Discriminating between Auditory and Motor Cortical Responses to Speech and Nonspeech Mouth Sounds

Zarinah K. Agnew, Carolyn McGettigan, and Sophie K. Scott

Abstract

■ Several perspectives on speech perception posit a central role for the representation of articulations in speech comprehension, supported by evidence for premotor activation when participants listen to speech. However, no experiments have directly tested whether motor responses mirror the profile of selective auditory cortical responses to native speech sounds or whether motor and auditory areas respond in different ways to sounds. We used fMRI to investigate cortical responses to speech and nonspeech mouth (ingressive click) sounds. Speech sounds activated bilateral superior temporal gyri more than other sounds, a profile not seen in motor and premotor cortices. These results suggest that there are qualitative differences in the ways that temporal and motor areas are activated by speech and click sounds: Anterior temporal lobe areas are sensitive to the acoustic or phonetic properties, whereas motor responses may show more generalized responses to the acoustic stimuli. ■

INTRODUCTION

Several recent theories of perceptual processing have identified a central role for motor representations in the recognition of action (Rizzolatti, Fogassi, & Gallese, 2001) and the use of simulation to guide perception (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) and as a basis for mirror responses in the human brain (Rizzolatti & Craighero, 2004). The motor theory of speech perception (Liberman & Mattingly, 1985) has been interpreted as requiring a central role for motor recruitment in speech perception (Galantucci, Fowler, & Turvey, 2006), and several studies have provided evidence for motor cortex activation during speech processing (Pulvermuller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004; Watkins, Strafella, & Paus, 2003; Fadiga, Craighero, Buccino, & Rizzolatti, 2002). However, motor cortex is activated by a wide range of complex sounds (Scott, McGettigan, & Eisner, 2009), and few studies have systematically attempted to whether motor and auditory responses to speech and other mouth sounds might differ (Wilson & Iacoboni, 2006).

Within the temporal lobes, responses lateral to primary auditory cortex respond to both acoustic modulations and acoustic–phonetic structure in speech (Scott, Rosen, Lang, & Wise, 2006; Scott, Blank, Rosen, & Wise, 2000), whereas responses in the STS and beyond are less sensitive to acoustic properties and more driven by the intelligibility of the speech (Scott et al., 2000). In contrast, support for sensory motor involvement in speech perception is provided by studies showing areas coactivated by speech production and perception in premotor cortex (Wilson et al., 2004) and by studies showing increased corticospinal excitability during processing of speech sounds (Fadiga et al., 2002).

Links between motor, somatosensory and acoustic processing have been suggested in the literature. For example, Nasir & Ostry (2009) have shown that subjects who adapt to jaw perturbations when producing speech also show perceptual adaptations, although the precise mechanisms underlying this are still unclear (Houde, 2009). However, the argument that motor cortex has an *essential* role for speech perception (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007) implies a sensitivity to acoustic–phonetic information, as is seen in auditory areas. In line with this, the place of articulation has been shown to be represented in both superior temporal cortex (Obleser, Lahiri, & Eulitz, 2004) and premotor cortex (Pulvermuller et al., 2006). What is harder to establish from these studies is the extent to which the neural responses are specific to speech sounds. It has been suggested that motor responses to perceptual events may reflect more general processes than those needed for object or event recognition (Heyes, 2010). A recent review of the functional neuroimaging (PET and fMRI) literature suggests that motor responses to acoustic stimuli are relatively generic, as opposed to more selective responses seen in auditory areas (Scott et al., 2009). It is thus possible that, unlike the dorsolateral temporal lobes, motor and premotor fields are more generally activated by mouth sounds rather than showing a speech-specific response.

We directly investigated the hypothesis that motor systems are central to the perception of speech by contrasting the neural responses to three kinds of auditory stimuli using fMRI. We used speech sounds, nonspeech mouth sounds, and signal-correlated noise (SCN) analogues of

University College London

both sound categories (in which only the amplitude envelope of the stimuli is preserved; Schroeder, 1968). SCN is a relatively complex auditory baseline, which has been used in several previous studies of auditory and motor responses to speech (Pulvermuller et al., 2006; Davis & Johnsrude, 2003; Mummery, Ashburner, Scott, & Wise, 1999).

We included both speech and nonspeech mouth sounds as they differ in their phonetic properties, while being produced by the same effectors. The nonspeech mouth sounds were four ingressive "click" sounds, which are phonemes in some African languages (e.g., Xhosa). These click sounds cannot be assimilated into English phonetic categories (Best, McRoberts, & Sithole, 1988), and English listeners do not show a right ear advantage for these sounds in dichotic listening paradigms (Best & Avery, 1999). We selected click sounds that are similar to some nonspeech sounds used by English speakers (e.g., similar to a "kissing" sound, a "tutting" sound, a "giddy-up" sound, and a "clopping" sound) and excluded less familiar click sounds, such as voiced nasals. The speech sounds were the unvoiced phonemes "t," "k," "f," and "ch" chosen so there was a mix of manner of articulation (two plosives, one fricative and one affricate) and place of articulation (one labio-dental, two alveolar, and one velar). Unvoiced phonemes were used to afford better matching with the ingressive click sounds, which are unvoiced. The sounds were presented with no associated voiced vowels to avoid the introduction of English phonemes into the "nonspeech" mouth sound category.

This aim of this experiment was to identify whether auditory cortical fields associated with speech perception (identified using a speech perception localizer) and motor and premotor cortical fields associated with speech-related orofacial movement (identified using a silent mouth movement localizer and also from previously reported premotor activation sites; Pulvermuller et al., 2006; Wilson et al., 2004) respond in a similar or a distinct manner to speech, ingressive click sounds, and SCN. We defined a speech-related neural response as one in which there was a preferential response to speech sounds relative to ingressive clicks and SCN. We also explored whether any cortical fields showed a common response to the speech and click sounds, relative to the SCN: Such a generalized response to speech and ingressive click sounds would follow the profile associated with a "voice"-selective response (Belin & Zatorre, 2003; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). We also identified any neural responses that were greater to the ingressive click sounds than to the speech and SCN sounds. Finally, we identified neural areas that showed a more generic response to the acoustic stimuli, being commonly activated by the speech, the click sounds and the SCN.

METHODS

Generation of Stimuli

The speech sounds were voiceless consonants comprising plosives (/t/, /k/), a fricative (/f/) and an affricate (/tʃ/; the phoneme at the start of "cheese"). The plosives (/t/, /k/) are non-continuants that is, are naturally produced with a short post-obstruent unvoiced airflow. The nonspeech mouth sounds comprised four ingressive click sounds: a dental click (/ǀ/), a post-alveolar click (/ǃ/), a lateral click (/ǁ/) and a bilabial click (/ʘ/). Respectively, these are similar to a "tutting" sound (generally written as "tsk-tsk" or "tut-tut"), a "clop," as in the clip-clop sound made when imitating a trotting horse, a "giddy-up" sound, the click sound made to indicate "get going" or "go faster" (e.g., when on a horse), and a "kissing" sound. These were all produced by a native speaker of British English. Thirty tokens of each sound were used in the experiment, and each token was presented once only (Figure 1).

Sounds were recorded using a solid state recorder (Edirol, R-09HR, Roland, Hosoe-cho, Hamamatsu, Japan) at 24 bits, 96 kHz, and saved as .wav files. The sound files were normalized to the same peak amplitude in Praat (Boersma & Weenink, 2010). Sounds were performed by a native British speaker who produced 30 tokens for each category of speech and ingressive click sound. SCN versions (Schroeder, 1968) were used as the baseline stimuli, and these were generated by multiplying the original waveforms with wide band noise between 50 Hz and 10 kHz.

Behavioral Testing

The stimuli were pretested to ensure that subjects could correctly categorize the sounds as speech or nonspeech. Eight subjects (five men, mean age = 25.7 years) listened to the same trains of sounds used in the fMRI section of this experiment before being asked to decide if the trains of sounds were speech or nonspeech sounds (60 trials in total, 30 speech, and 30 click trials). In a second pretest, the experiment was repeated with individual exemplars of each speech and ingressive sound (80 trials in total, each of the eight sounds was tested 10 times). In both tests, the same token was never presented more than once.

Subjects

Twenty-two healthy right-handed subjects (mean = 26.9 years, 11 men) participated in the present study. All were native English speakers, and we excluded any subjects who had experience with click languages (e.g., those having lived in South Africa). All gave informed consent according to the guidelines approved by the University College London Ethics Committee, who provided local ethics approval for this study.

fMRI

A 1.5-T Siemens system with a 32-channel head coil was used to acquire 183 T_2*-weighted EPI data ($3 \times 3 \times 3$ mm^3, repetition time = 10,000 msec, acquisition time = 3 sec,

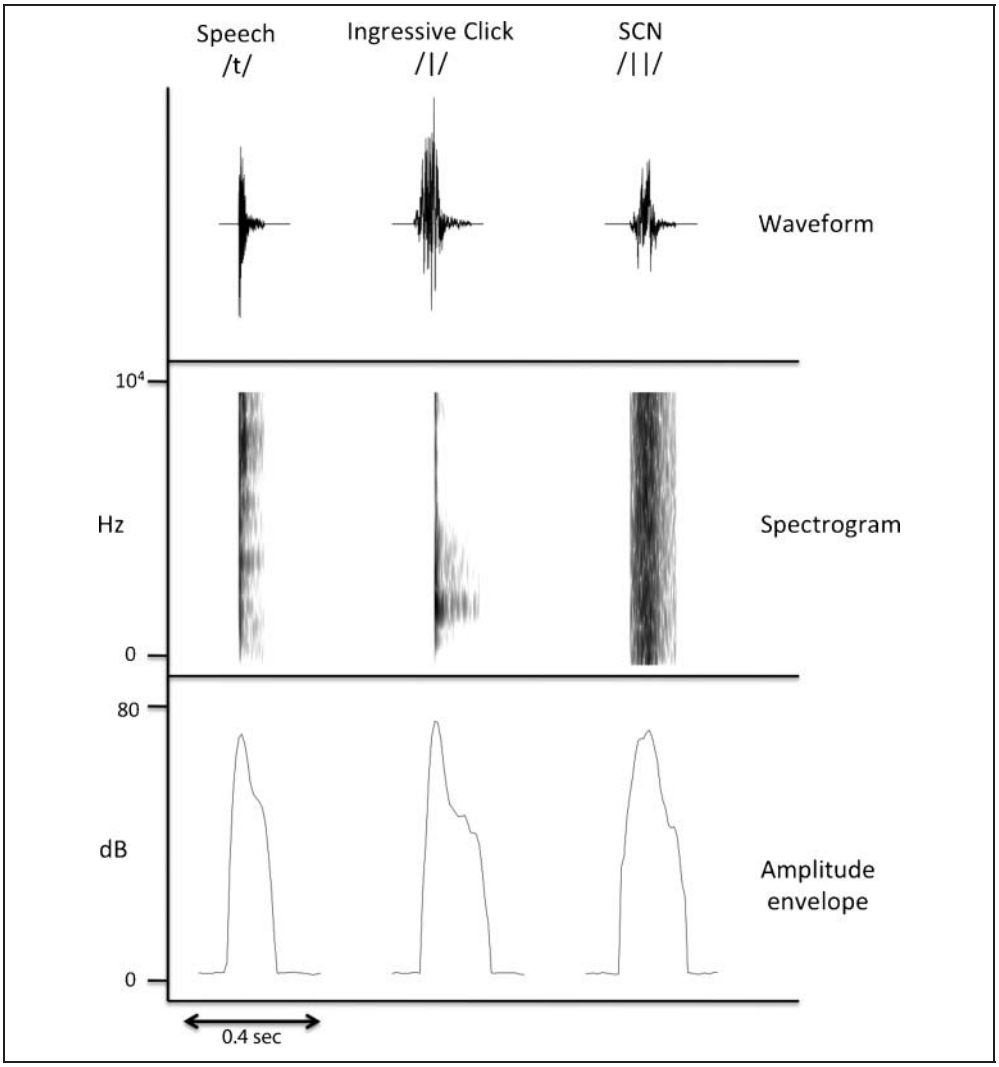

Figure 1. Speech, ingressive clicks, and SCN sounds share similar amplitude envelopes. Examples of tokens from the speech, ingressive click sounds, and SCN conditions used in the experiment. The top shows the waveform versions of the sounds, whereas the middle shows their spectrotemporal structure in the form of a spectrogram. The bottom shows the amplitude envelope, which describes the mean amplitude of the sounds over time. Note that the three tokens possess a similar amplitude envelope and that the SCN token has a much simpler spectral structure than the speech and click sounds (as shown in the spectrogram).

echo time = 50 msec, flip = 90°) using BOLD contrast. The use of a 32-channel head coil has been shown to significantly enhance signal-to-noise ratio for fMRI in the 1.5-T field (Parikh et al., 2011; Fellner et al., 2009). A sparse scanning protocol was employed to administer the auditory stimuli in the absence of scanner noise. The first two functional volumes were discarded to remove the effect of T_1 equilibration. High-resolution T_1 anatomical volume images (160 sagittal slices, voxel size = 1 mm^3) were also acquired for each subject. During the main experimental run, subjects lay supine in the scanner in the dark and were asked to close their eyes and listen to the sounds played to them. There was no task involved so as to avoid any form of motor priming that a response task, such as a button press, might entail (Figure 2).

Sounds for the main run and instructions for the localizer run were presented using MATLAB with the Psychophysics Toolbox extension (Brainard, 1997), via a Denon amplifier (Denon UK, Belfast, UK) and electrodynamic headphones (MR Confon GmbH, Magdeburg, Germany) worn by the participant. Instructions were projected from a specially configured video projector (Eiki International, Inc., Margarita, CA) onto a custom-built front screen, which the participant viewed via a mirror placed on the head coil.

Each trial was a train of four different speech or click sounds, lasting 3 sec (e.g., /t/–/k/–/t∫/–/f/). The order of sounds was randomized within trial and the ordering of sound category (speech, nonspeech, SCN) was randomized across trials. Across the whole experiment, none of the 30 recorded tokens of each speech/mouth sound were repeated. A ±500 msec onset jitter was used. This main run lasted approximately 30 min.

We carried out a separate localizer run to identify in each subject the cortical regions responsible for executing mouth movements and for speech perception. This employed a block design using a continuous acquisition protocol (repetition time = 3 sec). Subjects were cued via instructions on a screen to execute mouth movements (alternating lip and tongue movements) or to listen to sentences taken from the BKB list (Bench, Kowal, & Bamford, 1979). The baseline condition was silent rest. Each block

lasted 21 sec and was repeated four times. This localizer scan lasted approximately 11 min.

Preprocessing and Analyses

Functional data were analyzed using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK) running on Matlab 7.4 (Mathworks, Inc., Sherborn, MA). All functional images were realigned to the first volume by six-parameter rigid body spatial transformation. Functional and structural (T_1-weighted) images were then normalized into standard space using the Montreal Neurological Institute (MNI) template. Functional images were then co-registered to the T_1 structural image and smoothed using a Gaussian kernel of FWHM at 8 mm. The data were high-pass filtered at 128 Hz. First-level analysis was carried out using motion parameters as regressors of no interest at the single-subject level. A random-effects model was employed in which the data were thresholded at $p < .005$. Voxelwise thresholding was carried out at 30 voxels to limit potential Type I errors.

Individual contrasts were carried out to investigate the BOLD response to each condition minus the silent rest or SCN, Speech versus Clicks and Clicks versus Speech. These t contrasts were taken up to a second level model. A null conjunction was used to identify significantly active voxels common to more than one condition by importing contrasts at the group level (e.g., Speech > SCN and Clicks > SCN at a threshold of $p < .005$, cluster threshold of 10). Significant BOLD effects were rendered on a normalized template.

ROI analyses were carried out to investigate mean effect sizes in specific regions across all experimental conditions against a baseline condition using the MarsBar

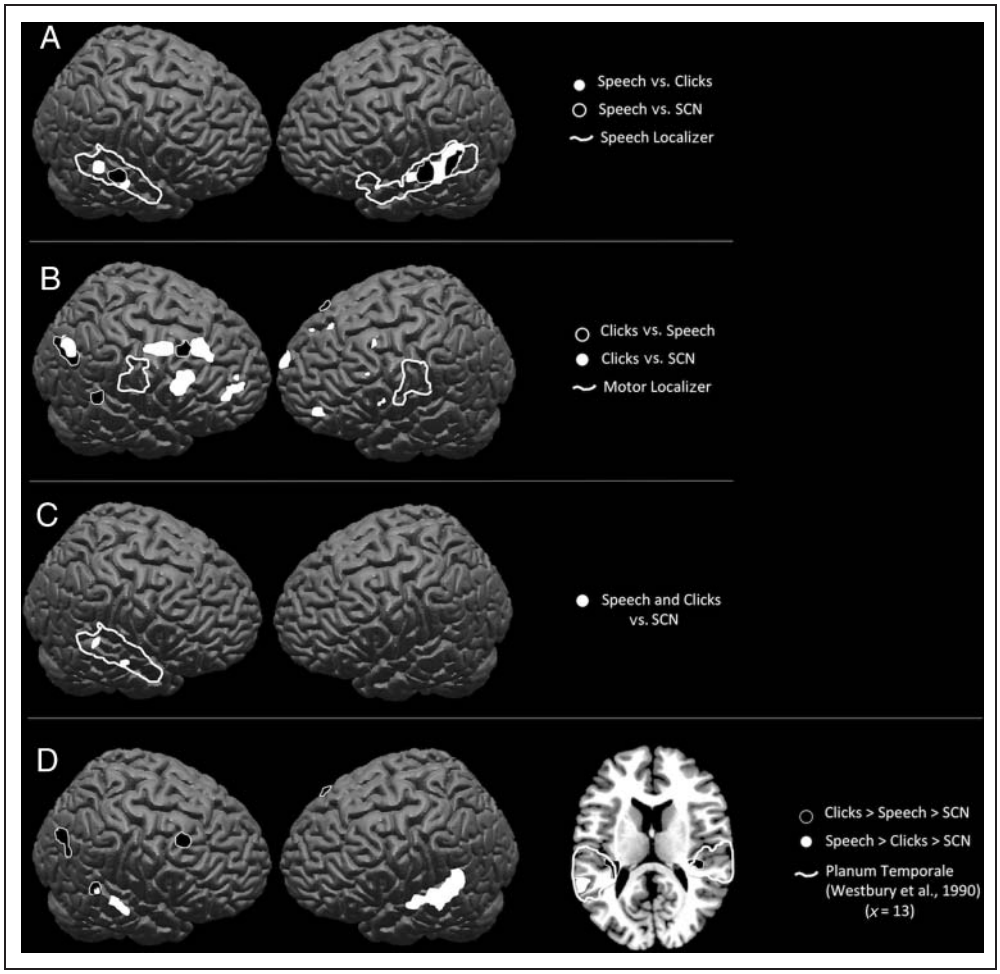

Figure 2. Perception of speech and ingressive click sounds is associated with increased activity in auditory regions. Perception of speech sounds compared with ingressive click sounds (A, white) was associated with increased BOLD activity in left middle and posterior STG ($p < .005$, cluster threshold = 30). Perception of speech sounds compared with SCN was associated with significant activity in the same regions but extending anteriorly in the left hemisphere (A, black) [Speech vs. SCN: −58 −48 19, −44 −6 −11, 62 −14 −4, 60 −34 6; Speech vs. Ingressive clicks: −66 16 0, 60 −20 −2, −68 −36 8, −22 −32 32]. These activations both lay within cortex identified as speech sensitive by an independent speech localizer run (A, white line). Listening to ingressive click sounds compared with speech sounds was associated with significant activity in prefrontal regions and right occipitoparietal cortex (B, black). [Ingressive clicks vs. SCN: 50 −60 28, −32 −34 8, −32 −20 −10, 42 26 50, 28 8 40, 64 −36 8; Ingressive clicks vs. Speech: 22 32 42, −30 58 0, 44 28 24, 40 10 46, 26 64 14, 44 −64 38]. Neither the comparison of click sounds to speech sounds or to SCN revealed significant activity in mouth motor regions identified by an independent motor localizer run (B, white line). (C) The common activity during the perception of both types of sounds compared with SCN in right STG ($p < .005$). These data indicate partially separate networks for processing of speech and ingressive click sounds whereby speech sounds are preferentially processed in left middle STG and ingressive click sounds are associated with increased activity in left posterior medial auditory areas known to comprise part of the dorsal "how" pathway. In contrast there is overlapping activity in right superior temporal cortex to both classes of sound. (D) Regions where there is a preferential response to speech in bilateral dorsolateral temporal lobes, with more extensive activation on the left. These activations were identified by the contrast [1 −0.01 −0.99, for Speech > Clicks > SCN, shown in white]. The same contrast for clicks [Clicks > Speech > SCN] did not reveal any effect in speech sensitive auditory areas in left temporal cortex (black).

toolbox that is available for use within SPM8 (Brett, Anton, Valabregue, & Poline, 2002). ROIs were created in three different ways:

1. A set of four 10-mm spherical ROIs were created from peak coordinates identified from separate motor and auditory localizer runs. These ROIs lay within left and right superior temporal gyri (STG) and within left and right mouth primary motor cortex (−60 −24 6, 72 −28 10, −53 −12 34, 64 0 28). Mean parameter estimates were extracted for speech and clicks compared with SCN. These are seen in Figure 3.
2. An additional set of 8-mm spherical ROIs were created from coordinates reported in two previous studies (Pulvermuller et al., 2006; Wilson & Iacoboni, 2006). These studies both reported significant activity in premotor regions during the perception of speech sounds (−62 −4 38, 56 −4 38, −54 −3 46, −60 2 25; Figure 4B). A diameter of 8 mm was chosen here to replicate the analyses done in these previous experiments. In these regions, mean parameter estimates were extracted for speech and clicks compared with SCN.
3. Finally, two cluster ROIs in ventral sensorimotor cortices were generated by the contrast of all sounds (speech, nonspeech, and SCN) over silent rest. This contrast identified a peak in ventral primary sensorimotor cortex in both hemispheres (Figure 4A). To allow statistical analyses of these data (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Vul, Harris, Winkleman, & Pashler, 2008), ROIs were created in an iterative "hold-one-out" fashion (McGettigan et al., 2011), in which the cluster ROIs for each individual participant were created from a group contrast of [All Sounds vs. Rest inclusively masked by the motor localizer] (masking threshold $p < .001$, cluster threshold = 30) from the other 21 participants. Mean parameter estimates were extracted for speech, clicks, and SCN compared with silent rest.

RESULTS

Subjects Correctly Categorize Speech and Ingressive Click Sounds

The average percentage of correctly classified trains of sounds was 96.3% (*SD* = 3.23%). For the trains of speech and click sounds, 95.56% (*SD* = 1.41%) and 97.04% (*SD* = 2.03%) of sounds were correctly categorized as speech or nonspeech sounds. A two-tailed *t* test showed that these scores were not significantly different ($p = .35$). In a second experiment, the same participants were tested on single sounds, rather than trains of sounds (one subject failed to fill in the form correctly so there were seven subjects for this assessment). A one-way ANOVA demonstrated that there were more within-category miscategorizations for the click sounds (miscategorizations of click sounds as other click sounds) than any other type of error ($p < .05$). Importantly, however, there was no significant difference ($p = .3$) between the number of speech sounds miscategorized as clicks (mean number of errors = 5.25 ± 5.06, $n = 40$) and vice versa (mean number of errors = 1.88 ± 1.08, $n = 40$). This confirms that the ingressive click sounds were not being systematically misclassified

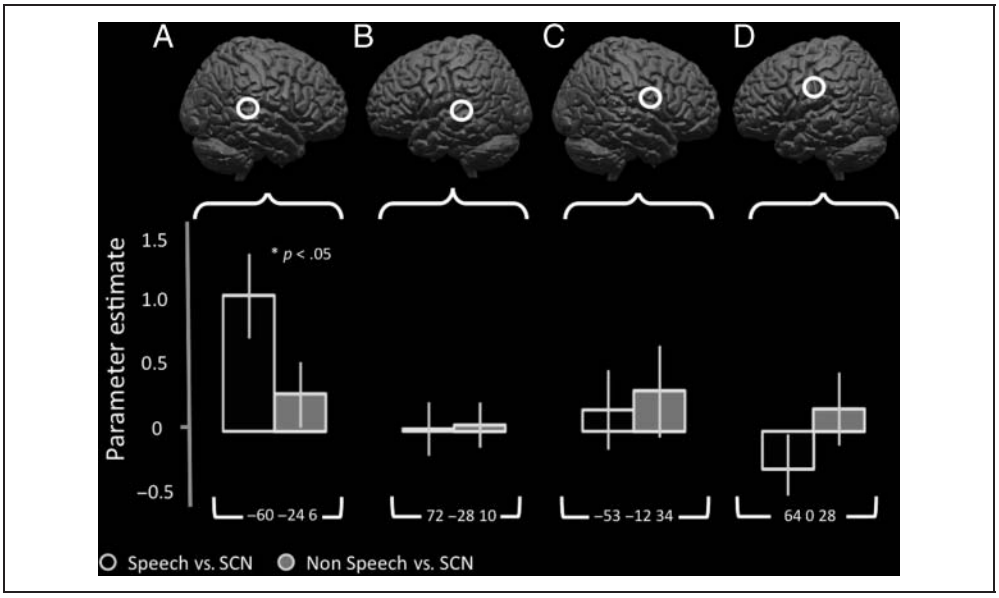

Figure 3. Left auditory areas preferentially encode speech sounds, but there is no speech specific activity in primary motor cortices. Parameter estimates for speech and ingressive click sounds compared with SCN were calculated within four ROIs generated from peak coordinates from an independent localizer. A and B display the left and right speech ROIs generated from the comparison of listening to sentences against a silent rest condition (FWE = 0.05, cluster threshold = 30) with the parameter estimates displayed below. C and D show the left and right mouth motor ROIs generated from alternating lip and tongue movements compared with silent rest (FWE = 0.05, cluster threshold = 30). Speech sounds were associated with significantly increased activity in left auditory cortex compared with ingressive click sounds. There was nonsignificant difference in levels of activity in right auditory cortex or in the mouth motor regions. In all three of these regions, there was a nonsignificant increase in activity for ingressive click sounds over SCN compared with speech sounds over SCN. Error bars indicate *SEM*.

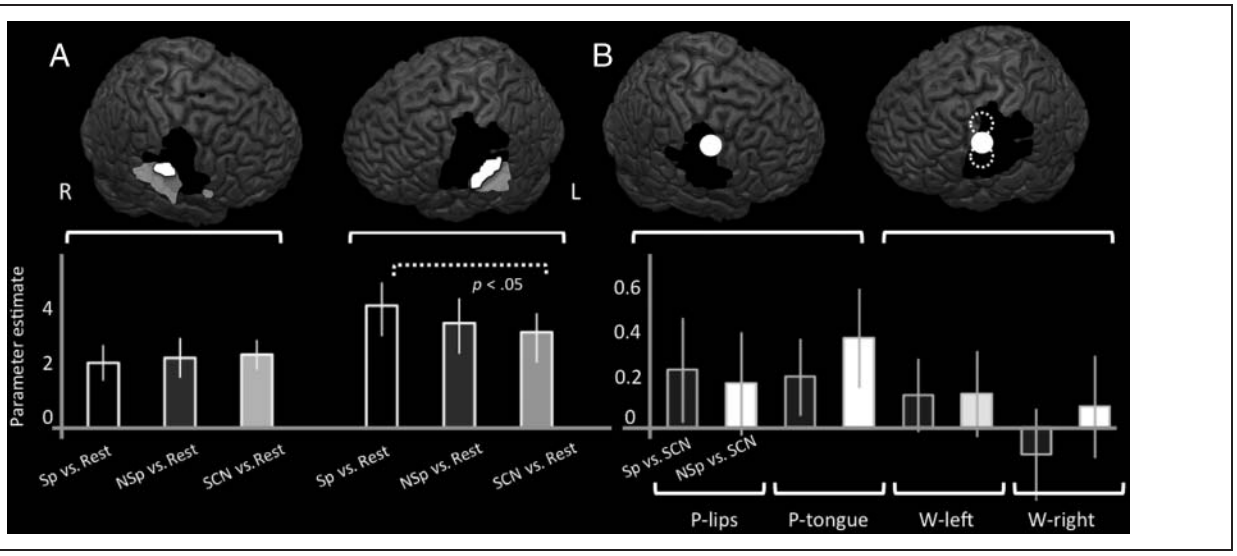

Figure 4. Auditory-sensitive sensorimotor regions do not discriminate between speech and ingressive click sounds. The whole-brain contrast of all sounds compared with rest revealed significant activity in bilateral auditory cortices and ventral sensorimotor cortices (A, transparent white). Using this contrast, masked inclusively by the motor localizer (A, black), cluster ROIs were generated in both left and right hemispheres (A, white). Mean parameter estimates were extracted for these two regions using an interactive "leave-one-out" approach (see Methods), and these are displayed in the bottom left. The only significant comparison was that of [Speech > Rest] compared with [SCN > Rest]; [Speech > Rest] compared with [Clicks > SCN] was not significantly different. To investigate whether there may be regions in premotor cortex that are specifically activated during the perception of speech compared with other sounds, we then generated 8-mm spherical ROIs on the basis of the coordinates reported in two previous studies; Wilson and Iacoboni (2006) represented in B by solid white circles (−62 −4 38 and 56 −4 38), and Pulvermuller et al. (2006) represented by dotted white lines in the left hemisphere involved in movement and perception of lip and tongue movements (−54 −3 46 and −60 2 25, respectively). Mean parameter estimates for these five regions are plotted below for speech sounds compared with SCN and for ingressive clicks compared with SCN. There were no significant differences in any of these regions between the mean response to speech sounds and ingressive clicks demonstrating that activity in these areas is not specific to speech sounds. This was also the case for all subpeaks identified by the motor localizer. Error bars indicate *SEM*.

as speech sounds more than the speech sounds were misclassified as click sounds.

Common and Separate Pathways for Processing Speech and Ingressive Click Sounds

The first analysis used a null conjunction (Nichols, Brett, Andersson, Wager, & Poline, 2005) to look at activity common to perception of speech and ingressive click sounds ($p < .005$, cluster threshold = 10). This revealed that activity in right posterior and middle STS was activated both by speech and mouth sounds, relative to SCN (Figure 2C).

The contrast of [Speech > SCN] showed activity in bilateral mid-STG/STS. Activity was more widespread on the left than on the right, extending into the anterior temporal lobe (Figure 2A). The contrast of Speech > Click sounds led to bilateral activation in mid STS with the main peak on the left (Figure 2A). These STG/STS regions have been consistently activated in response to perception of syllables, whole words, and sentences (Scott et al., 2000) and fall within regions of activity identified by a speech localizer (as described in the Methods section). These data indicate that, within bilateral STG/STS, there are regions that show significantly greater activity for unvoiced phonemes than for ingressive sounds over and above responses to the SCN baseline.

The contrast of [Clicks > SCN] was associated with activity in left posterior medial planum temporale, right dorsolateral pFC, and right parietal cortex. The contrast of [Clicks > Speech sounds] revealed more extensive activation in right dorsolateral pFC, bilateral frontal poles and right posterior parietal cortex (Figure 2B).

To identify the neural responses that would be greatest to speech, then to clicks, then to SCN, we entered contrasts of [Speech > Clicks > SCN]. The [Speech > Clicks > SCN] contrast revealed significant activity in bilateral medial and posterior STG, with a greater distribution of activity in the left (Figure 2D). We also ran a [Clicks > Speech > SCN] contrast, which revealed activity in medial planum temporale in both hemispheres.

ROI Analyses

To directly compare how motor cortices respond during perception of speech and ingressive click sounds, ROIs for speech and motor regions for control of lip and tongue movements were created from a separate motor output localizer run (see Methods). A motor ROI was created by the contrast of executing lip and tongue actions compared with silent rest at a group level (Figure 3C and D). This was associated with significant activity in lateral sensorimotor cortices in both hemispheres (FDR = 0.001, cluster threshold = 30). Areas of the brain involved in speech

perception were identified by comparing BOLD responses during auditory perception of sentences to silent rest (Figure 3A and B). This contrast revealed significant activity in widespread superior temporal lobes in both hemispheres and left inferior frontal gyrus (FDR = 0.001, k = 30). We created four spherical ROIs of 10-mm radius, centered around the peak of each of these contrasts in both hemispheres.

To investigate whether the mean parameter estimates extracted from the peaks in left and right temporal and frontal cortices responded differently to the speech and click sounds, we used a repeated measured ANOVA with two factors: "ROI" (four levels: left temporal, right temporal, left motor, and right motor) and "Condition" (two levels: [Speech vs. SCN] and [Clicks vs. SCN]). We found a significant interaction between the two factors, $F(1, 21) = 6.62, p < .05$. To investigate the possibility that this effect is driven by a Hemisphere × Condition interaction, separate 2 × 2 repeated measures ANOVAs were run for the left and right hemispheres. In neither of these were there any significant main effects. There was a significant interaction between condition and ROI in the left hemisphere ($F(1, 21) = 5.3, p < .05$) but no significant interaction in the right. Four t tests were carried out to compare the effect sizes in the contrasts of Speech > SCN with those for Clicks > SCN within each of the four ROIs. The only significant comparison was that of the [Speech vs. SCN] compared with [Clicks vs. SCN] in the left temporal region ($p < .05$).

Previous studies have reported premotor sites that are sensitive to perception of speech sounds (Pulvermuller et al., 2006; Wilson & Iacoboni, 2006). To investigate the activity of these premotor sites during perception of speech and ingressive click sounds, we created 8-mm ROIs at the two premotor peaks reported in these two studies, resulting in one left and right premotor ROI (W-1 and W-2, respectively) and two in mouth premotor cortex corresponding to lips and tongue movement (P-lips, P-tongue). These all lay close to or within our motor localizer. Mean parameter estimates were extracted for these four sites and are displayed in Figure 4B (bottom). We found no significant difference between the responses to the contrast [Speech > SCN] compared with [Ingressive clicks > SCN] for any of these regions.

Finally, to identify any general acoustic responses in motor cortices, we performed a whole-brain analysis of all sounds over silence (speech, ingressive clicks, and SCN over silent rest using a weighted t contrast of 1 1 1 −3). This revealed activity in bilateral superior temporal cortices extending dorsally on the right into sensorimotor cortex (Figure 4A, top). A plot of neural responses during all conditions in this sensorimotor peak showed a highly similar profile of activity across all three acoustic conditions (Figure 4A, bottom).

To assess this formally and in a statistically independent fashion, an iterative "leave-one-out" approach was taken to extract the data from this region in each subject. A second level model was created for all subjects bar one, for all possible configurations of subjects. In each case, [All sounds vs. Rest] inclusively masked by the motor localizer revealed bilateral peaks in ventral motor cortex within or neighboring frontal operculum (masking threshold $p <$ .001, cluster threshold = 30). These ventral motor clusters generated by each second level model were used as ROIs to extract mean parameter estimates for each subject using an independent mask so as to avoid the problem of "double dipping" (Kriegeskorte et al., 2009; Vul et al., 2008). The peak from each model was used for the extraction of parameter estimates for each subject (Figure 4A, bottom). A repeated measures ANOVA was run with two factors: "Hemisphere" (two levels) and "Condition" (three levels). We found a significant main effect of hemisphere, $F(1, 21) = 5.499, p < .05$, which reflects the far greater response in the left hemisphere. A significant Hemisphere × Condition interaction ($F(1, 21) = 17.304, p < .05$). Three planned paircomparisons were set up to explore potential difference between the conditions within each ROI. Using a corrected significance level of $p < .017$, the only significant comparison was that of [Speech > Rest] compared with [SCN > Rest]. The contrast of [Speech > Rest] > [Clicks > Rest] was not significant.

DISCUSSION

As would be predicted from previous studies (Scott et al., 2000, 2006; Davis & Johnsrude, 2003; Scott & Johnsrude, 2003; Binder et al., 2000), the dorsolateral temporal lobes show a preferential response to speech sounds, with the greater response on the left. In contrast, the neural response to both speech and ingressive click sounds in bilateral mouth motor peaks (identified using a motor localizer) did not differ from that to the baseline stimuli. This finding strongly suggests that motor and auditory cortices are differentially sensitive to speech stimuli, a finding that is difficult to reconcile with models that posit a critical role for motor representations in speech perception. The lack of a speech specific effect in motor and premotor fields is not because of lack of power, as we were able to identify, at a whole brain level, bilateral ventral sensorimotor responses in a contrast of both speech and click sounds over SCN. In a post hoc analysis, the activity in this region was significantly greater to speech than the SCN in the left hemisphere: The contrast of click sounds over SCN was not significant. These responses in ventral sensorimotor cortex could suggest a sensitivity to more generic aspects of processing mouth sounds, rather than a specific role in speech perception, as the Speech > Click sounds comparison was not significant in this analysis. Further investigation of this response will allow us to delineate the properties and possible functions of this activity.

Within the dorsolateral temporal lobes, there were common responses to speech and click sounds (over the SCN sounds), which converged in two separate peaks within right STS/STG, regions which have been linked to voice

processing (Belin et al., 2000). Selective responses to the speech sounds were in bilateral dorsolateral lobes, including widespread activity in left STG/STS as has been commonly found in studies using higher-order linguistic structure such as consonant–vowel syllables (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005), words (Mummery et al., 1999), and sentences (Scott et al., 2000). This is evidence that even very short, unvoiced speech sounds activate extensive regions of the auditory speech perception system, running into the auditory "what" pathway (Scott et al., 2009). In contrast, selective activation to the ingressive clicks compared with speech sounds was seen in left medial planum temporale when compared with speech sounds and bilateral medial planum temporale when compared with speech and SCN. Medial planum temporale has been implicated in sensorimotor processing of "doable" sounds (Warren, Wise, & Warren, 2005) or of processing of sounds that can be made by the human vocal tract (Hickok, Okada, & Serences, 2009) but is not selective to intelligible speech (Wise et al., 2001). Thus, although the speech sounds recruit a largely left-lateralized "what" stream of processing within the temporal lobes, the ingressive click sounds are more associated with the caudal "how" stream, possibly reflecting their lack of linguistic meaning.

In contrast to this pattern of responses in the temporal lobes, none of the motor peaks identified in the motor localizer showed a selective response to either category of mouth sounds, relative to the baseline. In an ROI analysis looking at previously investigated premotor regions (Pulvermuller et al., 2006; Wilson et al., 2004), we also found no significant difference between the response to the speech, ingressive click sounds and baseline stimuli. Furthermore, it was only in left ventral premotor cortex that we observed any difference between mouth sounds and the baseline condition. These peaks lie ventral to peaks associated with the control of articulation (Dhanjal, Handunnetthi, Patel, & Wise, 2008; Pulvermuller et al., 2006; Wilson & Iacoboni, 2006). Interestingly a few studies have reported similar activations during localization compared with recognizing of auditory stimuli (Maeder et al., 2001) and passive perception of sounds in controls and to a greater extent in a patient with auditory–tactile synaethesia (Beauchamp & Ro, 2008). Similar ventral sensorimotor peaks have been reported in a study specifically investigating controlled breathing for speech (Murphy et al., 1997).

The dissociation between the neural responses to speech and click sounds in the temporal lobes and motor cortex is strong evidence against an "essential" role for mouth motor areas in speech perception (Meister et al., 2007), and it has also been argued that the involvement of motor areas in perception may not specific to speech (Pulvermuller et al., 2006). Because we chose speech sounds and click mouth sounds, which are processed perceptually in very different ways by native English speakers (Best & Avery, 1999; Best et al., 1988), the neural systems *crucially* involved in speech perception would be expected to reflect this perceptual difference. If motor representations are not selectively involved in speech perception, this might implicate them in more general auditory and perceptual processes than those that are truly central to speech perception.

A previous study addressed the temporal lobe and motor response to non-native phonemes presented in a vowel–consonant–vowel context (Wilson & Iacoboni, 2006), finding an enhanced response to non-native speech sounds in all regions activated by the speech sounds in temporal and motor cortex, relative to rest. A key difference between that study (Wilson & Iacoboni, 2006) and the present study is the range of sounds employed: Wilson and Iacoboni (2006) used non-native sounds that can be assimilated into English phonemic categories (e.g., a voiceless retroflex post alveolar fricative, which is often confused with an English // or "sh") to ingressive click sounds, which are not confused with English speech sounds (Best & Avery, 1999; Best et al., 1988). Five times as many non-native sounds as native sounds (25 non-native, 5 native) were also presented, and this greater variation in the number and range of non-native sounds is likely to have lead to the greater overall activation seen to the non-native speech sounds.

TMS studies have indicated corticospinal excitability of motor cortex during speech perception. However, these studies have reported either no significant difference between the motor responses to speech and environmental sounds (Watkins et al., 2003) or have used overt tasks, such as lexical decision (Fadiga et al., 2002) and phoneme discrimination in noise (D'Ausilio et al., 2009; Meister et al., 2007), which encourage articulatory strategies (e.g., phoneme segmentation). These tasks may recruit motor regions as a function of the tasks, rather than due to basic speech perception mechanisms (McGettigan, Agnew, & Scott, 2010; Scott et al., 2009; Hickok & Poeppel, 2000). Indeed, a more recent TMS to left ventral premotor cortex specifically disrupted tasks requiring phoneme segmentation, but not tasks requiring phoneme or syllable recognition (Sato, Tremblay, & Gracco, 2009).

Conclusion

This is the first study to directly examine whether motor responses to speech sounds are truly selective for speech. In this functional imaging study, we compared the passive perception of phonemes and ingressive click sounds not processed as speech in monolingual English speakers. In contrast to the view that motor areas are "essential" (Meister et al., 2007) to speech perception, we found no evidence for a selective response in motor or premotor cortices to speech sounds. Indeed we found consistently similar responses to the speech sounds, the ingressive click sounds and the SCN stimuli in peaks taken from an independent mouth motor localizer and also in peaks taken from previous studies. These data demonstrate that

mouth motor or premotor areas do not have a speech-specific role in perception but may be involved in a more general aspect of auditory perception. Further work will be able to delineate what the functional role of such general auditory processing is in behavioral terms and how these motor systems interact with acoustic–phonetic systems in temporal lobe fields.

Reprint requests should be sent to Sophie K. Scott, Institute for Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, United Kingdom, or via e-mail: sophie.scott@ucl.ac.uk.

REFERENCES

Beauchamp, M. S., & Ro, T. (2008). Neural substrates of sound-touch synesthesia after a thalamic lesion. *Journal of Neuroscience, 28,* 13696–13702.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport, 14,* 2105–2109.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403,* 309–312.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13,* 108–112.

Best, C. T., & Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science, 10,* 65–70.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345–360.

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.

Boersma, P., & Weenink, D. (2010). *Praat, doing phonetics by computer (version 5.1.26)*. Retrieved 4 August 2010, from www.Praat.Org/.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). *Region of interest analysis using an SPM toolbox*. Paper presented at the International Conference on Functional Mapping of the Human Brain, Sendai, Japan.

D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19,* 381–385.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience, 23,* 3423–3431.

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

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience, 15,* 399–402.

Fellner, C., Doenitz, C., Finkenzeller, T., Jung, E. M., Rennert, J., & Schlaier, J. (2009). Improving the spatial accuracy in functional magnetic resonance imaging (fMRI) based on the blood oxygenation level dependent (BOLD) effect: Benefits from parallel imaging and a 32-channel head array coil at 1.5 tesla. *Clinical Hemorheology and Microcirculation, 43,* 71–82.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13,* 361–377.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119,* 593–609.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews, 34,* 575–583.

Hickok, G., Okada, K., & Serences, J. T. (2009). Area SPT in the human planum temporale supports sensory motor integration for speech processing. *Journal of Neurophysiology, 101,* 2725–2732.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131–138.

Houde, J. F. (2009). There's more to speech perception than meets the ear. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 20139–20140.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12,* 535–540.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

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.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage, 14,* 802–816.

McGettigan, C., Agnew, Z., & Scott, S. K. (2010). Are articulatory commands automatically and involuntarily activated during speech perception? *Proceedings of the National Academy of Sciences, U.S.A., 107,* E42.

McGettigan, C., Warren, J. E., Eisner, F., Marshall, C. R., Shanmugalingam, P., & Scott, S. K. (2011). Neural correlates of sublexical processing in phonological working memory. *Journal of Cognitive Neuroscience, 23,* 961–977.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17,* 1692–1696.

Mummery, C. J., Ashburner, 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.

Murphy, K., Corfield, D. R., Guz, A., Fink, G. R., Wise, R. J., Harrison, J., et al. (1997). Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology, 83,* 1438–1447.

Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 20470–20475.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25,* 653–660.

Obleser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience, 16,* 31–39.

Parikh, P. T., Sandhu, G. S., Blackham, K. A., Coffey, M. D., Hsu, D., Liu, K., et al. (2011). Evaluation of image quality of a 32-channel versus a 12-channel head coil at 1.5t for MR imaging of the brain. *American Journal of Neuroradiology*, *32,* 365–373.

Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 7865–7870.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169–192.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2,* 661–670.

Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language, 111,* 1–7.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735–1736.

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., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10,* 295–302.

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.

Vul, E., Harris, C., Winkleman, P., & Pashler, H. (2008). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4,* 274–290.

Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences, 28,* 636–643.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41,* 989–994.

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage, 33,* 316–325.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7,* 701–702.

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,* 83–95.