

Listening to talking faces: motor cortical activation during speech perception

Jeremy I. Skipper,^{a,b,*} Howard C. Nusbaum,^a and Steven L. Small^{a,b}

^aDepartment of Psychology, The University of Chicago, Chicago, IL 60637, USA

^bDepartment of Neurology and the Brain Research Imaging Center, The University of Chicago, MC 2030, 5841 South Maryland Avenue, Chicago, IL 60637, USA

Received 20 May 2004; revised 28 October 2004; accepted 3 November 2004

Available online 8 January 2005

Neurophysiological research suggests that understanding the actions of others harnesses neural circuits that would be used to produce those actions directly. We used fMRI to examine brain areas active during language comprehension in which the speaker was seen and heard while talking (audiovisual) or heard but not seen (audio-alone) or when the speaker was seen talking with the audio track removed (video-alone). We found that audiovisual speech perception activated a network of brain regions that included cortical motor areas involved in planning and executing speech production and areas subserving proprioception related to speech production. These regions included the posterior part of the superior temporal gyrus and sulcus, the pars opercularis, premotor cortex, adjacent primary motor cortex, somatosensory cortex, and the cerebellum. Activity in premotor cortex and posterior superior temporal gyrus and sulcus was modulated by the amount of visually distinguishable phonemes in the stories. None of these regions was activated to the same extent in the audio- or video-alone conditions. These results suggest that integrating observed facial movements into the speech perception process involves a network of multimodal brain regions associated with speech production and that these areas contribute less to speech perception when only auditory signals are present. This distributed network could participate in recognition processing by interpreting visual information about mouth movements as phonetic information based on motor commands that could have generated those movements.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Talking faces; Motor cortical activation; Speech perception

Introduction

When somebody speaks to us, how do we integrate observed lip and mouth movements with the ongoing process of speech perception? Recent neurophysiological evidence from nonhuman

primates suggests that regions traditionally considered to be responsible for motor planning and motor control play a role in perception and comprehension of action (Graziano and Gandhi, 2000; Romanski and Goldman-Rakic, 2002). In addition, some neurons with visual and/or auditory and motor properties in these regions discharge both when an action is performed and during perception of another person performing the same action (Gallese et al., 1996; Kohler et al., 2002; Rizzolatti et al., 1996). In the macaque brain, these neurons reside in area F5, which is the proposed homologue of Broca's area, the classic speech production region of the human (Rizzolatti et al., 2002).

The existence of these "mirror neurons" suggests a hypothesis that actions may be understood in part by the same neural circuits that are used in action performance. Indeed, although many neuroimaging studies of passive listening to auditory speech do not report activity in the motor system (for a review, see Small and Burton, 2001), several recent experiments that do not focus exclusively on passive auditory listening support this hypothesis. Specifically, perception has been shown to overlap production in motor areas using monosyllables (Wilson et al., 2004) and multisyllabic pseudo-words (Buchsbaum et al., 2001), and during story comprehension and verb generation (Papathanassiou et al., 2000). Furthermore, motor-evoked potentials increase in the tongue when listeners hear words that, if produced, would require tongue movement (Fadiga et al., 2002) and increase in the lips while listening to or watching speech relative to non-speech controls (Watkins et al., 2003).

Audiovisual speech perception might also be aided in the context of face-to-face interaction by activation of cortical areas that are more typically associated with speech production. Seeing facial motor behaviors corresponding to speech production (e.g., lip and mouth movements) might aid language understanding by recognition of the intended gesture within the motor system, thus further constraining possible interpretations of the intended message.

Indeed, though we are capable of comprehending auditory speech without any visual input, observation of articulatory movements produces significant effects on comprehension throughout the lifespan. Infants are sensitive to various characteristics of audiovisual speech (Kuhl and Meltzoff, 1982; Patterson and Werker, 2003). By adulthood, the availability of visual information

* Corresponding author. Department of Neurology and the Brain Research Imaging Center, The University of Chicago, MC 2030, 5841 South Maryland Avenue, Chicago, IL 60637, USA. Fax: +1 773 834 7610.

E-mail address: skipper@uchicago.edu (J.I. Skipper).

Available online on ScienceDirect (www.sciencedirect.com).

about speech production significantly enhances recognition of speech sounds in background noise (Grant and Seitz, 2000; Sumbly and Pollack, 1954) and improves comprehension even when the auditory speech signal is clear (Reisberg et al., 1987). Furthermore, incongruent audiovisual information can change the identity of a speech percept (Massaro, 1998; McGurk and MacDonald, 1976).

Our experience as talkers and as listeners may associate the acoustic patterns of speech with motor planning and proprioceptive and visual information about accompanying mouth movements and facial expressions. Thus, experience reinforces the relationships among acoustic, visual, and proprioceptive sensory patterns and between sensory patterns and motor control of articulation, so that speech becomes an “embodied signal” (Wilson, 2002), rather than simply an auditory signal. That is, perceptual information relevant to the phonetic interpretation of speech may derive partly from experience with articulatory movements that are generated by a motor plan during speech production. The mechanisms that mediate these associations could provide a neural account for some of the observed interactions between acoustic and visual information in speech perception that may not be apparent by studying acoustic speech perception alone.

The participation of multimodal brain areas critical for language production during audiovisual speech perception has not been fully explored. It may be that the observed effects on speech comprehension produced by observation of a speaker’s face involve only visual cortical areas or other multisensory areas (e.g., posterior superior temporal sulcus), and not areas involved in production. The evidence from nonhuman primates (Gallese et al., 1996; Kohler et al., 2002; Rizzolatti et al., 1996), however, suggests that production centers in concert with other brain regions are likely candidates for the neural structures mediating these behavioral findings.

Most neuroimaging studies of speech processing incorporating visible face movements have focused on the problem of determining specific sites of multisensory integration (Calvert et al., 2000; Mottonen et al., 2002; Olson et al., 2002; Sams et al., 1991; Surguladze et al., 2001) or sites important for speech (i.e., lip) reading (Calvert and Campbell, 2003; Calvert et al., 1997; Campbell et al., 2001; Ludman et al., 2000; MacSweeney et al., 2000, 2001, 2002a; Surguladze et al., 2001). Collectively, these experiments have yielded a fairly consistent result: integration, speech reading, and audiovisual speech perception produce activation of auditory cortices, most notably posterior superior temporal gyrus and superior temporal sulcus, consistent with the interpretation of this region’s role in biological motion perception (Allison et al., 2000). Though some studies have reported activation in motor and sensory areas important for speech production (e.g., MacSweeney et al., 2002b), there has not been much theoretical interpretation of these activations. This may be in part because some studies use tasks that require an explicit motor response (e.g., Calvert et al., 1997; MacSweeney et al., 2002b; Olson et al., 2002), which could limit the inferences that can be drawn about the role of these motor areas in perception (Small and Nusbaum, 2004). It would be surprising, however, if brain regions important for language production (e.g., Broca’s area and the precentral gyrus and sulcus) did not play a role in audiovisual speech perception, given the connectivity between frontal and parietal and superior temporal structures (Barbas and Pandya, 1989; Hackett et al., 1999; Petrides and Pandya, 1988, 2002; Romanski et al., 1999) and the multisensory sensitivity of frontal areas (Graziano and Gandhi, 2000; Kohler et al., 2002; Romanski and Goldman-Rakic, 2002) in nonhuman primates.

In the present study, we used fMRI to investigate whether audiovisual language comprehension activates a network of brain regions that are also involved in speech production and whether this network is sensitive to visual characteristics of observed speech. We also investigated whether auditory language comprehension alone, without visual information about the mouth movements accompanying speech production, would activate the same motor regions, as it has long been proposed that speech perception whether multimodal or unimodal occurs by reference to the speech production system (e.g., Liberman and Mattingly, 1985). Finally, we investigated whether the visual observation of the mouth movements accompanying connected discourse activate this network even without the auditory speech signal. In the latter condition, participants were not asked to try to understand the discourse by speech-reading nor would they be able to.

Materials and methods

Twelve participants were recruited from a student population. Three participants were not used in the analyses because of technical problems during the experimental session (i.e., head movement, stimulus presentation failure, and failure to complete the scanning sequence). The remaining nine participants (5 females; mean age = 25; SD = 8) were right handed as determined by the Edinburgh handedness inventory (Oldfield, 1971), had normal hearing, and normal uncorrected vision. The participants gave written consent, and the Institutional Review Board of the Biological Science Division of The University of Chicago approved the study.

In an audio-alone condition (A), participants listened to spoken stories. In an audiovisual condition (AV), participants watched and listened to high-resolution video clips of the storyteller, filmed from the neck up, telling the same stories. In a video-alone (V) condition, participants watched video clips of the storyteller telling these stories, but without the accompanying sound track. The stories were highly engaging, and participants were simply asked to attend to them. No overt motor response was required during scanning and participants were not explicitly asked to attempt to speech (i.e., lip) read in the V condition. In all, participants were presented 28 stories told by a single storyteller. Story duration ranged from 18 to 24 s. There were 24–30 s of rest (i.e., baseline) between each story during which participants remained fixated on the screen in front of them. The stories were repeated in each the AV, A, and V conditions and were counterbalanced. Each participant received a different ordering of the stimuli. Audio stimuli were delivered to participants at 85 dB SPL through headphones containing MRI-compatible electromechanical transducers (Resonance Technologies, Inc., Northridge, CA). Participants viewed stimuli through a mirror that allowed them to see a projection screen at the end of the scanning bed where video clips were shown.

After completion of the scanning session, participants were interviewed about the stimuli and they reported understanding and being engaged by the AV and A stories. They answered specific questions about the events that occurred in the stories though they were not instructed that they would be doing so. All participants accurately reported details from stories that they found interesting, indicating that they attended to the stories. No participants reported understanding any of the V stories though two participants reported seeing mouth movements that they thought they recognized as speech sounds.

Table 1
Regions of interest (ROIs) and their anatomical and functional characteristics

ROIs	Anatomical boundaries
<i>Frontal</i>	
F3o, pars opercularis of the inferior frontal gyrus	Brodmann's area 44
Anterior	anterior ascending ramus of the sylvian fissure
Posterior	precentral sulcus
Ventral	posterior horizontal ramus of the sylvian fissure
Dorsal	inferior frontal sulcus
F3t, pars triangularis of the inferior frontal gyrus	Brodmann's area 45
Anterior	a coronal plane defined as the rostral end of the anterior horizontal ramus of the sylvian fissure
Posterior	anterior ascending ramus of the sylvian fissure
Ventral	anterior horizontal ramus of the sylvian fissure
Dorsal	inferior frontal sulcus
PMd, dorsal precentral gyrus excluding the primary motor	Brodmann's area 6
Anterior	posterior border of the superior frontal sulcus
Posterior	anterior half of the precentral gyrus
Ventral	a plane dividing the dorsal 2/3 from the ventral 1/3 of the precentral gyrus
Dorsal	the most superior point of the precentral sulcus
PMv, ventral precentral gyrus excluding the primary motor	Brodmann's area 6
Anterior	precentral sulcus
Posterior	anterior half of the precentral gyrus
Ventral	posterior horizontal ramus of the sylvian fissure to the border with insula cortex
Dorsal	a plane dividing the dorsal 2/3 from the ventral 1/3 of the precentral gyrus
<i>Parietal</i>	
SGp-AG, posterior supramarginal gyrus-angular gyrus	Posterior Brodmann's area 40, 39
Anterior	a coronal plane defined as the endpoint of the sylvian fissure
Posterior	anterior occipital sulcus to a plane defined by a point where the parietooccipital sulcus cuts into the dorsal hemispheric margin
Ventral	superior temporal sulcus (upper bank) and anterior occipital sulcus
Dorsal	intraparietal sulcus
<i>Temporal</i>	
T1a, anterior superior temporal gyrus	Anterior Brodmann's area 22
Anterior	a coronal plane defined as the first coronal slice containing the temporofrontal junction
Posterior	Heschl's sulcus
Ventral	superior temporal sulcus (upper bank)
Dorsal	posterior horizontal ramus of the sylvian fissure
T1p, posterior superior temporal gyrus	Posterior Brodmann's area 22, 40, 41, 42
Anterior	Heschl's sulcus
Posterior	a coronal plane defined as the endpoint of the sylvian fissure

Table 1 (continued)

ROIs	Anatomical boundaries
<i>Temporal</i>	
Ventral	superior temporal sulcus (upper bank)
Dorsal	posterior horizontal ramus of the sylvian fissure
<i>Temporooccipital</i>	
TO2-OL	Brodmann's area 37, 17,18,19
TO2, temporooccipital part of middle temporal gyrus	
Anterior	a coronal plane defined as the endpoint of the sylvian fissure
Posterior	a coronal plane defined by a point where the parietooccipital sulcus cuts into the dorsal hemispheric margin
Ventral	superior temporal sulcus (lower bank) and anterior occipital sulcus
Dorsal	intraparietal sulcus
OL, occipital lobe	
Anterior	a coronal plane defined by a point where the parietooccipital sulcus cuts into the dorsal hemispheric margin
Posterior	hemispheric margin
Ventral	hemispheric margin
Dorsal	hemispheric margin

Scans were acquired on a 1.5-T scanner using spiral acquisition (Noll et al., 1995) with a standard head coil and volumetric T1-weighted scans (124 axial slices, $1.5 \times 0.938 \times 0.938$ mm resolution) were acquired to provide anatomical images on which landmarks could be found and on which functional activation maps could be superimposed. Twenty-four 6-mm spiral gradient echo T2* functional images were collected every 3 s in the axial plane. A total of 224 whole brain images were collected in each of four runs. Images were spatially registered in three-dimensional space by Fourier transformation of each of the time points and corrected for head movement, using the AFNI software package (Cox, 1996). Effective in-plane resolution was $1.875 \times 1.875 \times 6$ mm.

Functional imaging data was analyzed using multiple linear regression. Regressors were waveforms with similarity to the hemodynamic response, generated by convolving a gamma-variant function with the onset time and duration of the blocks of interest. There were three such regressors for each the AV, A, and V conditions. The remaining regressors for both groups were the mean, linear, and quadratic component of each of the functional runs. For stereotaxic group analyses, anatomical and functional images were interpolated to volumes with 2-mm^3 voxels, co-registered, converted to Talairach stereotaxic coordinate space (Talairach and Tournoux, 1988), and smoothed with a 4-mm Gaussian full-width half-maximum filter to decrease spatial noise. A voxel-wise mixed-effects two-factor analysis of variance (condition \times participant, where participants are considered a random sample) was applied to the normalized regression coefficients from the regression analysis. An activated voxel from the ANOVA was defined by an individual voxel probability less than 2.0×10^{-5} and a minimum cluster connection radius of 2.1 mm and cluster size of 10 μl . These thresholds were established using the AlphaSim component of AFNI and are based on 10,000 Monte Carlo simulations and cluster size thresholding, resulting in an overall corrected significance level of alpha less than 0.05. The same alpha level was used for statistical contrasts.

The stereotaxic group analysis is based on registering the different patterns of activity onto a single reference anatomy (Talairach and Tournoux, 1988). Despite its utility, it is useful to supplement such data with more fine-grained anatomical investigations. To this end, regions of interest (ROIs) were chosen and drawn onto each hemisphere of each participant's high-resolution structural MRI scan. These ROIs were adapted from an MRI-based parcellation system (Caviness et al., 1996; Rademacher et al., 1992). Table 1 describes the ROIs, their anatomical boundaries, and their functional properties. Some ROIs were chosen because they have been previously implicated in multimodal integration and speech production (see discussion). These were the pars opercularis of the IFG (F3o), pars triangularis of the IFG (F3t), the dorsal two-thirds (PMd) and ventral one-third (PMv) of the precentral gyrus, the posterior aspect of the STG and the upper bank of the STS (T1p), and the posterior aspect of the supramarginal gyrus and the angular gyrus (SGp-AG). We were particularly interested in F3o because the distribution of "mirror neurons" is hypothesized to be greatest in this area (Rizzolatti et al., 2002). Several regions were included in the analysis to control for the possibility that all differences between conditions were in the same direction and the same magnitude, thus diminishing the significance of any differences uncovered in the regions of interest. These regions were the anterior aspect of the STG/STS (T1a) and an ROI that encompassed the occipital lobe and temporal-occipital visual association cortex (including the lower bank of the posterior STS; TO2-OL). T1a was drawn with the hypothesis that activation in this area would be more closely associated with processing of connected discourse (Humphries et al., 2001) and therefore would not differ between the AV and A conditions. TO2-OL was drawn with the hypothesis that activity in this region would reflect visual processing and should not be active in the A condition. After delimiting these regions, we determined the total volume of activation and associated intensity within each ROI for each condition for each participant. We collected all significant voxels for each task compared to baseline, i.e., voxels exceeding the threshold of $z > 3.28$, $P < 0.001$ corrected as determined by a false-discovery rate algorithm using the 3dFDR component of AFNI. To determine the difference between conditions, we compared the volume and intensity of activity across participants for the AV and A or V conditions within each ROI using paired t tests correcting for multiple comparisons ($P < 0.004$ unless otherwise stated).

We also asked whether any of cortical areas that are activated by visual information are specifically activated by the phonetic information in mouth movements. In each video presentation, there were a variety of observable "nonlinguistic" (e.g., head nods, eye blinks) and "linguistic" (e.g., place of articulation) movements produced by the talker. Some of the latter conveyed phonetic feature information, though most mouth movements by themselves are not sufficient for phonetic classification. A subset of visual speech movements, however, called "visemes", have been shown to be sufficient (i.e., without the accompanying auditory modality) for phonetic classification (Goldschen, 1993; Jackson, 1988; Preminger et al., 1998). Do visemes, in contrast to other observable information about face and head movements in the AV stories, modulate activation in any of the ROIs we examined? If the motor system activity during AV stories is in service of understanding the speech, this activity should be modulated by visual information that is informative about phonetic features: Variation in the amount of visemes within a story should relate to the amount of observed

motor system activity. All stories were phonetically transcribed and grouped into quartiles according to the proportion of visemes relative to the total number of phonemes (visible and nonvisible) in each story. This proportion was 69.2%, 70.9%, 71.4%, and 75.3% for the first, second, third, and fourth quartiles, respectively. Stories in the first and fourth [$t(6) = 23.97$, $P < 0.00001$] and the first and third [$t(6) = 13.86$, $P < 0.00001$] quartiles significantly differed in the proportion of visemes, indicating that participants were exposed to differential visual information about phonetic production in these stories independent of the actual phonetic content. The volume and intensity of brain activity were compared in ROIs for the AV, A, and V condition between the first and second, third, or fourth viseme quartiles.

Results

Stereotaxic-based results

Relative to baseline (i.e., rest), the AV but not the A or V conditions activated a network of frontal brain regions (Table 2; Fig. 1). These areas include the inferior frontal gyrus [IFG; Brodmann areas (BA) 44 and 45, roughly corresponding to the pars opercularis and pars triangularis, respectively], the precentral gyrus and sulcus (BA 4 and 6), the postcentral gyrus, and the cerebellum. Of these regions, the A condition activated only a cluster in the anterior IFG (BA 45).

We were concerned that the high threshold used to correct for multiple comparisons could be responsible for the failure to observe activity in these areas in the A condition. Thus, we also examined activation patterns relative to baseline at a reduced threshold [$t(16) = 4$, single voxel $P = 0.0001$]. At this uncorrected threshold, the AV condition appeared to have more activation than the A condition in the right IFG (especially in BA 44) and aspects of the right precentral gyrus (BA 4 and 6). The mean Talairach coordinate in the right precentral gyrus for activation associated with the A condition was: $-49, 0, 39$. The AV condition and not the A condition activated bilateral aspects of more posterior STG/STS, left IFG (BA 44), the left precentral gyrus and sulcus (BA 4 and 6), and the right cerebellum. In addition, the V condition showed a more robust pattern of activation, including the IFG and precentral gyrus.

Relative to baseline, the AV condition appeared to activate more posterior aspects of the STG/STS (BA 22) relative to either the A or V conditions. That is, though both the AV and A conditions activated the STG/STS (BA 41/42/22) bilaterally, activation in the AV condition was more extensive and extended more posterior from the transverse temporal gyrus than activation in the A condition. The AV and V conditions activated visual cortices (BA 18, 19, 20, and 21) and the A condition did not. The V condition, however, only activated small clusters in the inferior occipital gyrus and the inferior temporal gyrus relative to baseline whereas the AV condition activated more extensive regions of occipital cortex as well as the left fusiform gyrus (BA 18).

In the statistical contrast of the AV and A conditions (AV–A; Table 3; Fig. 2), the AV condition produced greater bilateral activation in the IFG (BA 44, and 45), the precentral gyrus and sulcus (BA 4 and 6), the left cerebellum, the posterior superior temporal gyrus and sulcus, and visual areas. The A condition produced greater activation in the middle and superior frontal gyrus,

Table 2

Location, center of mass, and amount of cortical activity in significant [$t(16) = 7.2$, single voxel $P = 0.000002$, $P < 0.05$ corrected] clusters as shown by the audiovisual (AV), audio-alone (A), and video-alone (V) conditions relative to baseline for the group

	Brodmann area	Hemisphere	Audiovisual				Audio-alone				Visual-alone			
			Talairach coordinates			Cluster size	Talairach coordinates			Cluster size	Talairach coordinates			Cluster size
			x	y	z		x	y	z		x	y	z	
<i>Occipito-temporal</i>														
Fusiform gyrus	18	Left	21	91	−13	368								
Inferior temporal gyrus	21	Left	66	11	−18	16					66	11	−18	16
Inferior/middle occipital gyrus	19	Right	−44	70	−4	40								
Inferior/middle occipital gyrus	19	Right									−43	70	−4	16
Inferior temporal gyrus	20	Right									−43	12	−32	112
<i>Temporal and parietal</i>														
Superior temporal gyrus	21/22/41/42	Left	53	21	6	6552	54	15	5	2880				
Superior temporal gyrus	41	Left					36	32	16	72				
Superior temporal gyrus	13	Left					45	19	9	16				
Anterior superior temporal gyrus	21/38	Left	52	−7	−9	80								
Posterior superior temporal gyrus	22/39	Left	57	49	14	136								
Posterior superior temporal gyrus	22	Left	61	48	18	56	56	36	8	344				
Posterior superior temporal gyrus	22	Left					47	43	10	24				
Posterior superior temporal gyrus	22	Left	53	48	19	24								
Posterior superior temporal gyrus	22	Left	60	54	20	24								
Middle temporal gyrus	21	Left					51	7	−10	24				
Inferior parietal lobule	40	Left	52	45	22	24								
Superior temporal gyrus	21/22/41/42	Right	−54	18	3	5168	−55	15	5	2840				
Superior temporal gyrus	22	Right					−50	33	6	16				
<i>Frontal</i>														
Inferior frontal gyrus	9/44/45	Left	51	−19	21	1208								
Inferior frontal gyrus	9/45	Left					50	−20	22	440				
Inferior frontal gyrus	45	Left	46	−31	3	32								
Inferior frontal gyrus	46	Left	49	−26	14	16								
Precentral and postcentral gyrus	3/4/6	Left	53	10	45	104								
Precentral gyrus	6	Left	49	2	48	16								
Superior frontal gyrus	9	Left	8	−55	30	64								
Superior frontal gyrus	6	Left	5	−6	58	16								
Superior frontal gyrus	6	Left	7	−15	61	16								
Inferior frontal gyrus	44/45	Right	−55	−23	16	288								
Inferior frontal gyrus	45	Right	−54	−28	4	32								
<i>Extra-pyramidal</i>														
Thalamus/pulvinar		Left	10	29	0	160								
Putamen/lentiform nucleus		Left	26	−3	12	160	25	−4	11	24				
Cerebellum		Right	−24	72	35	136								

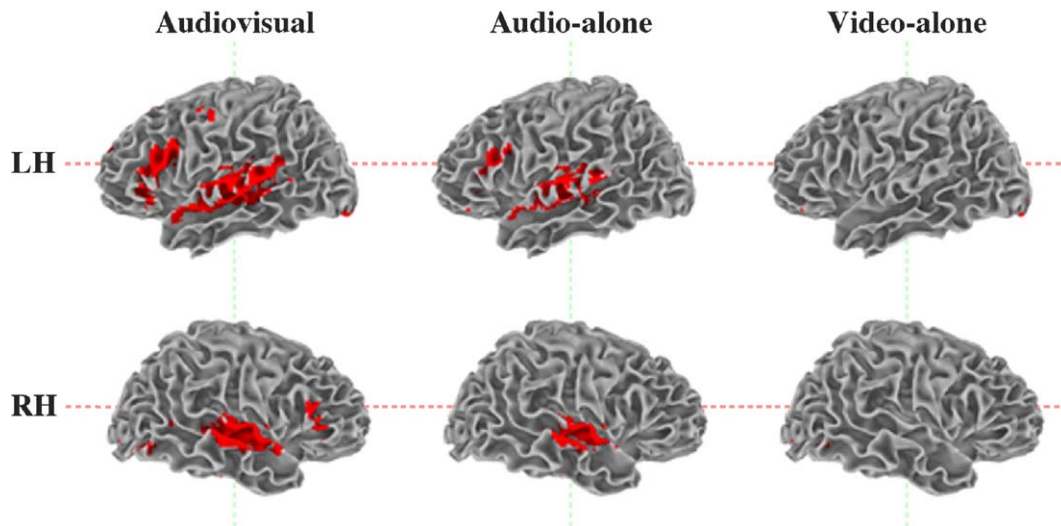


Fig. 1. Cortical activation produced by the audiovisual, audio-alone, and video-alone conditions relative to baseline for the group [$t(16) = 7.2$, single voxel $P = 0.000002$, $P < 0.05$ corrected]. Activation, in red, is projected onto the surface of the brain of a single person.

cingulate gyrus, and parietal lobules, including the postcentral gyrus. The contrast of the AV and V conditions (AV–V; Table 4; Fig. 2) showed that AV produced greater activation in most frontal areas with the exception of the MFG, and the right IFG (BA 44) for which V produced greater activation. V also produced greater activation bilaterally in the superior parietal lobule.

Region-of-interest-based results

AV relative to baseline differed from A relative to baseline in the total amount of activation in ROIs chosen because they have been implicated in multimodal integration and speech production. ROIs differed in total volume of activation in left PMd [$t(8) = 5.19$], right PMd [$t(8) = 3.70$], left F3o [$t(8) = 4.06$], left F3t [$t(8) = 3.54$], left T1p [$t(8) = 4.12$], and right T1p [$t(8) = 4.45$]. There was no significant difference in the right F3o, right F3t, and bilateral SGp-AG. The conditions differ in the intensity of activation in right F3t [$t(8) = 3.68$] and right T1p [$t(8) = 3.59$]. There were no significant differences in bilateral T1a. Finally, the AV and A conditions differed in the volume and intensity of activation in left TO2-OL [volume = $t(8) = 3.45$; intensity = $t(8) = 4.21$], and right TO2-OL [volume = $t(8) = 3.74$; intensity = $t(8) = 3.89$].

The intensity but not volume of activity significantly increased in the AV condition relative to baseline as a function of the number of visemes, in some of the ROIs. When comparing the first and fourth viseme quartiles intensity increased in the right T1p [$t(8) = 1.89$, $P < 0.05$] and right PMd [$t(8) = 2.81$, $P < 0.01$]. The first and third quartiles also differed for the AV condition in four areas, T1a [$t(8) = 2.55$, $P < 0.02$], left T1p [$t(8) = 3.38$, $P < 0.005$], right T1p [$t(8) = 4.26$, $P < 0.002$], and right PMd [$t(8) = 2.42$, $P < 0.02$]. There were no differences between the third and fourth quartiles. As for the A condition, none of the regions significantly differed when comparing the first to fourth quartiles. Only T1a [$t(8) = 2.79$, $P < 0.01$] and right SGp-AG [$t(8) = 2.21$, $P < 0.03$] differed for the A condition when comparing the first and third quartiles. There were no differences between the first and second quartiles for either the AV or A conditions. There were not differences between the first and second, third, or fourth quartiles for the V condition.

Discussion

The present results show that audiovisual language comprehension activates, in addition to commonly reported superior temporal regions, an extensive network that comprises the pars opercularis of the inferior frontal gyrus (roughly BA 44 of Broca's area), aspects of the precentral gyrus, including both premotor and primary motor cortices, the postcentral gyrus, and the right cerebellum. In these areas, there was little activation for either auditory comprehension of speech or seeing a talker produce speech. The audio-alone and video-alone conditions, however, showed activation of the inferior frontal gyrus and premotor cortex at a reduced threshold and the visual-alone condition activated the pars opercularis as shown by the AV–V contrast. In the discussion that follows, we argue that these brain areas are involved in both sensory and motor aspects of speech production.

Activation of these areas during audiovisual but not audio-alone language comprehension or to a much reduced extent during audio-alone language comprehension cannot be attributed simply to an overly conservative correction for multiple comparisons. Nor are these results likely attributable to differences in speech comprehensibility across conditions. Participants reported attending to and understanding details of the stories in both the AV and A conditions. In addition, previous research has shown that certain areas (e.g., the IFG) associated with language comprehension show greater activity with increasing difficulty of sentence comprehension (Just et al., 1996). If the A condition was more difficult to understand than the AV condition, then we would expect to see greater activity in these areas during the audio-alone condition, but we did not.

When considering the results of the audio-only condition, we attribute the lack or reduction of activity in cortical areas associated with speech production to the fact that under normal conditions listeners can process language solely in terms of its acoustic properties (cf. Klatt, 1979; Stevens and Blumstein, 1981) and thus may not need to recruit the motor system to understand speech (cf. Liberman and Mattingly, 1985). This aspect of our results is consistent with previous functional imaging studies in which passive listening to auditory stimuli does not reliably elicit Broca's area, premotor, or primary motor activation whereas overt phonetic

Table 3

Location, center of mass, and amount of cortical activity in significant [$t(16) = 7.2$, single voxel $P = 0.000002$, $P < 0.05$ corrected] clusters as shown by the audiovisual vs. audio-alone contrast for the group

	Brodmann area	Hemisphere	Audiovisual > audio-alone				Audio-alone > audiovisual			
			Talairach coordinates			Cluster size	Talairach coordinates			Cluster size
			x	y	z		x	y	z	
<i>Occipito-temporal</i>										
Cuneus	17/18	Left					17	71	22	288
Middle occipital gyrus	19	Left	49	70	−6	16				
Middle occipital gyrus	18	Left					11	91	17	40
Precuneus	31	Left					21	64	25	40
Precuneus	7	Left					13	43	47	16
Cuneus	18	Right					−18	86	19	120
Lingual gyrus	17/18/19	Left/Right	−11	76	−6	30,840				
Lingual gyrus	18	Right	−19	73	−10	96				
Middle occipital gyrus	37	Right	−57	62	−7	24				
Precuneus	7/18/19/30	Right					−1	66	23	63,960
Precuneus	7	Right					−29	49	50	736
<i>Temporal and parietal</i>										
Inferior parietal lobule	7	Left					32	45	51	960
Inferior parietal lobule	40	Left					44	35	47	24
Middle temporal gyrus	19/37/39	Left	49	67	10	920				
Postcentral gyrus	2/40	Left					46	31	36	304
Postcentral gyrus	40	Left					42	33	49	24
Postcentral gyrus	3	Left					16	36	65	880
Superior temporal gyrus	38	Left	36	−15	−20	976				
Superior temporal gyrus	22/40	Left	49	43	15	344				
Inferior parietal lobule	40	Right					−44	36	38	1560
Inferior parietal lobule	40	Right					−49	42	48	536
Middle temporal gyrus	21	Right	−60	57	0	56				
Postcentral gyrus	3	Right					−20	28	61	104
Superior parietal lobule	7	Right					−29	57	50	24
Superior parietal lobule	7	Right					−36	61	46	16
Supramarginal gyrus	40	Right					−56	44	37	72
<i>Frontal</i>										
Anterior cingulate	10/32	Left	4	−43	−6	1072				
Cingulate gyrus	24	Left					3	6	46	376
Cingulate gyrus	32	Left					4	−27	32	352
Inferior frontal gyrus	44/45	Left	46	−20	6	896				
Middle frontal gyrus	8/9	Left	48	−21	35	416				
Middle frontal gyrus	10	Left	42	−53	9	96				
Middle frontal gyrus	6	Left					20	15	60	600
Anterior cingulate	24	Right	−3	−34	−2	16				
Cingulate gyrus	31	Right					−7	29	43	664
Inferior frontal gyrus	47	Right	−27	−28	−10	448				
Inferior frontal gyrus	44/45	Right	−52	−24	16	2496				
Inferior frontal gyrus	44	Right	−48	−13	19	176				
Middle frontal gyrus	47	Right	−49	−39	−3	16				
Middle frontal gyrus	6	Right					−23	3	52	2488
Middle frontal gyrus	10	Right					−25	−53	10	1048
Precentral gyrus	4/6	Right	−49	1	44	416				
Precentral gyrus	6	Right	−48	−2	35	72				
Superior frontal gyrus	8	Right	−12	−50	38	368				
Superior frontal gyrus	6/8	Right	−9	−26	53	264				
Superior frontal gyrus	9	Right					−30	−33	33	2128
<i>Extra-pyramidal</i>										
Caudate and caudate body		Left	10	3	16	632				
Cerebellum		Left	3	48	−28	272				
Thalamus		Left	12	29	−1	1136				
Cerebellum		Right	−37	45	−35	288				
Cerebellum		Right	−1	44	−4	448				
Medial geniculum body		Right	−16	23	−3	2520				
Thalamus and medial dorsal nucleus		Right	−4	9	10	928				

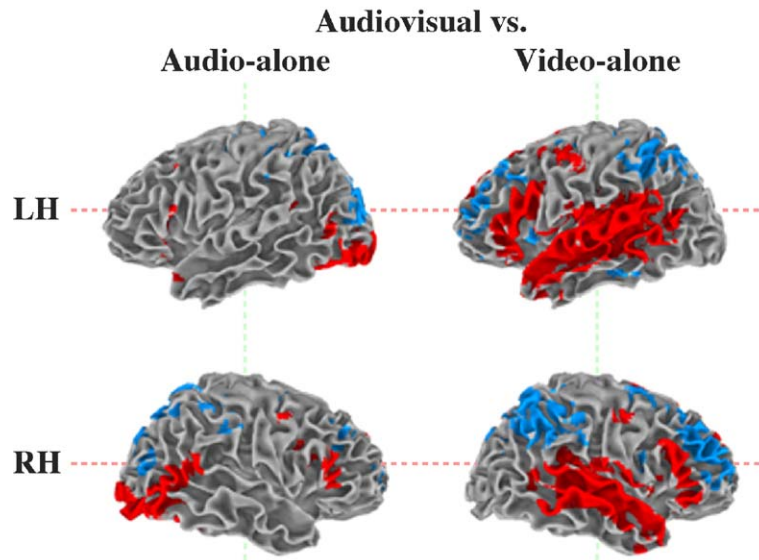


Fig. 2. Cortical activation produced by the contrast of the audiovisual condition with audio-alone and video-alone conditions for the group [$t(16) = 7.2$, single voxel $P = 0.000002$, $P < 0.05$ corrected]. Activation, in red, denotes that activity was higher in the audiovisual condition while blue denotes higher activity in the other conditions. Activation, in red, is projected onto the surface of the brain of a single person.

decisions among other overt tasks do (for a review, see Small and Burton, 2001). These overt linguistic decision tasks may engage parts of the brain involved in language production through covert rehearsal and/or working memory (e.g., Jonides et al., 1998; Smith and Jonides, 1999). In typical listening environments, however, the production system may only be weakly or transiently involved in auditory language comprehension, suggesting why it may not always be evident in imaging studies that rely on temporal averaging. Indeed, recent research shows that production-related areas can be engaged during auditory listening alone (Fadiga et al., 2002; Watkins et al., 2003; Wilson et al., 2004). Specifically, Wilson et al. (2004) have shown that passive listening to syllables activates a superior portion of ventral premotor cortex that overlaps production. This region corresponds to what we have called dorsal premotor cortex and which we have shown to be active at a reduced threshold in the auditory alone condition. Nonetheless, superior ventral premotor cortex is only a subset of the greater number of regions activated when the mouth can be seen. Our interpretation of these results is that audiovisual speech activates areas traditionally associated with both speech production and speech comprehension to encode observed mouth movements and to integrate them into the overall process of understanding spoken language. Below we show the productive nature of the cortical areas active in audiovisual speech perception and then relate these to a theoretical framework being developed to understand the functional neuroanatomy of speech perception (Hickok and Poeppel, 2000).

Broca's area

Broca's area was significantly active during both audiovisual and audio-alone language comprehension. This activity was primarily restricted to the pars triangularis (roughly BA 45) in the A condition though the pars opercularis (roughly BA 44) was active at a reduced uncorrected threshold. Broca's area is traditionally viewed as supporting a mechanism by which phonological forms are coded into articulatory forms (Geschwind, 1965). It is commonly activated during both overt and covert speech produc-

tion (Friederici et al., 2000; Grafton et al., 1997; Huang et al., 2001; Papathanassiou et al., 2000). Results of production studies, however, seem to suggest that Broca's area is not itself involved in controlling articulation per se (Bookheimer et al., 1995; Huang et al., 2001; Wise et al., 1999), but may be a "pre-articulatory" region (Blank et al., 2002). In support of this, naming is interrupted in fewer than 36% of patients stimulated at the posterior aspect of the inferior frontal gyrus (Ojemann et al., 1989). Furthermore, lesions restricted to Broca's area are clinically associated with Broca's aphasia for only a few days (Knopman et al., 1983; Masdeu and O'Hara, 1983; Mohr et al., 1978) and the role of Broca's area in producing Broca's aphasia is unclear (Dronkers, 1996, 1998). Further supporting the notion that Broca's area is not involved in controlling articulation per se is that activation in this area is not specific to oral speech, as Broca's area is activated during production of American Sign Language (Braun et al., 2001; Corina et al., 1999) and is activated by the observation and imitation of nonlinguistic but meaningful goal-directed movements (Binkofski et al., 2000; Ehrsson et al., 2000; Grezes et al., 1999; Hermsdörfer et al., 2001; Iacoboni et al., 1999; Koski et al., 2002). Nor does activation of Broca's area in nonlinguistic domains simply represent covert verbal coding of the tasks given to participants (Heiser et al., 2003).

This brief review suggests the Broca's area, though playing a role in speech production, is not simply a speech production area but rather, given its functional properties, is a general-purpose mechanism for relating multimodal perception and action. This review also suggests that we need to make finer distinctions in the functional neuroanatomy of Broca's area to understand its role in both speech comprehension and production. We distinguished between the pars triangularis (roughly BA 45) and the pars opercularis (roughly BA 44). We postulate that the common activation of the pars triangularis in both audiovisual and auditory language comprehension may reflect semantic or memory processing related to discourse comprehension in the two conditions (Devlin et al., 2003; Friederici et al., 2000; Gabrieli et al., 1998), and may not be related to cortical systems playing a role in speech

Table 4

Location, center of mass, and amount of cortical activity in significant [$t(16) = 7.2$, single voxel $P = 0.000002$, $P < 0.05$ corrected] clusters as shown by the audiovisual vs. video-alone contrast for the group

	Brodmann area	Hemisphere	Audiovisual > video-alone				Video-alone > audiovisual			
			Talairach coordinates			Cluster size	Talairach coordinates			Cluster size
			x	y	z		x	y	z	
<i>Occipito-temporal</i>										
Fusiform gyrus	20/30	Left	37	42	−13	448				
Fusiform/parahippocampal gyrus	20/37	Left	35	32	−19	232				
Inferior temporal gyrus	20	Left	45	3	−27	48				
Precuneus	7/31	Left	6	57	30	1288				
Cuneus	18	Left					11	84	20	64
Cuneus	19	Left					17	82	34	56
Middle temporal gyrus	20/37	Left					56	35	−14	848
Fusiform gyrus	37	Right	−48	53	−18	48				
Fusiform gyrus	20	Right	−37	11	−22	216				
Precuneus	31	Right	−10	53	33	280				
Cuneus	19	Right					−23	84	35	616
Precuneus	7/19/31/40	Right					−5	54	44	47,448
<i>Temporal and parietal</i>										
Postcentral gyrus	3	Left	44	23	40	88				
Superior temporal gyrus	20/21/22/38/39/40/41	Left	51	23	5	49,424				
Superior parietal lobule	7	Left					18	61	63	360
Inferior parietal lobule	40	Right	−60	44	22	48				
Superior temporal gyrus	20/21/22/38/39/40/41	Right	−52	18	2	33,992				
Postcentral gyrus	5/7	Right					−33	44	60	160
<i>Frontal</i>										
Cingulate gyrus	32	Left	18	−4	41	64				
Inferior frontal gyrus	9/44/45/47	Left	47	−20	14	12,136				
Precentral gyrus	4/6	Left	43	4	48	2112				
Superior frontal gyrus	8	Left	9	−29	54	472				
Superior frontal gyrus	6	Left	6	−8	60	1384				
Superior frontal gyrus	9	Left	2	−50	33	6856				
Middle frontal gyrus	6	Left					25	3	56	288
Insula	13	Left					35	−13	7	856
Middle frontal gyrus	11/47	Left					26	−37	−9	664
Middle frontal gyrus	9	Left					31	−41	27	8296
Inferior frontal gyrus	9/44/45/47	Right	−51	−23	13	5064				
Inferior frontal gyrus	47	Right	−34	−31	−5	264				
Insula	13	Right	−31	−2	14	440				
Medial frontal gyrus	6	Right	−8	12	65	224				
Medial frontal gyrus	6	Right	−10	−3	62	792				
Precentral gyrus	4	Right	−51	10	43	608				
Superior frontal gyrus	6	Right	−12	−27	51	560				
Cingulate gyrus	32	Right					−2	−29	27	10,032
Inferior frontal gyrus	9/44	Right					−49	−3	20	336
Middle frontal gyrus	8/9	Right					−33	−33	31	16,968
<i>Extra-pyramidal</i>										
Caudate and caudate head		Right					−11	−16	−4	528
Cerebellum		Left/right	−6	56	−33	12,624				

production or decoding mouth movements. As one moves more posterior along the gyrus toward the pars opercularis, however, functions tend to be more closely related to production (Petrides and Pandya, 2002).

Broca's area: pars opercularis

Results indicate that AV language comprehension specifically activates the dorsal aspect of pars opercularis. ROI analysis confirms that activation was truly in the opercular region of

individual participants. Pars opercularis contains Brodmann's cytoarchitectonic area 44, which is the suggested homologue of macaque inferior premotor cortex (area F5), a region containing mirror neurons that fire upon perception (i.e., hearing and/or seeing) and execution of particular types of *goal-directed* hand or mouth movements (Fadiga et al., 2000; Gallese et al., 1996; Kohler et al., 2002; Rizzolatti, 1987; Rizzolatti et al., 1996, 2002; Umiltà et al., 2001). Prior neuroimaging results suggest that the pars opercularis in the human has similar properties (Binkofski et al., 2000; Iacoboni et al., 1999; Koski et al.,

2002). Our results are consistent with these properties and with the claim that the dorsal aspect of the pars opercularis has more mirror-neuron like properties than the ventral portion, as the dorsal aspect is activated during both observation and imitation of goal-oriented actions whereas the more ventral portion is activated during imitation only (Koski et al., 2002; Molnar-Szakacs et al., 2004). The lack of activity in this region during the V condition relative to baseline may reflect the fact that participants did not have clear perceptual or linguistic goals while observing facial movements (which is a requisite property of mirror neuron activation as described above). That is, participants were not told to try to understand what the talker was saying nor were they trained to speech-read. The AV–V comparison, however, did show that aspects of the pars opercularis (BA 44) were more active in the V condition than in the AV condition in the right hemisphere. This may reflect that, though participants were not trained speech-readers, everyone speech-reads to some extent.

Precentral gyrus

The observed precentral gyrus and sulcus activity occurred reliably only in the AV condition though it could be observed at a lower threshold in the A and V conditions and therefore potentially due to Type I statistical error. This activation was primarily in the dorsal aspect (PMd) as determined by an arbitrary division of the dorsal 2/3 and ventral 1/3 of the gyrus. Note that if we had used a line at the superior aspect of the inferior frontal sulcus as our criteria the observed activation would fall on the boundary of these areas, perhaps residing more in ventral premotor. Nonetheless, activation included both premotor and nearby primary motor cortex and did not include classically defined frontal eye fields (Geyer et al., 2000). This activation encompassed the mouth area of primary motor cortex according to probabilistic maps (Fox et al., 2001). These activation patterns are consistent with the hypothesized role of this area in human speech perception and production (Wilson et al., 2004). Stimulation of the PMd region has been shown to disrupt vocalization (Ojemann, 1979) and to do so more consistently than stimulation of the inferior frontal gyrus (Ojemann et al., 1989). In addition, this region has been shown to be more consistently active than the pars opercularis during overt speech production (Huang et al., 2001; Wise et al., 1999).

We showed that activation in the right PMd region was modulated by the amount of viseme content for the AV condition and for no other conditions. The lack of modulation in this region for the A condition supports the inference that it was the viseme content and not the information in the stories per se that produced the effect. We believe that modulation of activity in this region by the viseme content of the stories reflects the role of PMd in sensory-motor integration—as the ambiguity of the audiovisual signal decreases, there is concomitant increase in PMd activity. This specificity may not be apparent in the pars opercularis because there the level of representation is more abstract. This result, however, needs to be replicated with more controlled stimuli as other linguistic factors could covary with viseme content.

Superior temporal gyrus and sulcus

The superior temporal gyrus and sulcus posterior to primary auditory cortex, anterior to the supramarginal and angular gyri,

were more active during the AV than in the A condition. Furthermore, during audiovisual comprehension, but not during the A or V conditions, activity in this region was significantly modulated by the amount of viseme content in the audiovisual stories, becoming more active as viseme content increased. Previous research has shown that damage to posterior superior temporal cortex results in a deficit in repeating speech (Anderson et al., 1999; Hickok, 2000) and stimulation of these sites results in speech production errors (Ojemann, 1979; Ojemann et al., 1989). In addition, neuroimaging studies have shown that this region is active during speech production and is also active during speech production in the absence of feedback from the speakers voice (Hickok et al., 2000). On the perceptual side, research indicates that the STS is activated by the observation of biologically relevant movements and by implied movements of the eyes, mouth, and hands (for a review, see Allison et al., 2000). In addition, this area is activated to a greater extent by linguistically meaningful facial movements than to facial movements not having linguistic meaning (e.g., Campbell et al., 2001). In the present study, the activation that was produced by the presence of visemes is consistent with the sensitivity of this region to biologically relevant movements and specifically to speech movements. In addition, our finding is consistent with the interpretation that this area is a site participating in the integration of seen and heard speech (Calvert et al., 2000; Sams et al., 1991). In sum, the posterior superior temporal gyrus and sulcus seem to participate in both speech perception and production as a cortical convergence zone (Damasio et al., 1990) having auditory, visual, and motor properties.

A network for audio–visual–motor integration

Taken together, we suggest that these areas form a network serving audio–visual–motor integration during language comprehension. This is an elaboration of the idea that there is a processing “stream” (e.g., posterior superior temporal and frontal cortices) involved in audio–motor integration that is active when a task, e.g., speech discrimination, requires explicit decisions about phonetic segments (Hickok and Poeppel, 2000, 2004). Audiovisual speech perception may represent one instance when this network is naturally active rather than driven by task demands. Typically activation of these regions occurs during tasks requiring explicit metalinguistic phonological judgments (Benson et al., 2001; Buchsbaum et al., 2001; Burton et al., 2000; Heim et al., 2003) or explicit articulation (Buchsbaum et al., 2001; Heim et al., 2003; Hickok et al., 2000; Paus et al., 1996a,b; Wise et al., 2001). During audiovisual language comprehension, however, this network appears to be sensitive to available cues in the environment and to use them to generate an action hypothesis regarding observed movements.

We propose an initial model of audio–visual–motor integration adapted from a previous model of motor control (Haruno et al., 2001; Wolpert and Kawato, 1998). This type of model has also been proposed to account for imitation using the properties of mirror neurons (Carr et al., 2003; Miall, 2003). Based on our results, we suggest that the audio–visual–motor integration network is composed of posterior superior temporal cortex, the superior portion of the pars opercularis, the dorsal (or superior ventral) aspect of premotor cortex, and adjacent motor cortex.

As noted, posterior aspects of superior temporal cortex are activated during multimodal integration, biological motion perception, speech perception, and speech production. Anatomically,

this region is connected to prefrontal cortices (Hackett et al., 1999; Petrides and Pandya, 1988; Romanski et al., 1999), which are in turn connect to motor cortices (Barbas and Pandya, 1989; Petrides and Pandya, 2002). One of these motor areas is the pars opercularis, which is activated by the same types of tasks that activate posterior superior temporal cortex. Based on their shared functional properties and connectivity, we suggest that these two areas form an interactive network. On a sensory-motor continuum, posterior superior temporal cortex is more closely associated with sensory perception whereas the pars opercularis is more closely associated with motor output (i.e., based on their respective neural contexts, McIntosh, 2000).

These two regions may interact to specify multimodal sensory representations of speech together with the motor commands that would result in these representations. Specifically, multimodal sensory representations derive from sound patterns and observed facial movements. These representations can be thought of as hypotheses—but not final interpretations (Nusbaum and Schwab, 1986) about the intended phonemes of the speaker. These hypotheses are also probabilistically associated with high-level (i.e., abstracted) motor commands that would have been used by the perceiver if s/he were producing those phonemes. Associations between sensory patterns associated with phonemes and the motor commands needed to produce them may be established over a lifetime of speech production.

Neurons in the pars opercularis, however, do not encode the actual dynamics of the movement or the effector required to perform a specific action (Rizzolatti et al., 2002). Rather, this encoding appears to be carried out by interaction between the pars opercularis and the dorsal (or superior ventral) aspect of the precentral gyrus and sulcus (i.e., premotor and motor cortex). That is, we suggest that these two areas encode the motor plans corresponding to the sensory-motor representations activated by the posterior superior temporal cortex and pars opercularis. This encoding of movement dynamics and effectors, also a necessary component of speech production, generates corollary estimates of its sensory consequences. This prediction can be used by the sensory system to help constrain interpretation of accumulating sensory information. One way this could occur is by shifting attention toward or away from a particular interpretation of the speech signal, a process that likely involves interaction between the posterior superior temporal cortex and/or the inferior parietal cortex through reafferent feedback from premotor or motor cortices.

By this model, audio-visual-motor integration occurs across several multimodal brain regions over time and depends on joint participation of a set of cortical areas that are typically viewed as part of a speech production motor network. This account is in contrast with previous views of audiovisual speech where integration occurs specifically in the posterior superior temporal gyrus or sulcus with no motor component.

Conclusions

In the present study, we have shown that language comprehension, in the context of the mouth and face movements involved in speech production, activates a network of brain regions involved in audio-visual-motor integration and speech production. This network is minimally activated during audio-alone language processing or during visual-alone viewing of oral-facial movements. Furthermore, this distributed network is

sensitive to visual information about the phonetic structure of the stories. The brain regions comprising this network include posterior superior temporal cortex, the superior portion of the pars opercularis, premotor cortex, adjacent motor cortex, somatosensory cortex, and the cerebellum. This result is consistent with recent findings in macaques (e.g., Rizzolatti et al., 2002) and humans (e.g., Iacoboni et al., 1998) about the role of motor areas in action understanding. It also extends the data on human action understanding to the realm of speech and language in the audiovisual environment (Rizzolatti and Arbib, 1998), an important goal-directed behavior. With regard to language comprehension, we suggest that this distributed network serves to represent the visual configuration of observed facial movements, the motor commands that could have been used to generate that configuration, and the associated expected auditory consequences of executing that hypothesized motor plan. It is possible that activity within this network mediates the improvement in understanding of speech gained with audiovisual presentations.

Acknowledgments

Many thanks to Bernadette Brogan, E Chen, Shahrina Chowdhury, Fred Dick, Petr Hlustik, Tom Hummer, Robert Lyons, Daniel Margoliash, Xander Meadow, Mom, David Poeppel, Ana Solodkin, Stephen Ufring, Virginie van Wassenhove, Ryan Walsh, and Nicholas “The Wheem” Wymbs and NIDCD: Language and Speech (Dr. Judith Cooper, Program Officer). This work was supported by NIH RO1-DC03378.

References

- Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4 (7), 267–278.
- Anderson, J.M., Gilmore, R., Roper, S., Crosson, B., Bauer, R.M., Nadeau, S., et al., 1999. Conduction aphasia and the arcuate fasciculus: a reexamination of the Wernicke–Geschwind model. *Brain Lang.* 70 (1), 1–12.
- Barbas, H., Pandya, D.N., 1989. Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 286 (3), 353–375.
- Benson, R.R., Whalen, D.H., Richardson, M., Swainson, B., Clark, V.P., Lai, S., et al., 2001. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang.* 78 (3), 364–396.
- Binkofski, F., Amunts, K., Stephan, K.M., Posse, S., Schormann, T., Freund, H.J., et al., 2000. Broca’s region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. *Hum. Brain Mapp.* 11 (4), 273–285.
- Blank, S.C., Scott, S.K., Murphy, K., Warburton, E., Wise, R.J., 2002. Speech production: Wernicke, Broca and beyond. *Brain* 125 (Pt. 8), 1829–1838.
- Bookheimer, S.Y., Zeffiro, T.A., Blaxton, T., Gaillard, W., Theodore, W., 1995. Regional cerebral blood flow during object naming and word reading. *Hum. Brain Mapp.* 3, 93–106.
- Braun, A.R., Guillemin, A., Hosey, L., Varga, M., 2001. The neural organization of discourse: an H₂ ¹⁵O-PET study of narrative production in English and American sign language. *Brain* 124 (Pt. 10), 2028–2044.
- Buchsbaum, B., 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.

- Burton, M.W., Small, S.L., Blumstein, S.E., 2000. The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12 (4), 679–690.
- Calvert, G.A., Campbell, R., 2003. Reading speech from still and moving faces: the neural substrates of visible speech. *J. Cogn. Neurosci.* 15 (1), 57–70.
- Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., et al., 1997. Activation of auditory cortex during silent lipreading. *Science* 276 (5312), 593–596.
- Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10 (11), 649–657.
- Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., et al., 2001. Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res. Cogn. Brain Res.* 12 (2), 233–243.
- Carr, L., Iacoboni, M., Dubeau, M.C., Mazziotta, J.C., Lenzi, G.L., 2003. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U. S. A.* 100 (9), 5497–5502.
- Caviness, V.S., Meyer, J., Makris, N., Kennedy, D.N., 1996. MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. *J. Cogn. Neurosci.* 8 (6), 566–587.
- Corina, D.P., McBurney, S.L., Dodrill, C., Hinshaw, K., Brinkley, J., Ojemann, G., 1999. Functional roles of Broca's area and SMG: evidence from cortical stimulation mapping in a deaf signer. *NeuroImage* 10 (5), 570–581.
- Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance NeuroImages. *Comput. Biomed. Res.* 29 (3), 162–173.
- Damasio, A.R., Damasio, H., Tranel, D., Brandt, J.P., 1990. Neural regionalization of knowledge access: Preliminary evidence. *Cold Spring Harbor Symposia on Quantitative Biology*. Plainview, New York, Cold Spring Harbor Laboratory Press. pp. 1039–1047.
- Devlin, J.T., Matthews, P.M., Rushworth, M.F., 2003. Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 15 (1), 71–84.
- Dronkers, N.F., 1996. A new brain region for coordinating speech articulation. *Nature* 384, 159–161 (November 14).
- Dronkers, N.F., 1998. Symposium: the role of Broca's area in language. *Brain Lang.* 65, 71–72.
- Ehrsson, H.H., Fagergren, A., Jonsson, T., Westling, G., Johansson, R.S., Forssberg, H., 2000. Cortical activity in precision- versus power-grip tasks: an fMRI study. *J. Neurophysiol.* 83 (1), 528–536.
- Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 2000. Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *Int. J. Psychophysiol.* 35 (2–3), 165–177.
- 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 (2), 399–402.
- Fox, P.T., Huang, A., Parsons, L.M., Xiong, J.H., Zamarippa, F., Rainey, L., et al., 2001. Location-probability profiles for the mouth region of human primary motor-sensory cortex: model and validation. *NeuroImage* 13 (1), 196–209.
- Friederici, A.D., Opitz, B., von Cramon, D.Y., 2000. Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types. *Cereb. Cortex* 10 (7), 698–705.
- Gabrieli, J.D., Poldrack, R.A., Desmond, J.E., 1998. The role of left prefrontal cortex in language and memory. *Proc. Natl. Acad. Sci. U. S. A.* 95 (3), 906–913.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. *Brain* 119 (Pt. 2), 593–609.
- Geschwind, N., 1965. Disconnection syndromes in animals and man. *Brain* 88, 237–294, 585–644.
- Geyer, S., Matelli, M., Luppino, G., Zilles, K., 2000. Functional neuroanatomy of the primate isocortical motor system. *Anat. Embryol. (Berl.)* 202 (6), 443–474.
- Goldschen, A.J., 1993. Continuous Automatic Speech Recognition by Lipreading. George Washington University.
- Grafton, S.T., Fadiga, L., Arbib, M.A., Rizzolatti, G., 1997. Premotor cortex activation during observation and naming of familiar tools. *NeuroImage* 6 (4), 231–236.
- Grant, K.W., Seitz, P.F., 2000. The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 108 (3 Pt. 1), 1197–1208.
- Graziano, M.S., Gandhi, S., 2000. Location of the polysensory zone in the precentral gyrus of anesthetized monkeys. *Exp. Brain Res.* 135 (2), 259–266.
- Grezes, J., Costes, N., Decety, J., 1999. The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain* 122 (Pt. 10), 1875–1887.
- Hackett, T.A., Stepniewska, I., Kaas, J.H., 1999. Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 817 (1–2), 45–58.
- Haruno, M., Wolpert, D.M., Kawato, M., 2001. Mosaic model for sensorimotor learning and control. *Neural Comput.* 13 (10), 2201–2220.
- Heim, S., Opitz, B., Muller, K., Friederici, A.D., 2003. Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Brain Res. Cogn. Brain Res.* 16 (2), 285–296.
- Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., Mazziotta, J.C., 2003. The essential role of Broca's area in imitation. *Eur. J. Neurosci.* 17 (5), 1123–1128.
- Hermesdörfer, J., Goldenberg, G., Wachsmuth, C., Conrad, B., Ceballos-Baumann, A.O., Bartenstein, P., et al., 2001. Cortical correlates of gesture processing: clues to the cerebral mechanisms underlying apraxia during the imitation of meaningless gestures. *NeuroImage* 14 (1 Pt. 1), 149–161.
- Hickok, G., 2000. Speech perception, conduction aphasia, and the functional neuroanatomy of language. In: Swinney, D. (Ed.), *Language and the Brain*. Academic Press, San Diego, pp. 87–104.
- Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn. Sci.* 4 (4), 131–138.
- Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92 (1–2), 67–99.
- Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A.K., Naeve-Velguth, S., Strupp, J.P., et al., 2000. A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci. Lett.* 287 (2), 156–160.
- Huang, J., Carr, T.H., Cao, Y., 2001. Comparing cortical activations for silent and overt speech using event-related fMRI. *Hum. Brain Mapp.* 15, 39–53.
- Humphries, C., Willard, K., Buchsbaum, B., Hickok, G., 2001. Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *NeuroReport* 12 (8), 1749–1752.
- Iacoboni, M., Woods, R.P., Mazziotta, J.C., 1998. Bimodal (auditory and visual) left frontoparietal circuitry for sensorimotor integration and sensorimotor learning. *Brain* 121 (Pt. 11), 2135–2143.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G., 1999. Cortical mechanisms of human imitation. *Science* 286 (5449), 2526–2528.
- Jackson, P.L., 1988. The theoretical minimal unit for visual speech perception: visemes and coarticulation. *Volta Rev.* 90 (5), 99–115.
- Jonides, J., Smith, E.E., Marshuetz, C., Koeppel, R.A., Reuter-Lorenz, P.A., 1998. Inhibition in verbal working memory revealed by brain activation. *Proc. Natl. Acad. Sci. U. S. A.* 95 (14), 8410–8413.
- Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., Thulborn, K.R., 1996. Brain activation modulated by sentence comprehension. *Science* 274 (5284), 114–116.

- Klatt, D.H., 1979. Speech perception: a model of acoustic–phonetic analysis and lexical access. In: Cole, R.A. (Ed.), *Perception and Production of Fluent Speech*. Lawrence Erlbaum Associates, Ltd., Hillsdale, NJ.
- Knopman, D.S., Selnes, O.A., Niccum, N., Rubens, A.B., Yock, D., Larson, D., 1983. A longitudinal study of speech fluency in aphasia: CT correlates of recovery and persistent nonfluency. *Neurology* 33 (9), 1170–1178.
- Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297 (5582), 846–848.
- Koski, L., Wohlschlaeger, A., Bekkering, H., Woods, R.P., Dubeau, M.C., Mazziotta, J.C., et al., 2002. Modulation of motor and premotor activity during imitation of target-directed actions. *Cereb. Cortex* 12 (8), 847–855.
- Kuhl, P.K., Meltzoff, A.N., 1982. The bimodal perception of speech in infancy. *Science* 218 (4577), 1138–1141.
- Lieberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. *Cognition* 21 (1), 1–36.
- Ludman, C.N., Summerfield, A.Q., Hall, D., Elliott, M., Foster, J., Hykin, J.L., et al., 2000. Lip-reading ability and patterns of cortical activation studied using fMRI. *Br. J. Audiol.* 34 (4), 225–230.
- MacSweeney, M., Amaro, E., Calvert, G.A., Campbell, R., David, A.S., McGuire, P., et al., 2000. Silent speech reading in the absence of scanner noise: an event-related fMRI study. *NeuroReport* 11 (8), 1729–1733.
- MacSweeney, M., Campbell, R., Calvert, G.A., McGuire, P.K., David, A.S., Suckling, J., et al., 2001. Dispersed activation in the left temporal cortex for speech-reading in congenitally deaf people. *Proc. R. Soc. Lond., B Biol. Sci.* 268 (1466), 451–457.
- MacSweeney, M., Calvert, G.A., Campbell, R., McGuire, P.K., David, A.S., Williams, S.C., et al., 2002a. Speechreading circuits in people born deaf. *Neuropsychologia* 40 (7), 801–807.
- MacSweeney, M., Woll, B., Campbell, R., McGuire, P.K., David, A.S., Williams, S.C., et al., 2002b. Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain* 125 (Pt. 7), 1583–1593.
- Masdeu, J.C., O'Hara, R.J., 1983. Motor aphasia unaccompanied by faciobrachial weakness. *Neurology* 33 (4), 519–521.
- Massaro, D.W., 1998. *Perceiving Talking Faces: from Speech Perception to a Behavioral Principle*. MIT Press, Cambridge, MA.
- McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. *Nature* 264 (5588), 746–748.
- McIntosh, A.R., 2000. Towards a network theory of cognition. *Neural Netw.* 13 (8–9), 861–870.
- Miall, R.C., 2003. Connecting mirror neurons and forward models. *NeuroReport* 14 (17), 2135–2137.
- Mohr, J.P., Pessin, M.S., Finkelstein, S., Funkenstein, H.H., Duncan, G.W., Davis, K.R., 1978. Broca aphasia: pathologic and clinical. *Neurology* 28 (4), 311–324.
- Molnar-Szakacs, I., Iacoboni, M., Koski, L., Mazziotta, J.C., 2004. Functional segregation within pars opercularis of the inferior frontal gyrus: Evidence from fMRI studies of imitation and action observation. *Cereb. Cortex*. Oct 28 [E-publication ahead of print].
- Mottron, R., Krause, C.M., Tiippana, K., Sams, M., 2002. Processing of changes in visual speech in the human auditory cortex. *Brain Res. Cogn. Brain Res.* 13 (3), 417–425.
- Noll, D.C., Cohen, J.D., Meyer, C.H., Schneider, W., 1995. Spiral k-space MRI imaging of cortical activation. *J. Magn. Reson. Imaging* 5, 49–56.
- Nusbaum, H.C., Schwab, E.C., 1986. The role of attention and active processing in speech perception. In: Schwab, E.C., Nusbaum, H.C. (Eds.), *Pattern Recognition by Humans and Machines: Volume 1. Speech Perception*. Academic Press, New York, pp. 113–157.
- Ojemann, G., 1979. Individual variability in cortical localization of language. *J. Neurosurg.* 50, 164–169.
- Ojemann, G., Ojemann, J., Lettich, E., Berger, M., 1989. Cortical language localization in left, dominant hemisphere: an electrical stimulation mapping investigation in 117 patients. *J. Neurosurg.* 71, 316–326.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Olson, I.R., Gatenby, J.C., Gore, J.C., 2002. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Brain Res. Cogn. Brain Res.* 14 (1), 129–138.
- Papathanassiou, D., Etard, O., Mellet, E., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2000. A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. *NeuroImage* 11 (4), 347–357.
- Patterson, M.L., Werker, J.F., 2003. Two-month old infants match phonetic information in lips and voice. *Dev. Sci.* 6 (2), 193–198.
- Paus, T., Marrett, S., Worsley, K., Evans, A., 1996a. Imaging motor-to-sensory discharges in the human brain: an experimental tool for the assessment of functional connectivity. *NeuroImage* 4 (2), 78–86.
- Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J., Evans, A.C., 1996b. Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur. J. Neurosci.* 8 (11), 2236–2246.
- Petrides, M., Pandya, D.N., 1988. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *J. Comp. Neurol.* 273 (1), 52–66.
- 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 (2), 291–310.
- Preminger, J.E., Lin, H., Payen, M., Levitt, H., 1998. Selective visual masking in speech reading. *J. Speech Lang. Hear. Res.* 41 (3), 564–575.
- Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., Caviness, V.S., 1992. Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging. *J. Cogn. Neurosci.* 4 (4), 352–374.
- Reisberg, D., McLean, J., Goldfield, A., 1987. Easy to hear but hard to understand: a lipreading advantage with intact auditory stimuli. In: Campbell, R. (Ed.), *Hearing by Eye: The Psychology of Lipreading*. Erlbaum, Hillsdale, NJ, pp. 97–114.
- Rizzolatti, G., 1987. Functional organization of inferior area 6. *Ciba Found. Symp.* 132, 171–186.
- Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. *Trends Neurosci.* 21 (5), 188–194.
- Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* 3 (2), 131–141.
- Rizzolatti, G., Fogassi, L., Gallese, V., 2002. Motor and cognitive functions of the ventral premotor cortex. *Curr. Opin. Neurobiol.* 12 (2), 149–154.
- Romanski, L.M., Goldman-Rakic, P.S., 2002. An auditory domain in primate prefrontal cortex. *Nat. Neurosci.* 5 (1), 15–16.
- Romanski, L.M., Bates, J.F., Goldman-Rakic, P.S., 1999. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 403 (2), 141–157.
- Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O.V., Lu, S.T., et al., 1991. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci. Lett.* 127 (1), 141–145.
- Small, S.L., Burton, M.W., 2001. Functional neuroimaging of language. In: Berndt, R.S. (Ed.), *Handbook of Neuropsychology*, second ed. Language and Aphasia, vol. 3. Elsevier, Amsterdam, pp. 335–351.
- Small, S.L., Nusbaum, H.C., 2004. On the neurobiological investigation of language understanding in context. *Brain Lang.* 89 (2), 300–311.
- Smith, E.E., Jonides, J., 1999. Storage and executive processes in the frontal lobes. *Science* 283 (5408), 1657–1661.
- Stevens, K.N., Blumstein, S.E., 1981. The search for invariant acoustic correlates of phonetic features. In: Eimas, P., Miller, J. (Eds.), *Perspectives on the Study of Speech*. Lawrence Erlbaum, Hillsdale, NJ.
- Sumby, W.H., Pollack, I., 1954. Visual contribution of speech intelligibility in noise. *J. Acoust. Soc. Am.* 26 (2), 212–215.
- Surguladze, S.A., Calvert, G.A., Brammer, M.J., Campbell, R.,

- Murphy, K., Giampietro, V., et al., 2001. Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiatry Res.* 106 (1), 1–14.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain: 3D Proportional System: An Approach to Cerebral Imaging*. Georg Thieme Verlag, New York, NY.
- Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., et al., 2001. I know what you are doing. a neurophysiological study. *Neuron* 31 (1), 155–165.
- Watkins, K.E., Strafella, A.P., Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41 (8), 989–994.
- Wilson, M., 2002. Six views of embodied cognition. *Psychon. Bull. Rev.* 9 (4), 625–636.
- Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7 (7), 701–702.
- Wise, R.J., Greene, J., Buchel, C., Scott, S.K., 1999. Brain regions involved in articulation. *Lancet* 353 (9158), 1057–1061.
- Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., Warburton, E.A., 2001. Separate neural subsystems within ‘Wernicke’s area’. *Brain* 124 (Pt. 1), 83–95.
- Wolpert, D.M., Kawato, M., 1998. Multiple paired forward and inverse models for motor control. *Neural Netw.* 11 (7–8), 1317–1329.