ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 32 (2006) 376 – 387

Clustered functional MRI of overt speech production

Peter Sörös,[a,b,*] Lisa Guttman Sokoloff,[c,d] Arpita Bose,[e] Anthony R. McIntosh,[f,g] Simon J. Graham,[a,f,h] and Donald T. Stuss[f,g,i]

[a]Imaging Research, Sunnybrook and Women's College Health Sciences Centre, Toronto, Ontario, Canada
[b]Toronto Rehabilitation Institute, Toronto, Ontario, Canada
[c]Communication Disorders, Baycrest, Toronto, Ontario, Canada
[d]Department of Speech–Language Pathology, University of Toronto, Ontario, Canada
[e]Department of Psychology, University of Windsor, Ontario, Canada
[f]The Rotman Research Institute, Baycrest, Toronto, Ontario, Canada
[g]Department of Psychology, University of Toronto, Ontario, Canada
[h]Department of Medical Biophysics, University of Toronto, Ontario, Canada
[i]Department of Medicine (Neurology, Rehabilitation Science), University of Toronto, Ontario, Canada

Received 27 October 2005; revised 21 February 2006; accepted 24 February 2006
Available online 2 May 2006

To investigate the neural network of overt speech production, event-related fMRI was performed in 9 young healthy adult volunteers. A clustered image acquisition technique was chosen to minimize speech-related movement artifacts. Functional images were acquired during the production of oral movements and of speech of increasing complexity (isolated vowel as well as monosyllabic and trisyllabic utterances). This imaging technique and behavioral task enabled depiction of the articulo-phonologic network of speech production from the supplementary motor area at the cranial end to the red nucleus at the caudal end. Speaking a single vowel and performing simple oral movements involved very similar activation of the cortical and subcortical motor systems. More complex, polysyllabic utterances were associated with additional activation in the bilateral cerebellum, reflecting increased demand on speech motor control, and additional activation in the bilateral temporal cortex, reflecting the stronger involvement of phonologic processing.

Keywords: Speech production; Articulation; Phonologic processing; Cerebellum; Superior temporal gyrus; fMRI; Clustered image acquisition

Introduction

The production of speech is a highly complex motor task that involves approximately 100 orofacial, laryngeal, pharyngeal, and respiratory muscles (Levelt, 1989). Precise and expeditious timing of these muscles is essential for the production of temporally complex speech sounds, which are characterized by transitions as short as 10 ms between frequency bands (Fitch et al., 1997) and an average speaking rate of approximately 15 sounds per second (Levelt, 1989). The neural basis of the exact and rapid coordination of these highly overlearned movements is not yet entirely clear (Munhall, 2001).

For years, the analysis of brain lesions and the correlation between lesion locations and behavioral deficits were the most important sources of knowledge about the mechanisms underlying speech production (Huber et al., 2000; Rorden and Karnath, 2004). The seminal lesion studies of Paul Broca indicated that the production of speech relies on the functional integrity of the left inferior frontal gyrus (Broca, 1861). The investigation of patients with stroke-related apraxia of speech (AOS) added further insights. AOS is conceptualized as a deficit of transforming phonologic plans and articulatory motor programs to appropriate speech movements (Darley et al., 1975; Croot, 2002). Recent research suggests that patients with AOS fail to retrieve the motor patterns essential for speech production (Aichert and Ziegler, 2004). MRI studies of these patients revealed lesions of and around the left inferior frontal gyrus, in particular Brocas area (Hillis et al., 2004), lesions of the left insula (Dronkers, 1996; Nagao et al., 1999), and of the basal ganglia (Peach and Tonkovich, 2004). Nonfluent progressive aphasia, presenting with apraxia of speech and syntactic deficits, is similarly associated with left inferior frontal and insular atrophy (Gorno-Tempini et al., 2004). However, the results of noninvasive neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), provide growing evidence that complex human skills are not located in highly specialized brain areas but are organized in networks connecting several different areas of both hemispheres instead (Sporns et al., 2004). Thus, a widespread network is most likely to underlie the

* Corresponding author. Imaging Research, Sunnybrook and Women's College Health Sciences Centre, 2075, Bayview Avenue Room SG28, Toronto, Ontario, Canada M4N 3M5.
E-mail address: peter.soros@swri.ca (P. Sörös).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.02.046

production of speech (Hickok, 2001), rather than isolated speech centers.

Characterization of speech production by fMRI has been complicated by motion-correlated head movements and by movements of the articulatory organs (Birn et al., 1999). Both motion inside the field of view (head movement) and motion outside the field of view (movement of the oral cavity, the sinuses, or the pharynx) (Yetkin et al., 1996) might cause magnetic field inhomogeneities masking brain activation or generating artifactual intensity changes. Previous studies have used inner speech (e.g., silent repetition of words or syllables) to overcome these speech-related artifacts (Wildgruber et al., 1996, 2001). Behavioral tasks involving inner speech usually do well minimizing task-related motion but face other disadvantages (Munhall, 2001). First, it is very difficult to monitor behavioral performance using covert speech production. This aspect would be less important in the present study of neurologically healthy adults but might be a major issue in experiments involving stroke patients. Second, speakers cannot hear their own voice while generating silent responses. Hearing ones own speech, however, is important for accurate speech motor control (Jones and Munhall, 2000). Third, different activation magnitudes have been observed in the cortical and subcortical portions of the speech motor system when comparing silent and overt word stem completion (Palmer et al., 2001; Rosen et al., 2000) or silent and overt production of monosyllabic or multisyllabic words (Shuster and Lemieux, 2005). These observations emphasize the importance of tasks involving overt utterances and of imaging techniques which are less susceptible to movement artifacts for the study of speech production.

For this study, we investigated overt nonlexical utterances using event-related fMRI with clustered image acquisition. A neuropsychological model of speech production includes at least two major cognitive processes, the assembling and the execution of a motor plan (Levelt, 1989). To separate these processes, subjects were asked to repeat acoustically presented sublexical speech sounds of different complexity and to perform nonverbal oral movements. Most previous studies investigated the production of lexical utterances. To minimize semantic and syntactic processing, sublexical speech was chosen for the present study. In addition, recently developed fMRI methodology, termed clustered volume (or image) acquisition (Edmister et al., 1999; Fu et al., 2002; Liebenthal et al., 2003; Ojanen et al., 2005; Rimol et al., 2005), compressed image acquisition (Abrahams et al., 2003) or sparse temporal sampling (Tanaka et al., 2000), enables improved investigation of movement- and speech-related brain activation. The principle underlying these techniques is that the entire brain volume is scanned in a fraction of the repetition time (TR), leaving an extended silent interval for auditory stimulation and speech production (Gracco et al., 2005). This technique is made possible by the difference between the rapid movements associated with speech production (Szirtes and Vaughan, 1977) and the comparatively slow rise of the hemodynamic response curve (Birn et al., 2004). Furthermore, with judicious timing of behavioral tasks, it is possible to separate the hemodynamic response associated with the auditory component of speech from the response associated with acoustic noise arising from the process of fMRI signal acquisition (Rimol et al., 2005).

This study has three goals: (i) to trace the distributed neural network of overt speaking; (ii) to characterize brain activation specific for speaking as compared to simple oral movements; and (iii) to characterize brain activation specific for speaking of polysyllabic sequences as compared to the production of an isolated vowel. It is hypothesized that speaking a less complex speech sound such as a single vowel activates a distributed motor network, similar to performing simple oral movements (Dresel et al., 2005). In addition, it is hypothesized that speech sounds of increasing complexity (monosyllabic consonant–vowel and trisyllabic consonant–vowel utterances) are associated with an increasing task demand and with the increased recruitment of additional brain regions, such as the left inferior frontal gyrus and the left anterior insula (Wise et al., 1999; Blank et al., 2002).

Methods

Participants

Blood oxygenation-level-dependent (BOLD) fMRI was acquired in 9 healthy volunteers (4 women, 5 men) with an average age of 26 years (range, 22–32). All participants were right-handed and, except one, native speakers of English. One volunteer's first language was German. This volunteer has lived in an English-speaking country for several years and used English as her primary language. The target speech sounds for the present study are common in both English and German. Volunteers were recruited with the help of the Rotman Research Institute volunteer database and by personal communication. The study was approved by Research Ethics Boards at Baycrest and at Sunnybrook and Women's College Health Sciences Centre, Toronto, Ontario, Canada. Informed consent for participation in the project was obtained from all subjects according to the Declaration of Helsinki.

Experimental tasks

Subjects were asked to repeat acoustically presented sublexical speech sounds of different complexity and to perform oral movements without vocalization. The required responses were the vowel (V) "ah", a consonant–vowel (CV) syllable (either "pa", "ka", or "ta"), a $C_1VC_2VC_3V$ utterance ("pataka"), and oral movements (opening the mouth or protruding the lips). Instructions were "say ah" (for the vowel condition), "say pa", "say ka", or "say ta" (for the CV condition), "say pataka" (for the $C_1VC_2VC_3$ V condition), "open your mouth", and "make a kiss" (for oral movements). Verbal instructions were transmitted through an fMRI compatible audio system with acoustically padded headphones to reduce acoustic fMRI noise by 25 dB (Silent Scan; Avotec, Stuart, FL, USA). All instructions were spoken by a speech–language pathologist in a sound-attenuated room, digitized at 22050 Hz and stored as a digital sound file. To avoid confusion with the English article "a", the long vowel "ah" was presented. Instructions were delivered at a constant onset-to-onset interstimulus interval of 10 s with the stimulation software Eprime 1.1 (Psychology Software Tools, Pittsburgh, PA, USA). Subjects were asked to perform the given task or to produce the required response immediately after the end of the instruction. Six experimental sessions were performed. Each session comprised 6 separate blocks of speech, 2 blocks of oral movement (50 s each), and 3 blocks of baseline (30 s). During the baseline, no verbal instructions were given, and no responses were performed. To minimize task-switching effects, a blocked