CHAPTER 2

Listening to speech activates motor areas involved in speech production

2.1 Abstract

To examine the role of motor areas in speech perception, we carried out a functional magnetic resonance imaging (fMRI) study in which subjects listened passively to monosyllables, and produced the same speech sounds. Listening to speech activated bilaterally a superior part of ventral premotor cortex largely overlapping a speech production motor area centered just posteriorly on the border of Brodmann areas 4a and 6, which we distinguished from a more ventral speech production area centered in area 4p. These findings support the view that the motor system is recruited in mapping the acoustic signal to a phonetic code.

2.2 Introduction

Language depends upon the maintanence of parity between acoustic and articulatory representations: there must be a common phonetic code (Liberman et al., 1967). Phonemes are organized in terms of distinctive features which are primarily defined by articulatory properties; for instance, /*p*/ is a voiceless bilabial stop consonant. If the

common phonetic code has motor properties, then the motor system could play a role in perceiving speech, as a transformation must be carried out from the acoustic signal to a phonetic representation. Recent work on mirror neurons has revitalized interest in the idea that motor areas are involved in perceptual processes (Rizzolatti & Craighero, 2004). However most neuroimaging studies of speech perception have focused on characterizing the strong responses consistently observed in the superior temporal lobe (Binder et al., 2000; Scott & Wise, 2004). Frontal areas have sometimes been reported to be activated by passive listening to speech (Binder et al., 2000; Benson et al., 2001), and are often responsive during audiovisual speech perception (Callan et al., 2003), but the potential motor properties of the areas found in these studies have not been investigated. While Broca's area is frequently implicated in studies involving phonological tasks or syntactic comprehension, it is presumed to be important for higher levels of linguistic processing (Bookheimer, 2002). Two studies employing transcranial magnetic stimulation (TMS) have shown facilitation of tongue (Fadiga et al., 2002) and lip (Watkins et al., 2003) muscles when subjects listened to speech, however the precise areas involved are not known due to limitations of spatial localization with TMS.

2.3 Materials and Methods

We carried out an fMRI experiment to examine whether passive listening to meaningless monosyllables would activate motor areas involved in producing speech. 10 subjects listened to 16-second blocks containing 23 repetitions of meaningless monosyllables. During the same scanning sessions, subjects were cued to produce the same syllables.

The motor tasks were performed for just 3 seconds; to avoid movement artifacts, we discarded volumes acquired during actual motor activity and analyzed subsequent volumes in which the delayed hemodynamic response occurred. 8 of the 10 subjects also listened to blocks of control nonspeech stimuli: a burst of white noise or a bell, and carried out a bimanual motor task.

2.3.1 Subjects and experimental design

After giving informed consent, 10 subjects (mean age: 27 years; 4 females) took part in the study. All participants had normal or corrected-to-normal vision. Two of the subjects were authors of the study; the remainder were naive as to the purpose of the study.

For 8 of the subjects, there were 3 runs, each of which was 606 seconds in duration. In each run, there were 12 blocks of auditory stimuli followed by 15 blocks in which participants were cued to perform motor tasks. The other 2 subjects (7 and 8) were run on a shorter version of the experiment which did not include nonspeech stimuli or finger movements. The descriptions of the experimental design below report the version of the experiment on which 8 subjects were run; the paradigm for the other 2 subjects was similar except that only a subset of the conditions were tested.

The auditory blocks were 16 seconds long and were separated by rest periods varying randomly in duration from 8 to 16 seconds with a mean of 12 seconds. During each block, 23 evenly spaced tokens of the same sound were presented. There were 4 sounds, thus 3 blocks of each per run, which were presented in random order. The sounds were (1) a male speaker producing the syllable /*pa*/; (2) a male speaker producing /*gi*/; (3) a

burst of white noise; (4) the sound of a bell. The actual vowels used were [ʌ] in /*pa*/ and [ɪ] in /*gi*/, therefore both stimuli were phonotactically possible English syllables, but due to the short vowels they were not phonotactically possible words. The syllables were recorded in a soundproof booth at UCLA, the white noise was generated with MATLAB (Mathworks, Natick, MA), and the bell sound was selected from a CD of environmental sounds. The sounds were edited with Audacity (http://audacity.sourceforge.net) in order to match them for duration and subjective loudness (determined in a norming study). The sounds were each 310 ms in duration and were presented using scanner-compatible headphones at a level chosen by each individual subject. Subjects were asked to choose a volume which was as loud as possible, so as to be heard over the scanner noise, without being uncomfortable. During the auditory component of each run, subjects maintained fixation on a white cross presented against a black background, which was either projected on a screen which they viewed through a mirror (subjects 1 through 8) or viewed through scanner-compatible goggles (subjects 9 and 10). Stimuli were presented with MATLAB using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

During the rest period following the final auditory block, the cross turned green to indicate to the subjects that motor blocks would begin in 10 seconds. There were three types of motor blocks, in which subjects were cued by means of visually presented commands ('say PA', 'say GI' and 'fingers') to carry out the following actions for three seconds: (1) produce the syllable /*pa*/ repeatedly; (2) produce the syllable /*gi*/ repeatedly; (3) alternate the thumbs back and forth between the four fingers on both hands simultaneously. In the linguistic conditions, subjects were instructed to move the jaw as

little as possible, so as to avoid movement artifacts. After 3 seconds, another cue ('STOP!'), displayed for 1 second, instructed subjects to stop moving. This was followed by rest periods of 12 to 16 seconds, with a mean of 14 seconds. The motor blocks were presented in a fixed sequence rather than randomly so that subjects could concentrate on producing the actions with minimal head movement without having to devote any attention to selecting the proper movement.

Subjects were instructed to remain absolutely still during the listening phase, and as still as possible during the motor phase. In particular, they were told to keep their face still. This was done in order to reduce the possibility of covert articulation during the listening blocks. Covert articulation is unlikely to provide an explanation for any motor cortical activity observed, as a prior study used electromyography (EMG) to demonstrate that participants' tongue muscles were absolutely relaxed as they listened to speech (Fadiga et al., 2002). Whereas covert articulation in reading and memorization tasks has been shown to lead to measurable phoneme-specific EMG responses (McGuigan & Winstead, 1974), similar EMG responses during listening to speech have only been reported in one extreme, unreplicated case (McGuigan, 1973) and are not generally observed (McGuigan, 1979), suggesting that covert articulation does not occur when subjects listen to speech.

2.3.2 Image acquisition

For the first 8 subjects, images were acquired on a 3 Tesla Varian scanner at the UCSD Center for Functional Magnetic Resonance Imaging. For subjects 9 and 10, images were

acquired on a 3 Tesla Siemens Allegra scanner at the Ahmanson-Lovelace Brain Mapping Center at UCLA. In each of the three runs, 305 functional volumes were acquired using a whole head EPI sequence (TR = 2.0 s, TE = 27.4 ms, flip angle = 90°, 30 axial slices with interleaved acquisition, 3.75 × 3.75 × 3.80 mm resolution, field of view = 240 × 240 × 114 mm). The first two volumes were discarded in order to allow the magnetization to reach steady state. For subjects run on the Varian scanner, a B_0 field map (a single set of multi-echo EPI images) was collected at the beginning of each scanning session and used to estimate the local B_0 field. This estimate was then used to correct displacements in the phase-encode direction (Reber et al., 1998). Following the last functional run, anatomical images were acquired using a magnetization-prepared rapid gradient echo (MPRAGE) sequence. For one subject, 88 consecutive volumes during the motor phase of the second run were excluded due to signal instability, but this still left ample data for mapping of motor areas.

2.3.3 Image analysis

Image analysis was carried out primarily with AFNI (Cox, 1996). After discarding the initial two volumes, each volume was registered to an image in the middle of the third run (toward the end of the listening phase), saving the 3 translation parameters and 3 rotation parameters to be used as regressors. The functional images were smoothed with a 4 mm FWHM Gaussian filter, then global signal intensity changes were calculated to be used as regressors. For subjects 1 to 8, the anatomical images were acquired in-plane with the functional images, so registration with the functional images was performed by

nudging anatomical images in the x and y planes by a few millimeters so as to obtain optimal overlap with the reference functional image. For subjects 9 and 10, the functional reference images were registered with the anatomical images using the FSL program FLIRT (Jenkinson & Smith, 2001).

A general linear model was fit to the concatenated data from the three runs at each voxel with the AFNI program 3dDeconvolve. The model contained 21 terms which modeled each of the 7 conditions at three different lags (1 to 3 TRs). This method makes no assumptions about the shape of the hemodynamic response, which allows for flexibility in fitting responses both across voxels and across conditions. In addition, there were 18 terms to take out slow drifts (6 for each of the 3 runs), 18 motion-related terms (6 for each of the 3 runs), and 3 global signal intensity terms (1 for each run). *F* tests for particular conditions versus baseline tested whether the sum of coefficients at the three lags differed from zero. Similarly, *F* tests involving multiple conditions tested whether the sums or differences of their summed coefficients differed from zero. No consistent differences were observed between activations for the syllables /*pa*/ and /*gi*/ in either listening or motor conditions, nor between the two nonspeech sounds, so these pairs of conditions were each collapsed together in subsequent analyses.

In order to reduce artifacts due to motion in the motor blocks, each pair of adjacent volumes (i.e. 4 s) during which speech production took place were excluded from analysis. Due to the lag of the hemodynamic response, this did not seriously affect power to detect motor-related activations, but it did avoid some artifactual activations, especially around the edges of the brain. Several studies have demonstrated the feasibility

of taking advantage of the fact that hemodynamic responses are delayed relative to movement-related artifacts in analyzing designs which entail task-correlated movement (Barch et al., 1999; Birn et al., 1999; Huang et al., 2001).

Anatomical images were registered to standard MNI space with FLIRT using 12-parameter affine transformations. Statistical images were resampled onto a 1 mm grid using trilinear interpolation. Clusters were defined as contiguous sets of voxels (with nearest neighbor connectivity) activated at $p < 10^{-4}$ uncorrected for listening tasks or $p < 10^{-12}$ for motor tasks. Clusters were considered significant if they were at least 300 mm^3 in volume, and only clusters of this size or larger appear in the figures. This would correspond to 5.6 original functional voxels in a brain the same size as the MNI template brain. Deactivations are not shown or reported. Centers of mass of activated clusters were calculated with the AFNI program 3dclust, based on fitted coefficients, which are linearly related to percent signal change. For speech production, several peaks are reported rather than centers of mass, since these clusters were found to contain multiple peaks in consistent locations. In the 4 (of 20) hemispheres in which two peaks could not be clearly identified, a ventral peak was selected deep in the CS, and a dorsal peak was selected around the border of the PrCG and CS, as these were the invariant locations of peaks in hemispheres where two peaks could be identified. Some subjects had additional peaks in dorsal premotor cortex. Speech production activations in inferior ventral premotor cortex were contiguous with the main PrCG/CS clusters in some subjects, but not in others, as has been previously reported (Birn et al., 1999).

Regions of interest (ROIs) for the time series shown in Figure 2.2a were defined for each subject. ROIs were defined as contiguous clusters of voxels, in original functional space, located in the PrCG or CS and responsive to listening to speech at $p < 10^{-4}$ or to listening to nonspeech sounds at $p < 10^{-4}$, with a minimum cluster size of 2 voxels (in original functional space). In most subjects, voxels responsive to nonspeech sounds were a subset of those responsive to speech. For each subject, mean hemodynamic responses were calculated for ROIs in the left and right hemispheres, then averaged together. Finally the time courses were averaged together across subjects, and the standard error of the mean calculated at each time point.

Likelihoods of centers of mass falling into particular Brodmann areas were calculated based on probabilistic cytoarchitectonic maps (Geyer et al., 1996; Geyer, 2004). For a given Brodmann area, these maps show for each voxel the percentage of subjects for whom that voxel lay in the given Brodmann area. All maps were smoothed with a 4 mm FWHM Gaussian filter. Probabilities were then read off the maps for several locations of interest from Table 2.1. Probabilities in the left and right hemispheres were similar for the locations reported. The color-coded map in Figure 2.2c was made by assigning opacity in each voxel based on the sum of the probabilities for the three areas, with maximum opacity obtained at 50%, then assigning the overlay color in RGB space by setting R, G and B values according to probabilities for areas 6, 4a and 4p respectively, with 50% probability corresponding to maximum intensity in each channel.

Percent signal changes by condition in voxels of interest were calculated as follows. Voxels of interest were selected for each individual subject: the center of mass for

listening to speech, and the BA 4a/6 and 4p peaks for speech production. Time courses were then extracted for these voxels, and mean hemodynamic responses calculated by averaging across blocks. Smoothed data were used so as to allow surrounding voxels some influence on the results. Peak signal change for listening blocks was defined as the average signal change between 4 and 10 seconds after stimulus onset, whereas for motor blocks, because of the shorter block length, the period between 4 and 8 seconds was used.

2.4 Results

In all 10 subjects, regions in the precentral gyrus (PrCG) extending into the anterior bank of the central sulcus (CS) were significantly activated by listening to speech, in comparison to rest (Figure 2.1). These activations were located primarily in the superior part of ventral premotor cortex (vPMC), extending toward primary motor cortex (see below). Activations were bilateral in 4 subjects, left-lateralized in 2 and right-lateralized in 4, however, at lower thresholds or with reduced minimum cluster sizes, responses could be seen to be bilateral for all subjects. This bilaterality is consistent on the one hand with the bilateral motor control of speech production (Fox et al., 2001), and on the other hand with the bilateral superior temporal responses regularly observed in speech perception studies (Scott & Wise, 2004). Other activated areas are reported in Table 2.1.

Areas in the PrCG and CS activated by production of the same syllables are shown in Figure 2.1 outlined in black (see also Table 2.1). Motor responses were bilateral

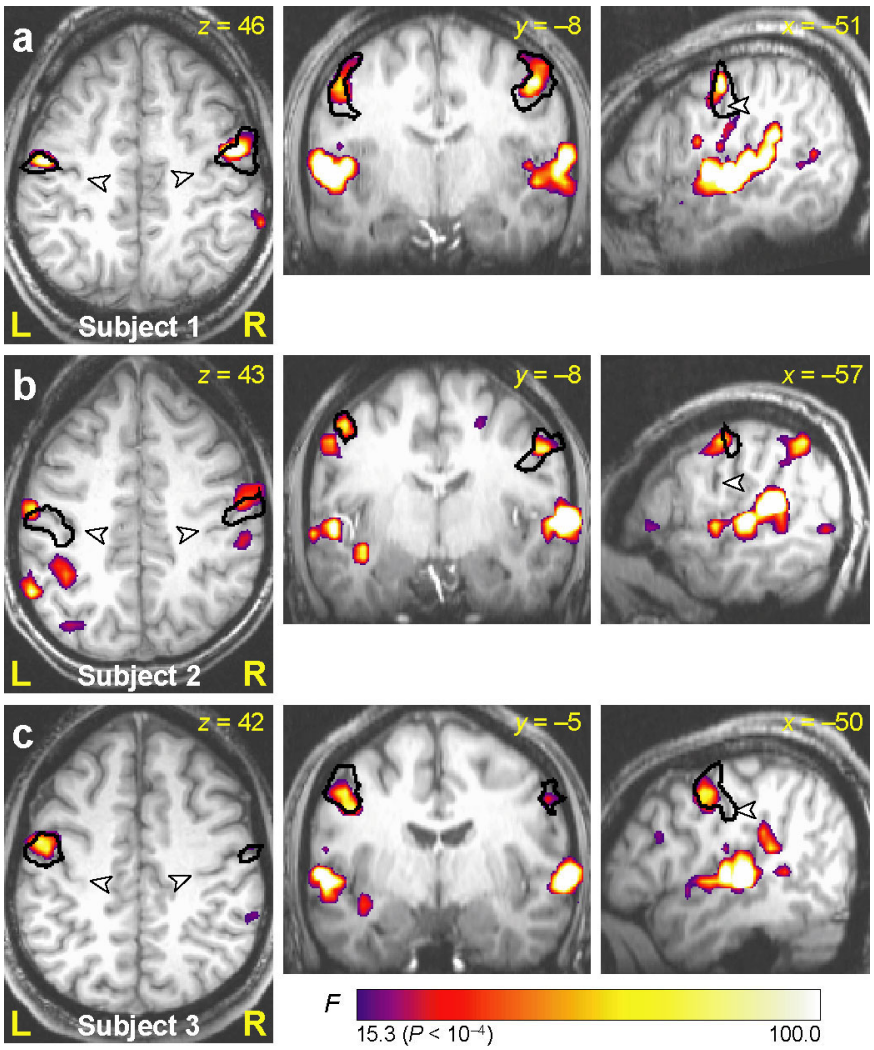

Figure 2.1 Areas activated by passive listening to meaningless monosyllables in three representative subjects. Individual activation maps were thresholded at $p < 10^{-4}$ (uncorrected) for listening conditions or $p < 10^{-12}$ for motor conditions, with a minimum cluster size of 300 mm^3. Mean MNI coordinates for centers of mass of areas activated by listening to speech were (–50, –6, 47) (left) and (55, –3, 45) (right). The black outlines show premotor and primary motor cortical activity while producing the same syllables. Other areas were also activated in the motor conditions, but are not shown. Arrowheads show the location of the central sulcus. Motor areas activated by both speech perception and production were observed in every subject. Robust responses in the superior temporal gyrus can also be observed in the coronal and sagittal views.

Table 2.1 Areas activated by listening to speech in 6 or more subjects, and PrCG/CS motor areas activated by producing speech or bimanual movement.

Area	Brodmann area(s)	Number of subjects	MNI coordinates (mm)			Mean extent (mm^3)
			x	*y*	*z*	
Listening to speech						
Left PrCG/CS	6, 4a	6	–50	–6	47	1516
Right PrCG/CS	6, 4a	8	55	–3	45	935
Left STG+	22, 41, 42	10	–54	–22	5	24025
Right STG+	22, 41, 42	10	59	–19	5	23501
Right SMG	40	8	53	–38	50	2469
Producing speech						
Left PrCG/CS	4a/6	10	–51	–11	46	7210
	4p		–45	–13	34	(total)
	6		–56	–4	22	
Right PrCG/CS	4a/6	10	56	–8	44	6227
	4p		48	–10	35	(total)
	6		60	0	20	
Moving fingers						
Left PrCG/CS	4, 6	10	–38	–22	58	19935
Right PrCG/CS	4, 6	10	39	–21	59	18176

Note. All 10 subjects had PrCG/CS activity for listening to speech in one or both hemispheres. The minimum cluster size for an area to count as activated was 300 mm^3, but clusters which did not meet the minimum cluster size were still used in calculating the mean coordinates and extents. For producing speech and moving the fingers, other areas were activated besides these premotor and primary motor areas, but they are not shown here. All coordinates refer to centers of mass (based on signal change) except those for producing speech which are voxels with peak signal change, since centers of mass cannot be readily calculated for overlapping areas. For listening to speech, the responses in the superior temporal gyrus and surrounding areas (STG+) in all 10 subjects reflect auditory and prelexical processing, and the activations in the right supramarginal gyrus (SMG) may relate to auditory attention. It is noteworthy that less consistent responses were also observed in several other motor-related areas: 5 subjects showed significant

activations in the left inferior frontal gyrus, pars opercularis, i.e. the posterior part of Broca's area; 4 subjects showed activity in this same region in the right hemisphere; and the supplementary motor area (SMA) was activated in 3 subjects, with a further 5 subjects showing similar SMA activations which did not reach the minimum cluster size.

---

in all subjects. Comparison of the regions activated by listening to and producing the syllables revealed substantial overlap for all subjects. Across subjects, 73 ± 7% of voxels in PrCG/CS regions activated by listening to speech were also activated by speech production. We defined regions of interest (ROIs) for each subject consisting of clusters of voxels in the PrCG or CS which were responsive to either listening to speech or listening to nonspeech sounds, and plotted the mean time course (Figure 2.2a). This plot demonstrates a robust response to speech production, confirming that this is a speech production area, and further shows that this region responds more strongly to speech than nonspeech sounds, though the response to nonspeech stimuli does exceed baseline.

Closer examination of speech production activations revealed several distinct peaks within each cluster. In 16 of 20 hemispheres we observed a ventral peak with $30 \leq z \leq 39$ and a dorsal peak with $40 \leq z \leq 50$ (except in 2 hemispheres where this peak was a few mm more dorsal). The ventral peaks were located deep in the central sulcus, whereas the dorsal peaks lay more laterally on the anterior lip of the sulcus (Figure 2.2b); activations always spanned the CS, but distinct sensory peaks were never observed. Based on probabilistic cytoarchitectonic maps (Geyer et al., 1996; Geyer, 2004), the ventral peaks were located in Brodmann Area (BA) 4p, whereas the dorsal peaks lay on the border of

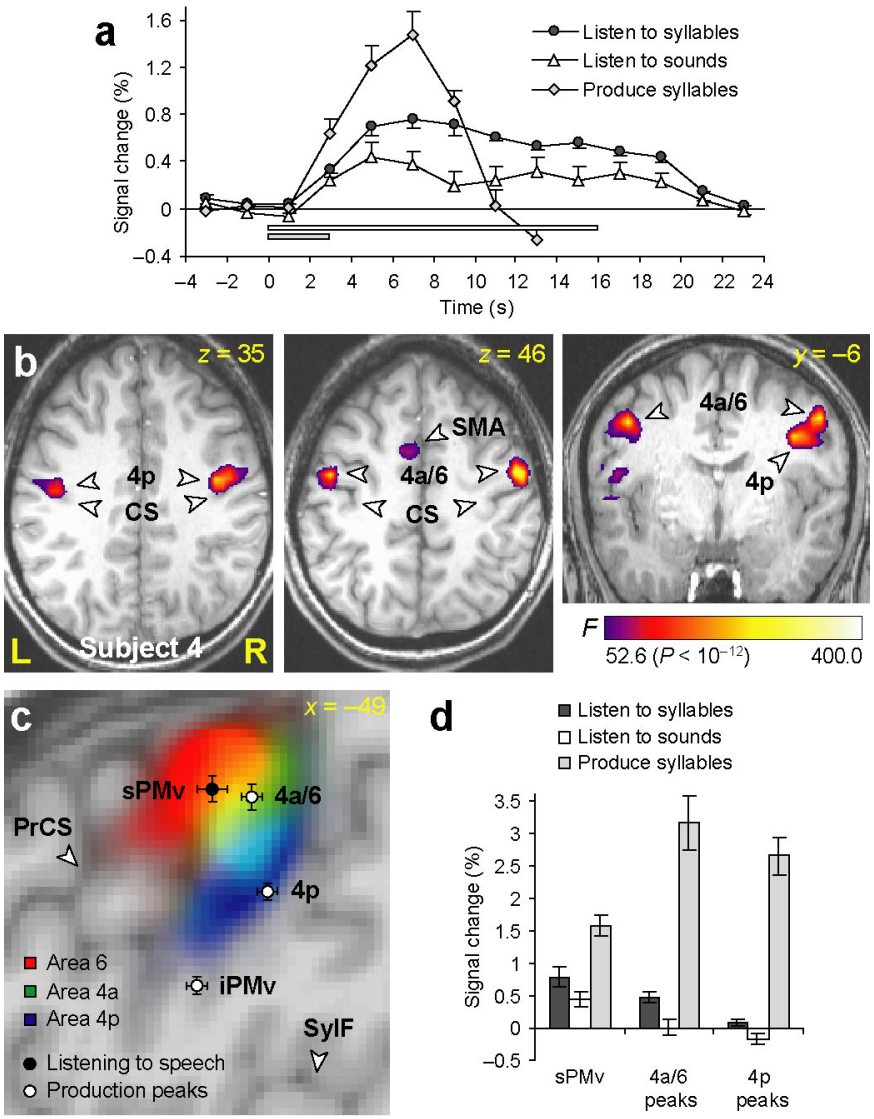

Figure 2.2 Characterization of the relationships between listening and motor areas. (**a**) Hemodynamic responses in PrCG/CS regions which were responsive to listening to either speech or nonspeech at $p < 10^{-4}$. The bars show the time periods during which auditory stimuli were presented (16 s) and during which subjects were producing speech (3 s). Error bars here and in other panels represent s.e.m. (**b**) Motor areas for speech production in BAs 4p and 4a/6 in a representative subject. In the coronal view, two distinct peaks can be seen in the right hemisphere, but in the left hemisphere, area 4p cannot be seen as it is posterior to the plane. Across subjects, the mean MNI coordinates of the 4p peaks were (–45, –13, 34) (left) and (48, –10, 35) (right), and of the 4a/6 peaks (–51, –11, 46) (left) and (56, –8, 44) (right). CS: central

sulcus; SMA: supplementary motor area. (**c**) The locations of listening and production peaks overlaid on probabilistic cytoarchitectonic maps of BAs 4a, 4p and 6 in the left hemisphere. The peaks were within 2 mm of this plane (x = –49) except for the BA 4p production peak which was 4 mm medial, and the inferior vPMC peak which was 7 mm lateral. A similar map for the right hemisphere appeared very similar. svPMC: superior ventral premotor cortex; ivPMC: inferior ventral premotor cortex; PrCS: precentral sulcus; Syl F: Sylvian fissure. (**d**) Maximum percent signal change by condition for three of the peaks from panel **c**.

BAs 4a and 6 (Figure 2.2c). Previous imaging studies of speech production have not distinguished two areas, and have generally reported group-averaged peak coordinates which lie between the two peaks we observed (Fox et al., 2001).

In relation to these peaks for speech production, the mean center of mass for listening to speech was located 4.5 ± 0.7 mm anterior to the BA 4a/6 production peak (p = 0.0005), but not significantly medial, lateral, superior or inferior to it (all ps > 0.05). This slightly anterior location means that it falls most likely in BA 6, in the superior part of vPMC (Rizzolatti & Craighero, 2004) (Figure 2.2c). Note that an inferior region of vPMC is also involved in speech production (Fox et al., 2001); it was activated bilaterally in all subjects and its location is shown in Figure 2.2c. However, this inferior region did not respond to listening to speech. We next examined listening responses in peak production voxels (Figure 2.2d). Peak BA 4a/6 voxels responded significantly to listening to speech, but not to nonspeech sounds, suggesting that the area activated by listening to speech may extend into the most anterior part of primary motor cortex. In contrast, peak BA 4p voxels did not respond in either listening condition. This functional distinction may be analogous to the recently reported greater involvement of BA 4a than BA 4p in motor imagery

(Ehrsson et al., 2003). The bimanual task allowed us to confirm that these motor areas are specific to the mouth. Finger/hand motor areas responsive to the bimanual task were located significantly medial, posterior and superior to the regions activated for listening to and producing speech (Table 2.1).

2.5 Discussion

These findings are consistent with the view that speech perception involves the motor system in a process of auditory to articulatory mapping in order to access a phonetic code with motor properties (Liberman et al., 1967). Whether the superior vPMC region we identified is necessary for normal speech perception is not known. Frontal lesions can severely compromise speech perception (Blumstein et al., 1977a, 1977b), however precisely which lesions lead to perception deficits is not clear. Besides responding robustly to speech, the superior vPMC region also exhibited diminished responses to nonspeech sounds. It has been argued that premotor cortex is involved in coding environmental features in a body-based but highly abstract form (Schubotz & Von Cramon, 2004). Premotor cortex may attempt to represent all auditory (and other sensory) stimuli in a body-referenced code, and it is possible that the level of activity may reflect the extent to which stimuli have clear motor correlates.