Perceptual Systems Controlling Speech Production

Novraj S. Dhanjal,¹ Lahiru Handunnetthi,¹ Maneesh C. Patel,² and Richard J. S. Wise¹

¹Division of Neuroscience and Mental Health and Medical Research Council Clinical Sciences Centre, Imperial College London, London W12 0NN, United Kingdom, and ²Department of Radiology, Imperial College Academic Health Sciences Centre, London W6 8RF, United Kingdom

It is proposed that the acquisition and maintenance of fluent speech depend on the rapid temporal integration of motor feedforward and polysensory (auditory and somatosensory) feedback signals. In a functional magnetic resonance imaging study on 21 healthy right-handed, English-speaking volunteers, we investigated activity within these motor and sensory pathways and their integration during speech. Four motor conditions were studied: two speech conditions (propositional and nonpropositional speech) and two silent conditions requiring repetitive movement of the principal articulators (jaw and tongue movements). The scanning technique was adapted to minimize artifact associated with overt speech production. Our result indicates that this multimodal convergence occurs within the left and right supratemporal planes (STPs), with peaks of activity at their posteromedial extents, in regions classically considered as unimodal auditory association cortex. This cortical specialization contrasted sharply with the response of somatosensory association cortex (SII), in which activity was suppressed during speech but not during the silent repetitive movement of the principal articulators. It was also clearly distinct from the response of lateral auditory association cortex, which responded to auditory feedback alone, and from that within a left lateralized ventrolateral temporal and inferior frontal system, which served lexical- and sentence-level language retrieval. This response of cortical regions related to speech production is not predicted by the classical model of hierarchical cortical processing, providing new insights into the role of the STP in polysensory integration and into the modulation of activity in SII during normal speech production. These findings have novel implications for the acquisition and maintenance of fluent speech.

Key words: speech; somatosensory; auditory; fMRI; parietal; multisensory

Introduction

Fluent speech requires fine control of the muscles of respiration, the larynx and the articulators. Their activity results in two modalities of feedback, auditory and somatosensory. During speech acquisition, the heteromodal feedback trains the motor system to generate sequences of sounds that match remembered auditory templates of fluent speech (Doupe and Kuhl, 1999). Even when trained, the system continues to match the anticipated with the actual consequence of speech to maintain fluency. These concepts are captured in the DIVA (Directions into Velocities of Articulators) computational model (Guenther et al., 2006). Cortical signals directed at bulbar and spinal motor neurons are copied to auditory and somatosensory “error maps.” Milliseconds later, reafferent discharges arrive at auditory and somatosensory “state maps.” Discrepancies between feedforward and feedback signal result in adjustments of subsequent motor output until the input to the error and state maps coincide.

Although the error and state maps in the DIVA model are hypothetical constructs, they are envisaged to be instantiated as neural structures, with the proposed location for the auditory maps in the caudal supratemporal plane (STP) and for the so-

matosensory maps in anterior parietal cortex. The alternative is a closer anatomical convergence. Studies on macaque monkeys have demonstrated that the caudomedial (CM) auditory belt, ventral to second-order somatosensory cortex (SII) in the parietal operculum, receives both auditory and somatosensory afferent projections (Smiley et al., 2007).

Previous human studies investigating sensory feedback associated with speech have investigated auditory processing alone. The usual finding is a relative suppression of activity within auditory cortex (Curio et al., 2000; Houde et al., 2002). This presents an alternative hypothesis about the behavioral consequences of an interaction between the expected and actual encoded sensory consequences of an action: as a process that differentiates between self-generated and externally generated events rather than monitoring and controlling a self-initiated action (Blakemore, 2003). It is possible that discrete neuronal populations serve both these functions. Thus, recent single-cell recordings in the nonhuman primate indicate that auditory neurons demonstrating suppression of activity in response to self-vocalizations remain sensitive to perturbations of the feedback auditory signal (Eliades and Wang, 2008).

We have demonstrated previously that the lateral temporal neocortex and the medial part of the planum temporale responds during overt speech production relative to a silent baseline condition (Blank et al., 2002). We (Wise et al., 2001) and others (Buchsbaum et al., 2001, 2005; Hickok et al., 2003; Hickok and Poeppel, 2007) have proposed that cortex at the caudal end of the lateral sulcus, in or close to the planum temporale, is a “sensorimotor interface” that supports speech production. In this func-

Received June 4, 2008; revised July 17, 2008; accepted Aug. 19, 2008.

N.S.D. is supported by grants from the Royal College of Physicians of London, the Dunhill Medical Trust, and the United Kingdom Medical Research Council.

Correspondence should be addressed to Dr. Novraj S. Dhanjal, Division of Neuroscience and Mental Health and Medical Research Council Clinical Sciences Centre, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, UK. E-mail: novraj.dhanjal@imperial.ac.uk.

DOI:10.1523/JNEUROSCI.2607-08.2008

Copyright © 2008 Society for Neuroscience 0270-6474/08/289969-07\$15.00/0

tional magnetic resonance imaging (fMRI) study, we wanted to characterize the functional response of this critical speech node in response to auditory and somatosensory feedback generated by the articulators. We addressed the specific hypothesis that “auditory” cortex within the caudal STP may respond to both the somatosensory and the auditory feedback associated with overt speech. Therefore, the study was designed to investigate whether there was a polymodal response within the STP that could be dissociated from the response within the parietal operculum (SII).

Materials and Methods

Participants and fMRI procedures. Twenty-one healthy, right-handed, native English speakers (eight females; median age, 26 years; range, 22–39 years) participated after giving informed written consent. Ethics approval was provided by the Hammersmith Hospital research ethics committee.

MR data were obtained on a Philips Intera 3.0 Tesla scanner using dual gradients, a phased array head coil, and sensitivity encoding with an undersampling factor of 2. A “sparse” fMRI design (Hall et al., 1999) was used to minimize movement- and respiratory-related artifact associated with speech studies (Gracco et al., 2005; Mehta et al., 2006). Tasks were performed over 7 s while an appropriate visual stimulus was displayed. The disappearance of that stimulus and the appearance of a fixation crosshair signaled to the subject to cease the task. One second later, data were acquired, followed by a further stimulus for the subject to commence an additional period of task performance.

Functional MR images were obtained using a T_2^* -weighted, gradient-echo, echoplanar imaging (EPI) sequence with whole-brain coverage (repetition time, 10.0 s; acquisition time, 2.0 s; echo time, 30 ms; flip angle, 90°). Thirty-two axial slices with a slice thickness of 3.25 mm and an interslice gap of 0.75 mm were acquired in ascending order (resolution, $2.19 \times 2.19 \times 4.0$ mm; field of view, $280 \times 224 \times 128$ mm). Quadratic shim gradients were used to correct for magnetic field inhomogeneities within the anatomy of interest. T_1 -weighted whole-brain structural images were obtained in all subjects.

Stimuli were presented using E-Prime software (Psychology Software Tools) run on an IFIS-SA system (In Vivo Corporation).

fMRI paradigm design. There were four active conditions: two overt speech conditions, propositional speech (“speech”) and counting aloud (“count”), and two silent conditions involving nonspeech movements of the principal articulators, jaw movements (“jaw”) and tongue movements (“tongue”). “Rest” was included as the nonmovement baseline condition. Trial onset was signaled by simple two-word written instructions displayed on a screen in front of the subject.

During speech trials, subjects were required to define simple, high-frequency nouns (e.g., “car”) selected from the Medical Research Council psycholinguistic database (Wilson, 1988). During count trials, subjects were required to count upward from one at a rate of approximately one number per second. In jaw trials, subjects were required to repetitively open and close their jaw, with their tongues resting in a neutral position at the floor of the mouth. During tongue trials, subjects were required to repetitively move the tongue from the floor of the mouth to the upper alveolar ridge of the hard palate and back, while keeping the jaw still. Specific instruction was given as to the exact movement required, which was practiced outside the scanner before starting the study. The subjects were trained to produce one movement per second.

During rest trials, subjects were instructed not to move their jaw or tongue, while breathing normally. Auditory output was recorded using an MR-compatible microphone attached to ear-defending headphones (MR Confon) to assess task performance in all conditions.

Trials were presented in pseudoblocks, each trial type being repeated twice before switching. There were two runs, each of 75 volumes, separated by acquisition of a high-resolution T_1 -weighted structural scan. The sequence in which trials occurred was different in each of the two runs, and the order of runs were randomized between subjects.

fMRI data analysis. Data was analyzed using SPM5 software (<http://www.fil.ion.ucl.ac.uk/spm>). Functional scans were initially realigned to

the first scan of the run, removing the effects of head movement between scans. The high-resolution T_1 -structural image was skull stripped using the brain extraction tool within MRIcro software (Smith, 2002) to remove non-brain matter and improve automatic segmentation to gray and white matter, before being coregistered to the mean functional image. The EPI images were then normalized into Montreal Neurological Institute (MNI) standard stereotactic space using parameters from segmentation of the T_1 -structural image. Smoothing of the normalized EPI images was performed using a 8 mm full-width at half-maximum Gaussian filter.

The initial analysis of the data were at the individual subject level, when individual design matrices were created, modeling each of the five experimental conditions. The movement parameters derived from the realignment stage were incorporated as nuisance variables. Contrast images were produced from these individual analyses for contrasts of interest, and these were used in the second-level, random-effects analysis.

For all contrasts of interest, the threshold for significance was set at $p < 0.01$ (except when stated otherwise), adjusted for multiple comparisons using the false discovery rate (FDR) correction (Genovese et al., 2002), with a cluster extent threshold of 10 voxels. Voxels that were common to two or more contrasts were determined by “inclusive masking,” using this same threshold when generating both the masking and reference contrasts. Region-of-interest (ROI) analysis was performed within the MarsBar toolbox of SPM5 (Brett et al., 2003), by producing spherical ROIs (4 mm radius) centered around functionally defined peaks of activity. The mean effect size in each of the active task conditions against rest was determined for each functionally defined ROI. These data were available for additional statistical analysis within SPSS.

The anatomical location of peaks of activity within premotor, motor, auditory, somatosensory, and cerebellar cortex was determined by referring to the Anatomical Toolbox (Eickhoff et al., 2005, 2006) in SPM5. The x , y , and z coordinates relate to the standard stereotactic space from MNI.

Results

Rates of speech production

During speech, the mean rate of syllable production was 19 syllables/7 s epoch (range, 5–35 syllables), and, during count, the rate was 13 syllables/7 s epoch (range, 6–26 syllables). The subjects had been pretrained to produce movements at ~ 7 syllables/7 s epoch during jaw and tongue.

Activity common to jaw, tongue, count, and speech

The distribution of common activity for all conditions (jaw, tongue, count, and speech), each relative to the rest condition, is illustrated in Figures 1 and 2. Each condition was separately contrasted with rest, and voxels common to all four contrasts were identified. Peaks and subpeaks of activity were located in the premotor [medial and lateral, Brodmann area (BA) 6], motor (BA 4), and somatosensory (BA 3) cortex of both hemispheres. The cerebellar peaks were located in left and right lobule VI (Schmahmann et al., 1999). The activated regions are summarized in the legend to Figure 1.

A number of studies (Wise et al., 1999; Carreiras et al., 2006; Riecker et al., 2006) have emphasized that the rate of single syllable or single word production influences activity in a number of speech-related neural structures. In the present study, the rates of articulatory movements across conditions were ranked: speech > count > jaw = tongue. Thus, the rate of articulatory movements and the linguistic and semantic complexity of the task were not orthogonal, and so an analysis based on correlating blood oxygenation level-dependent signal with rate was not appropriate for this study. Nevertheless, the confound of rate of production was taken into account when interpreting the results of contrasts between conditions.

We emphasize two novel findings. First, there was common

Figure 1. Statistical parametric maps. Sagittal (top), coronal (middle), and axial (bottom) views of the following analyses. Precise localization of all activated areas is detailed in supplemental Table 1 (available at www.jneurosci.org as supplemental material). **A**, A conjunction of activity of the speech, count, tongue, and jaw conditions, each contrasted with the rest condition. There is a distributed network of bilateral motor activity, involving the medial premotor cortex, the SMA (1), the lateral premotor and primary sensorimotor cortex (2), and cerebellum (3). In addition, there was common activity along the STP and involving the medial and lateral aspects of the planum temporale, particularly evident on the axial view (4). **B**, A conjunction of activity of speech and count contrasted separately with jaw and tongue. There is bilateral activity within lateral STG (5) lying in the region of auditory parabelt and within motor cortex (6). **C**, The somatotopy of tongue movement (FDR, $p < 0.05$): a contrast of tongue against jaw, demonstrating bilateral peaks (7) within primary motor cortex.

bilateral activity distributed along the length of the STP, normally considered to be unimodal auditory cortex, despite the jaw and tongue conditions being silent. The main peaks on the left and right were situated lateral and immediately anterior to primary auditory cortex. More caudal subpeaks were located over the most medial part of left and right planum temporale. However, activity spread from the fundus of each lateral sulcus [the location of somatosensory retroinsular cortex (Ri) in the nonhuman primate], across the mediolateral extent of the planum temporale, and along the STP to the planum polare, rostral to primary auditory cortex. Second, despite activity in bilateral primary sensorimotor cortex in all four conditions (confirmed by mapping the coordinates of the peaks of activity within the SPM5 Anatomical Toolbox), there was an absence of evidence for activity in the SII of either the left or right parietal operculum, although all four conditions will have generated somatosensory feedback.

Activity common to count and speech relative to jaw and tongue

The distribution of common activity for the count and speech conditions, relative to the jaw and tongue conditions, is illustrated in Figures 1 and 2 and summarized in the legend to Figure 1. Count and speech were separately contrasted with (jaw + tongue), and voxels common to each contrast were identified. The main finding was symmetrically distributed activity in the left and right superior temporal gyri (STG), centered on its lateral surface. In the nonhuman primate brain, this region is where

lateral belt areas merge with parabelt cortex. Neurons here respond most strongly to complex sounds, including conspecific vocalizations (Rauschecker and Tian, 2000). There was little or no response to the silent jaw and tongue conditions relative to rest (Fig. 2).

The count and speech conditions engaged laryngeal motor control and speech-related cortical control of breathing (Draper et al., 1959). In contrast, the metabolically determined respiratory cycle are mostly uninterrupted by jaw and tongue, and these conditions did not involve control of the vocal folds. The common primary sensorimotor activity for count and speech relative to jaw and tongue was located close to that previously described for speech production (Murphy et al., 1997). Tongue movements are necessarily associated with fixation of the mandible, into which the tongue is inserted, and so the contrast of the tongue with the jaw conditions best localized the sensorimotor somatotopy for the tongue (Fig. 1). The reverse contrast of the jaw and tongue conditions did generate activations within primary sensorimotor cortex, despite jaw fixation during tongue movements, a little medial and dorsal to that for the tongue (data not illustrated). These movement-specific sensorimotor peaks were embedded within a much larger and overlapping activation of bilateral sensorimotor cortex evoked by all four conditions.

Activity common to jaw and tongue relative to count and speech

Figure 3 demonstrates a symmetrical peri-Sylvian distribution of activity that was common to the jaw and tongue conditions and greater than that associated with the count and speech conditions. Jaw and tongue were separately contrasted with (count + speech), and voxels common to each contrast were identified. Separate clusters of activity were observed in the left and right parietal operculum. In addition, there was activity, symmetrically distributed between the hemispheres, in ventral lateral premotor cortex (BA 6), the operculum immediately ventral to the central sulcus and the adjacent insular cortex. The cluster within each parietal operculum lay predominantly within human homolog of monkey SII, within the region labeled OP1 (Eickhoff et al., 2006), but extending dorsally into the postcentral gyrus or sulcus. Although activity during the count condition was no different from the rest condition in either parietal operculum, activity was significantly less during the speech condition relative to rest (left: one-sample t test, $t_{(20)} = -3.9$, $p < 0.001$; right: one-sample t test, $t_{(20)} = -3.4$, $p < 0.01$). Therefore, the somatosensory feedback to SII during propositional speech appears to be suppressed rather than there being just an attenuated response relative to the other conditions that required movements of the articulators.

Figure 2. Three networks controlling speech production. **A**, Color-coded overlays of three separate analyses placed on sagittal T_1 -weighted brain slices taken from a single subject template. Distance from the midline is indicated in millimeters. Top row, Left; bottom row, right. **B**, Plots of mean effect size, with 95% confidence intervals, from specific regions of peak activity. In red is the conjunction of activity during the jaw, tongue, speech, and count conditions, each individually contrasted with the rest condition. In both cerebral hemispheres, common activity distributed along primary sensorimotor cortex within the central sulcus is contiguous with activity along the STP. **B1** shows the mean effect size at the peak of activity within the left and right medial planum temporale. In blue is the conjunction of activity in the speech and count conditions contrasted separately with the jaw and tongue conditions. **B2** shows the dissociation of activity between the overt speech conditions and the silent conditions involving movements of the jaw and tongue. In yellow is the contrast of the speech and count conditions, demonstrating lateral neocortical activity during propositional speech production. Activity was widely distributed along the length of the left STS, extending ventrally in the posterior left temporal lobe. Although there was bilateral activity, at the statistical threshold used there was evident asymmetry, left $\gg$ right. Activity was also extensively distributed throughout the left inferior frontal gyrus, including all of Broca's area. On the right, there was a small focus of activity in the caudal right lateral orbitofrontal cortex. **B3** shows the left lateralized response of the anterior inferior frontal gyrus to propositional speech. IFG, Inferior frontal gyrus.

Speech contrasted with count

This contrast of propositional with nonpropositional speech demonstrated widely distributed medial and lateral activity, corresponding to the activity observed in our studies of narrative speech production using positron emission tomography (PET) (Blank et al., 2002; Awad et al., 2007). The lateral neocortical activity is described here and is illustrated in Figure 2.

There were three features of note. The first was that activity associated with the speech condition alone spread into ventral left lateral temporal neocortex, and this extended in both caudal and rostral directions, including widespread activity throughout the extensive area 37 of Brodmann (that encompasses the caudal middle temporal, inferior temporal, and fusiform gyri). The second was that activity specific to the speech condition was observed in the left inferior frontal gyrus. The main peak was within the rostral part of Broca's area (BA 45), but there were widely

distributed subpeaks in caudal Broca's area (BA 44), the lateral orbitofrontal cortex just ventral to Broca's area, and the rostral insula. This activity was associated with activity in midline (lobule IV–V) and lateral right (Crus 1) cerebellum.

The third feature was that activity in the left inferior frontal gyrus appeared to be strongly left lateralized, in contrast to the symmetrically distributed activity within the STP and lateral STG. The left lateralization was confirmed by formal analysis. Data were extracted from an ROI centered on peak activity in the left BA 45 and the mirror voxel on the right (MNI coordinates: $x = -52$ and 52 , $y = 26$, $z = 18$). The mean effect sizes for the count and speech conditions, relative to the rest condition, were entered into a 2 (hemisphere) $\times$ 2 (condition) ANOVA. There was a weak main effect of hemisphere ($F_{(1,20)} = 6.4$; $p < 0.05$) and a strong main effect of condition ($F_{(1,20)} = 62.7$; $p < 0.001$), with a strong hemisphere $\times$ condition interaction ($F_{(3,60)} = 200.2$; $p <$

Figure 3. Activity relating to jaw and tongue movements. **A**, Color overlay of the conjunction of activity in the jaw and tongue conditions, each contrasted separately with the (speech + count) conditions. The overlay is placed on sagittal and coronal T₁-weighted brain slices taken from a single subject template. There was symmetrical activation in three areas: 1, lateral premotor cortex; 2, the Rolandic operculum at the most ventral extreme of the central sulcus, extending into the dorsal insular cortex; and 3, SII within the parietal operculum. **B**, The plots show mean activity across conditions at peak voxels in left (−64, −28, 30) and right (58, −18, 28) SII, with 95% confidence intervals, for all four conditions relative to the rest condition. There was significant activity during both the jaw and tongue conditions. Activity during the count condition was no different from that during the rest condition, whereas during the speech condition activity was significantly less (suppressed).

0.001). *Post hoc* paired *t* tests on the within-condition differences between the hemispheres demonstrated significance for both the count (right > left, $t_{(20)} = 4.3$; $p < 0.001$) and speech (left > right, $t_{(20)} = 7.1$; $p < 0.001$) conditions; that is, there was a reversal of interhemispheric asymmetry between the two conditions, with no between-condition difference in right BA 45 ($t_{(20)} = 1.4$; $p > 0.1$).

Figure 2 demonstrated that activity in right as well as left BA 37 was present, although apparently weaker on the right. A similar analysis on ROI data from left and right BA 37 (MNI coordinates: $x = -44$ and 44 , $y = -32$, $z = -20$) demonstrated a weak main effect of hemisphere ($F_{(1,20)} = 6.0$; $p < 0.05$), a strong main effect of condition ($F_{(3,60)} = 47.9$; $p < 0.001$), and a weak hemisphere $\times$ condition interaction ($F_{(3,60)} = 5.3$; $p < 0.05$). *Post hoc* paired *t* tests on the within-condition differences between the hemispheres demonstrated no hemisphere asymmetry for the count condition ($t_{(20)} = 1.5$; $p > 0.1$) but an asymmetry for the speech condition (left > right, $t_{(20)} = 2.8$; $p = 0.01$). Although, as predicted from the analysis within SPM5, there was a strong between-condition effect in the left hemisphere (speech > count, $t_{(20)} = 11.5$; $p < 0.001$), there was also a significant between-condition effect in the right hemisphere (speech > count, $t_{(20)} = 3.0$; $p < 0.01$). However, this was the result of suppression of regional activity during the count condition relative to the rest

condition (one-sample *t* test, $t_{(20)} = 2.3$; $p < 0.05$), whereas regional activity during the speech condition was no different from the rest condition (one-sample *t* test, $t_{(20)} < 0.1$; $p > 0.9$). These somewhat complex, and regionally rather different, interactions between activity during the count, speech, and rest conditions, all indicate that left, but not right, BA 45 and BA 37 responded to the speech condition.

Discussion

This fMRI study, using two overt speech conditions and silent nonspeech movements of two of the principal articulators, identified four separate subsystems within “unimodal” somatosensory and auditory cortical areas and heteromodal temporal cortex. The responses of these subsystems across conditions indicated their roles in the control of speech production. Three were symmetrically distributed between the hemispheres, in cortex adjacent to the lateral (Sylvian) sulcus. One additional, predominantly left-lateralized subsystem was identified, distributed along the superior temporal sulcus (STS) and part of ventrolateral temporal cortex and extending into the left inferior frontal gyrus.

Task-dependent propositional speech production, nonpropositional speech (counting), and movements associated with speech production (the mandibular cycle of jaw opening and closure and placing of the tongue tip on the upper alveolar ridge) all activated, as expected, premotor cortex (including the supplementary motor area), bilateral primary sensorimotor cortex, and bilateral cerebellar regions (Blank et al., 2002; Riecker et al., 2005;

Bohland and Guenther, 2006; Awad et al., 2007). Within motor cortex, partial somatotopy was evident for the cortical control of respiration and the larynx, and the tongue and jaw movements. A notable feature was the absence of common activity within SII, although all four tasks will have generated somatosensory feedback. Additional analysis demonstrated that, within SII, there was a strong response to repetitive tongue and jaw movements. In contrast, during counting, activity was no greater than during a rest state, and, during propositional speech, activity within SII was suppressed relative to rest. This same pattern of responses was evident in mid-insular (dysgranular cortex). It is known that the cortical motor area for the larynx in the monkey projects to insular cortex (Simonyan and Jürgens, 2005). Furthermore, dysgranular insular cortex connects directly with SII (Flynn et al., 1999). It appears that dysgranular insular cortex forms part of the distributed system monitoring articulatory movements, but, as in SII, the response is suppressed during normal speech.

Within a sensory cortical area, the relative contributions of feedforward predictive “corollary discharges” from motor regions and feedback sensory discharges cannot be determined with the temporal resolution afforded by fMRI. Nevertheless, the response within SII indicated processing of the somatosensory consequences of meaningless movements of the articulators, pro-

cessing that appeared absent during normal speech. Although these results do not exclude somatosensory self-monitoring during speech by a minority of neurons within SII, with suppression of activity in other neurons resulting in a null or negative response relative to the rest state, the profile of activity across conditions within SII contrasts sharply with that observed along the left and right STP. Despite an absence of auditory feedback during the nonspeech jaw and tongue movements, there was a common response to all four conditions. This activity encompassed the caudal STP (the planum temporale in the human), reaching medially to the depth of lateral sulcus, in which second-order auditory association cortex abuts second-order somatosensory cortex. The results from this study indicate that the STP responds to both the auditory and somatosensory consequences of movements of the articulators.

This observation is salient when interpreted in relation to the recent anatomical studies in the nonhuman primate, investigating the corticocortical (Smiley et al., 2007) and thalamocortical (Hackett et al., 2007) connections of the caudal STP. There are direct projections from primary sensory cortex to Ri, located at the fundus of the caudal lateral sulcus. Projections from Ri go to the medial part of the caudal STP, the so-called CM belt area. Area CM and the laterally adjacent caudolateral belt area are strongly connected. The result from the present study also accords with other studies, in both nonhuman primates and humans, reporting somatosensory-evoked responses within auditory cortex in response to sensory stimuli from the upper limb (Foxe et al., 2000, 2002; Schroeder et al., 2001; Fu et al., 2003; Kayser et al., 2005). Although these stimuli were externally generated, one study reported the effects on auditory neurons of discharges generated by self-initiated limb movements, consequent on trained responses to heard stimuli (Brosch et al., 2005).

A feature discordant with the recent nonhuman primate literature on the location of polymodal responses within unimodal auditory cortex is the extension of activity common across all four conditions into more rostral STP. One possibility is that the distribution of somatosensory afferents within the STP is more extensive in the human. The alternative is that the activity in the more rostral STP is driven predominantly by feedforward predictive corollary discharges (Paus et al., 1996). The latter explanation presupposes that movements of the articulators that are not intended by the subject to produce sounds nevertheless automatically generate corollary discharges to auditory cortex.

Specialization of the STP to bring into close anatomical proximity speech-related auditory and somatosensory feedback serves the purpose of closely matching the feedback signals in time. Movements of the articulators are rapid, and precise matching of sound and somatosensation in time will facilitate precision when learning to position the articulators to generate accurate speech sounds. One hypothesis is that a babbling infant activates both SII and the STP, but, as skill in speech is acquired, activity in SII becomes suppressed in favor of activity within the STP alone. A shift back to speech-related activity within SII may occur in the adult when learning a foreign language, speaking in the presence of an electronic delay in auditory feedback (Hashimoto and Sakai, 2003), and in a clinical setting, when lesions affecting the motor execution of articulation result in dysarthria. These hypotheses about speech-related activation of SII can be readily tested in future studies.

Recent single-cell recordings in nonhuman primates have presented a complex picture of auditory feedback processing in auditory cortex. The firing of many neurons in response to self-

vocalizations is suppressed, below baseline firing rates, in accordance with the human research on relative suppression of auditory cortical activity during speech. However, some of these neurons also demonstrate increased sensitivity to experimentally imposed perturbations of the auditory feedback experienced by the monkey, indicating a role in self-monitoring (Eliades and Wang, 2008). In the human, these neurons may concentrate over the lateral aspect of the left and right STG, which in the present study were areas that responded only when there was overt speech. At this site in the nonhuman primate, in which lateral belt cortex merges with parabelt cortex, neurons respond strongly to environmentally complex sounds, including externally generated conspecific vocalizations (Rauschecker and Tian, 2000). The response during overt speech observed in the present study, and in a previous PET study (Blank et al., 2002), indicates auditory processing alone in response to the feedback discharges or to a combination of feedback and corollary discharges.

It is probable that lexical retrieval corresponded to the activity that was observed in the more ventral left temporal neocortex and the inferior frontal gyrus (DeLeon et al., 2007). The demonstration that this signal was strongly left-lateralized accords with the lesion literature in terms of the localization of lexical functions during speech production. The activity throughout the caudal middle temporal, inferior temporal, and fusiform gyri (BA 37) was not observed in previous studies of speech production using PET (Blank et al., 2002; Matsumoto et al., 2004; Awad et al., 2007). Of the two factors, different scanning methodology and task, it would seem most likely that it is the task-dependent retrieval of word meaning (this study) (Vandenberghe et al., 1996) compared with free narrative speech during the recall of personal memories (the PET studies) that accounts for this difference in activity within left BA 37.

In conclusion, this study has provided new insight into the functioning of the temporoparietal junction during self-initiated speech production. The results accord with new evidence that cross-modal auditory and somatosensory processing occurs early in unimodal auditory association cortex. Previous studies predominantly investigated externally generated sensory experiences, and somatosensory signal was most commonly generated from the upper extremity. It can be conjectured that the simultaneous processing of the sounds and tactile sensations that accompany manipulating objects with the forepaws, as practiced by monkeys and apes, assists learned dexterity, which accords with the demonstration that auditory cues contribute to the response of mirror neurons (Kohler et al., 2002; Keysers et al., 2003). However, in the human, the acquisition and maintenance of fluent speech, which relies on the processing of self-generated sensations arising from the articulators, is of paramount importance. The results from the present study indicate that it is learning-related plasticity within the STP with suppression of the processing of feedback somatosensory information in the parietal operculum that supports this skill. The response of the STP during speech production was symmetrically distributed between the cerebral hemispheres. However, there is a caveat. The symmetrical physiological response cannot be used to infer symmetry of processing. The lateralization of propositional speech-related activity to higher-order lateral temporal and inferior frontal neocortex may modulate the function of the left STP, which in turn may result in asymmetrical processing of different components of the feedforward and feedback signals generated during propositional speech (Wise et al., 1999; Matsumoto et al., 2004).

References

- Awad M, Warren JE, Scott SK, Turkheimer FE, Wise RJS (2007) A common system for the comprehension and production of narrative speech. *J Neurosci* 27:11455–11464.
- Blakemore SJ (2003) Deluding the motor system. *Conscious Cogn* 12:647–655.
- Blank SC, Scott SK, Murphy K, Warburton E, Wise RJS (2002) Speech production: Wernicke, Broca and beyond. *Brain* 125:1829–1838.
- Bohland JW, Guenther FH (2006) An fMRI investigation of syllable sequence production. *Neuroimage* 32:821–841.
- Brett M, Anton JL, Valabregue R, Poline JB (2003) Region of interest analysis using an SPM toolbox [abstract]. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-ROM in *Neuroimage*, Vol. 16, No. 2, abstract 497.
- Brosch M, Selezneva E, Scheich H (2005) Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci* 25:6797–6806.
- Buchsbaum BR, Hickok G, Humphries C (2001) Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn Sci* 25:663–678.
- Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF (2005) Reading, hearing and the planum temporale. *Neuroimage* 24:444–454.
- Carreiras M, Mechelli A, Price CJ (2006) Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Hum Brain Mapp* 27:963–972.
- Curio G, Neuhold G, Numminen J, Jousmäki V, Hari R (2000) Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp* 9:183–191.
- DeLeon J, Gottesman RF, Kleinman JT, Newhart M, Davis C, Heidler-Gary J, Lee A, Hillis AE (2007) Neural regions essential for distinct cognitive processes underlying picture naming. *Brain* 130:1408–1422.
- Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567–631.
- Draper MH, Ladefoged P, Whitteridge D (1959) Respiratory muscles in speech. *J Speech Hear Res* 2:16–27.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335.
- Eickhoff SB, Amunts K, Mohlberg H, Zilles K (2006) The human parietal operculum. II. Stereotaxic maps and correlation with functional imaging results. *Cereb Cortex* 16:268–279.
- Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453:1102–1106.
- Flynn FG, Benson DF, Ardila A (1999) Anatomy of the insula: functional and clinical correlates. *Aphasiology* 13:55–78.
- Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res Cogn Brain Res* 10:77–83.
- Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM (2002) Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J Neurophysiol* 88:540–543.
- Fu KM, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2003) Auditory cortical neurons respond to somatosensory stimulation. *J Neurosci* 23:7510–7515.
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878.
- Gracco VL, Tremblay P, Pike B (2005) Imaging speech production using fMRI. *Neuroimage* 26:294–301.
- Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96:280–301.
- Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE (2007) Multisensory convergence in auditory cortex. II. Thalamic connections of the caudal superior temporal plane. *J Comp Neurol* 502:924–952.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223.
- Hashimoto Y, Sakai KL (2003) Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Hum Brain Mapp* 20:22–28.
- 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.
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci* 14:1125–1138.
- Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. *Neuron* 48:373–384.
- Keysers C, Kohler E, Umiltà MA, Nanetti L, Fogassi L, Gallese V (2003) Audiovisual mirror neurons and action recognition. *Exp Brain Res* 153:628–636.
- Kohler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297:846–848.
- Matsumoto R, Nair DR, LaPresto E, Najm I, Bingaman W, Shibusaki H, Lüders HO (2004) Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain* 127:2316–2330.
- Mehta S, Grabowski TJ, Razavi M, Eaton B, Bolinger L (2006) Analysis of speech-related variance in rapid event-related fMRI using a time-aware acquisition system. *Neuroimage* 29:1278–1293.
- Murphy K, Corfield DR, Guz A, Fink GR, Wise RJ, Harrison J, Adams L (1997) Cerebral areas associated with motor control of speech in humans. *J Appl Physiol* 83:1438–1447.
- Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236–2246.
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806.
- Riecker A, Mathiak K, Wildgruber D, Erb M, Hertrich I, Grodd W, Ackermann H (2005) fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology* 64:700–706.
- Riecker A, Kassubek J, Gröschel K, Grodd W, Ackermann H (2006) The cerebral control of speech tempo: opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. *Neuroimage* 29:46–53.
- Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, Petrides M (1999) Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage* 10:233–260.
- Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. *J Neurophysiol* 85:1322–1327.
- Simonyan K, Jürgens U (2005) Afferent cortical connections of the motor cortical larynx area in the rhesus monkey. *Neuroscience* 130:133–149.
- Smiley JF, Hackett TA, Ulbert I, Karmos G, Lakatos P, Javitt DC, Schroeder CE (2007) Multisensory convergence in auditory cortex. I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol* 502:894–923.
- Smith SM (2002) Fast robust automated brain extraction. *Hum Brain Mapp* 17:143–155.
- Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RSJ (1996) Functional anatomy of a common semantic system for words and pictures. *Nature* 383:254–256.
- Wilson M (1988) MRC psycholinguistic database: machine-usable dictionary, version 2.00. *Behav Res Methods Ins C* 20:6–10.
- Wise RJ, Greene J, Büchel C, Scott SK (1999) Brain regions involved in articulation. *Lancet* 353:1057–1061.
- Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural subsystems within “Wernicke’s area.” *Brain* 124:83–95.