NeuroImage 47 (2009) 314–325

Contents lists available at ScienceDirect

NeuroImage

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

Common neural substrates support speech and non-speech vocal tract gestures

Soo-Eun Chang, Mary Kay Kenney, Torrey M.J. Loucks, Christopher J. Poletto, Christy L. Ludlow *

Laryngeal and Speech Section, Medical Neurology Branch, NINDS/NIH, 10 Center Dr. MSC 1416 Building 10, Room 5D38, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Received 7 November 2008
Revised 23 February 2009
Accepted 11 March 2009
Available online 25 March 2009

Keywords:
Sensory–motor interaction
Auditory dorsal stream
Functional magnetic resonance imaging (fMRI)
Speech production
Speech perception
Non-speech

ABSTRACT

The issue of whether speech is supported by the same neural substrates as non-speech vocal tract gestures has been contentious. In this fMRI study we tested whether producing non-speech vocal tract gestures in humans shares the same functional neuroanatomy as non-sense speech syllables. Production of non-speech vocal tract gestures, devoid of phonological content but similar to speech in that they had familiar acoustic and somatosensory targets, was compared to the production of speech syllables without meaning. Brain activation related to overt production was captured with BOLD fMRI using a sparse sampling design for both conditions. Speech and non-speech were compared using voxel-wise whole brain analyses, and ROI analyses focused on frontal and temporoparietal structures previously reported to support speech production. Results showed substantial activation overlap between speech and non-speech function in regions. Although non-speech gesture production showed greater extent and amplitude of activation in the regions examined, both speech and non-speech showed comparable left laterality in activation for both target perception and production. These findings posit a more general role of the previously proposed "auditory dorsal stream" in the left hemisphere — to support the production of vocal tract gestures that are not limited to speech processing.

Published by Elsevier Inc.

Human speech involves precise, well-coordinated laryngeal and orofacial movements, likely dependent on neural networks encompassing frontal motor and temporoparietal auditory regions (Hickok and Poeppel, 2004). A common auditory dorsal pathway involving motor responses constrained by auditory experience has been proposed (Warren et al., 2005) that links the auditory processing of speech sounds with motor gestures, enabling accurate sound production (Hickok and Poeppel, 2007). Such auditory–motor interactions may support speech development in children, when speech motor gestures are tuned to, or guided by auditory speech targets (Hickok and Poeppel, 2004). The structures involved in the auditory dorsal stream, which is lateralized to the left hemisphere, may not be specialized for human speech but likely support other types of learned volitional vocal productions with auditory targets (Bottjer et al., 2000; Metzner, 1996; Pa and Hickok, 2008; Smotherman, 2007; Zarate and Zatorre, 2005).

Many studies have indicated that cerebral activation for speech perception can be distinguished from that for non-speech perception, particularly in the superior temporal regions (Benson et al., 2001; Binder et al., 2000; Liebenthal et al., 2005; Scott et al., 2000; Whalen et al., 2006). In some cases, the speech stimuli contained lexical–semantic information involving higher level language processing, greater in the left hemisphere. On the other hand the non-speech stimuli often did not involve vocal tract gestures and were either non-vocal simple tones, non-producible synthetic sounds or sounds from nature (Benson et al., 2001, 2006; Binder et al., 2000) rather than non-speech vocal tract gestures such as sigh, tongue click, and cry. In those instances, differences in brain activation found for speech and non-speech sound processing could have been because the non-speech stimuli did not contain oral motor or vocal targets, less likely to engage motor production circuits such as those involved in speech. One study did use vocally produced non-speech sounds and found that speech sounds activated most parts of the temporal lobe on both sides of the brain, while the right superior temporal lobe was activated to a greater degree by non-speech vocal sounds (i.e., sighs, laughs, cries) (Belin et al., 2002). In another study, however, when subjects performed sequence manipulation tasks with speech involving phoneme processing and non-speech involving oral sounds such as humming, comparable activation in the left posterior inferior frontal and superior temporal regions were found for both speech and non-speech (Gelfand and Bookheimer, 2003). Perhaps if non-speech vocal tract gestures involve segment sequencing, resulting in auditory and somatosensory feedback as is in the case of speech, they will activate comparable regions to speech processing.

Clinical lesion and intraoperative studies, as well as functional imaging studies have provided a wealth of data on neural structures supporting speech motor production. Apraxia of speech (AOS), characterized by difficulty in speech motor planning particularly for complex syllables, has been reported to result following damage to the anterior insula in the language-dominant hemisphere (Dronkers, 1996) as well as left-sided infarctions affecting blood supply to the middle cerebral artery, such as the posterior inferior frontal gyrus

* Corresponding author. Fax: +1 301 480 0803.
E-mail address: ludlowc@ninds.nih.gov (C.L. Ludlow).

1053-8119/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.neuroimage.2009.03.032

(Hillis et al., 2004). Speech execution in terms of rate, intonation, articulation, voice volume, quality, and nasality can be adversely affected in various dysarthrias, which can result from injuries to the basal ganglia (Schulz et al., 1999), thalamus (Ackermann et al., 1993; Canter and van Lancker, 1985), cerebellum (Kent et al., 1979), or cerebral cortex (Ozsancak et al., 2000; Ziegler et al., 1993).

Electrical stimulation of the exposed motor strip representation of face/mouth on either hemisphere controls vocalization (Penfield and Roberts, 1959), and stimulation of left inferior dorsolateral frontal structures can lead to speech arrest and inability to repeat articulatory gestures (Ojemann, 1994). Neuroimaging studies of normal speech motor control (Bohland and Guenther, 2006; Riecker et al., 2008; Soros et al., 2006; Wise et al., 1999), using a variety of speech tasks, have roughly converged on a "minimal network for overt speech production", including the "mesiofrontal areas, intrasylvian cortex, pre- and post-central gyrus, extending rostrally into posterior parts of the left inferior frontal convolution, basal ganglia, cerebellum, and thalamus" (Riecker et al., 2008).

One study found an opposite pattern of lateralization in the sensorimotor cortex during speech production and production of tunes (articulation constant; i.e., "la" while singing the melody), with the former eliciting predominantly left sided activity and the latter eliciting activity predominantly on the right (Wildgruber et al., 1996). Similarly in a follow-up study by the same group, opposite laterality effects were found when comparing speech and non-speech (singing) in the insula, motor cortex, and the cerebellum (Riecker et al., 2000). In these studies, however, it is still not clear whether singing or other non-speech gestures would be supported bilaterally or with right hemisphere dominance, different from left-lateralized speech production. This is because non-speech tasks such as singing melodies with a constant vowel or consonant–vowel syllable was not comparable to speech in the amount of sequencing required or the variety of vocal tract and oral gestures required for production.

In this study we sought to test whether volitional production of non-speech vocal tract gestures would be supported by comparable functional neuroanatomy as speech production. Non-speech production involved volitional vocal tract gestures such as whistle, cry, sigh, and cough, which have previously learned auditory targets, and require sensory–motor integration for accurate production as in the case of speech. We hypothesized that speech and non-speech would involve the same regions of activation when compared on whole brain analyses and in brain regions involved in speech. Second, we hypothesized no differences in laterality of activity. Third, we hypothesized that although non-speech targets would activate regions involved in the production of speech sounds, activation levels in these regions would be greater, as volitional production of non-speech oral–motor gesture sequences may be more novel and involve greater effort in producing the oral–motor gestures compared to speech.

Methods

Participants

The participants were 34 healthy adults (17 females) aged 18–57 (mean = 37 years), right handed on the Edinburgh handedness inventory (Oldfield, 1971), native English speakers, and scored within 1 standard deviation of the age-adjusted mean on speech, language, and cognitive testing. All subjects were free of communication, neurological or medical disorders, passed audiometric screening, and had normal structural MRI scans when examined by a radiologist. All subjects signed an informed consent form approved by the Internal

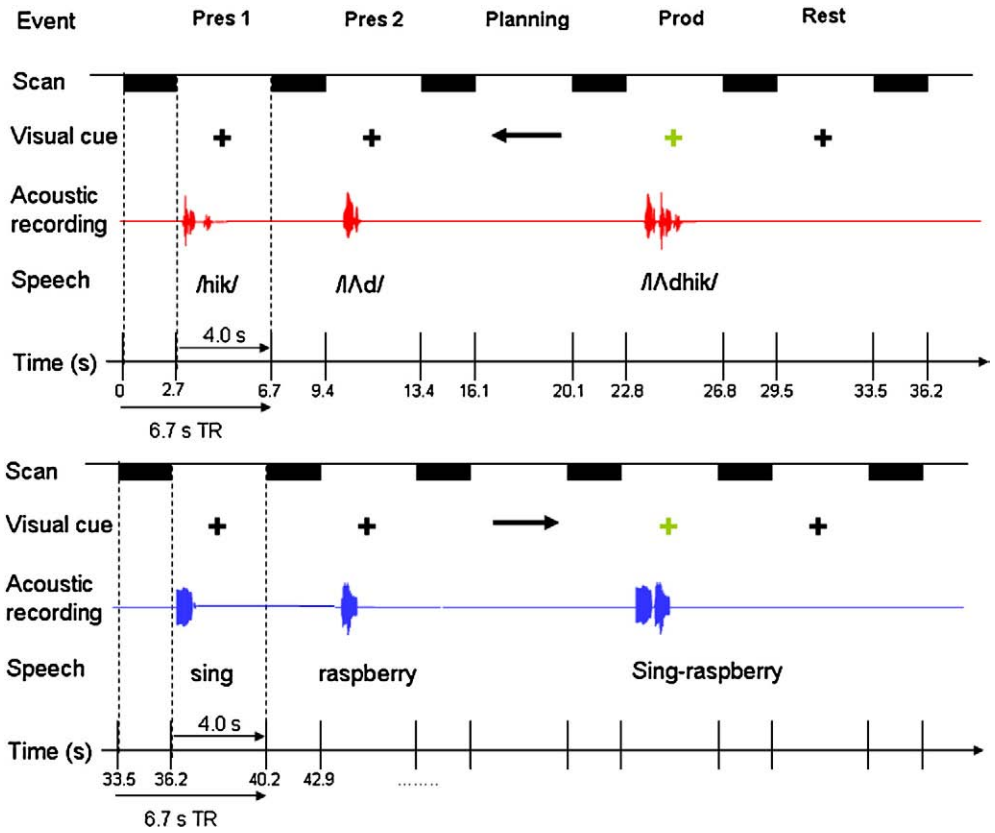

Fig. 1. Experiment outline. Here one speech trial (upper panel) and one non-speech trial (lower panel) are illustrated. Speech and non-speech trials were randomly presented. Each trial consisted of two target presentations (pres 1, pres 2), planning, production (prod), and rest, each presented/performed during a 4 second silent period, which was followed by 2.7 s of scanning. Note that only the first of the two responses associated with target presentation (scan following "pres 1") was used for perception analysis. See text for more detail on the experiment paradigm.

Review Board of the National Institutes of Neurological Disorders and Stroke. All were paid for their participation.

Procedure

Each trial of the experiment started with the presentation of pairs of either speech or non-speech targets, which required repetition (overt production) of the target after a delay period (Fig. 1). All the stimuli were previously recorded, using the same female speaker. Five different target pairs were randomly presented for the speech and non-speech conditions. The speech targets were pairs of meaningless consonant–vowel–consonant syllables /bem/-/dauk/, /hik/-/lʌd /, /saip/-/kuf/, /lok/-/chim/, and /raig/-/sot/, devoid of lexicality but following the rules of English phonology. Because our intention was to contrast brain activation for speech and non-speech vocal tract gestures, it was important to control for lexical/semantic differences. Therefore, we only used speech targets that did not have lexical/semantic reference.

We developed our non-speech gestures to include vocal tract gestures that involved sound targets but were devoid of phonological content. In addition we developed sequences of these gestures so that we had pairs of targeted vocal tract gestures parallel to the nonsense speech syllables. The non-speech targets were pairs of sounds of orofacial and vocal tract gestures: cough–sigh, sing ("/a/" on a tone)–raspberry, kiss–snort, laugh–tongue click, whistle–cry. All non-speech targets could be easily reproduced by each subject, yet involved complex oral motor sequencing, but without phonemic processing typical of speech processing. The non-speech and speech stimuli were similar in duration ($\bar{x}_{\text{speech}} = 820$ ms (SD = 136), $\bar{x}_{\text{non-speech}} = 916$ ms (SD = 142)) and root-mean-square power ($\bar{x}_{\text{speech}} = 0.15$ (SD = 0.04), $\bar{x}_{\text{non-speech}} = 0.12$ (SD = 0.07)), with no statistically significant difference ($p > 0.05$). The speech and non-speech targets did differ, however, in acoustic and motor complexity; speech included more transients and smaller articulatory gestures, and non-speech involved a greater variety of motor gestures, and included more forceful glottal attack (cry, cough) and lip closures (kiss), bilabial bursts (raspberries), tongue thrusts (tongue click), which possibly required more effort than articulating nonsense syllable sequences.

The target presentation phase was followed by a planning phase, when the subjects were visually cued that their upcoming production of the two stimuli should either be in the same order (right arrow), or in the reversed order (left arrow) from the presented pair. Subjects were instructed beforehand not to make any oral movements during this time period. This design separated motor planning from motor production, as the onset of production was signaled by a fixation cross replacing the arrow from the planning phase. The cross served as the "go" signal for subjects to produce the previously planned speech or non-speech response (Fig. 1).

Auditory and visual stimuli were delivered using Eprime software (version 1.2, Psychology Software Tools, Inc.) running on a PC, which synchronized each trial with functional image acquisition. Sound was delivered binaurally through MRI-compatible headphones (Silent Scan™ Audio Systems, Avotec Inc., Stuart, FL). The auditory stimuli were set at a comfortable volume level for each subject before the experiment and remained constant throughout the experimental runs. Subjects' productions were monitored and recorded using an MR-compatible microphone attached to the headphones (Silent Scan™ Audio Systems, Avotec Inc., Stuart, FL).

All subjects underwent a training session on the day of the experiment to familiarize them with the stimuli and tasks. Subjects were able to produce both speech and non-speech stimuli without difficulty. Ten speech and ten non-speech trials were randomly presented in each run, and a total of three runs were completed for each subject, resulting in 60 target presentation (only the first of the two presentation trials were taken for analysis), and 60 production responses; both containing 30 speech and 30 non-speech stimuli.

Image acquisition

All images were obtained from a 3.0 T GE Signa scanner equipped with a standard head coil. Subjects' head movements were minimized using padding and cushioning of the head inside the head coil. Gradient echo-planar pulse sequence was used for functional image acquisition (TE = 30 ms, TR = 6.7 s, FOV = 240 mm, 6 mm slice thickness, 23 contiguous sagittal slices). By using an event-related, sparse sampling design (Birn et al., 1999; Eden et al., 1999; Hall et al., 1999) the presentation of auditory stimuli, and the planning and production phases took place while the scanner was transiently silent before scanning 4 s later. Sparse sampling minimized scanner noise and movement related susceptibility artifacts. In this experiment, the scans were collected over 2.7 s within a TR of 6.7 s, leaving 4 s of silent period for auditory stimulus delivery and overt production (Fig. 1). High-order shimming before echo-planar image acquisition optimized the homogeneity of the magnetic field across the brain and minimized distortions. A high-resolution T1-weighted anatomical image was also acquired for registration with the functional data, using a 3D inversion recovery prepared spoiled gradient-recalled sequence (3D IR-Prep SPGR; TI = 450 ms, TE = 3.0 ms, flip angle = 12°, bandwidth = 31.25 mm, FOV = 240 mm, matrix 256×256 mm, 128 contiguous axial slices).

Data processing

Image preprocessing and all subsequent data analyses were carried out using Analysis of Functional Neuroimages (AFNI) software (Cox, 1996). The first four volumes were excluded from analysis to allow for initial stabilization of the fMRI signal. To correct for small head movements, each volume from the three functional runs were registered to the volume collected closest to the high-resolution anatomical scan using heptic polynomial interpolation. The percent signal change in each voxel was normalized by dividing the hemodynamic response amplitude at each time point by the mean amplitude of all the time points for that voxel from the same run, and multiplying by 100. These functional images from each run were then concatenated into one 3D + time file, and subsequently spatially smoothed using a 6 mm full-width half-maximum Gaussian filter.

The use of sparse sampling that captured only a narrow window near the peak of the hemodynamic response (HDR) ensured that task specific responses were sampled with minimal hemodynamic overlap. A rest period of 6.7 s with scanning preceded the first target presentation to further reduce any possible effects of motor planning and execution on the target presentation response. In addition, only data from the first of the two target presentation trials were used for target perception analysis so as to include primarily perception and not planning in the scanning during target perception.

The HDR for speech and non-speech planning responses, when modeled as a gamma variate function from visual cue onset, would have had negligible influence on the acquisition of the following production HDR, because data acquisition for production would have occurred at approximately 10 s into the HDR of planning, at which time the amplitude of the planning HDR was modeled to have been at 5% of the peak response. The visual cue for production was presented at the tail end of the planning HDR, and production occurred at an average of 500 ms after the cue, with average duration of 820 ms in speech and 916 ms for non-speech. One production HDR could be expected to return to baseline by approximately 12–13 s following visual cue to produce. There was no task following production, so the production HDR is likely to have had little if any influence on the following perception HDR.

During presentation of the auditory target, subjects not only perceived the stimulus but may have also engaged in non-vocal

silent rehearsal and short-term memory encoding. Therefore this was not solely a perception task. The subjects had to wait for the arrow onset approximately 4.7 s later to begin planning their production, as the arrow direction informed them of whether their upcoming production would have the same or a reversed order. The delay period also likely involved some short term memory encoding prior to production.

The amplitude coefficients for target perception (speech and non-speech) and production (speech and non-speech) for each subject were estimated using multiple linear regression. This created statistical parametric maps of t statistics for each of the linear coefficients. Statistical images were thresholded at $t>3.1$ ($p<0.01$, corrected). Correction for multiple comparisons was achieved using Monte Carlo simulations (program AlphaSim, part of AFNI), for which we selected a voxel-wise false-positive p threshold of 0.001 and a minimum cluster size of nine contiguous voxels (760 mm^3) to give a corrected p value of 0.01. Each individual's statistical map was transformed into standardized space (MNI 27 T1 weighted MRI from single subject) by using a 12 parameter affine registration.

Analyses

Comparisons of speech versus non-speech during target presentation and production stages

For group analyses, the t statistical maps of each condition were derived and entered into a mixed effects ANOVA, where task stage (target perception, production) and mode (speech versus non-speech) were fixed factors and subjects were a random factor. Contrasts between conditions of interest used pair-wise t-tests, resulting in statistical maps for each contrast. To identify overlapping and distinct regions of activation for speech and non-speech, in both the target presentation and production stages of the task, conjunction analyses (Friston et al., 1999; Nichols et al., 2005) were conducted based on the individual thresholded t statistical maps ($p<0.01$, corrected).

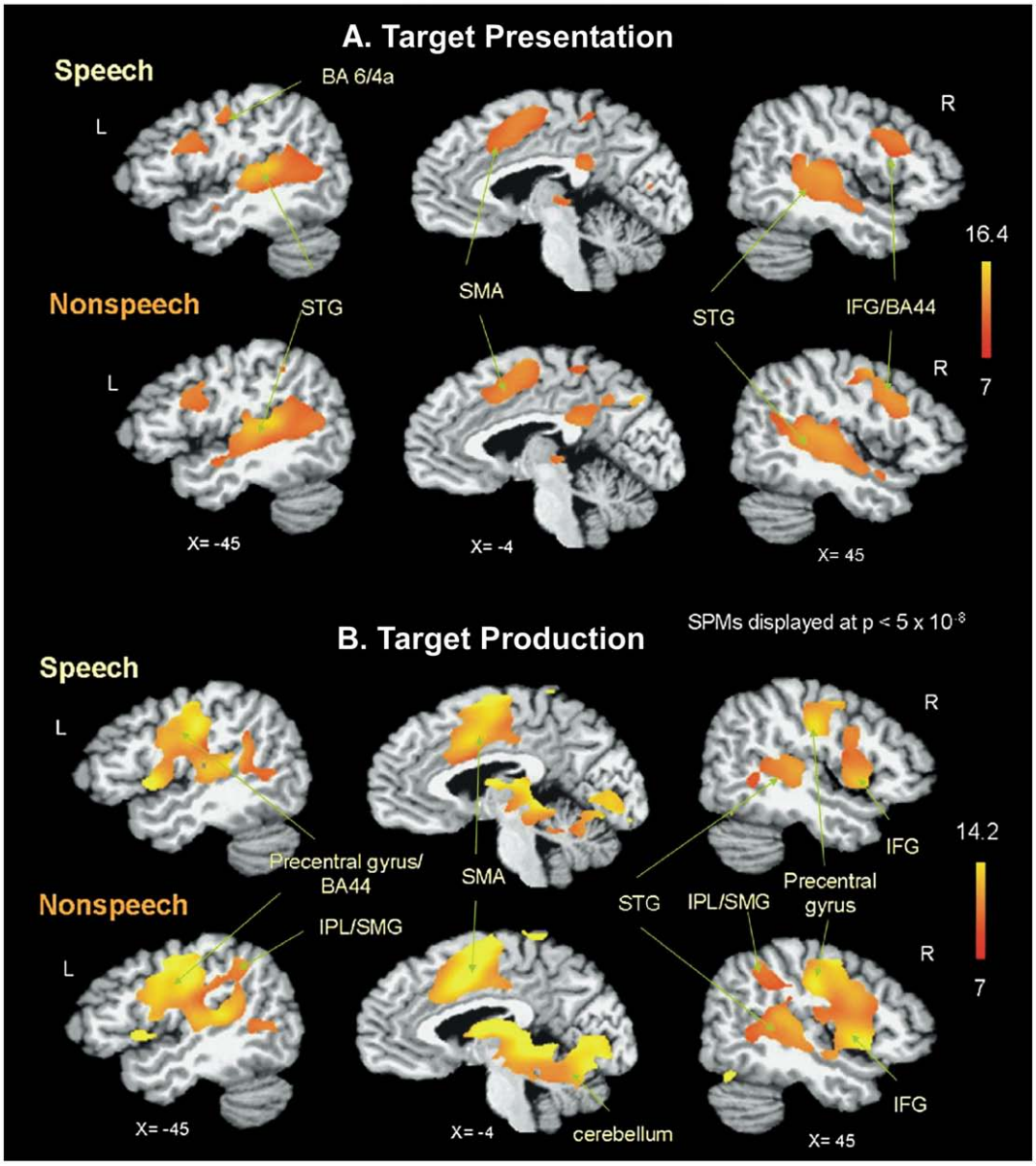

Fig. 2. Group main effects for task (A: target presentation, B: production) during speech and non-speech conditions. Speech and non-speech conditions resulted in comparable regions of activation, with differences primarily in the extent of activation. Non-speech conditions showed greater extent of activation than speech. The t statistical parametric maps were thresholded at $p=0.01$ (corrected for multiple comparisons). BA: Brodmann area, IFG: inferior frontal gyrus, IPL: inferior parietal lobule, SMA: supplementary motor area, SMG: supramarginal gyrus, STG: superior temporal gyrus.

Table 1
Regions activated for speech and non-speech target presentation.

Region	Approx BA	x	y	z	Peak t
Speech target presentation					
Left IFG	44	−42	11	25	11.62
Right IFG	45	45	14	26	12.32
Right paracentral lobule	4	4	−32	51	8.36
Left precentral gyrus	6	−40	−12	42	9.41
Right precentral gyrus	6	42	−5	46	6.85
Left SMA	6	−1	−2	53	13.57
Left cingulate gyrus	32	−3	21	34	6.75
Right cingulate gyrus	24	12	10	35	6.5
Left STG	22	−49	−29	3	14.98
Right STG	22	59	−26	4	17.98
Left caudate	N/A	−11	8	8	7.56
Right caudate	N/A	14	11	7	6.73
Left putamen	N/A	−25	3	8	6.94
Right putamen	N/A	21	12	7	7.33
Left thalamus	N/A	−2	−26	7	6.65
Right thalamus	N/A	9	−22	3	9.24
Right posterior cingulate	23	1	−36	24	9.04
Non-speech target presentation					
Left IFG	44	−40	5	25	9.43
Right IFG	44	44	12	25	10.33
Right paracentral lobule	4	7	−36	55	10.06
Left precentral gyrus	4a/6	−40	−9	47	7.9
Right precentral gyrus	6	51	−8	47	7.86
Left SMA	6	0	−2	53	13.6
Left STG	22	−45	−29	3	15.12
Right STG	42	66	−21	10	8.91
Right MTG	22	53	−37	3	13.36
Left IPL/SMG	40	−46	−46	47	6.54
Right IPL/SMG	40	52	−34	53	5.87
Left caudate	N/A	−12	2	13	7.04
Right caudate	N/A	15	9	6	6.68
Left putamen	N/A	−22	15	8	10.62
Right putamen	N/A	26	18	4	8.35
Left thalamus	N/A	−13	−30	14	7.38
Right thalamus	N/A	14	−24	−2	8.27
Right posterior cingulate	17	1	−57	12	5.93
Right precuneus	31	12	−47	34	9.56

t-scores of activation peaks for each anatomical region were thresholded at $t>3.6$, $p<0.01$ corrected. Results are reported for clusters exceeding 760 mm^3.

ROI analyses

We compared speech versus non-speech activation in regions encompassing those reported to be part of the speech production network (Bohland and Guenther, 2006; Guenther et al., 2006) (IFG (BA 44, 45), precentral motor (BA 4), STG, SMG). We additionally included those regions found to support speech motor processing in previous studies involving similar bi-syllabic non-sense speech production (Riecker et al., 2008) (SMA, sensorimotor (OP4), insula, and putamen). These ROIs were cytoarchitectonically defined, using atlas maps in standard space in the inferior frontal (BA 44, BA 45) (Amunts et al., 2004), sensorimotor (OP4, BA 4, supplementary motor area (SMA), preSMA) (Eickhoff et al., 2006; Zilles et al., 1995), and inferior parietal regions (supramarginal gyrus (SMG), angular gyrus) (Caspers et al., 2006), using maximum probability maps and macrolabel maps (Eickhoff et al., 2005) implemented in AFNI (Cox, 1996). These maps were not yet available for the posterior superior temporal gyrus (pSTG), insula, and putamen, so the Talairach daemon database (Lancaster et al., 2000) was used to define their regional boundaries. In addition, because the pSTG region including the planum temporale (PT) has high inter-subject variability, we manually edited the boundaries of the pSTG, so that its borders coincided from the posterior border of the first Heschl's gyrus (HG) (Heschl's sulcus) anteriorly, to the posterior ascending/descending rami posteriorly.

The ROIs were used as masks to extract two measures from each individual's standardized functional maps: the mean percent BOLD signal change values (relative to baseline rest) and mean percent volume of activation of voxels (thresholded at $t>3.3$, $p<0.01$, corrected). For each measure, a 4-way repeated measures ANOVA was used to examine the factors ROI (BA 44, BA 45, OP4, insula, putamen, BA 4, SMA, pSTG, angular gyrus, SMG), side (left, right), stage of task (target perception versus production), and mode (speech, non-speech) at $p=0.05$. If the contrast for speech versus non-speech or left versus right or their interactions were significant at $p<0.05$, then post hoc speech versus non-speech or left versus right or their interactions were tested across ROIs at $p=0.0045$ to correct for multiple comparisons.

Right–left comparisons

To assess functional laterality in brain activation for each task in each condition, a lateralization analysis (Husain et al., 2006) was performed to compare homologous left–right activation differences.

Table 2
Regions activated for speech and non-speech production.

Region	Approx BA	x	y	z	Peak t
Speech production					
Left IFG	44	−46	4	25	7.43
Right IFG	44	52	10	17	6.79
Left cingulate gyrus	24	−2	8	42	11.25
Right cingulate gyrus	24	5	11	40	12.17
Left insula	13	−46	−18	19	15.87
Right insula	44	48	6	3	13.19
Left precentral gyrus	4	−55	−15	38	13.49
Right precentral gyrus	4	48	−11	33	12.45
Left postcentral gyrus	43	−57	−10	23	12.49
Right postcentral gyrus	3b	56	−14	29	9.67
Left SMA	6	0	−3	44	15.78
Left STG	41	−38	−29	9	10.8
Right STG	22	60	−10	8	9.18
Left SMG	40	−34	−50	36	8.87
Right SMG	40	44	−47	42	6.67
Left caudate	N/A	−6	−1	10	7.39
Right caudate	N/A	13	−2	15	8.71
Left putamen	N/A	−15	6	8	7.63
Right putamen	N/A	20	8	2	8.12
Left thalamus	N/A	−12	−22	1	13.42
Right thalamus	N/A	13	−11	−1	11.57
Left cerebellum (VI)	N/A	−28	−65	−18	9.54
Right cerebellum (VI)	N/A	22	−68	−10	7.63
Non-speech production					
Left IFG	44	−54	4	29	9.6
Right IFG	44	54	10	27	9.48
Left MFG	9	−32	39	31	7.23
Right MFG	9	36	41	27	8.34
Left cingulate gyrus	24	−4	−5	51	13.08
Right cingulate gyrus	32	1	11	45	16.47
Left insula	N/A	−35	1	8	11.1
Right insula	13	49	0	1	10.16
Left precentral gyrus	4p	−54	−8	33	10.05
Right precentral gyrus	44	51	5	8	15.86
Left postcentral gyrus	43	−53	−18	20	16.17
Right postcentral gyrus	3b	51	−19	37	10.27
Left SMA	6	−3	−2	42	14.34
Right SMA	6	6	0	47	19.16
Left STG	41	−51	−20	6	9.93
Right STG	42	62	−26	17	10.23
Left IPL/SMG	40	−39	−48	40	10.61
Right IPL/SMG	40	58	−45	33	9.48
Left caudate	N/A	−13	−1	15	9.29
Right caudate	N/A	12	0	12	9.6
Left putamen	N/A	−20	0	7	7.85
Right putamen	N/A	22	2	7	5.88
Left thalamus	N/A	−17	−13	1	11.04
Right thalamus	N/A	13	−23	−2	13.54
Left cerebellum (VI)	N/A	−30	−65	−18	11.71
Right cerebellum (VI)	N/A	29	−75	−15	9.49

t-scores of activation peaks for each anatomical region were thresholded at $t>3.6$, $p<0.01$ corrected. Results are reported for clusters exceeding 760 mm^3.

Functional data from each subject were flipped along the *y* axis, and these maps of each condition were entered, along with the original data, into mixed effects ANOVA. Contrasts between the original and flipped functional data used one-way directional pair-wise *t*-tests, resulting in new sets of statistical maps that showed regions significantly more active for the left over the right hemisphere within each condition.

Results

Similarities between speech and non-speech during target presentation and production

Group analyses of the target presentation and production stages of the task showed similar BOLD responses during speech and non-speech (Fig. 2). Group main effects maps for the target presentation stage showed activation for both speech and non-speech in the auditory regions of STG/MTG and sylvian parietal temporal (Spt) region bilaterally, as well as regions associated with speech production including; inferior frontal gyrus, precentral gyri, SMA, precuneus, lentiform nucleus/putamen, and thalamus (Fig. 2A, Table 1).

Initially, we compared the BOLD response from the "repeat" trials with the "reverse" trials, and found no significant differences ($p>0.05$). Therefore the BOLD responses from both response types were combined for the production analysis. During the production stage, both speech and non-speech activated the bilateral pre- and postcentral gyri, middle and inferior frontal gyri, MTG, SMA, insula, cingulate cortex, SMG, lentiform nucleus, putamen, thalamus, and cerebellum, and the auditory regions of STG and Spt (Fig. 2B, Table 2). The extent of STG and Spt activation were greater on the right during the non-speech production condition (Fig. 2B).

A formal conjunction analysis showed substantial overlap between speech and non-speech conditions during both target presentation and production (Fig. 3). During target presentation, two regions were more active for speech compared to non-speech; the left inferior frontal gyrus (IFG) and the right insula/IFG. On the other hand, two other regions were more active for non-speech than for speech: the right and left supramarginal gyrus (SMG) and the right precentral gyrus (PrCGy) (Fig. 3A). During production, similar differences were noted; a region was more active in the right SMG during non-speech compared to speech (Fig. 3B). Some portions of the bilateral STG regions near STS appeared to be active for speech and not for non-speech.

Similar left laterality for speech and non-speech target presentation and production

Laterality analyses indicated that for both speech and non-speech, brain activation was significantly greater on the left during both target presentation and production tasks (Fig. 4). Both the pSTG and Spt regions were more active on the left during the perception and production of speech and non-speech targets.

ROI analyses also supported greater volume of activation and greater percent BOLD signal change on the left over the right in both speech and nonspeech. Repeated measures ANOVA revealed a significant ROI × side (left, right) interaction when examining percent volume of activation ($F_{10,330}=10.24$, $p<0.0005$) and percent signal change ($F_{10,330}=4.52$, $p<0.0005$). There was also significant ROI × mode (speech, non-speech) × side interaction for percent volume ($F_{10,330}=3.78$, $p<0.0005$) but not with percent signal change ($F_{10,330}=1.70$, $p=0.079$). The left volume of activation was greater than on the right during production in: OP4 ($F_{1,33}=32.16$, $p<0.0005$), pSTG ($F_{1,33}=11.51$, $p=0.002$) and SMG ($F_{1,33}=12.51$, $p=0.001$) (Fig. 5). In SMG, left sided percent volume for non-speech was greater than for speech during production ($F_{1,33}=11.74$, $p=0.002$) (Fig. 5B), and approached significance in perception ($F_{1,33}=8.38$, $p=0.007$), indicating that left laterality was greater during non-speech than speech in this region.

Comparisons between speech and non-speech in the extent of activation

To address this hypothesis, we contrasted speech and non-speech on a whole-brain analysis. During target presentation, non-speech showed greater activation than speech in the left inferior parietal region near SMG, right STG/MTG, the right middle frontal gyrus, right caudate, precuneus, and posterior cingulate gyrus (Fig. 6A, Table 1).

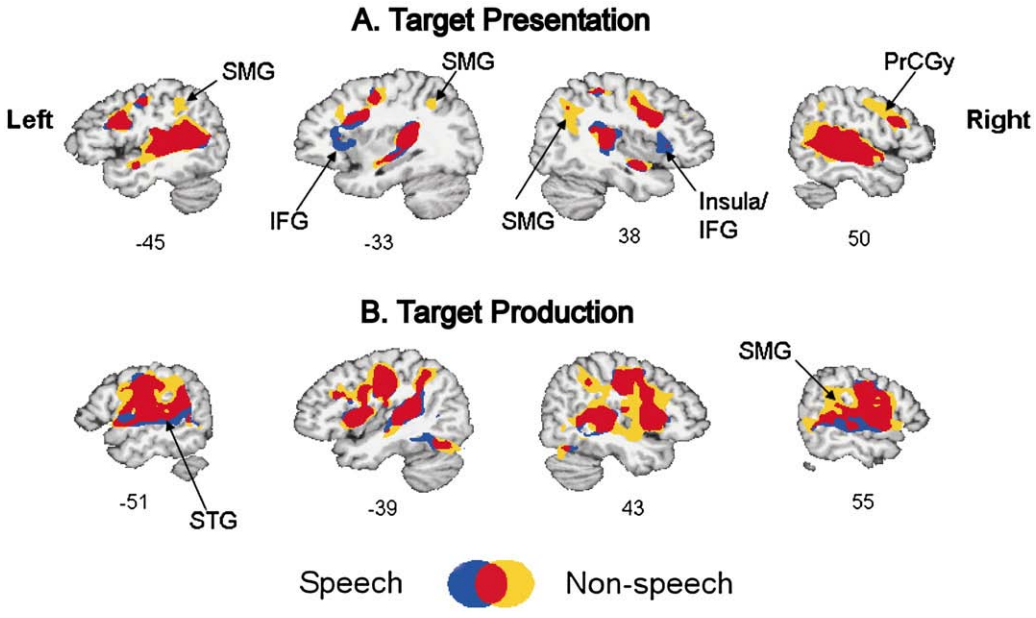

Fig. 3. Group conjunction maps showing overlapping regions activated for both speech and non-speech conditions (red), regions more specific to speech (blue), and non-speech (yellow). For display purposes, here each condition was thresholded at $t>6$ ($p<8.1\times10^{-7}$). IFG: inferior frontal gyrus, PrCGy: precentral gyrus, SMG: supramarginal gyrus, STG: superior temporal gyrus.

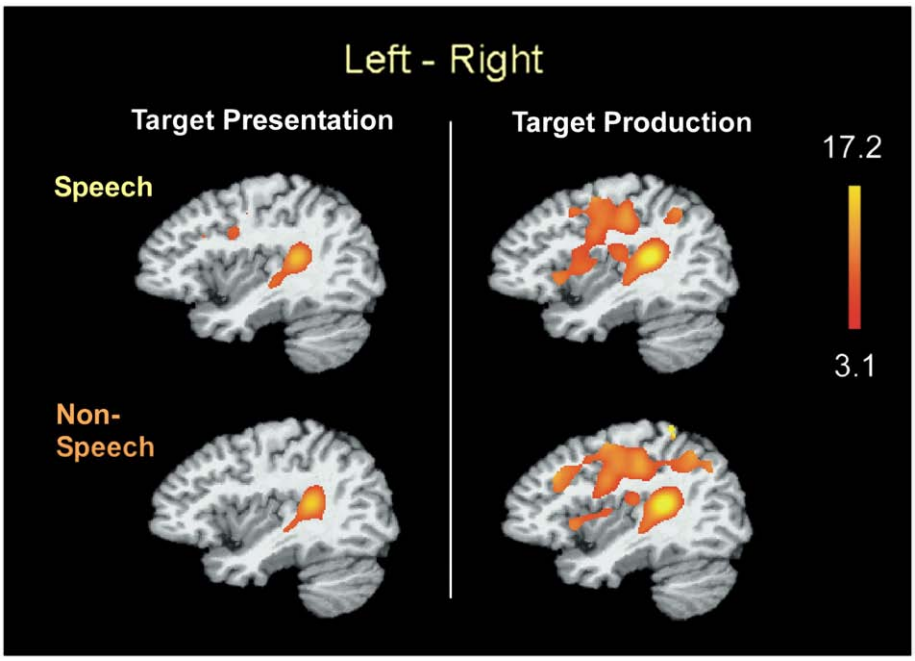

Fig. 4. Laterality analysis. Brain regions more active on the left hemisphere than the right hemisphere ($p = 0.01$, corrected) during target presentation and production. Speech and non-speech conditions activated comparable regions encompassing auditory dorsal stream structures with more left lateralization. The posterior superior temporal regions were consistently co-activated with left-bias for both target presentation and production stages, similar for speech and non-speech conditions.

No regions survived the threshold for being more active during speech target presentation than during non-speech target presentation.

During production, non-speech was more active than speech in: the bilateral precentral gyri/insula, inferior frontal gyri, bilateral inferior parietal lobule/SMG, thalamus, SMA, and the cerebellum (Fig. 6B, red–yellow, Table 2). Only the anterior cingulate cortex (ACC) and bilateral caudate were significantly more active during speech than during non-speech production (Fig. 6B, blue–light blue, Table 3). This difference could not be attributed to differences in reaction time (RT) of speech and non-speech production onsets: We measured RT offline for both speech and non-speech production onsets in a random sample of 19 subjects, based on digital recordings acquired during the whole experiment. The mean RT for speech production was 587 ms (SD: 187 ms) and the mean RT for non-speech was 564 ms (SD: 242 ms). A repeated measures ANOVA with speech and non-speech RTs as repeated factors showed that the two RT measures were not statistically different in any of the subjects examined ($F_{1,29} = 0.166$, $p = 0.687$).

ROI analyses also supported greater volume of activation and greater percent BOLD signal change during non-speech over speech during both target presentation and production in many of the ROIs examined. Repeated measures ANOVA revealed a significant ROI × mode (speech, non-speech) effect when examining percent volume of activation ($F_{10,330} = 4.09$, $p < 0.0005$) and percent signal change ($F_{10,330} = 2.57$, $p = 0.005$). There was also significant ROI × task (target presentation stage versus production) × mode effects for percent volume ($F_{10,330} = 5.67$, $p < 0.0005$) and percent signal change ($F_{10,330} = 3.47$, $p < 0.0005$).

When speech and non-speech were further compared within ROIs, non-speech target presentation was associated with significantly greater percent volume of activation than speech target presentation in the SMG ($F_{1,33} = 11.47$, $p = 0.002$). This approached significance in the BA 45 ($F_{1,33} = 9.07$, $p = 0.005$), angular gyrus ($F_{1,33} = 9.24$, $p = 0.005$), and pSTG ($F_{1,33} = 7.85$, $p = 0.008$) (Fig. 7A). No speech versus non-speech differences survived when measured using percent signal change.

During production, non-speech resulted in significantly greater percent volume of activation than speech in BA 44 ($F_{1,33} = 24.99$, $p < 0.0005$), OP4 ($F_{1,33} = 10.73$, $p = 0.002$), SMG ($F_{1,33} = 11.46$, $p = 0.002$), and the insula ($F_{1,33} = 17.24$, $p < 0.0005$), while BA 45 ($F_{1,33} = 9.11$, $p = 0.005$) approached significance (Fig. 7B). Non-speech production also had significantly greater percent signal changes than speech production in BA 44 ($F_{1,33} = 14.946$, $p = 0.001$), BA 45 ($F_{1,33} = 10.73$, $p = 0.002$), OP4 ($F_{1,33} = 9.78$, $p = 0.004$), SMG ($F_{1,33} = 10.11$, $p = 0.003$), preSMA ($F_{1,33} = 19.33$, $p < 0.0005$). This approached significance in the pSTG ($F_{1,33} = 8.73$, $p = 0.006$) (Fig. 7C).

Discussion

In this study we tested the idea of common neural substrates for target perception/encoding and production of speech and non-speech vocal tract gestures. The non-speech gestures used in this study, like speech, were easily producible in a consistent manner and had auditory and somatosensory targets linked to motor execution. This differentiates our non-speech gestures from other studies that have used either non-vocal (no phonation) oral gestures such as tongue movements (Salmelin and Sams, 2002) or non-vocal sounds such as tones (Benson et al., 2001; Binder et al., 2000). The perception and production of such non-vocal non-speech may have been less likely to have engaged the same neural substrates as those involved in speech, not because they were non-speech but because they did not involve vocal tract gestures to the same degree as the gestures used here. Further, in this study, both the speech and non-speech gestures required sequencing and neither the speech nor the non-speech conditions involved simple isolated gestures. The main difference for the non-speech gestures from speech used here was that they did not involve phonological processing. Despite this difference, regional functional activations for speech and non-speech target perception/encoding and production were similar, encompassing the bilateral IFG, STG, a superior temporal-parietal region (Spt), SMG, premotor regions, insula, subcortical areas (caudate, putamen, thalamus) and the cerebellum. Performance for both speech and non-speech tasks were associated with greater activation in the left hemisphere compared to the right, for both target perception/encoding and production.

Both our speech and non-speech tasks required motor productions that were linked to auditory and somatosensory targets, requiring sensory–motor mapping and were produced in a volitional manner but without communicative intent. However, they may have differed in

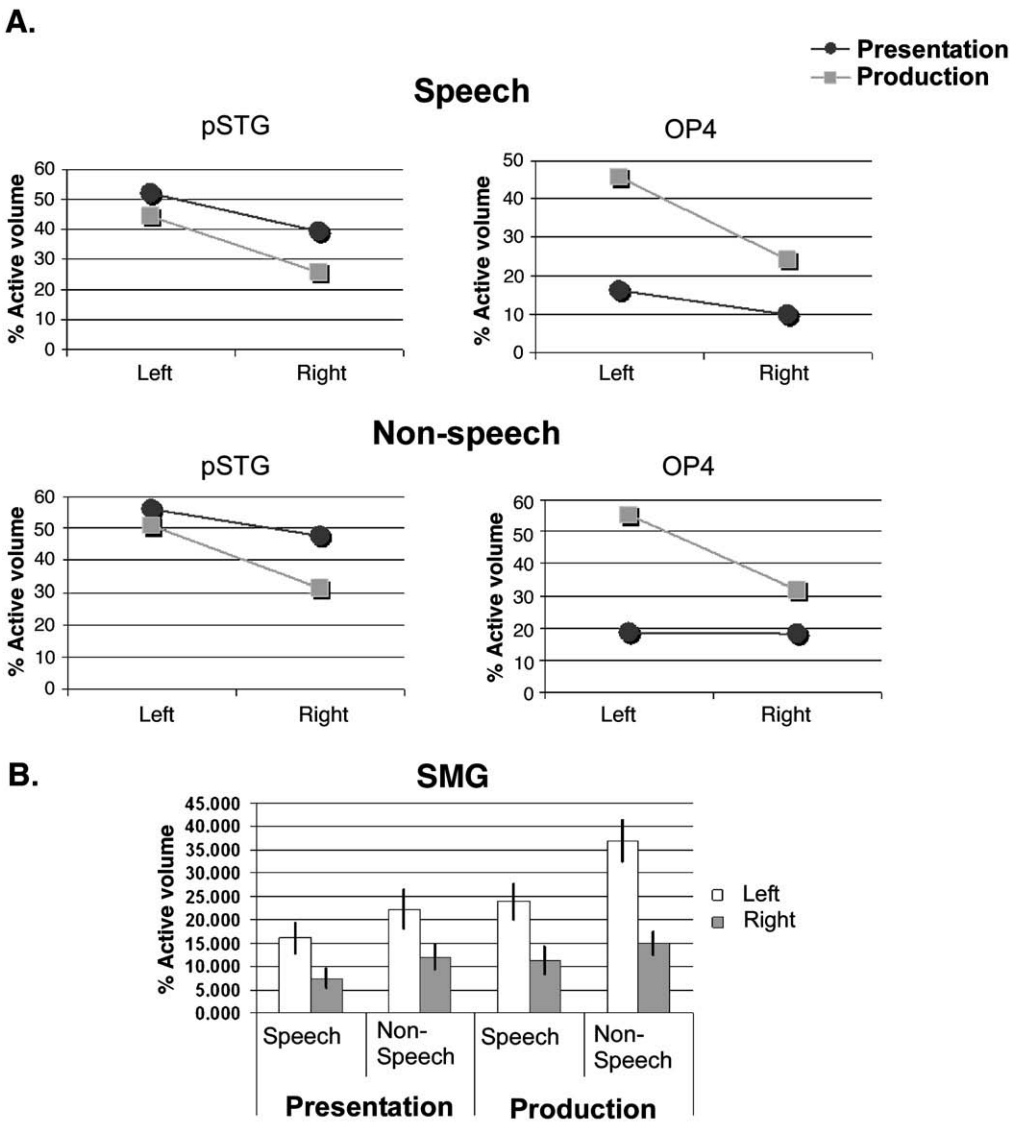

Fig. 5. (A) Mean volume of activation on the left and right hemispheres for target presentation and production stages on Speech and Non-speech shown for posterior superior temporal gyrus (pSTG) and OP4. In these ROIs, significant task (perception versus production) × side (left versus right) interactions were significant at $p = 0.01$. (B) Mean volume of activation on the left and right hemispheres for target presentation and production stages of task in the supramarginal gyrus (SMG). Here non-speech exhibited greater activation than speech during production (significant mode (speech versus non-speech) × task (perception versus production) × side (left versus right) interaction at $p = 0.01$). Error bars depict standard error of the mean.

complexity and familiarity, that is, the variety of gestures for speech articulation was less than those included in the oral non-speech gestures such as whistle or tongue click. Neither sets of gestures had semantic content, and it is unlikely that the nonsense speech syllables activated lexical representations (Binder et al., 2003; Vitevitch et al., 1999). However we cannot rule out that semantic representations may have been triggered by our non-speech gestures, such as cry or laugh. Given these potential differences, the regions activated for both speech and non-speech were remarkably similar, and underscore a strong common involvement of the same sensory–motor integration system. This system appears to support a larger domain of vocal tract gestures requiring sensory–motor mapping, and is not specialized to just the speech domain. These results are in agreement with recent studies by Hickok and colleagues who have suggested the auditory dorsal stream, and the posterior temporal–parietal region in particular, supports sensorimotor integration for not only speech but also non-speech (Hickok et al., 2003; Hickok and Poeppel, 2004, 2007; Pa and Hickok, 2008). They are also in agreement with studies that have examined perceptual discrimination of speech and non-speech sounds sharing similar temporal/acoustic characteristics and found that they activated overlapping regions (Joanisse and Gati, 2003; Zaehle et al., 2008).

Similar to what has been reported by other groups (Pulvermuller et al., 2006; Wilson et al., 2004), we also found motor area activation not only during production but also during target perception/encoding for both speech and non-speech gestures. Likely target presentation involved the perception as well as sub-vocal rehearsal of the oral–motor gestures for both speech and non-speech vocal tract gestures, and short-term memory encoding for the upcoming production stage. The regions that were active during target presentation were similar for vocal tract gestures and speech sounds, involving the ventral premotor, inferior frontal and motor regions in addition to the expected temporal auditory activations.

During the motor execution of both speech and non-speech vocal tract gestures, there was co-activation of motor, somatosensory, as well as auditory regions. Both speech and non-speech gestures were associated with activity in the IFG, ventral premotor areas, SMA, STG, insula, and SMG, cerebellum, and the basal ganglia, regions found to be active in other speech motor studies (Riecker et al., 2008).

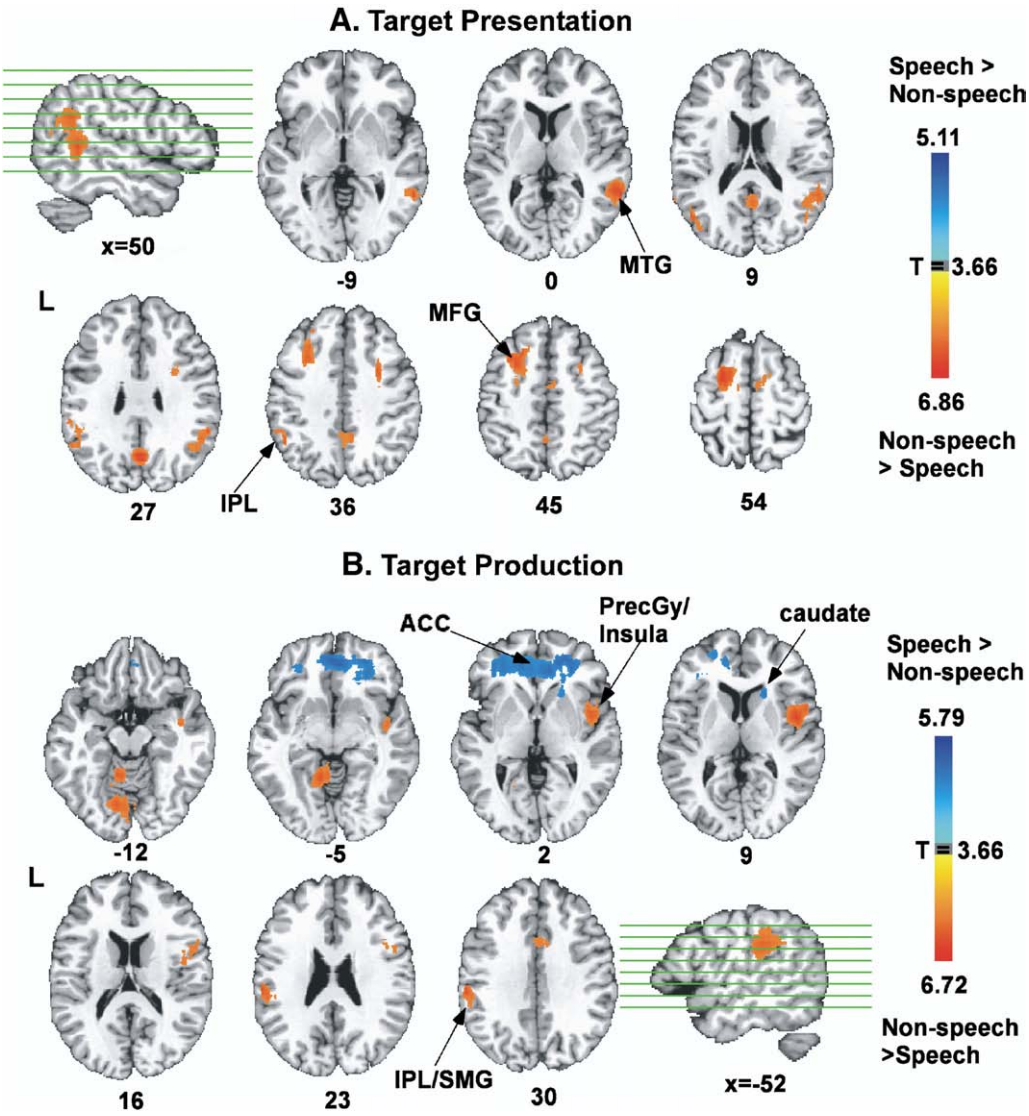

Fig. 6. Group contrasts between speech and non-speech conditions for target presentation and production. Regions colored red–yellow show areas more active during non-speech compared to speech, and regions colored blue–light blue show areas more active during speech compared to non-speech. All statistical maps were thresholded at $p = 0.01$ (corrected). ACC: anterior cingulate cortex, PrCGy: precentral gyrus, SMG: supramarginal gyrus, STG: superior temporal gyrus.

Table 3
Brain activation contrasts between speech and non-speech tasks.

Region	Approximate BA	*x*	*y*	*z*	*t*
Speech target presentation > non-speech target presentation					
No regions found significant					
Non-speech target presentation > speech target presentation					
Left mid. frontal gyrus	8	−29	17	41	6.86
Right mid. temporal gyrus	21	59	−47	8	6.49
Left inf. parietal lobule	40	−46	−49	39	6.01
Right precuneous	17/18/31	1	−62	26	5.82
Speech production > non-speech production					
Right anterior cingulate gyrus	32	4	40	−7	5.79
Right caudate	N/A	19	20	6	4.48
Non-speech production > speech production					
Left precentral gyrus/IFG	44	−55	2	12	4.53
Right precentral g./insula	44	45	−1	9	5.7
Cing g./SMA	6	0	1	46	6.72
Left inf. parietal lobule/SMG	40	−56	−25	26	5.74
Right SMG	40	47	−42	34	4.1
Left cerebellar culmen	N/A	−9	−46	−8	5.67

Although neither our speech nor non-speech gestures had lexical/semantic meaning associated with them, both involved volitional acts involving vocal tract gestures. Here the focus was on imitating an auditory target rather than on self-generation of a gesture to communicate affective or other information. Co-activation found in the premotor/frontal, as well as inferior parietal regions during perception as well as production of these gestures seems to parallel mirror neurons reported to be active during both action execution and action perception (Ferrari et al., 2003; Gallese et al., 1996; Rizzolatti et al., 1996; Rizzolatti and Sinigaglia, 2007). Of particular relevance to speech, the audiovisual mirror neurons found in the monkey F5, a Broca's area homologue (although some dispute this view, see Petrides et al., 2005), have been reported to discharge not just to the execution and observation of a specific action but also when this action can only be heard (Kohler et al., 2002). Although controversial, this area has been suggested to be a part of a mirror neuron system in humans, involved in the action production and action observation system. It has been proposed that this region, because of its capacity for supporting imitation, could have played a role in the evolution of speech (Rizzolatti and Arbib, 1998).

In the context of the putative mirror neuron system in humans, the neural pattern generated in the premotor areas during action

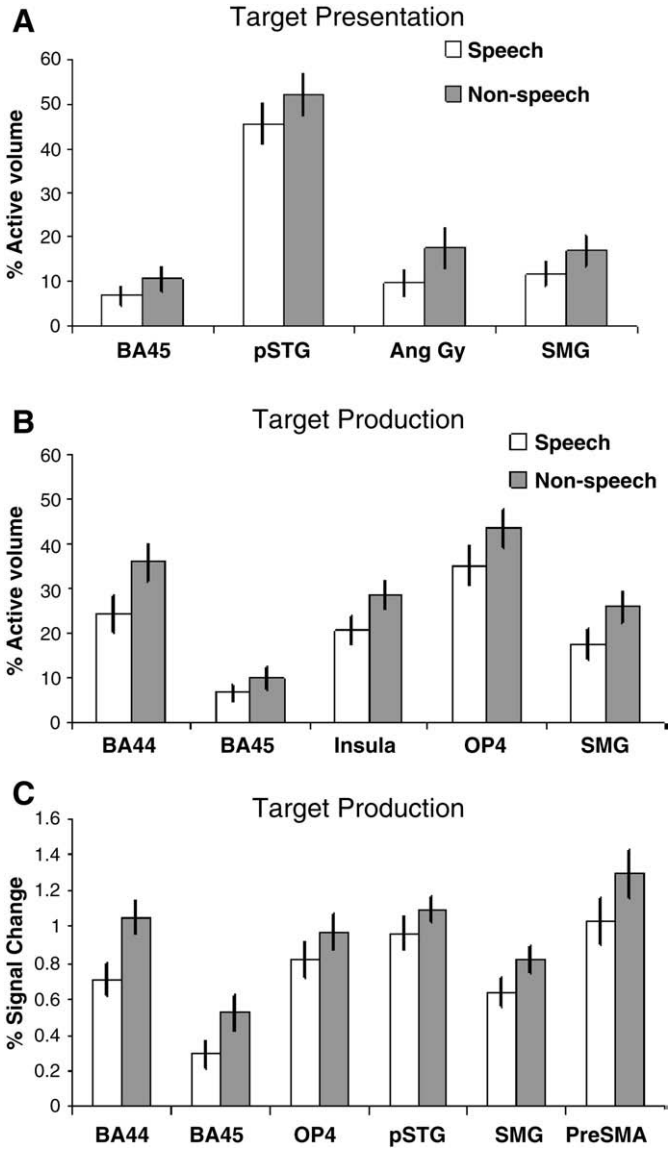

Fig. 7. Mean volume of activation across ROIs for speech versus nonspeech conditions for target presentation and production. The differences were all significant at $p = 0.01$. Error bars depict standard error of the mean.

recognition is similar to that generated to support production of that action (Kohler et al., 2002). Similarly, in the present data premotor regions were similarly active for perception and production regardless of speech or non-speech. This may be because even in the case of non-speech, these were produced involving actions that could be recognized from sound as well as produced, just like speech. Empirical findings of speech related motor activation during speech perception are easily found (Fadiga et al., 2002; Pulvermuller et al., 2006; Watkins et al., 2003; Wilson et al., 2004), and may reflect the involvement of regions suggested to have mirror neuron properties in humans for speech (posterior frontal/premotor) (Iacoboni and Mazziotta, 2007). Recent speech production models have also proposed that speech acquisition and production depend on imitative learning of speech through integrating action perception and production (Guenther, 2006; Hickok and Poeppel, 2007).

Laterality of activity during perception/presentation and production of targets were comparable for speech and non-speech, especially in the posterior temporal region pSTG and a sensorimotor region OP4. This suggests that vocal tract gestures with acoustic and somatosensory targets employ comparable neural substrates in the left dorsal stream regardless of whether they are speech or non-speech. The temporoparietal region in the present study that showed left laterality included the Spt region, argued to link sensory systems (whether auditory, somatosensory, or visual) with the motor effector, in this case the "vocal tract action system" (Pa and Hickok, 2008). Dhanjal et al. (2008) also showed that the Spt region was activated for speech as well as for non-speech tongue and jaw movements that result in somatosensory feedback (Dhanjal et al., 2008). This suggests that the Spt may not only be an auditory–motor integration area, but also a multisensory integration area for vocal tract gestures. Our finding of co-activation of this region during perception and production, for both speech and non-speech gestures, is in line with predictions that can be made on these previous studies; both sets of stimuli involved linking an auditory/somatosensory target presentation with vocal tract gestures.

Involvement of similar functional neuroanatomy for non-speech vocal tract gestures as for speech in humans may relate to previous findings suggesting that similar neural substrates underlie non-human primate calls, which also involve laryngeal and pharyngeal movement and sequencing. Monkeys have an architectonically comparable region to area 44 that controls orofacial muscle movement (Petrides et al., 2005), with cortico-cortical connections between the left temporal-parietal and frontal areas (Croxson et al.,

2005; Petrides and Pandya, 2002). Similar leftward asymmetries affect the planum temporale in monkeys (Gannon et al., 1998) and perisylvian homologues are activated in response to species-specific calls (Poremba et al., 2004).

The only difference between speech and non-speech processing was in the extent and amplitude of activation in regions within the shared neural network. Similarly, it has been shown that kinematically similar non-speech mouth movements elicit a higher level of activity in the motor cortex than speech movements (Saarinen et al., 2006) and is associated with spatially less focal activity (Salmelin and Sams, 2002) within the motor cortex. Because a greater extent and amplitude of response was seen even for kinematically similar non-speech gestures (Saarinen et al., 2006), the greater activation observed for our non-speech targets compared to speech may not be completely explained by the fact that non-speech required a greater variety of vocal tract/oral–motor gestures than used for speech targets.

Enhanced activation might be expected in auditory–motor regions during executions that are less familiar and less frequently produced, reflecting the need for active recruitment of regions to establish auditory–motor mapping. Enhanced activities in the premotor area, STG, PT, and cerebellum have been reported for non-native vowel contrasts (Callan et al., 2006). Similarly, non-native phonemes are associated with greater signal changes in speech regions, and increased signal changes occur in response to greater difficulties in production in the STG, insula, and Spt (Wilson and Iacoboni, 2006). We found heightened activation throughout the sensorimotor network for non-speech vocal tract gestures compared to speech, as would be expected for tasks if there was a less established feedforward system for motor output. Non-speech vocal tract gestures may have less well established auditory targets compared to speech.

In the present results, SMG was also more highly activated during non-speech compared to speech tasks, particularly on the left side. The left SMG may be an important region for integrating sounds to their articulator position information (Callan et al., 2006). In a speech computational model (DIVA; Directions into Velocities of Articulators) (Guenther, 2006), SMG is proposed as a "somatosensory error map", where the somatosensory target for a sound and the actual somatosensory state are compared. This may be parallel to what is proposed to occur in the posterior STG in this model, where expected and actual auditory consequences of a sound production are compared. Like the motor–auditory link, the motor–somatosensory link may be weaker for non-speech productions due to infrequent volitional production of these non-speech sounds. Hence, heightened activation in the SMG may reflect heightened need for somatosensory–motor integration to achieve the correct auditory–somatosensory target for non-speech production. For well-established skills such as speech, active somatosensory monitoring may not be required to the same degree as during less familiar sequences such as non-speech sequences. In fact, Dhanjal et al. (2008) showed that an area in the parietal operculum, SII (somatosensory association cortex), is less active during speech, compared to non-speech jaw and tongue movements, although both sets of tasks resulted in somatosensory feedback. This may reflect a greater reliance on conscious monitoring of the somatosensory feedback during non-speech tasks.

During target presentation, no areas were found more active during speech than non-speech, although with a more relaxed threshold we did see greater activation bilaterally in the STS regions during speech compared to non-speech. During production, the anterior cingulate cortex (ACC) and caudate nucleus were the only regions that were more active during speech than non-speech production. The ACC reportedly is involved in execution of appropriate verbal responses and suppression of inappropriate responses (Buckner et al., 1996; Paus et al., 1993). The caudate and basal ganglia have connections to the frontal cortical regions, and have been implicated as important when movement sequences need to be selected and initiated without external cues (Georgiou et al., 1994; Rogers et al., 1998). Perhaps the increased activation in the ACC and caudate during speech reflects the need for more precise movement and execution for speech.

There are several caveats to this study. Because target presentation was the first stage in the motor production task, subjects had to perceive the target and likely were involved in encoding and short-term rehearsal. This may explain the extensive neural overlap that occurred in brain regions active during the target presentation and production of speech and non-speech vocal tract gestures. This study did not use variable interstimulus intervals (ISIs), which could have allowed sampling of longer windows of hemodynamic responses. With variable ISIs, more extensive comparisons between speech and non-speech conditions over time might have been possible. Also, due to the low resolution of our functional scans, we may not have been able to capture small regions of activation that could have differentiated between speech and non-speech responses. Using multichannel MRI receivers and whole-brain surface coil arrays, one study showed that only with the high resolution and the increased signal to noise ratio, fine regions of modality specific responses could be captured using fMRI (Beauchamp et al., 2004). In the future, such advanced fMRI methods may allow for better elucidation of cortical regions that primarily process speech, non-speech, or both types of inputs.

In conclusion, we have shown overlapping sensory–motor responses during the target presentation and production of both speech and non-speech vocal tract gestures. We provide new data that supports the notion that the neural substrates involved in sensory to motor transformation in the left hemisphere are not specific to speech. Rather, these may have evolved for vocal communication in non-human primates and were subsequently adapted to support speech development in humans.

Acknowledgments

This research was supported by the Intramural Research Program in the National Institute of Neurological Disorders and Stroke, NIH. The authors wish to thank Richard Reynolds and Gang Chen for assistance during data analyses and Sandra Martin for conducting speech and language testing.

References

Ackermann, H., Ziegler, W., Petersen, D., 1993. Dysarthria in bilateral thalamic infarction. A case study. J. Neurol. 240, 357–362.
Amunts, K., Weiss, P.H., Mohlberg, H., Pieperhoff, P., Eickhoff, S., Gurd, J.M., Marshall, J.C., Shah, N.J., Fink, G.R., Zilles, K., 2004. Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic space — the roles of Brodmann areas 44 and 45. Neuroimage 22, 42–56.
Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A., 2004. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat. Neuroscience 7, 1190–1192.
Belin, P., Zatorre, R.J., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26.
Benson, R.R., Whalen, D.H., Richardson, M., Swainson, B., Clark, V.P., Lai, S., Liberman, A.M., 2001. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain Lang. 78, 364–396.
Benson, R.R., Richardson, M., Whalen, D.H., Lai, S., 2006. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. Neuroimage 31, 342–353.
Binder, J., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.
Binder, J.R., McKiernan, K.A., Parsons, M.E., Westbury, C.F., Possing, E.T., Kaufman, J.N., Buchanan, L., 2003. Neural correlates of lexical access during visual word recognition. J. Cogn. Neurosci. 15, 372–393.
Birn, R.M., Bandettini, P.A., Cox, R.W., Shaker, R., 1999. Event-related fMRI of tasks involving brief motion. Hum. Brain Mapp. 7, 106–114.
Bohland, J.W., Guenther, F.H., 2006. An fMRI investigation of syllable sequence production. Neuroimage 32, 821–841.
Bottjer, S.W., Brady, J.D., Cribbs, B., 2000. Connections of a motor cortical region in zebra pinches: relation to pathways for vocal learning. J. Comp. Neurol. 420, 244–260.
Buckner, R.L., Raichle, M.E., Miezin, F.M., Petersen, S.E., 1996. Functional anatomic studies of memory retrieval for auditory words and visual pictures. J. Neurosci. 16, 6219–6235.
Callan, A.M., Callan, D.E., Tajima, K., Akahane-Yamada, R., 2006. Neural processes involved with perception of non-native durational contrasts. NeuroReport 17, 1353–1357.

Canter, G.J., van Lancker, D.R., 1985. Disturbances of the temporal organization of speech following bilateral thalamic surgery in a patient with Parkinson's disease. J. Commun. Disord. 18, 329–349.
Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., Zilles, K., 2006. The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. Neuroimage 33, 430–448.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Croxson, P.L., Johansen-Berg, H., Behrens, T.E.J., Robson, M.D., Pinsk, M.A., Gross, C.G., Richter, W., Richter, M.C., Kastner, S., Rushworth, M.F.S., 2005. Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. J. Neurosci. 25, 8854–8866.
Dhanjal, N.S., Handunnetthi, L., Patel, M.C., Wise, R.J., 2008. Perceptual systems controlling speech production. J. Neurosci. 28, 9969–9975.
Dronkers, N.F., 1996. A new brain region for coordinating speech articulation. Nature 384, 159–161.
Eden, G.F., Joseph, J.E., Brown, H.E., Brown, C.P., Zeffiro, T.A., 1999. Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. Magn. Reson. Med. 41, 13–20.
Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.
Eickhoff, S.B., Grefkes, C., Zilles, K., Fink, G.R., 2006. The somatotopic organization of cytoarchitectonic areas on the human parietal operculum. Cereb. Cortex 16 (2), 254–267.
Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur. J. Neurosci. 15, 399–402.
Ferrari, P.F., Gallese, V., Rizzolatti, G., Fogassi, L., 2003. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. Eur. J. Neurosci. 17, 1703–1714.
Friston, K.J., Holmes, A.P., Price, C.J., Buchel, C., Worsley, K.J., 1999. Multisubject fMRI studies and conjunction analyses. Neuroimage 10, 385–396.
Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. Brain 119 (Pt 2), 593–609.
Gannon, P.J., Holloway, R.L., Broadfield, D.C., Braun, A.R., 1998. Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. Science 279, 220–222.
Gelfand, J.R., Bookheimer, S.Y., 2003. Dissociating neural mechanisms of temporal sequencing and processing phonemes. Neuron 38, 831–842.
Georgiou, N., Bradshaw, J.L., Iansek, R., Phillips, J.G., Mattingley, J.B., Bradshaw, J.A., 1994. Reduction in external cues and movement sequencing in Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 57, 368–370.
Guenther, F.H., 2006. Cortical interactions underlying the production of speech sounds. J. Commun. Disord. 39, 350–365.
Guenther, F.H., Ghosh, S.S., Tourville, J.A., 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 96, 280–301.
Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., Bowtell, R.W., 1999. Sparse' temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.
Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev., Neurosci. 8, 393–402.
Hickok, G., Buchsbaum, B., Humphries, C., Muftuler, T., 2003. Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J. Cogn. Neurosci. 15, 673–682.
Hillis, A.E., Work, M., Barker, P.B., Jacobs, M.A., Breese, E.L., Maurer, K., 2004. Re-examining the brain regions crucial for orchestrating speech articulation. Brain 127, 1479–1487.
Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., Horwitz, B., 2006. Neural bases of categorization of simple speech and nonspeech sounds. Hum. Brain Mapp. 27, 636–651.
Iacoboni, M., Mazziotta, J.C., 2007. Mirror neuron system: basic findings and clinical applications. Ann. Neurol. 62, 213–218.
Joanisse, M.F., Gati, J.S., 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. Neuroimage 19, 64–79.
Kent, R.D., Netsell, R., Abbs, J.H., 1979. Acoustic characteristics of dysarthria associated with cerebellar disease. J. Speech Hear Res. 22, 627–648.
Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846–848.
Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach atlas labels for functional brain mapping. Hum. Brain Mapp. 10, 120–131.
Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15, 1621–1631.
Metzner, W., 1996. Anatomical basis for audio–vocal integration in echolocating horseshoe bats. J. Comp. Neurol. 368, 252–269.
Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25, 653–660.
Ojemann, G.A., 1994. Cortical stimulation and recording in language. Academic Press, San Diego, CA.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Ozsancak, C., Auzou, P., Hannequin, D., 2000. Dysarthria and orofacial apraxia in corticobasal degeneration. Mov. Disord. 15, 905–910.
Pa, J., Hickok, G., 2008. A parietal-temporal sensory–motor integration area for the human vocal tract: evidence from an fMRI study of skilled musicians. Neuropsychologia 46, 362–368.
Paus, T., Petrides, M., Evans, A.C., Meyer, E., 1993. Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses — a positron emission tomography study. J. Neurophysiol. 70, 453–469.
Penfield, W., Roberts, L. (Eds.), 1959. Speech and brain mechanisms. Princeton University Press, Princeton.
Petrides, M., Pandya, D.N., 2002. Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. Eur. J. Neurosci. 16, 291–310.
Petrides, M., Cadoret, G., Mackey, S., 2005. Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. Nature 435, 1235–1238.
Poremba, A., Malloy, M., Saunders, R.C., Carson, R.E., Herscovitch, P., Mishkin, M., 2004. Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature 427, 448–451.
Pulvermuller, F., Huss, M., Kherif, F., Martin, F.M.D.P., Hauk, O., Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. Proc. Natl. Acad. Sci. U. S. A. 103, 7865–7870.
Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., Grodd, W., 2000. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. NeuroReport 11, 1997–2000.
Riecker, A., Brendel, B., Ziegler, W., Erb, M., Ackermann, H., 2008. The influence of syllable onset complexity and syllable frequency on speech motor control. Brain Lang. 107, 102–113.
Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends Neurosci. 21, 188–194.
Rizzolatti, G., Sinigaglia, C., 2007. Mirror neurons and motor intentionality. Funct. Neurol. 22, 205–210.
Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. Brain Res. Cogn. Brain Res. 3, 131–141.
Rogers, R.D., Sahakian, B.J., Hodges, J.R., Polkey, C.E., Kennard, C., Robbins, T.W., 1998. Dissociating executive mechanisms of task control following frontal lobe damage and Parkinson's disease. Brain 121, 815–842.
Saarinen, T., Laaksonen, H., Parviainen, T., Salmelin, R., 2006. Motor cortex dynamics in visuomotor production of speech and non-speech mouth movements. Cereb. Cortex 16, 212–222.
Salmelin, R., Sams, M., 2002. Motor cortex involvement during verbal versus non-verbal lip and tongue movements. Hum. Brain Mapp. 16, 81–91.
Schulz, G.M., Peterson, T., Sapienza, C.M., Greer, M., Friedman, W., 1999. Voice and speech characteristics of persons with Parkinson's disease pre- and post-pallidotomy surgery: preliminary findings. J. Speech Lang. Hear Res. 42, 1176–1194.
Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.
Smotherman, M.S., 2007. Sensory feedback control of mammalian vocalizations. Behav. Brain Res. 182, 315–326.
Soros, P., Sokoloff, L.G., Bose, A., McIntosh, A.R., Graham, S.J., Stuss, D.T., 2006. Clustered functional MRI of overt speech production. Neuroimage 32, 376–387.
Vitevitch, M.S., Luce, P.A., Pisoni, D.B., Auer, E.T., 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. Brain Lang. 68, 306–311.
Warren, J.E., Wise, R.J.S., Warren, J.D., 2005. Sounds do-able: auditory–motor transformations and the posterior temporal plane. Trends Neurosci. 28, 636–643.
Watkins, K.E., Strafella, A.P., Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41, 989–994.
Whalen, D.H., Benson, R.R., Richardson, M., Swainson, B., Clark, V.P., Lai, S., Mencl, W.E., Fulbright, R.K., Constable, R.T., Liberman, A.M., 2006. Differentiation of speech and nonspeech processing within primary auditory cortex. J. Acoust. Soc. Am. 119, 575–581.
Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., Grodd, W., 1996. Functional lateralization of speech production at primary motor cortex: a fMRI study. NeuroReport 7, 2791–2795.
Wilson, S.M., Iacoboni, M., 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. Neuroimage 33, 316–325.
Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neuroscience 7, 701–702.
Wise, R.J.S., Greene, J., Bu?chel, C., Scott, S.K., 1999. Brain regions involved in articulation. Lancet 353, 1057–1061.
Zaehle, T., Geiser, E., Alter, K., Jancke, L., Meyer, M., 2008. Segmental processing in the human auditory dorsal stream. Brain Res 1220, 179–190.
Zarate, J.M., Zatorre, R.J., 2005. Neural substrates governing audiovocal integration for vocal pitch regulation in singing. Ann. N. Y. Acad. Sci. 1060, 404–408.
Ziegler, W., Hartmann, E., Hoole, P., 1993. Syllabic timing in dysarthria. J. Speech Hear Res. 36, 683–693.
Zilles, K., Schlaug, G., Matelli, M., Luppino, G., Schleicher, A., Qu, M., Dabringhaus, A., Seitz, R., Roland, P.E., 1995. Mapping of human and macaque sensorimotor areas by integrating architectonic, transmitter receptor, MRI and PET data. J. Anat. 187 (Pt 3), 515–537.