Behavioral/Cognitive

Automatic Phoneme Category Selectivity in the Dorsal Auditory Stream

Mark A. Chevillet,[1,2,4] **Xiong Jiang,**[1] **Josef P. Rauschecker,**[2,3] **and Maximilian Riesenhuber**[1]

[1]Laboratory for Computational Cognitive Neuroscience, Department of Neuroscience, and [2]Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007, [3]Brain-Mind Laboratory, Department of Biomedical Engineering and Computational Science, Aalto University, 02150 Espoo, Finland, and [4]Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723

Debates about motor theories of speech perception have recently been reignited by a burst of reports implicating premotor cortex (PMC) in speech perception. Often, however, these debates conflate perceptual and decision processes. Evidence that PMC activity correlates with task difficulty and subject performance suggests that PMC might be recruited, in certain cases, to facilitate category judgments about speech sounds (rather than speech perception, which involves decoding of sounds). However, it remains unclear whether PMC does, indeed, exhibit neural selectivity that is relevant for speech decisions. Further, it is unknown whether PMC activity in such cases reflects input via the dorsal or ventral auditory pathway, and whether PMC processing of speech is automatic or task-dependent. In a novel modified categorization paradigm, we presented human subjects with paired speech sounds from a phonetic continuum but diverted their attention from phoneme category using a challenging dichotic listening task. Using fMRI rapid adaptation to probe neural selectivity, we observed acoustic-phonetic selectivity in left anterior and left posterior auditory cortical regions. Conversely, we observed phoneme-category selectivity in left PMC that correlated with explicit phoneme-categorization performance measured after scanning, suggesting that PMC recruitment can account for performance on phoneme-categorization tasks. Structural equation modeling revealed connectivity from posterior, but not anterior, auditory cortex to PMC, suggesting a dorsal route for auditory input to PMC. Our results provide evidence for an account of speech processing in which the dorsal stream mediates automatic sensorimotor integration of speech and may be recruited to support speech decision tasks.

Introduction

There remains an ongoing and contentious debate in speech processing concerning what role premotor cortex (PMC) might play in speech perception, a function that is largely attributed to the ventral auditory stream (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; DeWitt and Rauschecker, 2012). Although there is little doubt that PMC can be active while listening to speech (Wilson et al., 2004; Skipper et al., 2005; Turkeltaub and Coslett, 2010), recent studies providing evidence that disrupting PMC processing impedes performance on explicit phoneme category judgments (Meister et al., 2007; Sato et al., 2009) have rekindled the debate about whether PMC function in speech perception is essential (Meister et al., 2007; Pulvermüller and Fadiga, 2010) or not (Hickok et al., 2011; Schwartz et al., 2012). Yet, the aforementioned studies linking PMC to phoneme categorization performance used speech categorization tasks, making it difficult to disentangle PMC's contribution to perceptual versus decision processes (Binder et al., 2004; Holt and Lotto, 2010).

Current theories of speech processing have partially converged in proposing a nonperceptual role for the dorsal stream (Hickok et al., 2011; Rauschecker, 2011), proposing instead that the dorsal auditory stream likely mediates sensorimotor integration (i.e., mapping speech signals to motor articulations) and vice versa. Although this role could explain PMC activity during listening to speech, it is currently unknown whether the dorsal stream exhibits neural tuning reflecting this mapping.

Recent evidence that PMC activity is modulated by performance during a phoneme categorization task (Alho et al., 2012) have given rise to the hypothesis that the dorsal stream might be recruited in a task-specific manner to support speech categorization. Although these studies provide circumstantial support for this hypothesis, it is currently unknown whether the dorsal stream exhibits neural selectivity supportive of speech categorization, as it is difficult to infer neural tuning properties from traditional BOLD-contrast fMRI paradigms. Additionally, it is unknown whether dorsal-stream processing is automatic, a defining criterion for which is that processing take place even when it is unnecessary for the task at hand (Moors and De Houwer, 2006). Finally, although evidence suggests that dorsal-stream recruitment may vary across subjects (Alho et al., 2012; Szenkovits
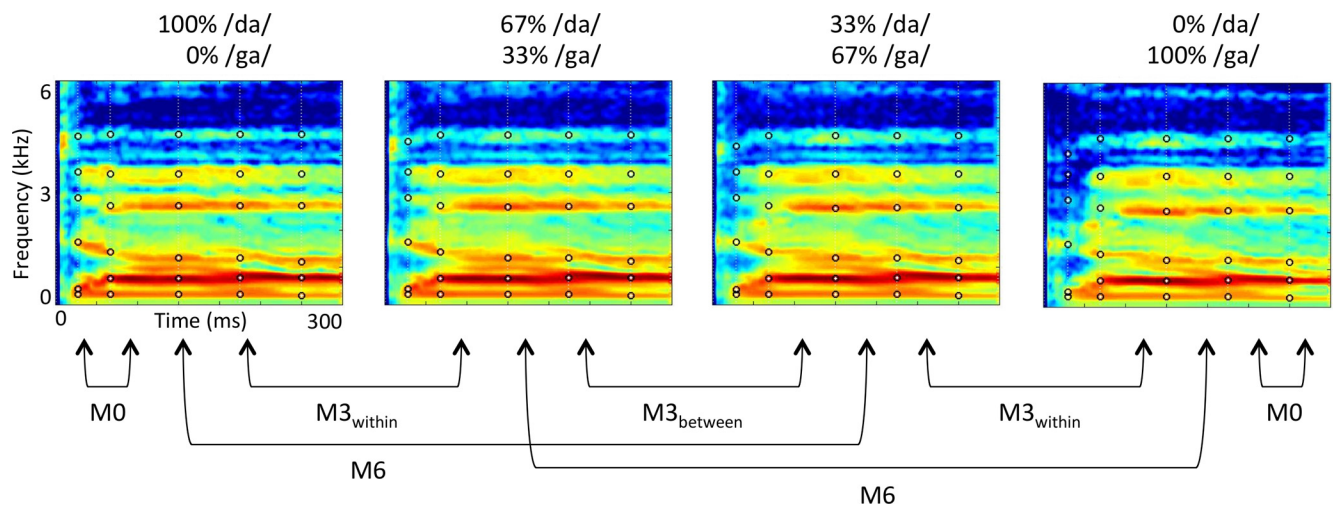
Figure 1. Selecting stimuli for fMRI conditions. The figure shows examples of auditory stimuli (displayed as spectrograms) and the pairing conditions used in the scanner. Phonetic continua were generated between two recorded phonemes using sound "morphing" software (STRAIGHT toolbox for MATLAB) for one male and one female voice. Using one morph line as example (male speaker), the figure shows the anchor points chosen for mapping from one stimulus to the other (○) and how stimulus pairs were constructed to probe acoustic-phonetic and phoneme category selectivity. Pairs are arranged to align with the category boundary as measured via discrimination for each individual subject (see Fig. 2), shown here at the 50% morph point between the two prototypes. Four different conditions were tested: M0, same category and no acoustic-phonetic difference; M3$_{within}$, same category and 33% acoustic-phonetic difference; M3$_{between}$, different category and 33% acoustic-phonetic difference; M6, different category and 67% acoustic-phonetic difference.

et al., 2012), it is currently unclear whether the underlying neural tuning properties predict behavioral performance.

Using fMRI rapid adaptation (fMRI-RA), we probed neural selectivity in human subjects listening to paired speech stimuli from a place-of-articulation continuum (/da/−/ga/). Subjects were asked to perform a difficult dichotic listening task for which phoneme category information was irrelevant. The following questions were addressed: (1) Would graded acoustic-phonetic differences between stimuli be detected in auditory cortical regions, especially in the left hemisphere? (2) Would phoneme category-selectivity be observed in the PMC despite it being task-irrelevant? Importantly, we asked whether the degree of category selectivity in PMC would be found to correlate with the sharpness of subjects' category boundary. The latter was measured outside the scanner after the imaging experiment. (3) In addition, structural equation modeling (SEM) was applied to reveal the source of auditory cortical input to PMC.

Materials and Methods

Participants. Sixteen subjects participated in this study (7 females, 18–32 years of age). The Georgetown University Institutional Review Board approval all experimental procedures, and all subjects gave written informed consent before participating. All subjects were right-handed, reported no history of hearing problems or neurological illness, and spoke American English as their native language. One subject (female) was excluded from the imaging study based on the behavioral pretest showing no evidence of categorical phoneme perception (see below). Imaging data from one additional subject (male) were excluded from subsequent analyses because of excessive head motion, leaving a total of 14 subjects in the imaging study.

Stimuli. To probe neuronal tuning for phonemes, we generated a place-of-articulation continuum between natural utterances of /da/ and /ga/ (Fig. 1) using the MATLAB toolbox STRAIGHT (MathWorks). STRAIGHT allows for finely and parametrically manipulating the acoustic and acoustic-phonetic structure of high-fidelity natural voice recordings (Kawahara and Matsui, 2003). Speech stimuli were recordings (44.1 kHz sampling frequency) of natural /da/ and /ga/ utterances taken from recordings provided by Shannon et al. (1999). Two phonetic continua (or "morphs") were generated at 0.5% intervals between /da/ and /ga/ prototypes: one for a male voice and one for a female voice. Morphed stimuli were generated up to 25% beyond each prototype (i.e., from −25% /ga/ to 125% /ga/), for a total of 301 stimuli per morph line. The stimuli created beyond the prototypes were

qualitatively assessed to ensure they were intelligible, and this assessment was then verified behaviorally in an identification experiment (described below), wherein subjects robustly labeled these stimuli appropriately. All stimuli were then resampled to 48 kHz, trimmed to 300 ms duration, and root-mean-square normalized in amplitude. A linear amplitude ramp of 10 ms duration was applied to sound offsets to avoid auditory artifacts. Amplitude ramps were not applied to the onsets, however, to avoid interfering with the natural features of the consonant sound.

Discrimination behavior. To identify participants' individual category boundaries while minimizing the risk that they would covertly categorize sounds in the scanner, participants completed a discrimination test before scanning. Participants' discrimination thresholds were measured at 10% intervals along each continuum, for stimuli displaced in each direction. The adaptive staircase algorithm QUEST (Watson and Pelli, 1983), implemented in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), was used to adjust the difference between paired stimuli based on subject performance. On each trial, participants heard two sounds and were asked to report as quickly and accurately as they could whether the two sounds were exactly the same, or in any way different. A maximum time of 3.0 s was allowed for a response before the next trial started. In half the trials, paired stimuli were identical and in the other half they were different. Of the different pairs, half were for displacement in one direction along the continuum, and half were in the opposing direction. Subjects completed 28 trials per condition, for 20 conditions (10% intervals from 0 to 90% with displacements toward 100%, and 10% intervals from 10 to 100% with displacement toward 0%), yielding 560 total trials.

Identification behavior. After subjects completed the fMRI experiment, they were asked to categorize the auditory stimuli along each continuum to confirm the location of participants' category boundary and to measure its sharpness. Categorization was tested at 10% intervals from −25% (i.e., 25% past /da/ away from /ga/) to 125% (25% past /ga/ away from /da/). On each trial, subjects were presented with a single sound and given up to 3 s to indicate as quickly and accurately as they could whether they heard /da/ or /ga/. Subjects completed 20 trials per condition for 15 conditions, for a total of 300 trials per morph. We then fit the resulting data with a sigmoid function to estimate the boundary location as well as boundary sharpness for each subject. The sigmoid was calculated according to the generalized logistic curve:

$$f(x) = \frac{1}{1 + e^{-(x-\alpha)/\beta}}$$

where x is the location along the morph line, α is the location of the boundary along the morph line, and $1/\beta$ is the steepness of the boundary (with larger values of $1/\beta$ resulting in sharper boundaries).

Event-related fMRI-RA experiment. After the phoneme identification screening, subjects participated in an fMRI-RA experiment. fMRI adaptation techniques have been developed to improve upon the limited ability of conventional fMRI to probe neuronal selectivity. Conventional fMRI takes average BOLD-contrast response to the stimuli of interest as an indicator of neuronal stimulus selectivity in a particular voxel. Although this is sufficient to assess whether neurons in the voxel respond to the stimuli of interest, a more detailed assessment of the degree of selectivity (and, of particular interest in our study, whether neurons show selectivity for acoustic features or for phoneme category labels) is complicated by the fact that the density of selective neurons as well as the broadness of their tuning contribute to the average activity level in a voxel. A particular mean voxel response could be obtained by a few neurons in that voxel, which each respond unselectively to many different sounds, or by a large number of highly selective neurons that each respond only to a few sounds, yet these two scenarios obviously have very different implications for neuronal selectivity in that particular region (Jiang et al., 2006). In contrast, the fMRI-RA approach is motivated by findings from inferotemporal cortex monkey electrophysiology experiments reporting that the second presentation of a stimulus (within a short time period) evokes a smaller neural response than the first (Miller et al., 1993). It has been shown that this adaptation can be measured using fMRI and that the degree of adaptation depends on stimulus similarity, with repetitions of the same stimulus causing the greatest suppression. We (Jiang et al., 2006, 2007) as well as others (Murray and Wojciulik, 2004; Fang et al., 2007; Gilaie-Dotan and Malach, 2007) have provided evidence that parametric variations in visual object parameters (shape, orientation, or viewpoint) are reflected in systematic modulations of the fMRI-RA signal and can be used as an indirect measure of neural population tuning (Grill-Spector et al., 2006). In fMRI-RA, two stimuli are presented in rapid succession in each trial, with the similarity between the two stimuli in each trial varied to investigate neuronal tuning along the dimensions of interest. In our experiment, in each trial, subjects heard pairs of sounds, each of which was 300 ms in duration, separated by 50 ms, similar to Joanisse et al. (2007) and Myers et al. (2009). There were five different conditions of interest, defined by acoustic-phonetic change and category change between the two stimuli: M0, same category and no acoustic-phonetic difference; M3$_{within}$, same category and 33% acoustic-phonetic difference; M3$_{between}$, different category and 33% acoustic-phonetic difference; M6, different category and 67% acoustic-phonetic difference; and null trials (Fig. 1). This made it possible to attribute possible signal differences between M3$_{within}$ and M3$_{between}$, which were equalized for acoustic-phonetic dissimilarity, to an explicit representation of the phonetic categories. Specifically, we predicted that brain regions containing category-selective neurons should show stronger activations in the M3between and M6 trials than in the M3within and M0 trials, as the stimuli in each pair in the former two conditions should activate different neuronal populations, whereas they would activate the same group of neurons in the latter two conditions. In this way, fMRI-RA makes it possible to dissociate phonetic category selectivity, which requires neurons to respond similarly to dissimilar stimuli from the same category as well as respond differently to similar stimuli belonging to different categories (Freedman et al., 2003; Jiang et al., 2007) from mere tuning to acoustic differences, where neuronal responses gradually drop off with acoustic dissimilarity, without the sharp transition at the category boundary that is a hallmark of perceptual categorization. Morph lines were also extended beyond the prototypes (25% in each direction) so that the actual stimuli used to create the stimulus pairs for each subject would span 100% of the difference between /da/ and /ga/ but could be shifted so that they centered on the category boundary for each subject, measured behaviorally before scanning. To observe responses largely independent of overt phoneme categorization, we scanned subjects while they performed an attention-demanding distractor task for which phoneme category information was irrelevant. Each sound in a pair persisted slightly longer in one ear than in the other ($\sim$30 ms between channels). The subject was asked to listen for these offsets

and to report whether the two sounds in the pair persisted longer in either the same or different ears. Subjects held two response buttons, and an "S" and a "D," indicating "same" and "different," respectively, were presented on opposing sides of the screen to indicate which button to press. Their order was alternated on each run, to disentangle activation resulting from categorical decisions from categorical motor activity (i.e., to average out the motor responses). The nature of this task required that subjects listen closely to all sounds presented. The average performance across subjects was 71%, indicating that the task was attention-demanding, minimizing the chance that subjects covertly categorized the stimuli in addition to doing the direction task. Importantly, as reaction time differences have been shown to cause BOLD response modulations (Sunaert et al., 2000; Mohamed et al., 2004) that could interfere with the fMRI adaptation effects of interest, reaction times on the dichotic listening task did not differ significantly across the four conditions of interest (ANOVA, $p > 0.17$). Images were collected for 6 runs, each run lasting 669 s. Trials lasted 12 s each, yielding 4 volumes per trial, and there were two silent trials (i.e., 8 images) at the beginning and end of each run. The first 4 images of each run were discarded, and analyses were performed on the other 50 trials (10 for each condition of interest). Trial order was randomized and counterbalanced using m-sequences (Buracas and Boynton, 2002), and the number of presentations was equalized for all stimuli in each experiment.

fMRI data acquisition. All MRI data were acquired at the Center for Functional and Molecular Imaging at Georgetown University on a 3.0-Tesla Siemens Trio Scanner using whole-head echo-planar imaging sequences (flip angle = 90°, TE = 30 ms, FOV = 205, 64 × 64 matrix) with a 12-channel head coil. A clustered acquisition paradigm (TR = 3000 ms, TA = 1500 ms) was used such that each image was followed by an equal duration of silence before the next image was acquired. Stimuli were presented after every fourth volume, yielding a trial time of 12.0 s. In all functional scans, 28 axial slices were acquired in descending order (thickness = 3.5 mm, 0.5 mm gap; in-plane resolution = 3.0 × 3.0 mm^2). After functional scans, high-resolution ($1 \times 1 \times 1$ mm^3) anatomical images (MPRAGE) were acquired. Auditory stimuli were presented using Presentation (Neurobehavioral Systems) via customized STAX electrostatic earphones at a comfortable listening volume ($\sim$65–70 dB) worn inside ear protectors (Bilsom Thunder T1) giving $\sim$26 dB attenuation.

fMRI data analysis. Data were analyzed using the software package SPM2 (http://www.fil.ion.ucl.ac.uk/spm/software/spm2/). After discarding images from the first 12 seconds of each functional run, echo-planar imaging images were temporally corrected to the middle slice, spatially realigned, resliced to 2 × 2 × 2 mm^3, and normalized to a standard MNI reference brain in Talairach space. Images were then smoothed using an isotropic 6 mm Gaussian kernel.

For whole-brain analyses, a high-pass filter (1/128 Hz) was applied to the data. We then modeled fMRI responses with a design matrix comprising the onset of predefined non-null trial types (M0, M3$_{within}$, M3$_{between}$, and M6) as regressors of interest using a standard canonical hemodynamic response function, as well as six movement parameters and the global mean signal (average over all voxels at each time point) as regressors of no interest. The parameter estimates of the hemodynamic response function for each regressor were calculated for each voxel. The contrasts for each trial type against baseline at the single-subject level were computed and entered into a second-level model (ANOVA) in SPM2 (participants as random effects) with additional smoothing (6 mm). For all whole-brain analyses, a threshold of at least $p < 0.001$ (uncorrected) and cluster-level correction to $p < 0.05$ using AlphaSim (Forman et al., 1995) was used unless otherwise mentioned.

Results
Behavior
Before scanning, subjects first participated in a behavioral experiment. The experiment served both as an initial screen to ensure subjects exhibited clear category boundaries for our stimuli as well as to locate the subject-specific category boundary for each of the two morph lines. To minimize the likelihood of subjects covertly categorizing the sounds during the scans, we did not have
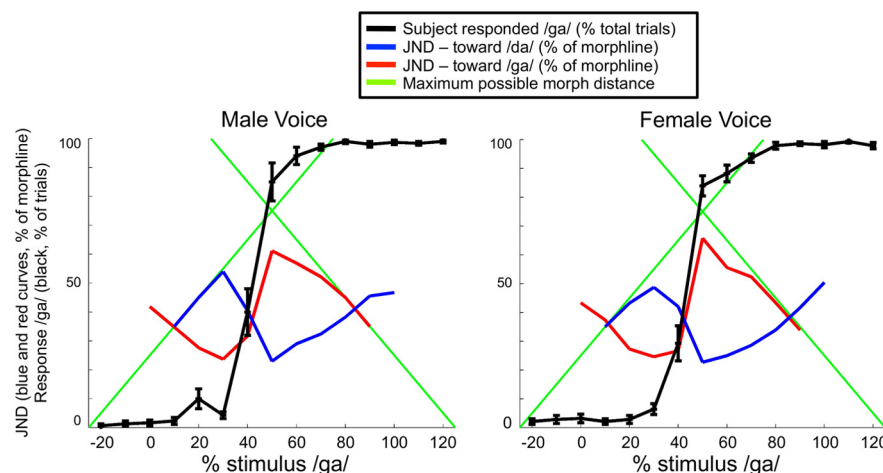
Figure 2. Behavioral discrimination performance outside the scanner. Before scanning, we measured subjects' JND at 10 percentage intervals along each morph line (one male voice, one female voice) in each direction: toward /da/ (blue) and toward /ga/ (red). The halfway point between the minima for these two curves accurately predicts the category boundary, measured directly after scanning (black; see Materials and Methods). Given that the morph lines have finite extents (25% past each endpoint, or 150% total, see Materials and Methods), the green lines indicate the maximum possible difference at each morph step. Thus, the fact that, in each direction, the JND decreases beyond the category boundary is an artifact that simply reflects the maximum possible difference for those steps. Error bars indicate SEM.

them overtly categorize the sounds before scanning. Instead, subjects were asked to report whether pairs of sounds were exactly the same or in any way different. Discrimination was tested at 10% intervals along each morph line, and the morph distance between paired sounds was varied according to the adaptive staircase algorithm QUEST (Watson and Pelli, 1983) in each direction of the morph line, similar to the experiments by Kuhl (1981). This allowed us to measure the just-noticeable difference (JND) at each location (for each morph direction), which should have its minimum value at the category boundary. In all but one subject, we observed clear minima in the JND for morph differences in each direction. The subject with no detectable boundary was excluded from further participation and analyses. The boundary in all other subjects was inferred to be halfway between the two smallest JND measurements (one in each direction). After scanning, subjects participated in a second behavioral experiment in which they were asked to overtly categorize each sound to explicitly confirm the location of the category boundary. Average discrimination and categorization results for the group (Fig. 2) indicate a clear correspondence between the JND minima and the explicit category boundary.

fMRI-RA

To probe neuronal selectivity using fMRI, we used an event-related fMRI-RA paradigm (Kourtzi and Kanwisher, 2001; Jiang et al., 2006, 2007) in which a pair of sounds of varying acoustic-phonetic similarity and category similarity was presented in each trial while subjects performed the dichotic listening task. To identify the subset of voxels that exhibited adaptation to our phoneme stimuli, we created a mask using the contrast of M6 > M0 (Table 1). As a control, we performed the inverse contrast of M0 > M6, which identified no selective clusters even at the very permissive threshold of $p < 0.1$ (corrected) anywhere in the brain. To locate cortical regions sensitive to acoustic-phonetic differences regardless of category label, we examined the results of the contrast $M6 + M3_{between} + M3_{within} > 3 * M0$ within the M6 > M0 mask. This contrast yielded two significant clusters (Fig. 3) in the left hemisphere; the anterior middle temporal gyrus (aMTG; MNI

coordinate, −56, −6, −22) and the posterior superior temporal gyrus (pSTG; MNI coordinate, −56, −50, 8). We then extracted the percentage signal change for each condition from each of these clusters. *Post hoc* paired t tests revealed that $M3_{within}$, $M3_{between}$, and M6 conditions were each significantly higher than the M0 condition in both the aMTG and pSTG ($p < 0.0028$, one-tailed).

Phoneme category selectivity in left PMC

We then examined the $M3_{between} > M3_{within}$ contrast on the activation masked by M6 > M0 (see above) to locate cortical regions explicitly selective for the phoneme categories. This contrast yielded only one significant cluster (Fig. 4), in the posterior portion of the left middle frontal gyrus (MNI coordinate, −54, 4, 44), corresponding to PMC (Brodmann area 6). *Post hoc* paired t tests showed that $M3_{between}$ and M6 conditions were each significantly higher than both the M0 and $M3_{within}$ conditions ($p < 0.007$, one-tailed), with no significant differences between M0 and $M3_{within}$ or between $M3_{between}$ and M6 ($p > 0.28$, two-tailed).

Functional connection between pSTG and PMC

To determine whether the category signals observed in PMC are functionally linked to either the anterior or posterior speech selectivity clusters identified in auditory cortex, we examined the functional connectivity between each acoustic-phonetic region of interest (ROI) (i.e., aMTG and pSTG) and the phoneme category-selective ROI in PMC. For this analysis, we extracted the measured BOLD time course in each of these group-defined ROI and calculated their pairwise correlation coefficients in each subject. A strong correlation was found between aMTG and pSTG (mean, $r = 0.41$), and the distribution of r values was found to be significantly different from zero ($p < 0.0001$). Similarly, connectivity was detected between pSTG and PMC (mean, $r = 0.28$, $p < 0.01$). However, correlations between aMTG and PMC were not significant (mean, $r = 0.11$, $p > 0.3$). Testing for partial correlations, controlling for the strong correlation between pSTG and aMTG, further supported the existence of a dorsal-stream connection from pSTG to PMC (mean, $r = 0.29$, $p < 0.005$), and no ventral stream connection from aMTG to PMC (mean, $r = 0.015$, $p > 0.8$). These results suggest that the category selectivity in PMC arises from processing in the dorsal stream (via pSTG) rather than the ventral stream, probably because of a direct projection from posterior auditory areas (e.g., pSTG) to premotor areas (e.g., PMC). To test this hypothesis, we used SEM, with the SEM toolbox (Steele et al., 2004), to assess the connectivity among the three brain regions. Using the pairwise correlation coefficients between the three regions (see above), we tested all 60 "probable" candidate models of connectivity among the three regions ($60 = 4^3$-4, after assuming that any one of the three regions is connected to at least one of the other two regions). The goodness of fit between the model and data was tested according to the standardized root mean square residual (SRMR), which can have values between 0 and 1 and where SRMR < 0.05 indicates a particularly good fit (Steele et al., 2004). The best-fitted structural equation model is shown in Figure 5 (SRMR:

Table 1. Cortical regions in which adaptation was observed for the contrast with greatest acoustic-phonetic dissimilarity (M6 vs M0)

Region	mm³	Z_E	MNI coordinates		
			X	Y	Z
L superior temporal gyrus	2564	4.54	−58	−50	10
		4.26	−44	−52	12
		3.79	−50	−54	22
R superior temporal gyrus	312	4.07	54	−44	14
R cerebellum	196	3.85	32	−64	−26
R superior parietal lobule	480	3.81	32	−64	48
		3.39	26	−68	54
		3.16	42	−66	42
L cerebellum	216	3.73	−10	−76	−32
R frontal lobe	276	3.72	44	4	38
L inferior temporal gyrus	284	3.66	−54	−6	−24
L middle temporal gyrus		3.42	−60	−8	−10
R brainstem	264	3.61	14	−14	−8
R supramarginal gyrus	224	3.52	44	−40	32
R inferior parietal lobule	152	3.38	42	−54	46
L precuneus	84	3.38	−10	−46	48

L, Left; R, right.

0.0246 ± 0.0193; mean ± SD). This model includes a bidirectional connection between aMTG and pSTG and a unidirectional connection from pSTG to PMC, suggesting that the PMC selectivity we observe via fMRI-RA can be attributed to the dorsal, rather than the ventral, auditory pathway. Additionally, the identified dorsal connection was found to be directional, with signals propagating from pSTG to PMC.

PMC signals predict behavior

The phoneme category selectivity identified in PMC raises the possibility that this area might be recruited when subjects are asked to explicitly categorize morphed phoneme stimuli. In a separate experiment, after the scanning, we therefore determined whether the category selectivity observed in left premotor cortex was correlated to subjects' ability to categorize the morphed phoneme stimuli. More specifically, we predicted that subjects with larger signal differences between the M3$_{between}$ and M3$_{within}$ conditions, indicative of stronger release from adaptation for stimuli in different phoneme categories relative to stimuli of the same acoustic-phonetic dissimilarity but belonging to the same category, would exhibit better categorization performance. To test this, we extracted the M3$_{between}$-M3$_{within}$ signal values for each subject from the group-defined PMC cluster and tested for correlations against the sharpness of the behavioral category boundary for each subject. Although there appeared to be a strong correlation for most of the subjects, there were several subjects ($n = 4$) for which M3$_{between}$-M3$_{within}$ was zero, making the correlation for the group insignificant. We hypothesized that this was attributable to intersubject variability in functional localization, and we refined our analysis by defining subject-specific PMC ROI. For each individual subject, we searched the M6 + M3$_{between}$ > M3$_{within}$ + M0 contrast map and identified activity peaks within a 20 mm radius of the group-defined PMC cluster. In cases where multiple peaks were detected, the one closest to the group-defined PMC cluster was chosen. We then centered a 5 mm spherical ROI on each of these coordinates and extracted the percentage signal change for each testing condition. Figure 6 shows the signal difference M3$_{between}$-M3$_{within}$ (in percentage signal change) measured in each subjects' PMC ROI against the sharpness of their behavioral category boundary, revealing a significant correlation between these measures ($r = 0.63$, $p < 0.014$).

Discussion

Although current theories of speech processing (Hickok et al., 2011; Rauschecker, 2011) implicate the dorsal auditory stream in sensorimotor integration—mapping speech sounds to their associated motor articulations, and vice versa— contentious debate still remains about whether the dorsal stream is involved in speech "perception" (decoding of speech sounds and subsequent mapping of these acoustically variable sounds to their invariant representations). Recent reports of correlative (Osnes et al., 2011; Alho et al., 2012; Szenkovits et al., 2012) and causal (Meister et al., 2007; Sato et al., 2009) evidence of PMC involvement in speech processing have given rise to the hypothesis that the dorsal stream might be recruited in a task-specific manner to support speech categorization tasks. Although providing circumstantial support for this hypothesis, these studies do not address whether or not PMC exhibits neural tuning properties that might be relevant to a speech categorization task.

What neural tuning properties are exhibited by PMC?

In the present fMRI-RA study, we detected selectivity for phoneme category in left PMC (Fig. 4). Although previous studies have suggested that the dorsal stream may be recruited to perform speech judgments (at least under adverse listening conditions), our present results provide insight into why such recruitment may be beneficial; in the case of phoneme categories, PMC may provide additional information (potentially useful for disambiguating similar words when sensory information is degraded). Whereas phoneme-category tuning was observed in left PMC, graded acoustic-phonetic selectivity was observed in left aMTG and left pSTG (Fig. 3). Although not tested in the current experiment, such neural tuning properties are likely useful in tasks requiring fine discriminations between speech sounds. Additional experiments would add valuable insight into why this tuning profile was observed in both anterior (ventral) and posterior (dorsal) regions and how the function of both regions differs (for a potential explanation, see Obleser et al., 2007).

Is PMC tuning the result of dorsal- or ventral-stream processing?

In our ROI and connectivity analyses, we identified a left-lateralized network in which connectivity was directional from pSTG to PMC. Similar directional connectivity, in that case between anatomically defined ROIs in planum temporale and premotor cortex, was observed in a previous study (Osnes et al., 2011). We interpret our finding to reflect processing within the dorsal stream. This interpretation is consistent with recent electroencephalography data (De Lucia et al., 2010), reporting selectivity for conspecific vocalizations in humans at 169–219 ms after stimulus in pSTG, and then later at 291–357 ms in PMC. Whereas Osnes et al. (2011) also found connectivity from middle superior temporal sulcus to PMC, however, we found no connectivity between aMTG and PMC. We interpret our finding to indicate that our difficult dichotic listening task (71% mean accuracy) was effective at distracting subjects from the task-irrelevant phoneme-category information. Our findings are also consistent with the sensorimotor integration function putatively assigned to the dorsal auditory pathway, which critically involves an internal modeling process (Hickok et al., 2011; Rauschecker, 2011), similar to that described in motor control theory (Wolpert et al., 1995). This internal model is thought to map motor articulations to the speech sounds that result (i.e., a "forward model") and can be inverted to map a speech sound to the motor articulations likely to have caused it (an "inverse model"). We interpret the category
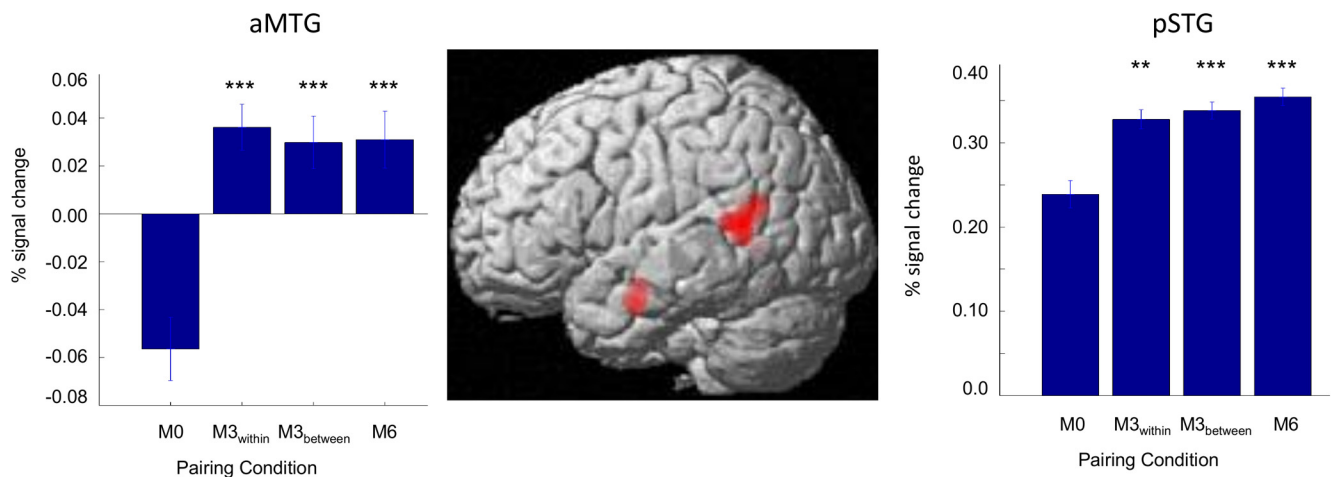
Figure 3. Neural selectivity for acoustic-phonetic features in left auditory cortex. Whole-brain contrasts of all different-sound pairs with greater signal than same-sound pairs ($M6 + M3_{between} + M3_{within} > M0$), masked by expected adaptation ($M6 > M0$), reveals only two significant clusters ($p < 0.05$, corrected for cluster-extent): left aMTG (MNI $-56 -6 -22$, etc.) and left pSTG (MNI $-56 -50\ 8$), shown in the middle in red, on a standard reconstructed cortical surface. Extracting percentage signal change for each testing condition in these clusters shows full release from adaptation for all different-sound pairs (i.e., $M3_{within}$, $M3_{between}$ and M6, shown to the right in blue bar graphs). Asterisks indicate significant differences from the M0 condition (**$0.001 \leq p < 0.01$, ***$p < 0.001$, *post hoc* paired *t* test). Error bars indicate SE.
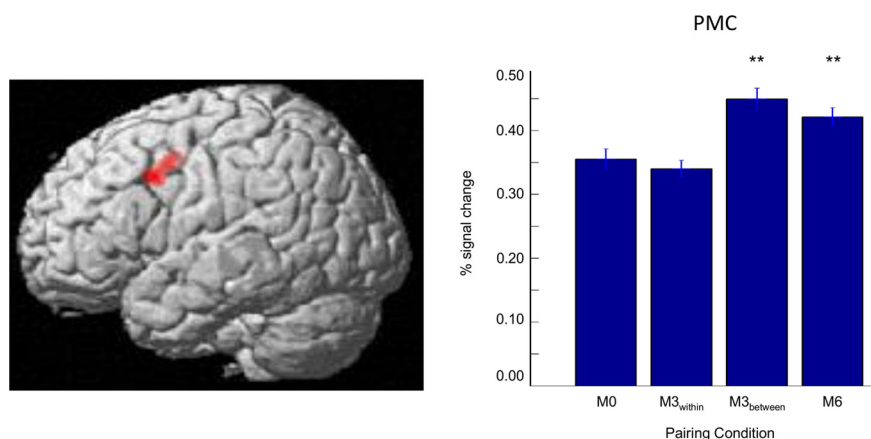

Figure 4. Neural selectivity for phonetic categories in left premotor cortex. Whole-brain contrasts of categorically different pairs with greater signal than categorically identical pairs with equal acoustic-phonetic difference ($M3_{between} > M3_{within}$), masked by expected adaptation ($M6 > M0$), reveals only one significant cluster ($p < 0.05$, corrected for cluster extent), in left middle frontal gyrus ($-54\ 4\ 44$), shown on the left in red on a standard reconstructed cortical surface. Extracting percentage signal change for each testing condition in this cluster shows no release from adaptation for same-category pairs (i.e., M0 and $M3_{within}$, shown to the right in blue bar graphs) and full release for different-category pairs (i.e., $M3_{between}$ and M6). Asterisks indicate significant differences from the M0 condition (**$p < 0.01$, *post hoc* paired *t* test). Error bars indicate SE.

Figure 5. Structural equation modeling shows connectivity of pSTG and PMC ROI. We used the SRMR (with SRMR < 0.05 indicating a good fit) to evaluate all probable models ($60 = 4^3$-4, with a prior assumption that any of the three regions could be connected with at least one other region). The best-fitted model is shown here, along with the mean connectivity "strength" ($\pm$SD). This model is a good fit for nearly all individual subjects: the SRMR from 13 of 14 subjects is <0.05 and 0.082 in the remaining subject.

selectivity observed in PMC to reflect the categorical sensorimotor representation of this inverse model.

Is dorsal stream phoneme-category selectivity automatic or task-specific?

We observed phoneme-category selectivity in PMC despite subjects never being informed of, and actively being distracted from, the phoneme-category information present in the stimuli. Our dichotic listening task proved to be difficult for subjects, who attained only 71% mean accuracy. Although we interpret this behavioral performance to indicate that the dichotic listening task was effective at distracting subjects from the task-irrelevant phoneme-category information, we cannot be certain that this manipulation of subject attention was 100% effective (i.e., subjects may have still attended to some degree to phoneme-category information). Nonetheless, our observation suggests that this process fits at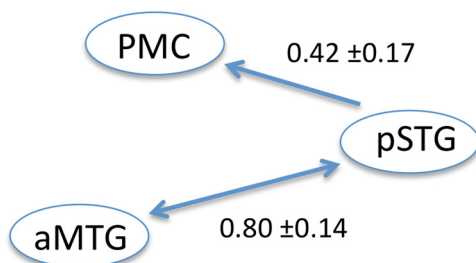 least one defining criterion for an automatic process, in that it is goal-independent, being performed even when it is irrelevant for the proximal goal (Moors and De Houwer, 2006), in our case, localization and not categorization. This suggests that this mapping takes place regardless of whether the resulting category-label information is task-relevant, which is compatible with general theories of cognitive processing implicating the premotor cortex in automatic processing in a variety of tasks (Ashby et al., 2007; D'Ostilio and Garraux, 2011), including other sensorimotor tasks, such as handwriting (Ashby et al., 2007; D'Ostilio and Garraux, 2011).

Do neural tuning properties in dorsal stream predict subject performance?

We found that subjects with more category-selective activity in PMC also exhibited sharper category boundaries. Thus, our findings provide novel evidence that the degree of category tuning
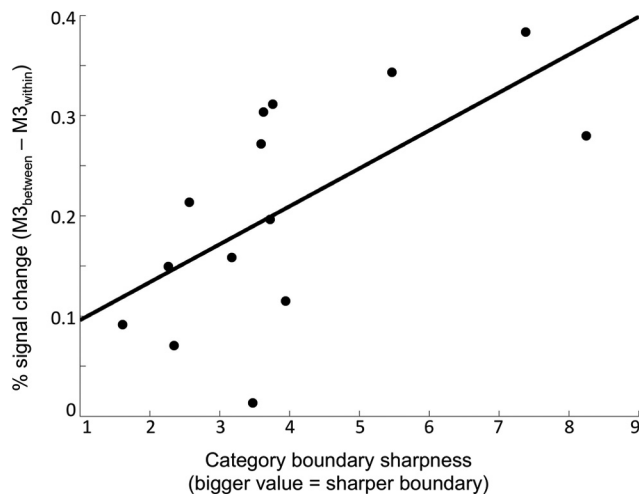
Figure 6. Phonetic selectivity in PMC predicts behavioral categorization ability. PMC ROI were defined for each subject ($n = 14$) by detecting the peak of the M6 + M3$_{between}$ > M3$_{within}$ + M0 contrast within a 20 mm radius of the group-defined PMC cluster. We then calculated the BOLD-contrast response difference between the M3$_{within}$ and M3$_{between}$ conditions for each morph line and subject (y-axis) and plotted this index against the sharpness of the category boundary measured behaviorally outside the scanner ($1/\beta$, see Materials and Methods; x-axis). Black line represents the regression line ($r = 0.63$, $p < 0.014$).

exhibited by PMC can account for behavioral performance in phoneme categorization, further supporting the idea that this representation might be usefully recruited to assist in speech categorization tasks. One possibility for this correlation is that category selectivity in PMC might merely be a corollary of earlier categorization processes somewhere else in the brain that may not have been detected by our fMRI-RA paradigm. Previous work has also reported category effects for speech in posterior STG (Liebenthal et al., 2010) and supramarginal gyrus (Raizada and Poldrack, 2007). However, a causal contribution of representations in PMC to phoneme categorization, at least in some circumstances, is supported by recently reported results from transcranial magnetic stimulation studies, where temporary disruption of the PMC has been found to interrupt (Meister et al., 2007; Möttönen and Watkins, 2009) or delay (Sato et al., 2009) phoneme category processing. Our results are also in agreement with a recent study using magnetoencephalography that found a correlation of left PMC responses with behavioral accuracy in a phoneme categorization task (Alho et al., 2012; Szenkovits et al., 2012).

Why did we not find tuning for phoneme category in the ventral auditory stream?

We found evidence of graded acoustic-phonetic selectivity in an anterior region of auditory cortex, which has also been reported by other studies (Frye et al., 2007; Toscano et al., 2010). This result is consistent with studies reporting phoneme selectivity in the ventral stream (Liebenthal et al., 2005; Leaver and Rauschecker, 2010; DeWitt and Rauschecker, 2012). One previous study (using a similar adaptation design as the present one and stimuli paired along a phonetic continuum) even claimed evidence for phonetic category selectivity in the left anterior superior temporal sulcus/middle temporal gyrus (Joanisse et al., 2007), which contrasts with the present results. A difficulty comparing the results of the two studies is that Joanisse et al. (2007) presented the stimuli used in their "within-category" and "same-category" conditions substantially more often than those used in

the "between-category" condition. This imbalance in stimulus presentations would be expected to cause increased adaptation for the "within" and "same-category" stimuli relative to the "between-category" stimuli (Grill-Spector et al., 1999), possibly producing the appearance of category selectivity in the adaptation paradigm despite an underlying graded noncategorical representation.

Two other recent studies have observed prefrontal correlates of phoneme category even when subjects performed noncategorical tasks, such as volume- or pitch-difference detection (Myers et al., 2009; Lee et al., 2012). We think that these results are compatible with ours; although the volume- and pitch-difference detection tasks may not require category information, it is unclear whether subjects covertly categorized the sounds. We interpret the lack of any ROI in prefrontal regions in our study, expected during explicit categorization (Binder et al., 2004; Husain et al., 2006), to indicate that our difficult dichotic listening task (71% mean accuracy) was effective at distracting subjects from the task-irrelevant phoneme-category information.

In conclusion, we have provided novel evidence in support of the hypothesis that dorsal-stream regions, specifically left PMC, are recruited to support speech categorization tasks. Whereas previous evidence was circumstantial, we provide evidence that PMC exhibits neural tuning properties useful for a phoneme-categorization task, that this processing may take place automatically (making the information available for recruitment on demand), and that tuning in this representation can account for differences in behavioral performance. It would be useful for future experiments to consider whether the role of PMC in categorization extends to speech sounds that are judged less categorically than stop consonants, such as vowels (Schouten et al., 2003), and perhaps even nonspeech stimuli. Indeed, evidence suggests that the same process of matching sensory inputs with their motor causes may also occur for other kinds of "doable sounds" outside the speech domain (Rauschecker and Scott, 2009). There is indeed experimental evidence for this operation to include musical sounds, tool sounds, or "action sounds" in more general terms (Hauk et al., 2004; Lewis et al., 2005; Zatorre et al., 2007).

References

Alho J, Sato M, Sams M, Schwartz JL, Tiitinen H, Jääskeläinen IP (2012) Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. Neuroimage 60: 1937–1946. CrossRef Medline

Ashby FG, Ennis JM, Spiering BJ (2007) A neurobiological theory of automaticity in perceptual categorization. Psychol Rev 114:632–656. CrossRef Medline

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295–301. CrossRef Medline

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. CrossRef Medline

Buracas GT, Boynton GM (2002) Efficient design of event-related fMRI experiments using M-sequences. Neuroimage 16:801–813. CrossRef Medline

De Lucia M, Clarke S, Murray MM (2010) A temporal hierarchy for conspecific vocalization discrimination in humans. J Neurosci 30:11210–11221. CrossRef Medline

DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A 109:505–514. CrossRef Medline

D'Ostilio K, Garraux G (2011) Automatic stimulus-induced medial premotor cortex activation without perception or action. PloS One 6:e16613. CrossRef Medline

Fang F, Murray SO, He S (2007) Duration-dependent FMRI adaptation and

distributed viewer-centered face representation in human visual cortex. Cereb Cortex 17:1402–1411. CrossRef Medline

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33:636 –647. CrossRef Medline

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 23:5235–5246. Medline

Frye RE, Fisher JM, Coty A, Zarella M, Liederman J, Halgren E (2007) Linear coding of voice onset time. J Cogn Neurosci 19:1476–1487. CrossRef Medline

Gilaie-Dotan S, Malach R (2007) Sub-exemplar shape tuning in human face-related areas. Cereb Cortex 17:325–338. CrossRef Medline

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24:187–203. CrossRef Medline

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23. CrossRef Medline

Hauk O, Johnsrude I, Pulvermüller F (2004) Somatotopic representation of action words in human motor and premotor cortex. Neuron 41:301–307. CrossRef Medline

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402. CrossRef Medline

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69: 407– 422. CrossRef Medline

Holt LL, Lotto AJ (2010) Speech perception as categorization. Atten Percept Psychophys 72:1218–1227. CrossRef Medline

Husain FT, Fromm SJ, Pursley RH, Hosey LA, Braun AR, Horwitz B (2006) Neural bases of categorization of simple speech and nonspeech sounds. Hum Brain Mapp 27:636–651. CrossRef Medline

Jiang X, Rosen E, Zeffiro T, Vanmeter J, Blanz V, Riesenhuber M (2006) Evaluation of a shape-based model of human face discrimination using FMRI and behavioral techniques. Neuron 50:159–172. CrossRef Medline

Jiang X, Bradley E, Rini RA, Zeffiro T, Vanmeter J, Riesenhuber M (2007) Categorization training results in shape- and category-selective human neural plasticity. Neuron 53:891–903. CrossRef Medline

Joanisse MF, Zevin JD, McCandliss BD (2007) Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. Cereb Cortex 17: 2084–2093. CrossRef Medline

Kawahara H, Matsui H (2003) Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. Proc ICASSP'2003 I:256–259.

Kuhl PK (1981) Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. J Acoust Soc Am 70:340–349. CrossRef

Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30:7604–7612. CrossRef Medline

Lee YS, Turkeltaub P, Granger R, Raizada RD (2012) Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. J Neurosci 32:3942–3948. CrossRef Medline

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25:5148–5158. CrossRef Medline

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631. CrossRef Medline

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR (2010) Specialization along the left superior temporal sulcus for auditory categorization. Cereb Cortex 20:2958–2970. CrossRef Medline

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M (2007) The essential role of premotor cortex in speech perception. Curr Biol 17:1692–1696. CrossRef Medline

Miller EK, Gochin PM, Gross CG (1993) Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. Brain Res 616:25–29. CrossRef Medline

Mohamed MA, Yousem DM, Tekes A, Browner N, Calhoun VD (2004)

Correlation between the amplitude of cortical activation and reaction time: a functional MRI study. AJR Am J Roentgenol 183:759–765. Medline

Moors A, De Houwer J (2006) Automaticity: a theoretical and conceptual analysis. Psychol Bull 132:297–326. CrossRef Medline

Möttönen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. J Neurosci 29: 9819–9825. CrossRef Medline

Murray SO, Wojciulik E (2004) Attention increases neural selectivity in the human lateral occipital complex. Nat Neurosci 7:70–74. CrossRef Medline

Myers EB, Blumstein SE, Walsh E, Eliassen J (2009) Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci 20: 895–903. CrossRef Medline

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple stages of auditory speech perception reflected in event-related FMRI. Cereb Cortex 17:2251–2257.

Osnes B, Hugdahl K, Specht K (2011) Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. Neuroimage 54:2437–2445. CrossRef Medline

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437– 442. CrossRef Medline

Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci 11:351–360. CrossRef Medline

Raizada RD, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. Neuron 56:726–740. CrossRef Medline

Rauschecker JP (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hearing Res 271:16–25. CrossRef Medline

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724. CrossRef Medline

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

Schouten B, Gerrits E, Hessen AV (2003) The end of categorical perception as we know it. Speech Commun 41:71–80. CrossRef

Schwartz JL, Basirat A, Ménard L, Sato M (2012) The Perception-for-Action-Control Theory (PACT): a perceptuo-motor theory of speech perception. J Neurolinguistics 25:336–354. CrossRef

Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X (1999) Consonant recordings for speech testing. J Acoust Soc Am 106:71–74. CrossRef Medline

Skipper JI, Nusbaum HC, Small SL (2005) Listening to talking faces: motor cortical activation during speech perception. Neuroimage 25:76–89. CrossRef Medline

Steele JD, Meyer M, Ebmeier KP (2004) Neural predictive error signal correlates with depressive illness severity in a game paradigm. Neuroimage 23:269–280. CrossRef Medline

Sunaert S, Van Hecke P, Marchal G, Orban GA (2000) Attention to speed of motion, speed discrimination, and task difficulty: an fMRI study. Neuroimage 11:612–623. CrossRef Medline

Szenkovits G, Peelle JE, Norris D, Davis MH (2012) Individual differences in premotor and motor recruitment during speech perception. Neuropsychologia 50:1380–1392. CrossRef Medline

Toscano JC, McMurray B, Dennhardt J, Luck SJ (2010) Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. Psychol Sci 21:1532–1540. CrossRef Medline

Turkeltaub PE, Coslett HB (2010) Localization of sublexical speech perception components. Brain Lang 114:1–15. CrossRef Medline

Watson AB, Pelli DG (1983) QUEST: a Bayesian adaptive psychometric method. Percept Psychophys 33:113–120. CrossRef Medline

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

Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. Science 269:1880–1882. CrossRef Medline

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music:

auditory-motor interactions in music perception and production. Nat Rev Neurosci 8:547–558. CrossRef Medline