ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 36 (2007) 131–143

Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions

Torrey M.J. Loucks, Christopher J. Poletto, Kristina Simonyan, Catherine L. Reynolds, and Christy L. Ludlow*

Laryngeal and Speech Section, Medical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

Received 23 August 2006; revised 22 December 2006; accepted 19 January 2007
Available online 12 March 2007

Phonation is defined as a laryngeal motor behavior used for speech production, which involves a highly specialized coordination of laryngeal and respiratory neuromuscular control. During speech, brief periods of vocal fold vibration for vowels are interspersed by voiced and unvoiced consonants, glottal stops and glottal fricatives (/h/). It remains unknown whether laryngeal/respiratory coordination of phonation for speech relies on separate neural systems from respiratory control or whether a common system controls both behaviors. To identify the central control system for human phonation, we used event-related fMRI to contrast brain activity during phonation with activity during prolonged exhalation in healthy adults. Both whole-brain analyses and region of interest comparisons were conducted. Production of syllables containing glottal stops and vowels was accompanied by activity in left sensorimotor, bilateral temporoparietal and medial motor areas. Prolonged exhalation similarly involved activity in left sensorimotor and temporoparietal areas but not medial motor areas. Significant differences between phonation and exhalation were found primarily in the bilateral auditory cortices with whole-brain analysis. The ROI analysis similarly indicated task differences in the auditory cortex with differences also detected in the inferolateral motor cortex and dentate nucleus of the cerebellum. A second experiment confirmed that activity in the auditory cortex only occurred during phonation for speech and did not depend upon sound production. Overall, a similar central neural system was identified for both speech phonation and voluntary exhalation that primarily differed in auditory monitoring.

Keywords: Phonation; Exhalation; Speech; fMRI

Phonation for human speech is a highly specialized manner of laryngeal sensorimotor control that extends beyond generating the sound source of speech vowels and semi-vowels (e.g. /y/ and /r/). Laryngeal gestures for voice onset and offset distinguish voiced and unvoiced consonant pairs through precise timing of voice onset (Raphael et al., 2006). Brief interruptions in phonation by glottal stops (/ʔi/) mark word and syllable boundaries through hyper-adduction of the vocal folds to offset phonation. Each of these vocal fold gestures, in addition to rapid pitch changes for intonation, requires precisely controlled variations in intrinsic laryngeal activity (Poletto et al., 2004). To understand the neural control of phonation for speech, laryngeal control must be investigated independent of the oral articulators (lips, tongue, jaw). Several neuroimaging studies of phonation for speech included oral articulatory movements with phonation, which can limit the examination of laryngeal control independently from oral articulatory function (Huang et al., 2001; Schulz et al., 2005). We expect that the neural control over repetitive laryngeal gestures for speech phonation will show a greater hemodynamic response (HDR) response in the left relative to right inferolateral sensorimotor areas (Wildgruber et al., 1996; Riecker et al., 2000; Jeffries et al., 2003). On the other hand, the neural control for vocalizations that are not specific to speech, such as whimper or prolonged vocalization, will show a more bilateral distribution (Perry et al., 1999; Ozdemir et al., 2006).

Phonation for speech also involves volitional control of respiration as subglottal pressure throughout exhalation must be controlled to initiate and maintain vocal fold vibration (Davis et al., 1996; Finnegan et al., 2000). To better understand the central neural control of laryngeal gestures for speech, the contribution of volitional control over respiration must also be examined. Previous neuroimaging findings of volitional respiratory control, however, have been inconsistent. In one positron emission tomography (PET) study (Ramsay et al., 1993), brain activity for volitional exhalation involved similar cortical and subcortical regions as those reported for voiced speech (Riecker et al., 2000; Turkeltaub et al., 2002; Bohland and Guenther, 2006). During exhalation, activity increased in the left inferolateral frontal gyrus (IFG) where the laryngeal motor cortex is thought to be located (Ramsay et al., 1993). On the other hand, functional MRI (fMRI) studies of volitional exhalation found increases in more dorsolateral sensorimotor regions, potentially corresponding to diaphragm and chest wall control, but not in the laryngeal motor regions (Evans et al.,

* Corresponding author.
E-mail address: LudlowC@ninds.nih.gov (C.L. Ludlow).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.01.049

1999; McKay et al., 2003). The neural correlates for voluntary exhalation need to be distinguished to understand the neural control of phonation for speech.

The two aims of the current study were to identify the central organization of phonation for speech and to contrast phonation for speech with volitional exhalation using event-related fMRI. We hypothesized that brain activity for glottal stops plus vowels produced as syllables involves a left predominant sensorimotor system as suggested by the results of Riecker et al. (2000) and Jeffries et al. (2003) for simple speech tasks. We also expected that this left predominant system for laryngeal syllable production would differ from upper airway control for prolonged exhalation, which may involve a more bilateral and more dorsolateral sensorimotor system.

Phonation and exhalation differ in auditory feedback because phonation produces an overt sound heard by the speaker while exhalation is quiet. To test whether auditory feedback affects the HDR during phonation for speech, we manipulated auditory feedback during phonation and whimper (a non-speech task with similar levels of auditory feedback). If no differences were found between phonation for speech and whimper then it would be the auditory feedback which could explain a greater auditory response during phonation than during exhalation. On the other hand, if there were greater responses in the auditory area during phonation when compared to whimper, it might be because the participants attend to a greater degree to their own voice during a phonatory task and than during the non-speech task of whimper. Once masking noise was applied we expected that phonation and whimper should not differ in their auditory responses because the participants could not hear their own productions. Therefore, we predicted that there would be differences in responses in auditory areas between the phonated and whimper productions in the normal feedback condition but not during the masked condition.

Materials and methods

Experiment 1: identifying and contrasting of the neural control of phonation for speech and voluntary exhalation using event-related fMRI

Subjects

Twelve adults between 23 and 69 years participated in the study (mean—35 years, s.d.—12 years, 7 females). All were right-handed, native English speakers. None had a history of neurological, respiratory, speech, hearing or voice disorders. Each had normal laryngeal structure and function on laryngeal nasoendoscopy by a board certified otolaryngologist. All provided written informed consent before participating in the study, which was approved by the Internal Review Board of the National Institute of Neurological Disorders and Stroke.

Phonatory and respiration tasks

The two phonation tasks involved: (1) repetitions of the syllable beginning with a glottal stop requiring vocal fold hyper-adduction /ʔi/ and followed by a vowel requiring precise partial opening of the vocal folds to allow vibration /i/ (similar to the "ee" in "see). This syllable was produced reiteratively (/ʔiʔiʔiʔi/) at a rate similar to normal speech production (between 3 and 5 repetitions per second); and (2) the onset and offset of vocal fold vibration for the same vowel /i/ which was prolonged for up to 4.5 s. The phonation tasks were chosen to minimize the linguistic and oral requirements of the task and focus on voice onset and offset using laryngeal gestures involved in speech, i.e., glottal stops and vowels.

The exhalation task involved voluntarily prolonged oral exhalation produced in quiet after a rapid inhalation. The subjects were trained to prolong the exhalation for the same duration as the phonatory tasks but to maintain quiet exhalation not producing a sigh or a whisper. The exhalation productions were monitored throughout the study to assure that participants did not produce a sigh or whisper. A quiet rest condition was used as the control condition.

On each phonation or exhalation trial, the subject first heard an auditory example of the target for 1.5 s presented at a comfortable loudness level through MRI compatible headphones (Avotec, Stuart, FL, USA). The exhalation example was amplified to the same intensity as the phonation example to serve as a clearly audible task cue, although subjects were trained to produce quiet exhalation. The acoustic presentation was followed at 3.0 s by the onset of a visual cue (green arrow) to produce the target. The subjects were trained to begin phonating or exhaling after the arrow onset and to stop immediately following offset of the arrow 4.5 s later or 7.5 s into the trial. A single whole-brain volumetric scan was then acquired over the next 3 s (0.3 s silent period +2.7 s TR) resulting in a total trial duration of 10.5 s (Fig. 1). The subjects were not provided with any instructions for the quiet rest condition, except that on certain trials there would not be an acoustic cue, although the visual arrow was presented. A fixation cue in the form of a cross was presented throughout the experiment except when the arrow was shown. Phonation was recorded through a tubing-microphone system, placed on the subject's chest approximately 6 in. from the mouth (Avotec, Stuart, FL, USA). An MRI compatible bellows system placed around the subject's abdomen recorded changes in respiration.

Each experiment consisted of 45 repetitive syllable trials, 45 continuous phonation trials, 90 exhalation trials and 120 rest trials distributed evenly over 5 functional scanning runs giving 60 trials per run. Two initial scans at the beginning of each run, to reach homogenous magnetization, were not included in the analyses. The stimuli were randomized so each subject received a different task order. Prior to the fMRI scanning session, each subject was trained to produce continuous and repeated syllable phonations and prolonged exhalation for 4.5 s.

Functional image acquisition

To minimize head movements during syllable that could produce artifacts in the blood oxygenation level dependent (BOLD) signal, a vacuum pillow cradled the subject's head. Additionally, an elastic headband placed over the forehead provided tactile feedback and a slight resistance to shifts in head position. A sparse sampling approach was used; the scanner gradients were turned off during stimulus presentation and production so the task could be performed with minimal background noise (Birn et al., 1999; Hall et al., 1999). A single echo-planar imaging volume (EPI) of the whole brain was acquired starting 4.8 s after the onset of the production phase with the following parameters: TR—2.7 s, TE—30 ms, flip angle—90°, FOV—240 mm, 23 sagittal slices, slice thickness—6 mm (no gap). Because the subjects took between 0.5 and 1.0 s to initiate the task following the visual cue, the acquisition from 4.0 to 6.7 s occurred during the predicted peak of the HDR for speech based on previous studies (Birn et al., 1999, 2004; Huang et al., 2001; Langers et al.,