Functional Overlap between Regions Involved in Speech Perception and in Monitoring One's Own Voice during Speech Production

Zane Z. Zheng, Kevin G. Munhall, and Ingrid S. Johnsrude

Abstract

■ The fluency and the reliability of speech production suggest a mechanism that links motor commands and sensory feedback. Here, we examined the neural organization supporting such links by using fMRI to identify regions in which activity during speech production is modulated according to whether auditory feedback matches the predicted outcome or not and by examining the overlap with the network recruited during passive listening to speech sounds. We used real-time signal processing to compare brain activity when participants whispered a consonant–vowel–consonant word ("Ted") and either heard this clearly or heard voice-gated masking noise. We compared this to when they listened to yoked stimuli (identical recordings of "Ted" or noise) without speaking. Activity along the STS and superior temporal gyrus bilaterally was significantly greater if the auditory stimulus was (a) processed as the auditory concomitant of speaking and (b) did not match the predicted outcome (noise). The network exhibiting this Feedback Type × Production/Perception interaction includes a superior temporal gyrus/middle temporal gyrus region that is activated more when listening to speech than to noise. This is consistent with speech production and speech perception being linked in a control system that predicts the sensory outcome of speech acts and that processes an error signal in speech-sensitive regions when this and the sensory data do not match. ■

INTRODUCTION

In the study of human communication, the relationship between speech perception and speech production has long been controversial (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). That the two sides of spoken language are linked somehow is not in dispute—prelingual deafness without immediate cochlear implantation severely hinders the development of normal speech in children (Schauwers et al., 2004; Geers, Nicholas, & Sedey, 2003), and hearing impairment in adults affects aspects of speech production such as the control of fundamental frequency and intensity (Cowie, Douglas-Cowie, & Kerr, 1982). Talking, like all motor control, must require sensory feedback to maintain the accuracy and stability of movement (Levelt, 1983). However, the nature of the link between perception and action in speech, its cognitive structure and its neural organization, are not well understood. Whether the same neural system subserves auditory perception both for speech comprehension and for the purpose of articulatory control is not known.

Clinical data indicate that if auditory feedback is eliminated, speech fluency is affected (Schauwers et al., 2004; Geers et al., 2003; Cowie et al., 1982), and if feedback is delayed, fluency is disrupted (Fukawa, Yoshioka, Ozawa, & Yoshida, 1988; Howell & Powell, 1987; Siegel, Schork, Pick, & Garber, 1982; Mackay, 1968; Black, 1951). Given the rapidity of speech, it would be difficult for acoustic feedback to contribute to control as part of a servomechanism (Lashley, 1951) because the articulation of speech sounds would be finished by the time acoustic feedback could be processed. Alternatively, feedforward mechanisms, capable of predictive motor planning, have been proposed (Kawato, 1989). Such "internal models" of the motor system (Jordan & Rumelhart, 1992) are hypothesized to learn through sensory (e.g., auditory) feedback and thus involve representations of actions and their consequences.

One source of evidence for internal models in speech comes from auditory perturbation studies, in which talkers are observed to alter their speech production in response to altered acoustic feedback. For example, if the fundamental frequency of auditory feedback is shifted or if frequencies of vowel formants are altered, talkers accommodate rapidly as if to normalize their production so that it is acoustically closer to a desired output (Chen, Liu, Xu, & Larson, 2007; Purcell & Munhall, 2006a, 2006b; Jones & Munhall, 2000; Houde & Jordan, 1998; Elman, 1981). Such studies yield two important conclusions. First, that subjects rapidly compensate for the perturbation in subsequent trials indicates the presence of an error detection and correction mechanism (e.g., Jones & Munhall, 2000; Houde & Jordan, 1998). Second, the compensation persists for a short period after acoustic feedback is returned to normal (e.g., Purcell & Munhall, 2006a, 2006b), suggesting that the

Queen's University, Ontario, Canada

© 2009 Massachusetts Institute of Technology

error detection mechanism is producing learning that tunes the speech motor controller.

The neural network supporting this acoustic error detection and correction has been investigated in both nonhuman primates and humans. A recent study in marmosets reveals that vocalization-induced suppression within neurons in the auditory cortex is markedly reduced (firing rate increases) if feedback is altered by frequency shifting (Eliades & Wang, 2008). Recent functional neuroimaging studies in humans reveal increases in activity in the superior temporal areas when vocalization with altered feedback is compared with a condition with normal feedback (Tourville, Reilly, & Guenther, 2008; Christoffels, Formisano, & Schiller, 2007; Fu et al., 2006; Hashimoto & Sakai, 2003). This signal increase has been interpreted as evidence for an error signal that could be used to tune ongoing speech production (Guenther, 2006).

The engagement of the superior temporal region in processing speech during listening has been extensively documented. Superior temporal activation is observed for processing of speech stimuli (see Table 1) when compared with noise (Obleser et al., 2006; Rimol, Specht, Weis, Savoy, & Hugdahl, 2005; Jancke, Wustenberg, Scheich, & Heinze, 2002; Binder et al., 1994; Zatorre, Evans, Meyer, & Gjedde, 1992), tones (Burton & Small, 2006; Ashtari et al., 2004; Poeppel et al., 2004; Zaehle, Wustenberg, Meyer, & Jancke, 2004; Joanisse & Gati, 2003; Specht & Reul, 2003; Benson et al., 2001; Vouloumanos, Kiehl, Werker, & Liddle, 2001; Binder et al., 2000; Burton, Small, & Blumstein, 2000; Binder, Frost, Hammeke, Rao, & Cox, 1996; O'Leary et al., 1996), and other nonspeech sounds (Benson, Richardson, Whalen, & Lai, 2006; Uppenkamp, Johnsrude, Norris, Marslen-Wilson, & Patterson, 2006; Meyer, Zysset, Yves von Cramon, & Alter, 2005; Gandour et al., 2003; Thierry, Giraud, & Price, 2003; Belin, Zatorre, & Ahad, 2002; Giraud & Price, 2001). Recent functional imaging studies in which speech and nonspeech conditions are very closely matched acoustically also elicit signal change in this region, suggesting that it is the processing of speech *qua* speech and not the acoustic structure of speech that is responsible for the signal change (Desai, Liebenthal, Waldron, & Binder, 2008; Mottonen et al., 2006; Scott, Rosen, Lang, & Wise, 2006; Uppenkamp et al., 2006; Dehaene-Lambertz et al., 2005; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Narain et al., 2003; Scott, Blank, Rosen, & Wise, 2000).

Despite the evidence for the involvement of the superior temporal region in the processing of both error signal and speech, whether and how these two functions are linked has never been investigated in the same context. In the experiment presented here, we examined whether the same regions within subjects are processing the speaker's own auditory feedback as well as subserving the perception of speech. We argued that simply observing a change in auditory cortical activity during altered, compared with normal, feedback is not sufficient evidence of an error detection mechanism. The characteristics of a speech error signal include that it be generated uniquely during speech production and not during listening. Accordingly, it is best identified as an interaction, where the same acoustic difference between altered and normal feedback in the context of talking provokes a different pattern of activity than when the identical stimuli are presented in the context of listening. In addition, we will examine whether this error detection mechanism recruits regions that are also active—albeit in a different way—when listeners passively hear speech.

Accordingly, we will compare speech production to passive listening, with identical acoustic stimuli present in both conditions. On production trials, participants either hear their own voice as normal (unaltered feedback) or hear voice-gated, signal-correlated noise (manipulated feedback). On listening trials, participants hear either recordings of their own voice or noise yoked to previous production trials. We identified brain regions that respond more to noise feedback than unaltered feedback during speech production (consistent with an error signal), but that are not sensitive to speech during passive listening, or that respond more to unaltered feedback than to noise feedback during passive listening. By assessing the interaction between speaking condition and feedback, we will examine whether the same brain regions contribute to speech perception for the purpose of ongoing articulatory control and for the purpose of comprehension.

METHODS

Subjects

Written informed consent was obtained from 21 individuals (mean age = 23 years, range = 18–45 years, 16 women). All were right-handed, without any history of neurological or hearing disorder, and spoke English as their first language. Each participant received $15 to compensate them for their time. Procedures were approved by the Queen's Health Sciences Research Ethics Board.

Procedure and Functional Image Acquisition

We adopted a 2 × 2 factorial design (four experimental conditions) and a low-level silence/rest control condition. The four experimental conditions were as follows: *production-clear*—producing whispered speech ("Ted") and hearing this through headphones; *production-masked*—producing whispered speech ("Ted") and hearing voice-gated, signal-correlated masking noise (Schroeder, 1968), which is created by applying the amplitude envelope of the utterance to white noise; *listen-clear*—listening to the stimuli of production-clear trials (without production); and *listen-masked*—listening to the stimuli of production-masked trials (without production). Whispered speech was used to minimize bone conducted auditory feedback (Barany, 1938) and to make sure

Table 1. The Activation Peaks within the Left Temporal Lobe Reported in Previously Published Neuroimaging Studies in which Listening to Speech Is Compared with Listening to a Nonspeech Stimulus

		Coordinates				
Authors	*Contrast*	*x*	*y*	*z*	*Area*	*Task*
Binder et al. (1996)	Monosyllabic words vs. pure tones	−56	−12	−2	L STS	Passive
Mummery, Ashbumer, Scott, and Wise (1999) (PET)	Bisyllabic nouns vs. signal correlated noise	−54	−28	+4	L MTG	Passive
Binder et al. (2000)	Monosyllabic words vs. pure tones	−52	−42	+6	L STS	Passive
Scott et al. (2000) (PET)	Speech + Vco vs. Rsp + RVco	−66	−12	−12	L STS	Passive
Benson et al. (2001)	Nonsense syllables vs. tonal nonspeech	−57	−42	+12	L STG	Passive
Belin et al. (2002)	Speech sounds vs. nonspeech vocalizations	−62	−40	+10	L MTG	Passive
Jancke et al. (2002)	CV syllables vs. white noise	−64	−16	0	L STG	Passive
Narain et al. (2003)	Speech + Vco vs. Rsp + RVco	−52	−54	+14	L STG	Passive
Specht and Reul (2003)	Mono/bisyllabic words vs. pure tones	−63	−12	−6	L STS	Passive
Rimol et al. (2005)	CV syllables vs. noise (white, pink, brown)	−59	−27	−2	L MTG	Passive
Uppenkamp et al. (2006)	Natural and synthetic vowels vs. nonspeech sounds	−66	−20	0	L STS	Passive
Scott et al. (2006) (PET)	Vco vs. Vco + RVco	−60	−44	+10	L STS	Passive
Present study	"Ted" vs. white noise	−66	−36	+3	L MTG	Passive
Zatorre et al. (1992) (PET)	CVC syllables vs. white noise	−58	−21	+8	L STG	Active
O'Leary et al. (1996) (PET)	CV syllables vs. pure tones	−57	−31	+7	L STG	Active
Burton et al. (2000)	CVC syllables vs. pure tones	−64	−23	+6	L STG	Active
Giraud and Price (2001) (PET)	Words + nonwords vs. environmental sounds + white noise	−70	−38	+6	L STG	Active
Vouloumanos et al. (2001)	Sine wave (speech) vs. pure tones	−60	−16	−8	L MTG	Active
Gandour et al. (2003)	Consonant syllables vs. hums	−54	−54	+10	L MTG	Active
Joanisse and Gati (2003)	CV syllables vs. pure tones	−56	−22	+3	L STG	Active
Liebenthal, Binder, Piorkowski, and Remez (2003)	Sine wave (speech) vs. sine wave (nonspeech)	−51	−30	+19	L STG	Active
Thierry et al. (2003) (PET)	Spoken words vs. environmental sounds	−60	−40	0	L STG	Active
Ashtari et al. (2004)	Phonemes vs. pure tones	−63	−36	+3	L MTG	Active
Poeppel et al. (2004) (PET)	CV syllables vs. frequency modulated tones	−62	−36	+4	L STG	Active
Zaehle et al. (2004)	CV syllables vs. pure tones	−66	−15	0	L STG	Active
Dehaene-Lambertz et al. (2005)	Sine wave (speech) vs. sine wave (nonspeech)	−56	−40	0	L STS	Active
Liebenthal et al. (2005)	CV syllables (phonemic) vs. CV syllables (nonphonemic)	−60	−8	−3	L STS	Active
Meyer et al. (2005)	Streams of syllables vs. sequences of laugh bursts	−59	−13	+6	L STG	Active
Benson et al. (2006)	Sine wave speech vs. nonspeech + chord progressions	−64	−32	−8	L STS	Active
Burton and Small (2006)	CVC syllables vs. pure tones	−61	−12	−1	L STG	Active
Mottonen et al. (2006)	Sine wave (speech) vs. sine wave (nonspeech)	−61	−39	+2	L STS	Active
Obleser et al. (2006)	German vowels vs. band-pass noise	−63	−2	−2	L STG	Active
Desai et al. (2008)	Sine wave (speech) vs. sine wave (nonspeech)	−51	−43	+5	L STS	Active

The contrast for the present study listed in the table represents (listen-clear − listen-masked). "Passive" means speech perception in a passive listening mode. "Active" means speech perception with active tasks involved, such as detection or discrimination of stimulus items. The previous studies are classified by the nature of their task (passive or active) and listed in chronological order within each task category. These studies are all fMRI studies except for the studies specified as PET, which indicates positron emission tomography. CV = consonant–vowel; CVC = consonant–vowel–consonant; Vco = noise-vocoded speech; Rsp = spectrally rotated normal speech; RVco = spectrally rotated noise-vocoded speech.

that noise would effectively mask speech (Houde & Jordan, 2002).

fMRI data were collected on a 3-T Siemens Trio MRI system, using a rapid sparse-imaging procedure (Orfanidou, Marslen-Wilson, & Davis, 2006). To hear responses and to minimize acoustic interference, we presented stimuli and recorded responses (i.e., a single trial occurred) in a 1400-msec silent period between successive 1600-msec scans (EPI; 26 slices; voxel size = 3.3 × 3.3 × 4.0 mm). A high-resolution T1-weighted MPRAGE structural scan was also acquired on each subject.

In the two production conditions, participants spoke into the optical microphone (Phone-Or; Magmedix, Fitchburg, MA), and their utterances were digitized at 10 Hz with 16-bit precision using a National Instruments PXI-6052E input/output board. Real-time analysis was achieved using a National Instruments PXI-8176 embedded controller. Processed signals were converted back to analog by the input/output board at 10 kHz with 16-bit precision and played over high-fidelity magnet-compatible headphones (NordicNeuroLab, Bergen, Norway) in real time. The processing delays were short enough (iteration delay less than 10 msec) that they would not be noticeable to listeners. The processed signals were also recorded and stored on a computer to be used for the yoked trials of the listen-only conditions. Sound was played at a comfortable listening level (approximately 85 dB; see Figure 1A).

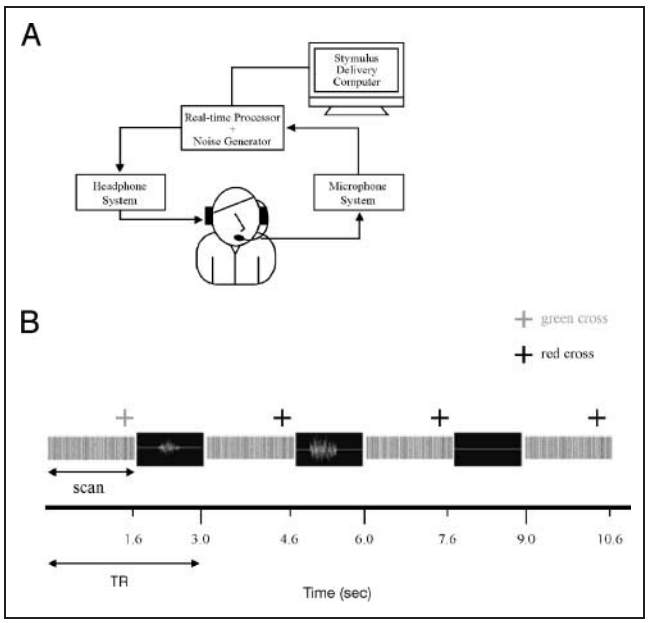

Figure 1. (A) Schematic diagram of the hardware used for the experiment. (B) Schematic diagram of three trials from a functional run. The trials are (in order) production-clear, listen-masked, and rest. The cross appeared on the screen 200 msec before scan offset and cued the subject to either whisper "Ted" (light cross) or remain silent (dark cross). The 1600-msec long scans were separated by the 1400-msec long silent periods permitting speaking and listening.

In the production-clear condition, the whispered utterance was simply digitized and recorded and played out unaltered. The masking noise in the production-masked condition was produced by applying the amplitude envelope of the utterance to uniform Gaussian white noise so that the resulting noise would have the envelope of the original speech signal or signal-correlated noise (Schroeder, 1968). The masking noise was also temporally gated with the onset and offset of speaking. All our subjects reported in the exiting interviews that they could only hear masking noise and not their own speech through the headphones when they were speaking in the production-masked condition.

Participants were scanned in three functional runs, each lasting 9 min and comprising 180 trials; 36 of each of the five conditions. Conditions were presented in pseudorandom order, with the limitation that transitional probabilities were approximately equal and all five conditions were presented once in a block of five trials (conditions could repeat at the transition from one block of five trials to the next). The stimuli for listening trials were taken from the production trials in the preceding block, except in the first run when the stimuli for listening trials were from the production trials in the same block. Trials were 3000 msec long, comprising a 1400-msec period without scanning, followed by a 1600-msec whole-brain EPI acquisition (see Figure 1B). Each trial began with a fixation cross appearing in the middle of a black screen (viewed through mirrors placed on the head coil and in the scanner bore) 100 msec before the offset of the previous scan; this signaled the beginning of the trial, and the color of the cross indicated whether the volunteer should whisper "Ted" (if green) or remain silent (if red). Of the five conditions types, the two production conditions were cued with a green cross, and the two listening conditions and the rest condition were cued with a red cross (see Figure 1B). Pilot testing in five subjects revealed that it took volunteers at least 200 msec to respond to the green or red cue, so we could present it 100 msec before the end of the scan and still ensure that subjects (on production trials) would not commence speaking during the scan, thereby using our 1400-msec silent period effectively.

Subjects practiced the whispering-on-cue task before scanning commenced, and we monitored their behavior during scanning: Subject's vocal production and auditory feedback signal were segregated into two different channels and therefore monitored and recorded simultaneously. Their performance in each trial was monitored in real time in the control room for possible errors. In addition, we also inspected the recordings of both production and auditory feedback in each trial afterward to ensure that every incorrect trial was identified and properly counted. The auditory feedback was not picked up by the microphone, and thus there was no danger of the recorded utterances being masked by it. Sessions in which the error rate exceeded 5% were excluded from analysis, as explained in the Results section.

Analysis

SPM2 (www.fil.ion.ucl.ac.uk/spm/spm2.html) was used for data analysis and visualization. Data were first realigned, within subjects, to the first true functional scan of the session (after discarding two dummy scans), and individual's structural image was coregistered to the mean fMRI image. The coregistered structural was spatially normalized to the ICBM 152 T1 template, and the realigned functional data were normalized using the same deformation parameters. The fMRI data were then smoothed using a Gaussian kernel of 10 mm (FWHM).

Data from each subject were entered into a fixed-effects general linear model using an event-related analysis procedure (Josephs & Henson, 1999). Four event types were modeled for each run. We included six parameters from the motion correction (realignment) stage of preprocessing as regressors in our model to ensure that variability due to head motion was properly accounted for. We chose the hemodynamic response function as the basis function. High-pass filter (cutoff = 128 sec) and AR(1) correction for serial autocorrelation were applied. Contrast images assessing main effects, simple effects, and interactions were created for each subject, and these were entered into random-effects analyses (one-sample t tests) comparing the mean parameter-estimate difference over subjects to zero. Clusters were deemed significant if they exceeded a statistical threshold of $p < .05$ after correction for multiple comparisons at the cluster level.

RESULTS

Behavior

Errors occurred when volunteers (a) spoke after a red cross, (b) remained silent after a green cross, or (c) spoke so quietly that the gated noise was not triggered on production-masked trials. We excluded any trial in which an error occurred. In addition, we excluded from analysis any run in which errors exceeded 5% of the trials; this happened in three runs—one run from each of three individuals. Performance exceeded 95% correct in all remaining runs. For the three individuals with missing data, contrast images were computed at the single-subject level from the remaining two runs, and these were included in the random-effects analyses across subjects.

Functional Data

We analyzed main effects of our two factors (production vs. listening and speech feedback vs. noise feedback) and interactions between these factors at the group level. When interactions were significant, we also analyzed simple effects. We started by reporting interactions because they are the main focus of the study and because their presence influences the interpretation of the main effects.

Interactions

(Production-masked − Production-clear) versus (Listen-masked − Listen-clear)

We reasoned that regions involved in processing auditory feedback during talking ought to exhibit a greater increase in activity for noise compared with normal speech, when these are heard as the auditory concomitants of one's own utterances compared with when one is simply listening to them, without talking. We observed such an interaction in the posterior superior temporal gyrus (STG) bilaterally (see Table 2 and Figure 2). To better understand how differences among conditions produced this significant interaction, we explored the simple effects that constitute this interaction. We observed that, within the bilateral STG regions where this interaction yielded significant activity, activation for production-masked trials was significantly greater than for production-clear trials. In the left hemisphere, there was a significant cluster (see Table 3) in the posterior STG, extending into the middle temporal gyrus (MTG) and posteriorly into the supramarginal gyrus. In the right hemisphere, there was one cluster in the STG. The contrast of listen-clear − listen-masked yielded activation that was strongly lateralized to the left hemisphere, with a peak cluster in the left MTG extending into the middle STS and a cluster in the right anterior STG (see Table 4). The left MTG activation peak for this contrast is also in the neighborhood of areas shown in previous studies that contrasted speech stimuli with nonspeech sounds (see Table 1 and Figure 3). The overlap between regions exhibiting a significant Feedback Type × Production/Perception interaction and regions sensitive to speech (activated more by speech than noise during passive listening) is shown in Figure 4. As can be seen, a region of the STG/MTG is sensitive to speech and exhibits increased activation during production when masking noise is heard.

(Production-clear − Production-masked) versus (Listen-clear − Listen-masked)

The opposite interaction contrast revealed areas where normal feedback yielded more activation than masking noise during talking compared with passive listening. These regions included the left inferior frontal gyrus, the left

Table 2. Areas in which the Difference in Activity between the Production-masked and Production-clear Conditions Is Significantly Greater Than the Difference in Activity between Listen-masked and Listen-clear Conditions

x	*y*	*z*	*t Statistic*	*No. Voxels*	*Anatomical Area*
−66	−45	15	6.50	293	L Posterior STG
54	−18	−6	5.88	347	R STG

Cluster peaks are reported if they exceeded a statistical threshold of $p < .05$ after correction for multiple comparisons at the cluster level, except for Table 6 where we used a more stringent threshold of $p < 10^{-11}$.

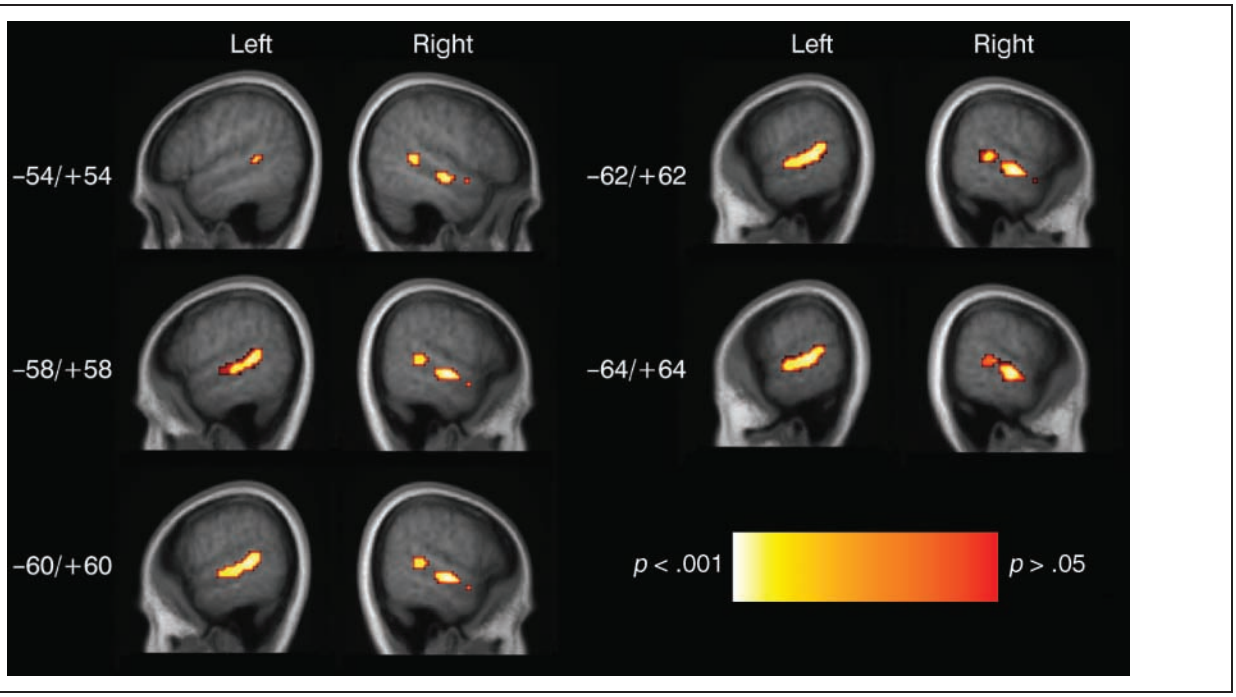

Figure 2. Areas in which the difference in activity between the production-masked and production-clear conditions is significantly greater than the difference in activity between listen-masked and listen-clear conditions. Results are shown at $p < .001$, uncorrected.

superior parietal lobule, the left anterior cingulum, the left middle occipital gyrus, the left caudate, and the right fusiform (see Table 5). Again, to understand this interaction, we explored the simple effects. The contrast of production-clear − production-masked yielded no significant activation. However, the contrast of listen-masked − listen-clear revealed significant clusters in regions highlighted by the interaction (see Table 5). Thus, the interaction appears to be due to a greater difference in these areas when passively listening to noise compared with speech, relative to hearing these sounds while talking.

Main Effects

Production versus Listening: (Production-clear + Production-masked) versus (Listen-clear + Listen-masked)

Production–listening activated an extremely large region involving both hemispheres centered on the left inferior frontal gyrus. Much greater activity during production than listening is consistent with a number of previous studies of speech motor control (Dhanjal, Handunnetthi, Patel, & Wise, 2008; Riecker et al., 2005; Blank, Scott, Murphy, Warburton, & Wise, 2002; Wise, Greene, Buchel, & Scott, 1999; Wildgruber, Ackermann, Klose, Kardatzki, & Grodd, 1996). For example, Wilson, Saygin, Sereno, and Iacoboni (2004) threshold activation maps at $p < 10^{-4}$ for listening conditions and at $p < 10^{-12}$ for speech production conditions to achieve comparable levels of activity. This suggests that speech production must have yielded much more activation relative to listening in their study. When we increased the threshold to $p < 10^{-11}$, we observed clusters in the left inferior frontal gyrus, left postcentral gyrus, and right thalamus (see Figure 5 and Table 6), consistent with previous studies investigating vocal production (Moser et al., 2009; Riecker, Brendel, Ziegler, Erb, & Ackermann, 2008; Kleber, Birbaumer, Veit, Trevorrow, & Lotze, 2007). The reverse contrast, in which activity during production conditions was subtracted from that during listening conditions, did not reveal any significant activation.

Speech versus Noise: (Production-clear + Listen-clear) versus (Production-masked + Listen-masked)

We did not observe significant activation at the whole-brain level when comparing speech with noise. This is probably due to the strong Feedback Type × Production/Perception

Table 3. Areas That Show Increased Activation for Production-masked Relative to Production-clear

x	*y*	*z*	*t Statistic*	*No. Voxels*	*Anatomical Area*
−63	−45	18	5.97	248	L Posterior STG
51	−54	21	5.62	118	R STG

Table 4. Areas That Are More Activated for Listen-clear Than for Listen-masked

x	*y*	*z*	*t Statistic*	*No. Voxels*	*Anatomical Area*
60	−15	−6	4.88	138	R STG
−66	−36	3	4.73	162	L MTG

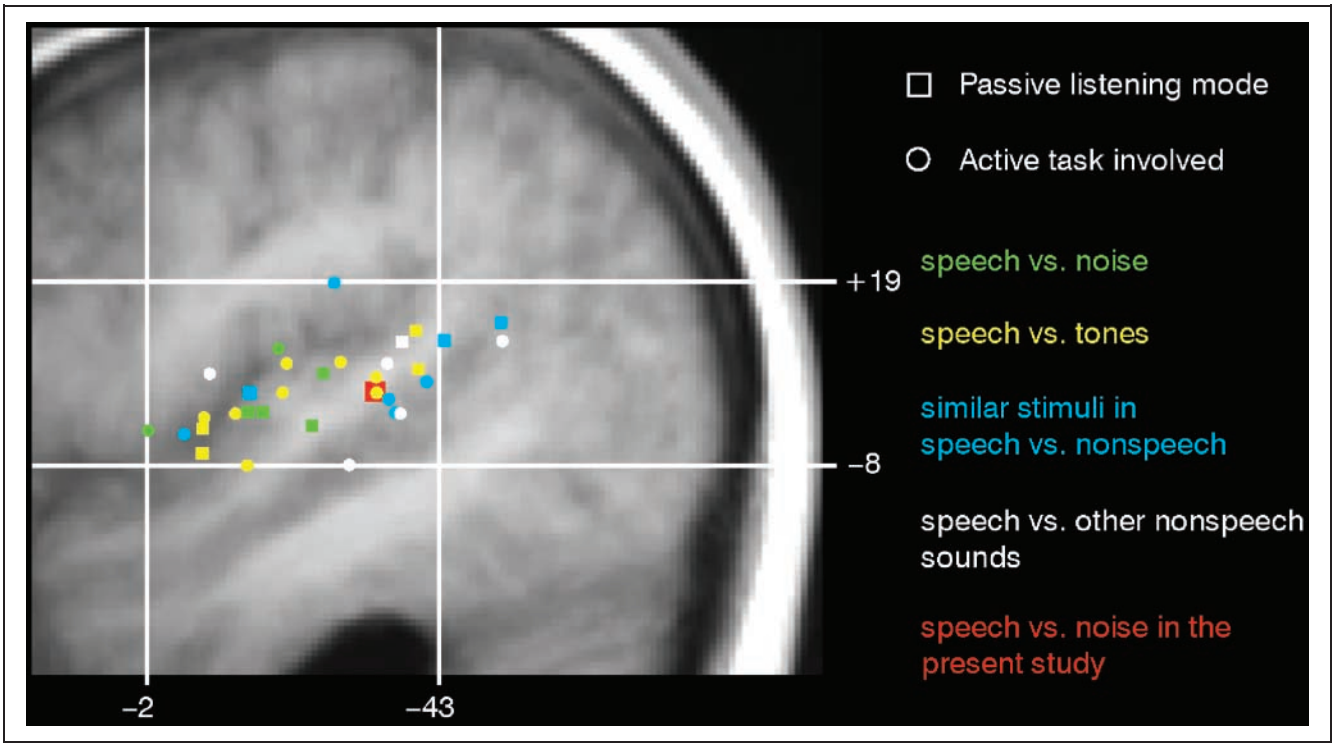

Figure 3. The location of peak activity in the left temporal lobes from the studies listed in Table 1. We used squares and circles to differentiate studies exploring speech perception under passive listening mode from those using active tasks such as target detection or discrimination. Green: studies comparing speech stimuli with noise. Yellow: studies contrasting speech with tones. Blue: studies adopting similar or identical stimuli both for the speech condition and nonspeech condition. White: studies comparing speech with other nonspeech sounds such as environmental sounds, laugh bursts, or hums. Red: the location of the left MTG peak revealed by the contrast of listening-clear versus listening-masked in the present study. The sagittal slice shown is at $x = -58$.

interaction in speech-sensitive regions described earlier; the effect of this crossover interaction is to markedly attenuate the main effect, rendering it not significant. As noted above, during passive listening, signal increases in response to speech as opposed to noise were statistically significant in a left MTG region that many other studies have identified as speech sensitive.

Noise versus Speech: (Production-masked + Listen-masked) versus (Production-clear + Listen-clear)

Brain regions more responsive to noise than to speech include the left inferior frontal gyrus, the left postcentral gyrus, the left putamen, the left cerebellum, and the right supramarginal gyrus (see Table 7). Significant activation in auditory regions for this contrast was not observed.

DISCUSSION

In the bilateral STG, greater activity is elicited during speech production when the auditory concomitants of one's own speech are masked with noise compared with when they are heard clearly. These regions, together with more inferior STS/MTG regions, also exhibit greater activity when listening to speech compared with noise, consistent with them being speech sensitive. That the same STG/MTG region is recruited both for the perception of speech and for processing an error signal during production when the predictive signal and auditory concomitant do not match implies that speech-sensitive regions are implicated in an on-line articulatory control system.

The signal modulation that we observed here was an increase for masked speech compared with clear speech during production but not listening. Although the relationship between hemodynamic response and neural activity is uncertain, this pattern is consistent with neurophysiological work, which demonstrates a release from neural suppression when feedback is altered during production. For example, Eliades and Wang (2008) have observed that, in marmosets, the majority of auditory cortex neurons exhibited suppression during vocalization. However, this vocalization-induced suppression of neural activity was significantly reduced when auditory feedback was altered during vocal production but not during passive listening (Eliades & Wang, 2005, 2008). The attenuation of suppression implies changes in balance of excitatory and inhibitory processes, which would affect regional metabolic energy demands and alter vascular response (Logothetis, 2008). This could in principle manifest as changes in the blood oxygen level that would be detectable as fMRI BOLD signal. Our observation of activity in the bilateral posterior STG in response to altered feedback in production is consistent with such a pattern of neural activity. Models of speech

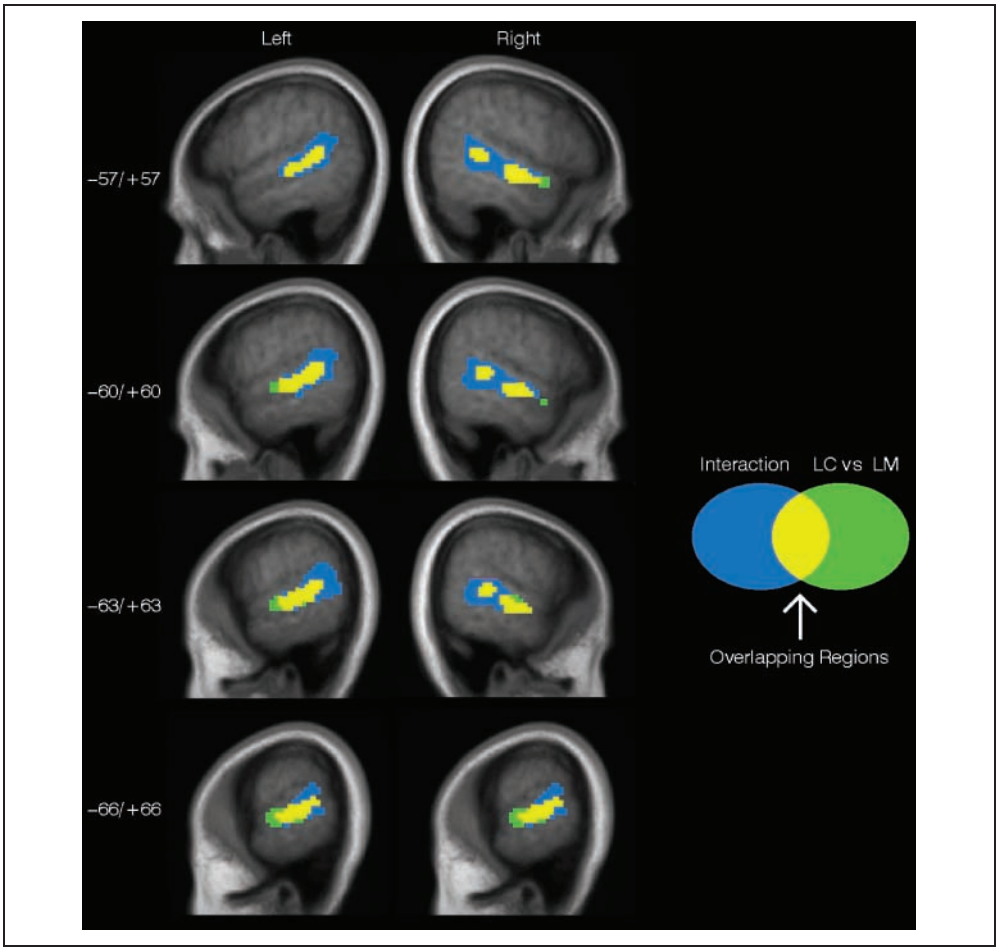

Figure 4. Overlap between areas that are speech sensitive and areas supporting feedback processing. Regions that are activated more by speech than by noise during passive listening are shown in green. Areas in which the difference in signal between masked and clear speech is significantly greater in production than in listening conditions (interaction) are shown in red. Overlapping areas are shown in yellow. Results are shown at $p < .001$, uncorrected.

motor control (e.g., Guenther, 2006) suggest that vocal motor centers generate specific predictions about the expected sensory consequences of articulatory gestures, which are compared with the actual sensory outcome. Elevated activity in the STG in the presence of mismatched auditory feedback may reflect the involvement of this region in an error detection mechanism when vocalization occurs (Bays, Flanagan, & Wolpert, 2006; Sommer & Wurtz, 2006; Matsuzawa, Matsuo, Sugio, Kato, & Nakai, 2005; Blakemore, Wolpert, & Frith, 1998).

The observed bilateral STG regions (and in particular the left posterior STG) in the interaction contrast are in accordance with a number of previous studies of speech processing (see Table 1), including studies in which speech and nonspeech stimuli are matched acoustically [e.g., Dehaene-Lambertz et al., 2005: (−60, −24, +4); Mottonen et al., 2006: (−61, −39, +2); Narain et al., 2003: (−52, −54, +14); and Scott et al., 2006: (−60, −44, +10)]. This lends support to the idea that this area is sensitive, if not specific, for speech sounds.

In the macaque monkey, the extreme capsule (EmC) interconnects the rostral part of the lateral and medial frontal cortex with the midpart of the STG and the cortex of the STS. These frontal regions are connected with inferior parietal cortex through the middle longitudinal fasciculus (MdLF; Makris & Pandya, 2009; Petrides & Pandya, 1988, 2006). This EmC–MdLF pathway, spanning frontal–parietal–temporal cortex, has been suggested to play a crucial role in language functions (Makris & Pandya, 2009; Schmahmann et al., 2007). Another fiber system, the superior longitudinal fasciculus–arcuate fasciculus, courses between the caudal-dorsal pFC and the caudal part of the

Table 5. Areas in which the Difference in Activity between the Production-clear and the Production-masked Conditions Is Significantly Greater Than the Difference in Activity between Listen-clear and Listen-masked Conditions

x	*y*	*z*	*t Statistic*	*No. Voxels*	*Anatomical Area*
−39	−18	18	6.26	91	L Inferior frontal gyrus
36	−84	18	6.02	180	R Middle occipital gyrus
−27	0	24	5.34	38	L Caudate nucleus
30	−54	−9	5.02	70	R Fusiform gyrus
−18	−63	54	4.85	73	L Superior parietal lobule
−18	39	12	4.77	51	L Anterior cingulate
−30	−96	21	4.38	36	L Middle occipital gyrus

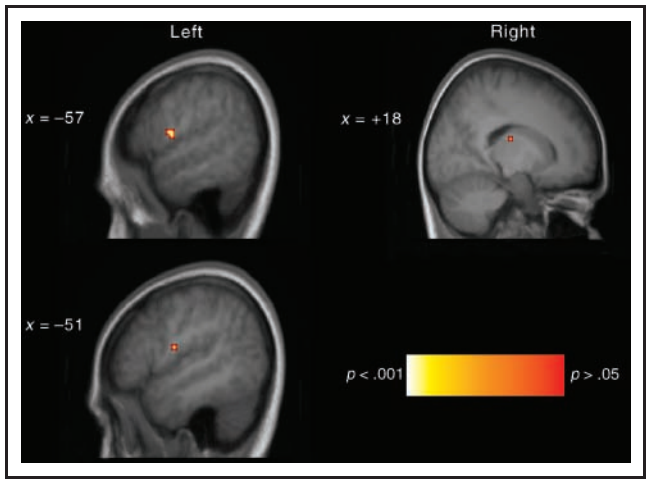

Figure 5. Areas that are activated more for production than for listening conditions. Results are shown at $p < 10^{-11}$, uncorrected.

STG. This pathway has been implicated in sensory-motor processing (Makris & Pandya, 2009). A recent study using diffusion tensor imaging-based tractography in humans has found that sublexical repetition of speech is mediated by the superior longitudinal fasciculus–arcuate fasciculus pathway, whereas auditory comprehension is subserved by the EmC–MdLF system (Saur et al., 2008). Therefore, the anatomical location of the posterior STG demonstrating signal increases for masked speech over clear speech during production, and the opposite during passive listening (i.e., the interaction), is such that it may be involved in both of these processing pathways, thereby serving both functions of speech processing and sensory-motor integration (Okada & Hickok, 2006; Buchsbaum et al., 2005; Warren, Wise, & Warren, 2005; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Okada, Smith, Humphries, & Hickok, 2003; Scott & Johnsrude, 2003).

A number of previous neuroimaging studies have reported bilateral STG activation in response to manipulated auditory feedback compared with normal feedback during reading aloud (Christoffels et al., 2007; Toyomura et al., 2007; Fu et al., 2006; Hashimoto & Sakai, 2003; McGuire, Silbersweig, & Frith, 1996). One feature shared by our study and these studies is that we all manipulated auditory feedback in a way that caused it to be different from what was expected. What distinguishes our study is that we have crossed the feedback-type manipulation common to these studies (normal vs. altered feedback) with a task manipulation (production vs. listening). It is precisely the specificity of this interaction that indicates that activity in brain regions involved in speech perception is modulated by speech production.

Table 6. Areas That Are More Activated for Production (Production-clear + Production-masked) Than for Listening (Listen-clear + Listen-masked)

x	*y*	*z*	*t Statistic*	*No. Voxels*	*Anatomical Area*
−57	9	9	14.55	4	L Inferior frontal gyrus
−51	−6	15	14.26	1	L Postcentral gyrus
18	−18	18	13.87	2	R Thalamus

Cluster peaks are reported if they exceeded a statistical threshold of $p < 10^{-11}$ after correction for multiple comparisons at the cluster level.

Table 7. Areas That Are More Responsive to Production-masked + Listen-masked Than to Production-clear + Listen-clear

x	*y*	*z*	*t Statistic*	*No. Voxels*	*Anatomical Area*
57	−39	24	6.53	284	R Supramarginal gyrus
−24	−6	18	5.13	30	L Putamen
−51	−6	9	5.39	25	L Inferior frontal gyrus
−57	−24	27	4.76	63	L Postcentral gyrus
−18	−60	−24	4.62	20	L Cerebellar hemisphere

We used whispered speech to control the bone-conducted acoustic signals in the masking noise condition more effectively (Houde & Jordan, 2002; Paus, Perry, Zatorre, Worsley, & Evans, 1996). Previous studies indicate that the auditory regions activated by whispered speech are very similar to those activated by vocalized speech, although the level of activation could be somewhat different (Haslinger et al., 2005; Schulz, Varga, Jeffires, Ludlow, & Braun, 2005). This suggests that the use of whispered speech did not qualitatively affect our results.

Vocal production invariably entails both auditory and somatosensory goals (Nasir & Ostry, 2008). We did not manipulate somatosensory signals due to the complexity of such experimental setups (e.g., Tremblay, Shiller, & Ostry, 2003). In addition, because speech production in our study only involves repetitively vocalizing a single consonant–vowel–consonant word ("Ted"), what we have observed is undoubtedly an underestimate of the functional network subserving the speech production and speech perception. Future work could investigate how speech production and perception are coupled using somatosensory perturbations and more naturalistic speech stimuli.

Conclusions

We have observed enhanced activity in the STG region bilaterally during speech production when the auditory feedback and the predicted auditory consequences of speaking do not match. The same region is sensitive to speech during listening. This suggests a self-monitoring/feedback system at work, presumably involved in controlling on-line articulatory planning. Furthermore, the network supporting speech perception appears to overlap with this self-monitoring system, in the STG at least, highlighting the intimate link between perception and production.

Acknowledgments

The authors thank the referees for their valuable comments. This work was supported by an R-01 operating grant from the U.S. National Institutes of Health, NIDCD grant DC08092 (K. M.), and by grants from the Canadian Institutes of Health Research (I. J.), the Natural Sciences and Engineering Research Council of Canada (I. J. and K. M.), the Ontario Ministry of Research and Innovation, and the Queen's University (I. J. and Z. Z.). I. J. was supported by the Canada Research Chairs Program.

Reprint requests should be sent to Zane Z. Zheng, Centre for Neuroscience Studies, Queen's University, 62 Arch Street, Kingston, Ontario, Canada K7L 3N6, or via e-mail: 5zz2@queensu.ca.

REFERENCES

Ashtari, M., Lencz, T., Zuffante, P., Bilder, R., Clarke, T., Diamond, A., et al. (2004). Left middle temporal gyrus activation during a phonemic discrimination task. *NeuroReport, 15,* 389–393.

Barany, E. (1938). A contribution to the physiology of bone conduction. *Acta Oto-Laryngologica, 26,* 1–228.

Bays, P. M., Flanagan, J. R., & Wolpert, D. M. (2006). Attenuation of self-generated tactile sensations is predictive, not postdictive. *PLoS Biology, 4,* e28.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13,* 17–26.

Benson, R. R., Richardson, M., Whalen, D. H., & Lai, S. (2006). Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *Neuroimage, 31,* 342–353.

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 and Language, 78,* 364–396.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10,* 512–528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain, 119,* 1239–1247.

Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., et al. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology, 35,* 662–672.

Black, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders, 16,* 56–60.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience, 1,* 635–640.

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain, 125,* 1829–1838.

Buchsbaum, B. R., Olsen, R. K., Koch, P. F., Kohn, P., Kippenhan, J. S., & Berman, K F. (2005). Reading, hearing, and the planum temporale. *Neuroimage, 24,* 444–454.

Burton, M. W., & Small, S. L. (2006). Functional neuroanatomy of segmenting speech and nonspeech. *Cortex, 42,* 644–651.

Burton, M. W., Small, S. L., & Blumstein, S. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience, 12,* 679–690.

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of the Acoustical Society of America, 121,* 1157–1163.

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping, 28,* 868–879.

Cowie, R., Douglas-Cowie, E., & Kerr, A. G. (1982). A study of speech deterioration in post-lingually deafened adults. *Journal of Laryngology and Otology, 96,* 101–112.

Dehaene-Lambertz, G., Pallier, C., Semiclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage, 24,* 21–33.

Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience, 20,* 1174–1188.

Dhanjal, N. S., Handunnetthi, L., Patel, M. C., & Wise, R. J. (2008). Perceptual systems controlling speech production. *Journal of Neuroscience, 28,* 9969–9975.

Eliades, S. J., & Wang, X. (2005). Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex, 15,* 1510–1523.

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature, 453,* 1102–1106.

Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *Journal of the Acoustical Society of America, 70,* 45–50.

Fu, C. H., Vythelingum, G. N., Brammer, M. J., Williams, S. C., Amaro, E., Jr., Andrew, C. M., et al. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex, 16,* 969–977.

Fukawa, T., Yoshioka, H., Ozawa, E., & Yoshida, S. (1988). Difference of susceptibility to delayed auditory feedback between stutterers and nonstutterers. *Journal of Speech and Hearing Research, 31,* 475–479.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X. J., et al. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping, 20,* 185–200.

Geers, A. E., Nicholas, J. G., & Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear and Hearing, 24,* 46S–58S.

Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience, 13,* 754–765.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39,* 350–365.

Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. *Human Brain Mapping, 20,* 22–28.

Haslinger, B., Erhard, P., Dresel, C., Castrop, F., Roettinger, M., & Ceballos-Baumann, A. O. (2005). "Silent event-related" fMRI reveals reduced sensorimotor activation in laryngeal dystonia. *Neurology, 65,* 1562–1569.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music and working memory in area Spt. *Journal of Cognitive Neuroscience, 15,* 673–682.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279,* 1213–1216.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech: I. Compensation and adaptation. *Journal of Speech, Language, and Hearing Research, 45,* 295–310.

Howell, P., & Powell, D. J. (1987). Delayed auditory feedback with delayed sounds varying in duration. *Perception and Psychophysics, 42,* 166–172.
Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage, 15,* 733–746.
Joanisse, M. F., & Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage, 19,* 64–79.
Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America, 108,* 1246–1251.
Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16,* 307–354.
Josephs, O., & Henson, R. N. (1999). Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 354,* 1215–1228.
Kawato, M. (1989). Motor theory of speech perception revisited from the minimum torquechange neural network model. In *8th Symposium on Future Electron Devices*, Tokyo, Japan, pp. 141–150.
Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage, 36,* 889–900.
Lashley, K. S. (1951). The problem of serial order in behaviour. In L. A. Jefree (Ed.), *Cerebral mechanisms in behaviour*. New York: Wiley.
Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41–104.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.
Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. T. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience, 15,* 549–558.
Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex, 15,* 1621–1631.
Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453,* 869–878.
Mackay, D. G. (1968). Metamorphosis of a critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. *Journal of the Acoustical Society of America, 43,* 811–821.
Makris, N., & Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Structure & Function, 213,* 343–358.
Matsuzawa, M., Matsuo, K., Sugio, T., Kato, C., & Nakai, T. (2005). Temporal relationship between action and visual outcome modulates brain activation: An fMRI study. *Magnetic Resonance in Medical Sciences, 4,* 115–121.
McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain, 119,* 907–917.
Meyer, M., Zysset, S., Yves von Cramon, D., & Alter, K. (2005). Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Cognitive Brain Research, 24,* 291–306.
Moser, D., Fridriksson, J., Bonilha, L., Healy, E. W., Baylis, G., Baker, J. M., et al. (2009). Neural recruitment for the production of native and novel speech sounds. *Neuroimage, 46,* 549–557.
Mottonen, R., Calvert, G. A., Jaaskelainen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage, 30,* 563–569.
Mummery, C. J., Ashbumer, J., Scott, S. K., & Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustical Society of America, 106,* 449–457.
Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex, 13,* 1362–1368.
Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience, 11,* 1217–1222.
Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., et al. (2006). Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping, 27,* 562–571.
Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language, 98,* 112–117.
Okada, K., Smith, K. R., Humphries, C., & Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *NeuroReport, 14,* 2323–2326.
O'Leary, D. S., Andreason, N. C., Hurtig, R. R., Hichwa, R. D., Watkins, G. L., Ponto, L. L., et al. (1996). A positron emission tomography study of binaurally and dichotically presented stimuli: Effects of level of language and directed attention. *Brain and Language, 53,* 20–39.
Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of Cognitive Neuroscience, 18,* 1237–1252.
Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience, 8,* 2236–2246.
Petrides, M., & Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *Journal of Comparative Neurology, 273,* 52–66.
Petrides, M., & Pandya, D. N. (2006). Efferent association pathways originating in the caudal prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology, 498,* 227–251.
Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., & Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia, 42,* 183–200.
Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *Journal of the Acoustical Society of America, 120,* 966–977.
Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America, 119,* 2288–2297.
Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language, 107,* 102–113.
Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., et al. (2005). fMRI reveals two distinct cerebral

networks subserving speech motor control. *Neurology, 64,* 700–706.
Rimol, L. M., Specht, K., Weis, S., Savoy, R., & Hugdahl, K. (2005). Processing of sub-syllabic speech units in the posterior temporal lobe: An fMRI study. *Neuroimage, 26,* 1059–1067.
Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 18035–18040.
Schauwers, K., Gillis, S., Daemers, K., De Beukelaer, C., De Ceulaer, G., Yperman, M., et al. (2004). Normal hearing and language development in a deaf-born child. *Otology and Neurotology, 25,* 924–929.
Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., et al. (2007). Association fibre pathways of the brain: Parallel observations from diffusion spectrum imaging and autoradiography. *Brain, 130,* 630–653.
Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735–1736.
Schulz, G. M., Varga, M., Jeffires, K., Ludlow, C. L., & Braun, A. R. (2005). Functional neuroanatomy of human vocalization: An H215O PET study. *Cerebral Cortex, 15,* 1835–1847.
Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400–2406.
Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences, 26,* 100–107.
Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *Journal of the Acoustical Society of America, 120,* 1075–1083.
Siegel, G. M., Schork, E. J., Jr., Pick, H. L., Jr., & Garber, S. R. (1982). Parameters of auditory feedback. *Journal of Speech and Hearing Research, 25,* 473–475.
Sommer, M. A., & Wurtz, R. H. (2006). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature, 444,* 374–377.
Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: An auditory rapid event-related fMRI task. *Neuroimage, 20,* 1944–1954.
Thierry, G., Giraud, A. L., & Price, C. J. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron, 38,* 499–506.
Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage, 39,* 1429–1443.
Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., et al. (2007). Neural correlates of auditory feedback control in human. *Neuroscience, 146,* 499–503.
Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature, 423,* 866–869.
Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage, 31,* 1284–1296.
Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13,* 994–1005.
Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neuroscience, 28,* 636–643.
Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., & Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: An fMRI study. *NeuroReport, 7,* 2791–2795.
Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7,* 701–702.
Wise, R. J. S., Greene, J., Buchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet, 353,* 1057–1061.
Zaehle, T., Wustenberg, T., Meyer, M., & Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fMRI study. *European Journal of Neuroscience, 20,* 2447–2456.
Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256,* 846–849.