]

[a] Department of Rehabilitation Sciences, College of Health Sciences, University of Kentucky, Lexington, KY, USA
[b] Department of Behavioral Sciences, College of Medicine, University of Kentucky, Lexington, KY, USA

ARTICLE INFO

Article history:
Received 19 April 2010
Received in revised form 15 July 2010
Accepted 6 August 2010

Keywords:
Acoustic
Somatosensory
Laryngeal
Speech
Neuroimaging

ABSTRACT

Functional magnetic resonance imaging (fMRI) was used to investigate cortical activity related to differential control of the human phonatory system during a sentence production task. Our focus in this report was on activation of the temporo-parietal (TP) junctional region, suggested by recent models in speech production/perception to play a critical role between self-generated acoustic and associated somatosensory inputs related to the consequences of speech. Thirteen healthy participants produced multiple trials of phonetically balanced sentences during each of three performance conditions: “covert”, “whisper” and “voice”. An event-related, sparse sampling fMRI method was used to examine TP activity associated with sentence production during each condition. Results demonstrated differential responsiveness in the TP region bilaterally as a function of task conditions, with covert production generating the highest level of TP activation. These results suggest that the TP region is differentially responsive to phonation-specific production variables. Our finding that covert production instead of overt voicing resulted in the greatest activity in TP is consistent with recent reports demonstrating TP activation related to temporal ordering judgments and task-dependent memory use.

Considering speech production involves respiratory, phonatory, resonance and articulatory subsystems working in dynamic coordination with audition, the role and contribution of each subsystem to the overall neural control of speech has yet to be disambiguated. The coordinated subsystems of speech production are in fact, experimentally divisible and amenable to individual analyses with limited confounding influence from remaining production components [26]. For those with interest in the neural substrate of human phonatory control, the ability to isolate the contribution of the larynx during speech may allow for a better understanding of the contribution of phonation-related neural activity during speech.

The laryngeal system has been described as a functional microcosm of the vocal tract [21] whose actions require substantial sensorimotor control and afferent monitoring to achieve the precise changes in pitch and intensity necessary during functional phonatory behaviors [3,22]. Low-threshold mechanoreceptive endings encode the dynamics of vocal fold behavior [1], providing critical movement-related somatosensory feedback that is in temporal register with acoustic feedback generated during phonation [2,28,33]. In turn, the laryngeal motor control system is responsive to this barrage of sensory feedback during phonation.

Current theoretical models of speech production/perception (see [16,17]) suggest a critical role for comparative and integrative neural operations of sensorimotor and acoustic inputs during speech [5,16,17]. One cortical region suggested to play such a role is the zone in and around the temporo-parietal (TP) junction, a multimodal association area in the inferior parietal cortex, extending into the adjacent superior temporal gyrus. Geschwind [15] and Damasio and Damasio [11] implicated the involvement of the temporo-parietal region during speech as early as 1965 and 1980, respectively. More recently, Caplan et al. [6] found evidence among stroke patients for involvement of the posterior supramarginal gyrus and parietal operculum in acoustic–phonetic processing. Celsis et al. [7] identified the involvement of this same region in healthy volunteers asked to detect changes in phonological units during auditory tasks. The left posterior Sylvian fissure on the boundary of the parietal and temporal lobe has also shown evidence of auditory–motor responsivity and integration [19]. Hence, the TP region may form an important point of convergence for auditory and somatosensory inputs during speech production.

Given that the laryngeal system is the dominant source of sound during speech and that laryngeal control requires exquisite sensorimotor regulation, we consider the TP region to be a strong candidate to assess the effects of production conditions that manipulate acoustic output and sensorimotor regulation. We chose to

* Corresponding author at: Department of Rehabilitation Sciences, College of Health Sciences, University of Kentucky, 900 S. Limestone, Wethington 120-F, Lexington, KY 40536-0200, USA. Tel.: +1 859 218 0523.
E-mail address: richard.andreatta@uky.edu (R.D. Andreatta).

0304-3940/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2010.08.017

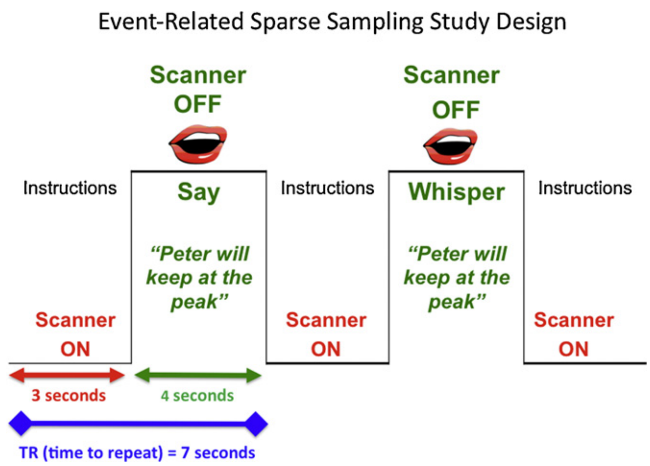

Fig. 1. Sparse sampling study paradigm.

focus our investigation on the TP region specifically because of its theoretical relevance to the integration of both acoustic and somatosensory inputs during speech [16,17]. The purpose of this report was to document changes in TP activity related to differential voluntary control of the phonatory system during a sentence production task under functional magnetic resonance imaging (fMRI). Manipulation of laryngeal-related acoustic and sensorimotor activity was achieved by requiring participants to overtly speak, whisper and covertly read a series of sentences during fMRI. It was hypothesized that TP activity would be sensitive to adjustments in the levels of laryngeal-related acoustic and sensorimotor activity during and across each of the production conditions for our sentence task.

Thirteen healthy, right-handed, native English speakers (4 male, 9 female), 22–57 years participated in the study. Participants had no history of neurological conditions or hearing loss. Participants underwent a videostroboscopic examination of the vocal folds prior to the fMRI procedure to ensure normal laryngeal structure and function. This study was approved by the Institutional Review Board at the University of Kentucky. All participants signed a written informed consent form prior to participation.

Participants performed multiple trials of six phonetically balanced sentences from the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) [20] during each of three different performance conditions: "covert", "whisper" and "voice". Subjects were pre-trained on condition features and the meaning of instructions that would be visually provided in the scanner. Conditions were defined as: (1) "voice"—reading the sentence aloud at a normal and comfortable pitch and loudness, (2) "whisper"—reading the sentence aloud, yet with no voicing, and (3) "covert"—participants were instructed to read the sentences silently "in their heads" (as if reading aloud), without any speech-related movement or voicing. We confirmed that each participant understood our instructions by having each practice the three conditions outside of the scanner.

An event-related, sparse sampling fMRI design was used in this study to allow for the production of task conditions in the absence of gradient noise (Fig. 1) [31,32]. Participants received visual instructions for each trial projected onto a mirror on the head coil. A screen providing the condition cue for the upcoming trial was presented for 3 s. The next screen provided the target sentence and the participant was required to produce the sentence in either a covert, voiced or whispered manner. The sentence task occurred during the silent off-period of the sparse sampling sequence, a provision that reduced the potential for motion-related MR artifact [4]. Each trial period (scanner off) was jittered 3.5–4.5 s to better capture the hemodynamic response peak. Total subject time within the scanner was 35–40 min.

With the absence of gradient noise during task production, participants were capable of self-monitoring their sentence productions during the whisper and overt voicing conditions without the need for headphones. Participants were free to self-correct their speech. Participant's compliance with condition parameters was externally monitored by randomly sampling productions over the audio–visual interface within the scanner suite. Qualitatively, participants were compliant with all performance parameters.

Task conditions were designed to produce concurrent and graded sensorimotor and acoustic changes during sentence production. For the "voice" condition, phonation and articulatory acoustics were typical and contained all phonetic acoustic cues for all sound classes. For the "whisper" condition, the acoustic output was substantially degraded (no phonation, loss of voiced-consonant feature, added airstream turbulence). Lastly, the "covert" condition produced no acoustic output of any kind (laryngeal and/or supralaryngeal in origin). Overall, the production conditions can be viewed as existing on a continuum of laryngeal engagement, with the "voice" state at one end (most engaged) and the "covert" state at the other (least engaged). Acoustic features for the sentences produced under each condition were considered unique for each condition allowing for interpretation of TP activity related to both global acoustic features and peripheral sensorimotor performance. Commonalities across conditions related to language were subtracted using a comparative analysis between conditions, allowing us to identify any difference in cortical activity as a function of task demand.

Thirty trials for each condition of "covert", "whisper", and "voice" and 60 rest trials were distributed over three functional runs for each subject. Sentences were placed in an ABBA order, while the task conditions (including rest) were pseudo-randomized using a sequence provided by the Analysis of Functional Neuroimaging (AFNI) software [9]. All participants received the same order of trials to maintain task presentation consistency.

Movement artifact was minimized during signal acquisition by stabilizing the skull with memory-foam against the head coil, and using an event-related, sparse sampling approach. Scanner gradients were turned off during speech production and turned on for 3 s afterward during the visual instruction phase for the upcoming trial. Whole brain volumetric scans of blood oxygenation level dependent (BOLD) activity for the previous sentence production was collected during the scanner-on period. Head movement was not measured because of the methodology's strength in controlling for motion-related artifact. Functional data were T2* weighted echo-planar images. A single echo-planar imaging (EPI) volume was acquired with a TR = 7.0 s. A high-resolution, 3D anatomic image was acquired using a sagittal T1 weighted (MP-RAGE) sequence (TR = 2100 ms, TE = 2.93 ms, TI = 1100 ms, flip angle = 12°, FOV = 192 mm × 224 mm × 256 mm, with 1 mm isotropic voxels). The following parameters were applied during volume acquisition: TR = 2.5 s; TP = 156; TE = 30 ms; flip angle = 81°; 39 axial slices; 224 mm × 224 mm FOV (field of view); slice thickness = 3.5 mm; 64 × 64 matrix (yielding 3.5 mm × 3.5 mm × 3.5 mm voxels); bandwidth = 2056 Hz/Px.

Image processing and analyses was conducted using AFNI [10]. After pre-processing, structural 3D data were transformed into Talairach space using AFNI [10]. Following exclusion of the first few functional volumes due to T1 saturation effects, timing differences between slices due to acquisition order were adjusted with sync interpolation. fMRI data were motion adjusted to the image collected nearest in time to the structural image, and smoothed (4 mm FWHM). Each voxel was normalized within each functional run to yield activation measures expressed as "percent change" from baseline. For each subject, the voxel time series for each trial type ("covert", "whisper" and "voice") were estimated using AFNI. The general linear model for event-related fMRI was used to esti-