Published in final edited form as:
J Cogn Neurosci. 2011 December ; 23(12): 4038–4047. doi:10.1162/jocn_a_00106.

Discriminating between auditory and motor cortical responses to speech and non-speech mouth sounds

Z.K. Agnew[1], **C. McGettigan**[1], and **S.K. Scott**[1,*]

[1]Institute for Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR

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 non-speech 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/phonetic properties while motor responses may show more generalised responses to the acoustic stimuli.

Keywords

speech; motor cortex; auditory cortex; movement; fMRI

Introduction

Several recent theories of perceptual processing have identified a central role for motor representations in the recognition of action (Rizzolatti, Fogassi, & Gallese, 2001), 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 (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Pulvermuller et al., 2006; Watkins, Strafella, & Paus, 2003; Wilson, Saygin, Sereno, & Iacoboni, 2004). 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, Blank, Rosen, & Wise, 2000; Scott, Rosen, Lang, & Wise, 2006), while responses in the superior temporal sulcus (STS) and beyond are less sensitive to acoustic properties, and more driven by the

[*]**Corresponding author:** Institute for Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR (sophie.scott@ucl.ac.uk, +44 (0)207 679 1144).

Europe PMC Funders Author Manuscripts

intelligibility of the speech (Scott et al., 2000). In contrast, support for sensory-motor involvement in speech perception is provided by studies showing areas co-activated 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 and Ostry (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 et al., 2007) implies a sensitivity to acoustical-phonetic information, as is seen in auditory areas. In line with this, 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 (positron emission tomography and functional magnetic resonance imaging) 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 functional magnetic resonance imaging (fMRI). We used speech sounds, non-speech mouth sounds, and signal correlated noise (SCN) analogues of both sound categories (in which only the amplitude envelope of the stimuli is preserved (Schroeder, 1968)). Signal correlated noise is a relatively complex auditory baseline, which has been used in several previous studies of auditory and motor responses to speech (Davis & Johnsrude, 2003; Mummery, Ashburner, Scott, & Wise, 1999; Pulvermuller et al., 2006).

We included both speech and non-speech mouth sounds as they differ in their phonetic properties, while being produced by the same effectors. The non-speech 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 non-speech 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, one velar). Unvoiced phonemes were used to afford better matching with the 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 ‘non-speech’ mouth sound category.

This aim of this experiment was to identify whether auditory cortical fields associated with speech perception (identified using a speech perception localiser) and motor and premotor cortical fields associated with speech-related orofacial movement (identified using a silent mouth movement localiser, 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 signal correlated noise. We defined a speech-related neural response as one in which there was a preferential response to speech sounds relative to ingressive clicks and signal correlated noise. We also explored whether any cortical fields showed a common response to the speech and click sounds, relative to the signal correlated noise: 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 signal correlated noise 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 signal correlated noise.

Materials and Methods

Generation of stimuli

The speech sounds were voiceless consonants comprised of plosives (/t/, /k/), a fricative (/f/) and an affricate (/t∫/) (the phoneme at the start of 'cheese'). The plosives (/t/, /k/) are non-continuants i.e. are naturally produced with a short post-obstruent unvoiced airflow. The non-speech mouth sounds comprised four ingressive click sounds: a dental click (/|/), a post-alveolar click (/!/), a lateral click (/||/) and a bilabial click (/⊙/). 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.

Sounds were recorded using a solid state recorder (Edirol, R-O9HR) 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 thirty tokens for each category of speech and ingressive click sound. Signal correlated noise 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.

Behavioural testing

The stimuli were pre-tested to ensure that subjects could correctly categorise the sounds as speech or non-speech. Eight subjects (5 male, mean age 25.7 yrs) 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 non-speech sounds (60 trials in total, 30 speech and 30 click trials). In a second pre-test, the experiment was repeated with individual exemplars of each speech and ingressive sound (80 trials in total, each of the 8 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 male) 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 UCL Ethics Committee who provided local ethics approval for this study.

fMRI

A 1.5 Tesla Siemens system with a 32 channel head coil was used to acquire 183 T_2*-weighted echo-planer images (EPI) data ($3 \times 3 \times 3mm^3$, TR/TA/TE/flip 10,000 ms/3s/50 ms/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 Tesla field (Fellner et al., 2009; Parikh et al., 2011). A sparse scanning protocol was employed in order to administer the auditory stimuli in the absence of scanner noise. The first two functional volumes were discarded in order to remove the effect of T_1 equilibration. High resolution T_1 anatomical volume images (160 sagittal slices, voxel size $1mm^3$) 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.

Sounds for the main run and instructions for the localiser run were presented using MATLAB with the Psychophysics Toolbox extension (Brainard, 1997), via a Denon amplifier (Denon UK, Belfast, UK) and electrodynamic headphones worn by the participant (MR Confon GmbH, Magdeburg, Germany). Instructions were projected from a specially-configured video projector (Eiki International, Inc., Rancho Santa 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 seconds (e.g. /t/ - /k/ - / t∫/ - /f/). The order of sounds was randomized within trial and the ordering of sound category (speech, non-speech, SCN) was randomized across trials. Across the whole experiment, none of the 30 recorded tokens of each speech/mouth sound was repeated. A ± 500ms onset jitter was used. This main run lasted approximately 30 minutes.

We carried out a separate localizer run in order 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 (TR=3secs). 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 seconds and was repeated 4 times. This localizer scan lasted approximately 11 minutes.

Pre-processing 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 coregistered to the T_1 structural image and smoothed using a Gaussian kernel of full width half medium (FWHM) 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 $p<0.005$. 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 signal correlated noise, Speech Vs Clicks and Clicks Vs 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 $p<0.005$, cluster threshold of 10. Significant BOLD effects were rendered on a normalized template.

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

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 10mm 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 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 to signal correlated noise. These are seen in Figure 3.
2. An additional set of 8mm 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 8mm 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 to signal correlated noise.
3. Finally, two cluster ROIs in ventral sensorimotor cortices were generated by the contrast of all sounds (speech, non-speech, SCN) over silent rest. This contrast identified a peak in ventral primary sensorimotor cortex in both hemispheres (Figure 4a). In order to allow statistical analyses of these data (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Vul, Harris, Winkleman, & Pashler, 2008), regions of interest 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 V Rest inclusively masked by the Motor localiser] (masking threshold $p<0.001$, cluster threshold = 30) from the other 21 participants. Mean parameter estimates were extracted for speech, clicks and SCN compared to silent rest.

Results

Subjects correctly categorize speech and ingressive click sounds

The average percentage of correctly classified trains of sounds was 96.3% (+/− 3.23 SD). For the trains of speech and click sounds, 95.56 (+/− 1.41 SD) and 97.04 (+/−2.03 SD) of sounds were correctly categorized as speech or non-speech sounds. A two-tailed t-test showed that these scores were not significantly different ($p=0.35$). 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 7 subjects for this assessment). A one-way ANOVA demonstrated that there were more within-category miscategorisations for the click sounds (miscategorisations of clicks sounds as other click sounds) than any other type of error ($p<0.05$). Importantly, however, there was no significant difference ($p=0.3$) between the number of speech sounds miscategorised 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 clicks sounds were not being systematically misclassified 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 ($p<0.005$, cluster threshold, 10). This revealed that activity in right posterior and middle superior temporal sulcus (STS) was activated both by speech and mouth sounds, relative to signal correlated noise (SCN), (Figure 2c).

The contrast of [Speech >SCN] showed activity in bilateral mid superior temporal gyri/sulci. 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 localiser (as described in the Methods section). These data indicate that within bilateral superior temporal gyri (STG)/STS, there are regions that show significantly greater activity for unvoiced phonemes than for ingressive sounds, over and above responses to the signal correlated noise baseline.

The contrast of [Clicks > SCN] was associated with activity in left posterior medial planum temporale, right dorsolateral prefrontal cortex and right parietal cortex. The contrast of [Clicks > Speech sounds] revealed more extensive activation in right dorsolateral prefrontal cortex, 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.

Region of interest analyses

In order to directly compare how motor cortices respond during perception of speech and ingressive click sounds, regions of interest (ROI) for speech and motor regions for control of lip and tongue movements were created from a separate motor output localiser run (see Methods). A motor ROI was created by the contrast of executing lip and tongue actions compared to 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 10mm radius, centered around the peak of each of these contrasts in both hemispheres.

In order 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: ‘Region of Interest’ (four levels: left temporal, right temporal, left motor and right motor) and ‘Condition’ (two levels: [Speech Vs SCN] and [Clicks V SCN]). We found a significant interaction between the two factors, $F(1, 21)=6.62$, $p<0.05$. In order to investigate the possibly 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 ($F(1, 21)=5.3$, $p<0.05$) 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 to [Clicks V SCN] in the left temporal region ($p<0.05$).

Previous studies have reported premotor sites that are sensitive to perception of speech sounds (Pulvermuller et al., 2006; Wilson et al., 2004). In order to investigate the activity of these premotor sites during perception of speech and ingressive click sounds, we created

8mm 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 (lower panel). We found no significant difference between the responses to the contrast [speech >SCN] compared to [ingressive clicks > SCN] for any of these regions.

Finally, in order 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 panel). 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, lower panel).

In order 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 Localiser revealed bilateral peaks in ventral motor cortex within or neighbouring frontal operculum (masking threshold $p<0.001$, cluster threshold = 30). These ventral motor clusters generated by each second level model, were used as regions of interest 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 lower panel). A repeated measures ANOVA was run with two factors: 'Hemisphere' (two levels) and 'Condition' (three levels). We found a significant main effect of hemisphere, $F(1, 21)=5.499$ $p<0.05$ which reflects the far greater response in the left hemisphere. A significant Hemisphere × Condition interaction ($F(1, 21)=17.304$, $p<0.05$). Three planned pairwise comparisons were set up to explore potential difference between the conditions within each ROI. Using a corrected significance level of $p<0.017$, 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 (Binder et al., 2000; Davis & Johnsrude, 2003; Scott et al., 2000; Scott & Johnsrude, 2003; Scott et al., 2006), 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 localiser), 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 which 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 due to a 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 signal correlated noise. 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 signal correlated noise sounds) which converged in two separate peaks within right superior temporal sulcus/gyrus, 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 to speech sounds was seen in left medial planum temporale when compared to speech sounds, and bilateral medial planum temporale when compared to speech and SCN. Medial planum temporale has been implicated in sensorimotor processing of 'do-able' 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 while 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 localiser showed a selective response to either category of mouth sounds, relative to the baseline. In a region of interest 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 to 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., 2006), and it has also been argued that the involvement of motor areas in perception may not specific to speech (Pulvermuller et al., 2006). Since we chose speech sounds and click mouth sounds which are processed perceptually in very different ways by native English speakers (Best et al, 1988, Best and Avery, 1999), 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 (VCV) 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 et al. (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 /ʃ/

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.

Transcranial magnetic stimulation (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, Wilson, Deblieck, Wu, & Iacoboni, 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 (Hickok & Poeppel, 2000; McGettigan, Agnew, & Scott, 2010; Scott et al., 2009). 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 clicks 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/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 behavioural terms, and how these motor systems interact with acoustic-phonetic systems in temporal lobe fields.

References

Beauchamp MS, Ro T. Neural substrates of sound-touch synesthesia after a thalamic lesion. J Neurosci. 2008; 28(50):13696–13702. [PubMed: 19074042]

Belin P, Zatorre RJ. Adaptation to speaker’s voice in right anterior temporal lobe. Neuroreport. 2003; 14(16):2105–2109. [PubMed: 14600506]

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature. 2000; 403(6767):309–312. [PubMed: 10659849]

Bench J, Kowal A, Bamford J. The bkb (bamford-kowal-bench) sentence lists for partially-hearing children. Br J Audiol. 1979; 13(3):108–112. [PubMed: 486816]

Best CT, Avery RA. Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. Psychological Science. 1999; 10(1):65–70.

Best CT, McRoberts GW, Sithole NM. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by english-speaking adults and infants. J Exp Psychol Hum Percept Perform. 1988; 14(3):345–360. [PubMed: 2971765]

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 2000; 10(5):512–528. [PubMed: 10847601]

Boersma P, Weenink D. Praat, doing phonetics by computer (version 5.1.26). 2010 retrieved 4 august 2010 from http://www.Praat.Org/.

Brainard DH. The psychophysics toolbox. Spatial Vision. 1997; 10(4):433–436. [PubMed: 9176952]

Brett, M.; Anton, JL.; Valabregue, R.; Poline, JB. Region of interest analysis using an spm toolbox; Paper presented at the International Conference on Functional Mapping of the Human Brain; Sendai, Japan. 2002;

D'Ausilio A, Pulvermuller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. The motor somatotopy of speech perception. Curr Biol. 2009; 19(5):381–385. [PubMed: 19217297]

Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. J Neurosci. 2003; 23(8):3423–3431. [PubMed: 12716950]

Dhanjal NS, Handunnetthi L, Patel MC, Wise RJ. Perceptual systems controlling speech production. J Neurosci. 2008; 28(40):9969–9975. [PubMed: 18829954]

Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: A tms study. European Journal of Neuroscience. 2002; 15(2):399–402. [PubMed: 11849307]

Fellner C, Doenitz C, Finkenzeller T, Jung EM, Rennert J, Schlaier J. 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. Clin Hemorheol Microcirc. 2009; 43(1):71–82. [PubMed: 19713602]

Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. Psychon Bull Rev. 2006; 13(3):361–377. [PubMed: 17048719]

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. Brain. 1996; 119(Pt 2):593–609. [PubMed: 8800951]

Heyes C. Where do mirror neurons come from? Neurosci Biobehav Rev. 2010; 34(4):575–583. [PubMed: 19914284]

Hickok G, Okada K, Serences JT. Area spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol. 2009; 101(5):2725–2732. [PubMed: 19225172]

Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 2000; 4(4):131–138. [PubMed: 10740277]

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8(5): 393–402. [PubMed: 17431404]

Houde JF. There's more to speech perception than meets the ear. Proc Natl Acad Sci U S A. 2009; 106(48):20139–20140. [PubMed: 19934047]

Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E. Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. J Neurosci. 2003; 23(29):9541–9546. [PubMed: 14573533]

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: The dangers of double dipping. Nature Neuroscience. 2009; 12(5):535–540.

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21(1): 1–36. [PubMed: 4075760]

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cereb Cortex. 2005; 15(10):1621–1631. [PubMed: 15703256]

Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Clarke S. Distinct pathways involved in sound recognition and localization: A human fmri study. Neuroimage. 2001; 14(4): 802–816. [PubMed: 11554799]

McGettigan C, Agnew Z, Scott SK. Are articulatory commands automatically and involuntarily activated during speech perception? Proceedings of the National Academy of Sciences of the United States of America. 2010; 2010

McGettigan C, Warren JE, Eisner F, Marshall CR, Shanmugalingam P, Scott SK. Neural correlates of sublexical processing in phonological working memory. J Cogn Neurosci. 2011; 23(4):961–977. [PubMed: 20350182]

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Curr Biol. 2007; 17(19):1692–1696. [PubMed: 17900904]

Mummery CJ, Ashburner J, Scott SK, Wise RJ. Functional neuroimaging of speech perception in six normal and two aphasic subjects. J Acoust Soc Am. 1999; 106(1):449–457. [PubMed: 10420635]

Murphy K, Corfield DR, Guz A, Fink GR, Wise RJ, Harrison J, Adams L. Cerebral areas associated with motor control of speech in humans. J Appl Physiol. 1997; 83(5):1438–1447. [PubMed: 9375303]

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM. Defining a left-lateralized response specific to intelligible speech using fmri. Cereb Cortex. 2003; 13(12):1362–1368. [PubMed: 14615301]

Nasir SM, Ostry DJ. Auditory plasticity and speech motor learning. Proc Natl Acad Sci U S A. 2009; 106(48):20470–20475. [PubMed: 19884506]

Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the minimum statistic. Neuroimage. 2005; 25(3):653–660. [PubMed: 15808966]

Obleser J, Lahiri A, Eulitz C. Magnetic brain response mirrors extraction of phonological features from spoken vowels. J Cogn Neurosci. 2004; 16(1):31–39. [PubMed: 15006034]

Parikh PT, Sandhu GS, Blackham KA, Coffey MD, Hsu D, Liu K, Sunshine JL. Evaluation of image quality of a 32-channel versus a 12-channel head coil at 1.5t for mr imaging of the brain. AJNR Am J Neuroradiol. 2011; 32(2):365–373. [PubMed: 21163877]

Pulvermuller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y. Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci U S A. 2006; 103(20):7865–7870. [PubMed: 16682637]

Rizzolatti G, Craighero L. The mirror-neuron system. Annual Review of Neuroscience. 2004; 27:169–192.

Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci. 2001; 2(9):661–670. [PubMed: 11533734]

Sato M, Tremblay P, Gracco VL. A mediating role of the premotor cortex in phoneme segmentation. Brain Lang. 2009; 111(1):1–7. [PubMed: 19362734]

Schroeder MR. Reference signal for signal quality studies. Journal of the Acoustical Society of America. 1968; 44:1735–1736.

Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 2000; 123(Pt 12):2400–2406. [PubMed: 11099443]

Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 2003; 26(2):100–107. [PubMed: 12536133]

Scott SK, McGettigan C, Eisner F. A little more conversation, a little less action--candidate roles for the motor cortex in speech perception. Nat Rev Neurosci. 2009; 10(4):295–302. [PubMed: 19277052]

Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. J Acoust Soc Am. 2006; 120(2): 1075–1083. [PubMed: 16938993]

Vul E, Harris C, Winkleman P, Pashler H. Puzzlingly high correlations in fmri studies of emotion, personality, and social cognition. Perspectives on Psychological Science. 2008; 4(3):274–290.

Warren JE, Wise RJ, Warren JD. Sounds do-able: Auditory-motor transformations and the posterior temporal plane. Trends Neurosci. 2005; 28(12):636–643. [PubMed: 16216346]

Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia. 2003; 41(8):989–994. [PubMed: 12667534]

Wilson SM, Iacoboni M. Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. Neuroimage. 2006; 33(1):316–325. [PubMed: 16919478]

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. Nat Neurosci. 2004; 7(7):701–702. [PubMed: 15184903]

Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. Separate neural subsystems within ‘wernicke’s area’. Brain. 2001; 124(Pt 1):83–95. [PubMed: 11133789]

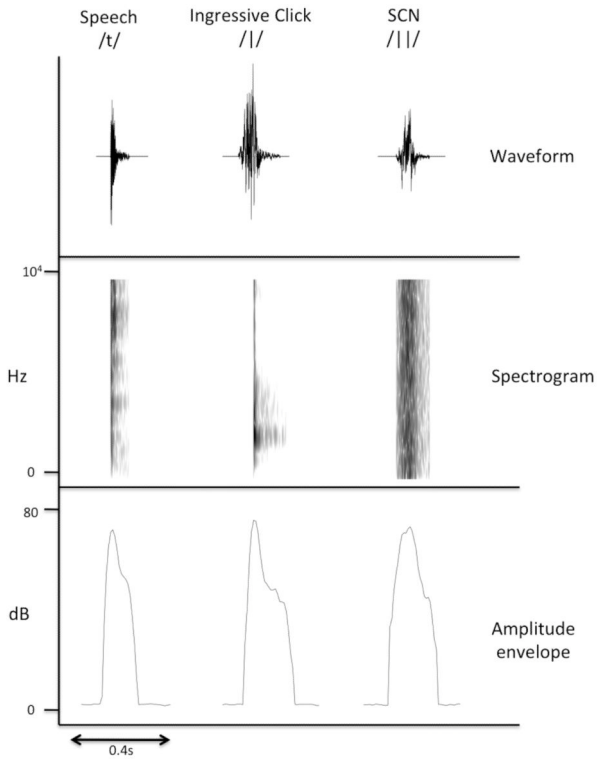

Figure 1. Speech, ingressive clicks and signal correlated noise sounds share similar amplitude envelopes

Example tokens from the speech, ingressive click sounds and signal correlated noise conditions used in the experiment. The upper panel shows the waveform versions of the sounds, while the middle panel shows their spectrotemporal structure in the form of a spectrogram. The bottom panel 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 signal correlated noise token has a much simpler spectral structure than the speech and click sounds (as shown in the spectrogram).

Europe PMC Funders Author Manuscripts

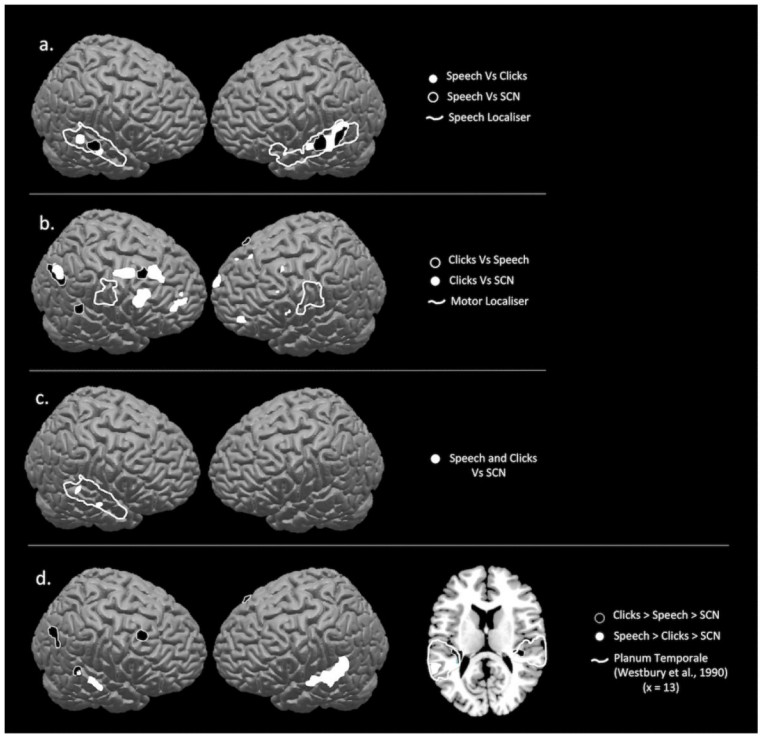

Figure 2. Perception of speech and ingressive click sounds is associated with increased activity in auditory regions

Perception of speech sounds compared to ingressive click sounds (Figure a, white) was associated with increased BOLD activity in left middle and posterior superior temporal gyrus ($p<0.005$, cluster threshold=30). Perception of speech sounds compared to signal correlated noise was associated with significant activity in the same regions but extending anteriorly in the left hemisphere (Figure 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 (Figure a, white line). Listening to ingressive click sounds compared to speech sounds was associated with significant activity in prefrontal regions and right occipitoparietal cortex (Figure 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 signal correlated noise revealed significant activity in mouth motor regions identified by an independent motor localizer run (Figure b, white line). Figure c shows common activity during the perception of both types of sounds compared to signal correlated noise in right superior temporal gyrus ($p<0.005$). 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. Figure d shows 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).

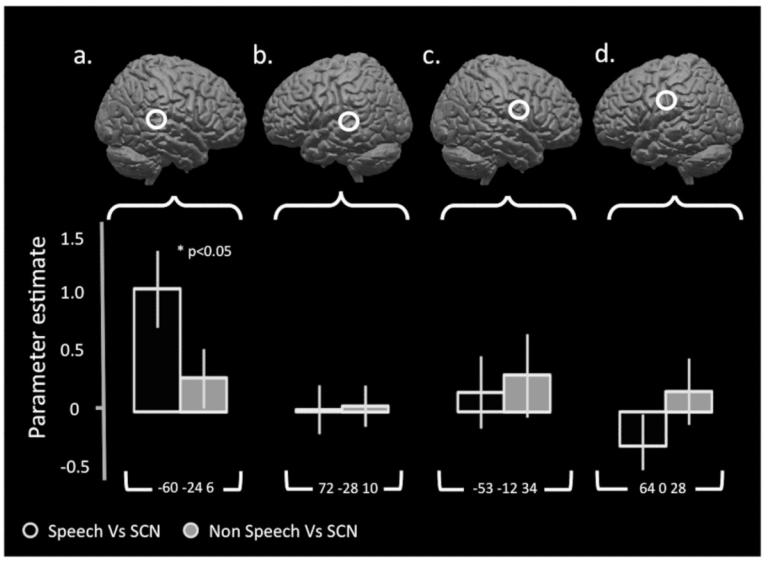

Figure 3. Left auditory areas preferentially encode speech sounds but there is no speech specific activity in primary motor cortices

Parameter estimates for speech and ingressive click sounds compared to signal correlated noise were calculated within four regions of interest generated from peak coordinates from an independent localizer. Figures a and b display the left and right speech regions of interest 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. Figures c and d shows the left and right mouth motor regions of interest generated from alternating lip and tongue movements compared to silent rest (FWE 0.05, cluster threshold=30). Speech sounds were associated with significantly increased activity in left auditory cortex compared to ingressive click sounds. There was non-significant difference in levels of activity in right auditory cortex or in the mouth motor regions. In all three of these regions there was a non-significant increase in activity for ingressive click sounds over signal correlated noise (SCN) compared to speech sounds over SCN. Error bars indicate standard error of the mean.

Europe PMC Funders Author Manuscripts

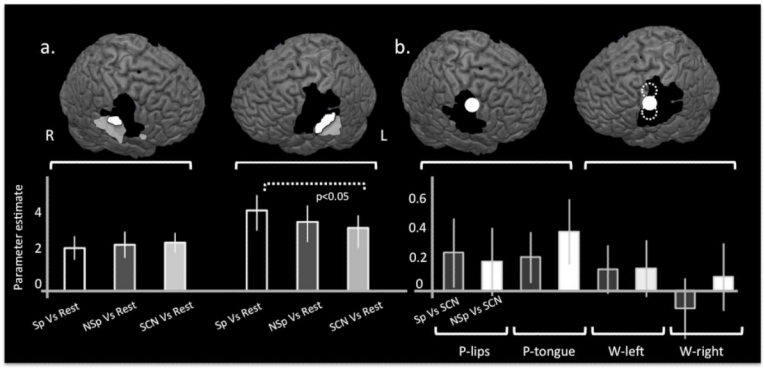

Figure 4. Auditory sensitive sensorimotor regions do not discriminate between speech and ingressive click sounds

The whole brain contrast of all sounds compared to rest revealed significant activity in bilateral auditory cortices and ventral sensorimotor cortices (Figure a, transparent white). Using this contrast, masked inclusively by the motor localiser (Figure a, black), cluster regions of interest were generated in both left and right hemispheres (Figure 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 panel of Figure 4.. The only significant comparison was that of [Speech > Rest] compared to [SCN > Rest]; [Speech>Rest] compared to [ClicksrSCN] was not significantly different. In order to investigate whether there may be regions in premotor cortex that are specifically activated during the perception of speech compared to other sounds, we then generated 8mm spherical regions of interest based on the coordinates reported in two previous studies; Wilson et al., (2006) represented in Figure 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 to SCN and for ingressive clicks compared to 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 localiser. Error bars indicate standard error of the mean.