

## Brain Mechanisms Implicated in the Preattentive Categorization of Speech Sounds Revealed Using fMRI and a Short-Interval Habituation Trial Paradigm

Marc F. Joanisse<sup>1</sup>, Jason D. Zevin<sup>2</sup> and Bruce D. McCandliss<sup>2</sup>

<sup>1</sup>Department of Psychology and Program in Neuroscience, The University of Western Ontario, London, Ontario N6A 5C2, Canada and <sup>2</sup>Sackler Institute for Developmental Psychobiology, Weill-Cornell Medical College, NY, USA

**A hallmark of categorical perception is better discrimination of stimulus tokens from 2 different categories compared with token pairs that are equally dissimilar but drawn from the same category. This effect is well studied in speech perception and represents an important characteristic of how the phonetic form of speech is processed. We investigated the brain mechanisms of categorical perception of stop consonants using functional magnetic resonance imaging and a passive short-interval habituation trial design (Zevin and McCandliss 2005). The paradigm takes advantage of neural adaptation effects to identify specific regions sensitive to an oddball stimulus presented in the context of a repeated item. These effects were compared for changes in stimulus characteristics that result in either a between-category (phonetic and acoustic) or a within-category (acoustic only) stimulus shift. Significantly greater activation for between-category than within-category stimuli was observed in left superior sulcus and middle temporal gyrus as well as in inferior parietal cortex. In contrast, only a subcortical region specifically responded to within-category changes. The data suggest that these habituation effects are due to the unattended detection of a phonetic stimulus feature.**

**Keywords:** categorical perception, fMRI adaptation, middle temporal gyrus, speech perception, superior temporal gyrus

### Introduction

Speech perception in humans involves processing rapidly changing and fast-fading acoustic events. Despite the complexity of speech, listeners quickly and effortlessly decode and categorize auditory phonetic information. How this is accomplished is the subject of much theoretical debate. In particular, a great deal of research has centered on establishing the degree to which speech perception involves domain-general versus domain-specific processes (Lieberman and Mattingly 1985; Kluender and Lotto 1999). This debate has extended to neuroimaging research, where studies have sought to identify speech-selective brain regions (for review, see Price et al. 2005). Of special interest is the extent to which neural regions are specialized for processing phonetic information. The phonetic form of speech consists of an abstract mapping between acoustic events that are perceived and the articulatory gestures that are used to produce them. Thus, neuroimaging studies of speech have examined neural activity when participants listen to speech stimuli compared with activity evoked by acoustically similar nonspeech sounds (Demonet et al. 1992; Zatorre et al. 1996; Binder et al. 2000; Burton 2001; Vouloumanos et al. 2001; Joanisse and Gati 2003). Studies of this kind have identified

regions that yield greater activation when participants actively process or passively attend to nonspeech sounds.

One challenge with comparing speech and nonspeech has been the difficulty in adequately equating the 2 types of stimuli. With respect to acoustic parameters, speech contains unique acoustic cues that distinguish it from other kinds of natural sounds (both nonspeech vocalizations and environmental sounds). For instance, speech is characterized by rapid temporal information, including transitions between periods of high and low acoustic energy and changes in frequency over very short periods. Speech also has intrinsic semantic content, given that it (usually) signals meaning. In contrast, nonspeech sounds are arguably more semantically sparse, in that they transmit a much more limited set of semantic information. This imbalance might make nonspeech sounds inadequate for use as a baseline in identifying regions of phonetic processing. To offset acoustic imbalances between speech and nonspeech, recent studies have made considerable efforts to create auditory stimuli that are closely equated with speech with respect to temporal and spectral acoustic content (Belin et al. 2000; Scott et al. 2000; Vouloumanos et al. 2001; Joanisse and Gati 2003; Liebenthal et al. 2005). Such studies have identified subregions of left superior temporal gyrus and sulcus (STG and STS) and middle temporal gyrus (MTG) that show specific activation for speech sounds. There is also some evidence that left inferior frontal gyrus (IFG) plays a greater role in speech versus nonspeech, especially with respect to tasks that involve fine-grained analysis of a sound's phonological structure (Zatorre et al. 1996; Burton et al. 2000; Burton 2001; Blumstein et al. 2005). Nevertheless, it remains unclear whether such differences are simply due to the experiential and semantic differences between speech and nonspeech sounds. For instance, some studies have identified significant overlap in STG/STS activation for speech and nonspeech sounds when both signals are closely matched for rapid temporal acoustic cues and when semantically neutral tasks are used (Jäncke et al. 2002; Joanisse and Gati 2003; for further discussion, see Price et al. 2005).

In the current study, we take a different approach to studying phonetic versus nonphonetic processing of speech sounds, which effectively sidesteps the issue of comparing speech stimuli to a nonspeech baseline. The study takes advantage of an important characteristic of phonetic processing that speech tends to be perceived categorically. That is, listeners accurately perceive acoustic differences that signal phonetically relevant distinctions but ignore those that are phonetically irrelevant. (Note that this phenomenon is especially strong in consonants, whereas vowels tend to be perceived in a more continuous or noncategorical fashion, Fry et al. 1962.) For

example, stop consonant voicing is signaled by the voice onset time (VOT) parameter, which is the temporal lag between the release of a stop consonant and the onset of voicing of the following vowel. Stops with VOTs of 20 ms or less tend to be perceived as voiced (e.g., /g/), whereas those with VOTs of 30 ms or more are perceived as voiceless (e.g., /k/; Liberman et al. 1958). Accordingly, listeners are more sensitive to a 20-ms VOT difference that crosses a phonetic boundary (e.g., 10 vs. 30 ms VOT, which are perceived as /g/ and /k/, respectively) than one that does not cross a boundary (0 vs. 20 ms VOT, perceived as 2 instances of /g/). Many types of acoustic cues serve to differentiate phonetic distinctions, for instance, the frequency and duration of steady-state formants and formant transitions.

The extent to which such an acoustic shift is perceived as a phonetic change also depends on the listener's native language. For instance, the /l/ and /r/ sounds are easily discriminated by native English speakers. However, Japanese speakers are very poor at making this same discrimination because it is not phonetically contrastive in their native language (Oyama 1978; McCandliss et al. 2002). A number of studies have taken advantage of this to examine the neural basis of phonetic versus nonphonetic perception. Callan et al. (2004) used functional magnetic resonance imaging (fMRI) to study the identification of r-l as well as vowels in groups of native Japanese and English speakers. They observed that both temporal and inferior parietal regions show stronger responses for discrimination of native contrasts. Jacquemot et al. (2003) also used fMRI to assess neural activation in Japanese speakers and French speakers as they actively discriminated native versus nonnative contrasts, in this case long versus short vowels (a distinction that is phonetic in Japanese, but nonphonetic in French) and open versus closed syllables (which are distinct in French but not Japanese). They too found regions of left temporal and inferior parietal cortices that responded specifically to phonetically contrastive distinctions.

Some have suggested that phonetic perception occurs even in very early processing mechanisms within STG. For instance, one study using magnetoencephalography (MEG) contrasted passive listening to acoustic and phonetic oddball stimuli within a continuous stream of syllables (Phillips et al. 2000). They found stronger mismatch field responses to phonetic shifts, replicating a number of previous electrophysiological and MEG investigations. In addition, magnetic source localization suggested the generator of this effect was located within auditory cortex in STG. This finding contrasts somewhat with the above-mentioned fMRI studies suggesting instead that such a region lies outside auditory cortex within areas of STS and MTG.

Interestingly, not all studies support the assertion that temporal regions are specifically involved in phonetic processing. Dehaene-Lambertz et al. (2005) used fMRI to investigate active discrimination of phonetic versus nonphonetic speech contrasts for sine wave speech analogue stimuli. In their study, a region of STG/STS showed greater activation for speech compared with nonspeech; however, it did not differ with respect to phonetically contrastive and noncontrastive differences. They did identify a parietal region (left supramarginal gyrus [SMG]) that did show such an effect, however. Finally, other studies have suggested that left IFG also has a role in phonetic perception. Blumstein et al. (2005) found that left IFG activation was significantly modulated by phonetic category goodness of fit during an overt identification task using a VOT continuum. In contrast, temporal and inferior parietal regions

showed selectivity only to end point stimuli. They interpret these data as indicating that frontal regions are used in coordinating the executive functions engaged by phonetic processing.

What is interesting about these and other studies of categorical perception is that they all identified a broad network of perisylvian regions that are engaged by auditory speech; however, only a subset of these regions appear to be specifically sensitive to phonetic distinctions, localized either in superior temporal (STG/STS) or in inferior parietal cortex (e.g., SMG), or both. A concern with these experiments is the degree to which the observed neural activation is the result of effortful processes engaged by the types of tasks and stimuli being used. Studies of speech sound categorization thus far have used overt task paradigms that require participants to discriminate, categorize, or otherwise monitor for specific characteristics of speech sounds. Task-dependent processing could lead to activation in regions that are not actually engaged during day-to-day speech perception. Moreover, because between-category distinctions are by definition more easily discriminated, presumed differences in phonetic versus nonphonetic discriminations could stem from an imbalance in the overt response being generated in either condition. Thus, although such results are informative with respect to understanding how individuals perform overt perception tasks of this type—which are after all the source of much of our understanding of categorical perception—the top-down processing that is necessarily involved in these tasks makes it difficult to determine the locus of bottom-up phonetic processing. A somewhat broader concern is that these tasks are also not ideal for studying speech perception in populations such as children, individuals with language impairments, as well as second language learners; these groups might tend to draw on different cognitive skills—or even the same cognitive skills but to different degrees—to perform speech perception tasks compared with neurologically healthy adults listening to speech in their first language.

In a recent study, Zevin and McCandliss (2005) took a different approach of using a passive listening paradigm to study discrimination of unattended stimuli. Their study adapted the mismatch negativity procedure more commonly used in electroencephalography (EEG) and MEG research (Dehaene-Lambertz 1997; Näätänen et al. 1997; Dehaene-Lambertz and Baillet 1998) to identify neural correlates of speech discrimination. The mismatch component is observed when a repeated token (a standard) changes to an oddball token (a deviant). Zevin and McCandliss adapted this paradigm to event-related fMRI by using short-interval habituation trials. Of interest was whether specific regions would show activation responsive to small changes in acoustic information that result in a change in phonetic category (in this case, the syllables /ra/ and /la/). Stimuli were presented in trains of 4 syllables, comprising either standard trials where the same stimulus was repeated 4 times or deviant trials in which the fourth stimulus was phonetically different from the first 3. Results revealed regions of left STG and SMG that responded more strongly to deviant trials. These habituation effects appear to be due to assemblies of cells that show an adaptation response to the repetition of a stimulus and consequently a recovery from adaptation when the stimulus changes (Grill-Spector and Malach 2001).

In the present experiment, we used this short-interval habituation paradigm along with high-field event-related fMRI to investigate mechanisms specifically related to detecting and discriminating phonetic information in auditory speech. One

question that arises from the Zevin and McCandliss (2005) study is the extent to which the habituation effects they observed were attributable to nonphonetic differences between the naturally produced stimuli they used. To address this, we compared habituation effects for both phonetically relevant and irrelevant acoustic changes within a single acoustic continuum of speech sounds, ranging from /ga/ to /da/. We hypothesized that we would observe stronger dishabituation for deviant derived from a different phonetic category than the standard compared with a nonphonetic difference that spanned an equivalent distance in the continuum. As in several EEG and MEG mismatch studies (Dehaene-Lambertz 1997; Sharma and Dorman 1999; Phillips et al. 2000; Dehaene-Lambertz et al. 2005; Sittiprapaporn et al. 2005), we expected that a phonetic difference would yield distinct effects from an acoustically equivalent nonphonetic difference—in one instance, a stronger mismatch response was observed for a smaller between-category acoustic change than for a larger within-category change (Näätänen et al. 1997). The contribution of fMRI is the ability to closely identify neural regions that are involved in processing this distinction, especially during passive perception.

## Methods

### Subjects

Participants were 10 neurologically healthy right-handed adults (7 female, 3 male) recruited from The University of Western Ontario community. All were native English speakers, and none reported significant experience in a second language. Mean age was 25.1 years (range 22–31 years). Each participant gave written informed consent before participating and was paid for their participation. Testing protocols were reviewed and approved by The University of Western Ontario Office of Research Ethics.

### Stimuli and Procedures

Auditory stimuli consisted of synthetic speech syllables created using a digital implementation of the Klatt cascade/parallel formant synthesizer (Klatt 1980). Waveforms were 155 ms in duration, digitized at 11 025 Hz (16-bit quantization). A continuum of 8 consonant-vowel syllables, ranging from /ga/ to /da/, was created by manipulating the onset consonant's  $F_2$  transition between 1640 and 1703 Hz and the  $F_3$  transition from 2100 to 2803 Hz. Both transitions were manipulated in evenly spaced increments (7 Hz increments for  $F_2$  and 78 Hz increments for  $F_3$ ; also see Appendix). All other parameters were held constant between stimuli. Pilot data indicated that naive adult listeners reliably categorized the first 4 stimuli in the continuum (i.e.,  $F_3 = 2100$ – $2412$  Hz) as /ga/ and the remaining stimuli (i.e.,  $F_3 = 2568$ – $2802$ ) as /da/. We selected 3 items in the continuum to serve as stimuli in this experiment:  $ga_1$ , which was the first item in the continuum ( $F_2$ : 1640 Hz,  $F_3$ : 2100 Hz);  $ga_4$  was the fourth step in the continuum and was also perceived as “ga” ( $F_2$ : 1661 Hz,  $F_3$ : 2334 Hz); and  $da_7$ , which was the seventh step and was perceived as “da” ( $F_2$ : 1682 Hz,  $F_3$ : 2568 Hz).

Each trial in the experiment consisted of the presentation of a train of 4 syllables with a 50-ms interstimulus interval, for a total of 770 ms (Fig. 1A). There were 3 conditions, which were differentiated by the relationship between the first 3 syllables to the fourth syllable in the train (Fig. 1B). In the REP condition, the  $ga_1$  syllable was repeated 4 times. In the 2 “deviant” conditions, a repeated syllable was played 3 times followed by a fourth stimulus that was acoustically different from it. In the “between-category deviant” (BW) condition, the  $ga_4$  syllable was presented 3 times, followed by the  $da_7$  syllable. In the “within-category deviant” (WN) condition, the  $ga_1$  syllable was played 3 times followed by  $ga_4$ . Note that both the BW and WN conditions involved a shift of the same distance along the continuum (i.e., the onset frequency of  $F_2$  and  $F_3$  differed by the same extent in both cases); as well, the direction of this change was constant across the 2 conditions, such that formant values increased in both conditions. Critically,

however, in the BW condition, the final token crossed a phonetic boundary, whereas in the WN condition, it did not.

The shift of 4 steps along the continuum represented the minimal acoustic shift necessary to produce a phonetic distinction in the BW condition. The magnitude of the shift was kept small in order to explore the sensitivity of the habituation paradigm. The previous Zevin and McCandliss (2005) study used end point stimuli that represented a very obvious difference and could have led to a “pop-out” effect that directed participants’ attention to the deviant trials. By employing smaller acoustic changes, the present stimuli would minimize any such effects.

Stimulus trains were played during silent intervals between scans (see below), using an fMRI-compatible headset, which along with foam earplugs also served to attenuate scanner noise. Participants’ heads were stabilized with foam padding. Each scanning run consisted of 7 repetitions of each trial type (BW, WN, REP) plus 7 silent (SIL) trials in which no stimulus was played. Trials were presented in pseudorandom order at 12-s intervals. Each participant was scanned over a minimum of 5 functional runs. Additional runs were acquired if time and participant comfort permitted, such that we obtained 10 runs for participants 3–8 and 9 runs for participants 9–10. (Note that a random-effects model was used to calculate groupwise statistics. As a consequence, differing the number of samples over participants was unlikely to skew the results.) Because we were interested in preattentive auditory perception effects, participants were told not to attend to the auditory stimuli. To help, participants also watched a feature movie of their choice during scanning, presented silently with subtitles.

### Imaging Acquisition and Analysis

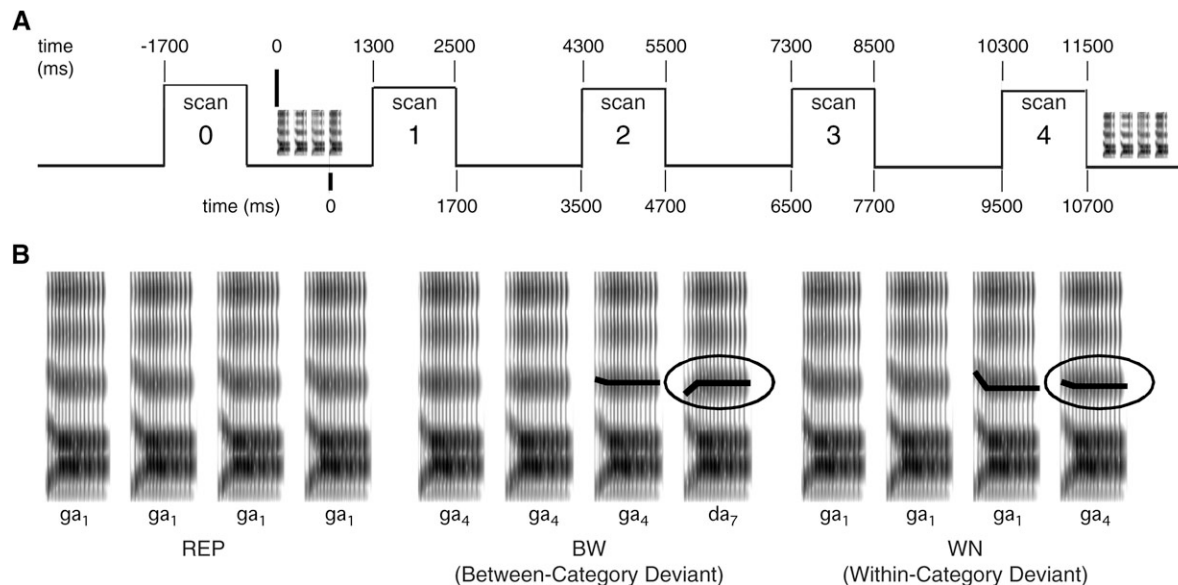
MRI images were acquired at 4 T using a Varian/Siemens scanner equipped with a hybrid quadrature head coil for signal transmission and reception. Robust automated shimming technique using arbitrary mapping acquisition parameters (RASTAMAP) automatic shimming was performed at the start of the scanning session to optimize magnetic field homogeneity (Klassen and Menon 2004). A total of 113  $T_2^*$ -weighted volumes were acquired in each functional run using a 2-shot navigator-corrected spiral pulse sequence for blood oxygen-dependent (BOLD) imaging: volume acquisition time = 1200 ms, echo time (TE) = 15 ms, Flip = 60°. An additional 4 volumes were acquired but discarded at the start of each run to permit  $T_2^*$  signal levels to stabilize prior to stimulus presentation. A clustered volume acquisition procedure was used (Talavage et al. 1999), which permitted the presentation of stimuli during the 1800-ms silent periods between scans, thus yielding an effective repetition time (TR) of 3000 ms (Fig. 1).

Stimuli were presented after every fourth volume. To avoid masking effects, a 500-ms gap was left between the end of the scan and the start of the stimulus train. Each volume consisted of eleven  $64 \times 64$  transverse slices (voxel size:  $3 \times 3 \times 3$  mm, 192 mm field of view), with the centermost slice aligned with the sylvian fissure. This slice plan captured the STG and MTG and STS and middle temporal sulcus, the IFG, the inferior parietal lobe, as well as portions of the occipital lobes (Fig. 2). Note that this slice prescription ignores portions of the inferior temporal lobe as well as the superior portion of the frontal, parietal, and occipital lobes, which are of less interest in the present study. It instead permitted us to acquire high-resolution functional volumes in a short period of time, minimizing the relative duration of scanning versus silent periods. Whole-head anatomical scans were acquired within session using a  $T_1$ -weighted 3-dimensional (3D) spiral pulse sequence:  $256 \times 256 \times 120$  voxels ( $0.75 \times 0.75 \times 1.5$  mm), TE = 3 ms, TR = 50 ms, inversion time = 1300 ms.

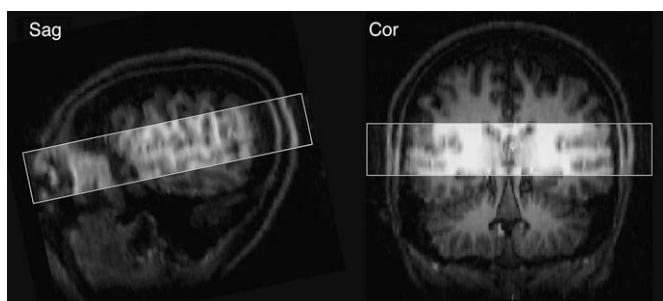
Preprocessing and statistical analyses were performed using Brain-Voyager QX (Brain Innovation B.V., Maastricht, The Netherlands, Max Planck Society). Scans were preprocessed as follows: functional scans were aligned to anatomical scans, resampled to  $1 \text{ mm}^3$  resolution, and transformed to the stereotaxic space of Talairach and Tournoux (1988). Rigid-body motion correction was applied and indicated no movement greater than one voxel in any direction for any participant. Spatial smoothing was performed in 3D using an 8-mm full-width, half-maximum Gaussian kernel. Temporal smoothing was performed using a high-pass filter that removed any low-frequency drift or oscillations in the data (cutoff frequency of 3 cycles per run).

Statistical analyses were conducted by convolving the onset of each stimulus train (REP, BW, and WN) with a hemodynamic response





**Figure 1.** Schematic diagram of the stimulus presentation paradigm. (A) Timing of scans relative to stimulus presentation, taking as the stimulus onset either the onset of speech sounds (upper numbers) or the onset of the fourth stimulus that distinguishes the deviant trials (lower numbers). (B) Spectrograms of REP, WN (within category), and BW (between category) stimulus trains. The cue that differentiated the deviant from the preceding standards is the onset frequency of the second and third formants (third formant difference is emphasized and circled). In the WN condition, the standard and deviant are acoustically different ( $ga_1$  and  $ga_4$ ), but both are perceived as / $ga$ /. In contrast, the BW condition uses the  $ga_4$  and  $da_7$  syllables, such that the standard is perceived as / $ga$ / and the deviant is perceived as / $da$ /.



**Figure 2.** Slice plan of functional scans for a representative participant. A 33-mm-thick volume was acquired (rectangular region), with the center slice aligned along the sylvian fissure. The narrow slice prescription permitted the acquisition of high-resolution images within a relatively short volume acquisition time (1200 ms).

predictor (Delta: 2.5, Tau: 1.25). A random-effects general linear model (GLM) was used to identify neural regions significantly correlated with each predictor at a voxelwise statistical threshold of  $t = 4.0$ ,  $P < 0.003$ . We ran a correction performance analysis (AlphaSim, Forman et al. 1995) to determine an appropriate contiguity threshold to control for multiple comparisons. This implemented a 10 000-iteration Monte Carlo simulation of the functional acquisition volume taking into account spatial smoothing and voxelwise statistical threshold parameters. Given the discrete integer nature of the contiguity threshold, we adopted the highest corrected alpha level falling below  $P < 0.05$ . This resulted in a 4 contiguous voxel threshold corresponding to an alpha level of  $P < 0.002$ .

The resulting GLM was used to obtain 3 statistical contrast maps: 1) SPCH > SIL contrasted the combined 3 speech conditions (SPCH) with the no stimulus (SIL) condition trials to identify regions associated with presentation of the auditory stimuli compared with baseline. 2) BW > WN compared the BW trials with the 2 speech conditions that did not involve a phonetic change. 3) WN > REP examined the extent to which a within-category change differed from the no-change (REP) condition. This contrast sought to identify any regions showing a strictly acoustic effect. Note the BW condition was excluded from the negative portion of this contrast, on the logic that it represented both an acoustic and phonetic change, and including it could have obscured a pure acoustic habituation effect.

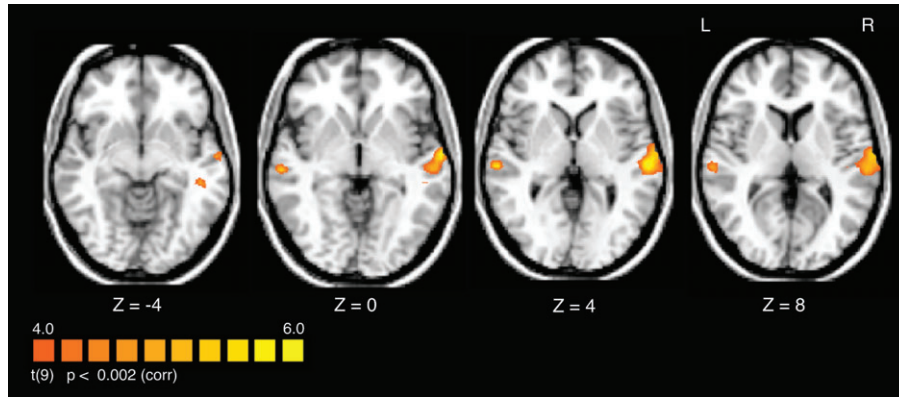
### Behavioral Posttest

Overt behavioral discrimination data were also acquired on the stimulus trains from 8 of the fMRI participants (2 participants were not available for the posttest). The purpose of this was to examine whether passive neural responses to these stimuli are congruent with a more meta-linguistic task requiring overt change detection. Testing took place a minimum of 6 months after scanning to assure that familiarity with the stimuli did not artificially influence task performance. Participants were instructed to listen to each syllable train over headphones and judge whether the final syllable was identical to the first 3. Responses were indicated on paper by circling SAME or DIFFERENT for each trial. Each REP, BW, and WN stimulus train was played 4 times each in pseudorandom order.

### Results

#### SPCH > SIL

The first contrast examined brain regions that showed a significant contrast between all speech stimuli and the silent (no stimulus) condition (Fig. 3, Table 1). Statistical maps were obtained by contrasting the no-change (REP), BW, and WN trials versus the SIL baseline trials. We observed large regions of bilateral activation in STG and STS. Activation in STG extended into lateral portions of A1 cortex in both the left and right hemispheres, assessed using anatomical constraints described in Penhune et al. (1996) and Rademacher et al. (2001). Activation also extended somewhat to the inferior plane of right SMG. This bilateral pattern of activation is consistent with what was observed in a prior experiment using a similar paradigm (Zevin and McCandliss 2005). The extent of activation in STG appears to be slightly greater here than in the previous study, however. In particular, we observed significant activation in A1 cortex, whereas this was not observed as strongly in the earlier study. As we discuss further below, this could be due to the use of a higher magnetic resonance imaging field strength, shorter volume acquisition time, and somewhat quieter spiral



**Figure 3.** SPCH > SIL contrast: fMRI activation of the combined speech conditions (REP, BW, WN) compared with SIL trials during which no stimulus was played. Activation was observed in the posterior portions left and right superior temporal gyri and sulci (STG, STS), including primary and secondary auditory regions. Statistical maps are overlaid on a Talairach-transformed  $T_1$ -weighted anatomical scan.

**Table 1**

Regions of significant activation for each contrast performed

Region	Talairach coordinates			Size (mm <sup>3</sup> )	Mean <i>t</i> value
	<i>x</i>	<i>y</i>	<i>z</i>		
<b>SPCH &gt; SIL</b>					
L STG/STS (BA 22)	-58	-21	4	724	4.60
R STG (BA 41)	58	-18	6	3941	4.72
R STG/STS/SMG (BA 22)	41	-23	14	146	4.16
R MTG (BA 21/22)	46	-31	-3	222	4.42
<b>BW &gt; WN</b>					
L STS/MTG (BA 21)	-66	-26	7	223	4.52
L STS/MTG (BA 21)	-64	-25	-7	141	4.41
L AG/SMG (BA 39/40)	-47	-63	39	478	5.12
<b>BW &gt; WN + REP</b>					
L STS/MTG (BA 21)	-67	-27	8	222	4.76

Statistical threshold was set at  $t = 4.0$ , corrected to  $\alpha = 0.002$  (voxelwise  $P = 0.003$ , minimum cluster size: 108 mm<sup>3</sup>). L, left; R, right; BA, Brodmann area.

pulse sequence. We also noted a larger extent of voxels in the right hemisphere.

### Habituation Effects

We analyzed the full data set using 3 contrasts to examine responses specific to phonetic processing. The BW > WN contrast isolated brain regions that showed increased activation when a series of repeated stimuli were followed by a stimulus from a phonetically different category. This contrast controls for the effect of acoustic differences between the repeated and change stimuli, so that observed responses might be interpreted as indicating areas that are specifically sensitive to phonetic contrast. All significant clusters are listed in Table 1. We observed significant effects in 2 regions of left STG: one region along the lateral plane of Heschl's gyrus adjacent to A1 (illustrated in Fig. 4A) and another more medial cluster located anterior to Heschl's gyrus. This contrast also revealed a significant cluster of activated voxels in a posterior portion of left parietal cortex extending from the SMG to the border of the angular gyrus (AG). Additionally, we examined the difference between the BW condition and the 2 conditions in which no phonetic change was present (BW > WN + REP). This analysis again identified significant STS/MTG activation (Table 1); however, the SMG activity was weaker and failed to reach statistical threshold at both the voxelwise and cluster size

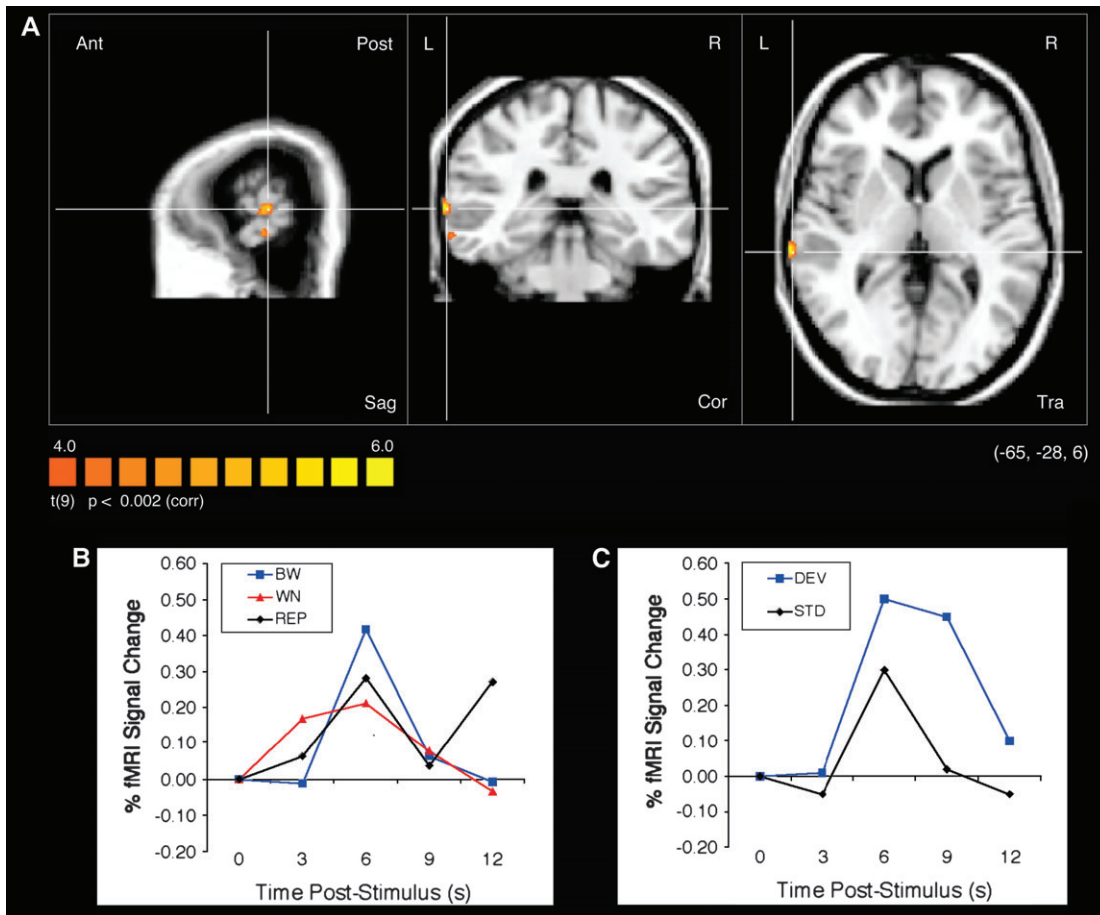
thresholds (i.e., a 20 mm<sup>3</sup> cluster in this region was significant at  $t_9 = 3.20$ ,  $P < 0.01$ ).

The WN > REP contrast was also performed to identify brain regions demonstrating a habituation response to a purely acoustic change in stimulus properties. Recall that the WN condition involved a stimulus shift but not a change in phonetic category. No cortical region achieved statistical threshold for this contrast. We did observe a small cluster of significant voxels in a subcortical region (head of the left caudate nucleus; Talairach coordinates: -7, 11, 1; cluster size: 13 mm<sup>3</sup>) that reached significance in uncorrected voxelwise analyses but did not meet the cluster size threshold.

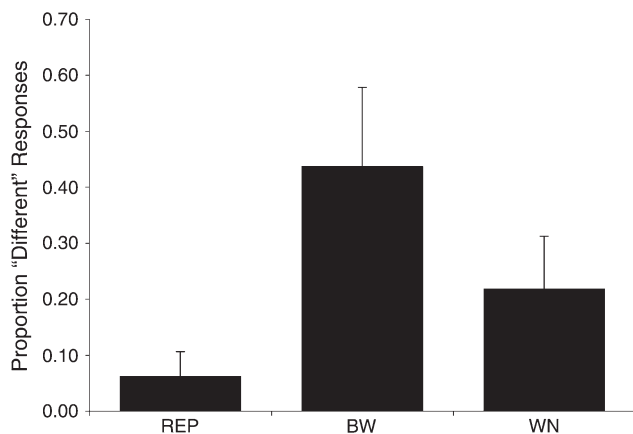
To further examine the nature of the neural responses in areas identified in the BW > WN analyses, we also conducted region of interest (ROI) analyses within the left superior/middle temporal and inferior parietal regions. Mean subjectwise GLM beta weights were calculated for significant voxels within these 2 anatomical ROIs (i.e., all voxels that reached threshold for the BW > WN contrast, above). Paired-samples *t*-tests were then used to compare average beta weights of each condition. In the STG/MTG region, the BW condition yielded significantly greater beta values than both the WN condition,  $t_9 = 5.68$ ,  $P < 0.001$ , and the REP condition,  $t_9 = 2.56$ ,  $P < 0.05$ . The WN and REP conditions did not differ significantly from each other. A somewhat different pattern was observed in the inferior parietal ROI, however. Here, beta values were significantly different for the BW versus WN conditions,  $t_9 = 6.88$ ,  $P < 0.001$ ; however, the BW condition did not differ significantly from the REP condition. In addition, the REP condition exhibited significantly greater beta values than the WN condition ( $t_9 = 2.54$ ,  $P < 0.05$ ).

### Behavioral Posttest

Discrimination performance is plotted in Figure 5 and indicated higher rates of "different" responses for the BW condition compared with both the REP and WN conditions. Although there was a strong bias to respond "same" for all stimulus conditions, the distribution of "different" responses was significantly different from a uniform distribution,  $\chi^2_2 = 11.55$ ,  $P < 0.005$ . This supports our assumption that categorical perception effects would lead to better discriminability in the BW condition. The generally low discrimination rates likely reflect the choice of using minimal acoustic shifts in the BW and WN stimulus trains. This was done to minimize any attentional



**Figure 4.** BW > WN contrast. (A) Region of lateral left temporal cortex exhibiting habituation effects for a between-category change in a consonant sound. (B) Event-related curves showing the time course and magnitude of fMRI activation for the conditions in the experiment. (C) Event-related curve from Zevin and McCandliss (2005) obtained from left STG showing greater activation levels for a deviant condition compared with a repeated standard.



**Figure 5.** Overt discrimination of stimuli in the fMRI experiment obtained from 8 participants a minimum 6 months after scanning. Participants heard the same 4-syllable stimulus trains as in scanning but were asked to explicitly identify whether the fourth stimulus was different from the first 3.

pop-out effect these shifts might evoke, which could have led to unintended explicit processing. Note that the relatively low discrimination rates observed in the behavioral data are still consistent with what other studies have found for discrimination of synthesized speech involving small acoustic differences (Sussman 1991; Joanisse and Gati 2003; Serniclaes et al. 2005).

## Discussion

The finding that many speech sounds are categorically perceived has generated considerable interest in speech research. It has helped inform theories as to the cognitive and neural mechanisms involved in speech perception (Liberman and Mattingly 1985; Massaro 1997) and has also been used to study how infants acquire and “lose” phonetic distinctions as a result of exposure to their native language (Werker and Tees 1984; Kuhl et al. 1992; Jusczyk 1997). Finally, some children with language and reading disorders show poor categorical perception of speech sounds, indicating an underlying phonological processing deficit (Godfrey et al. 1981; Werker and Tees 1987; Joanisse et al. 2000).

Previous neuroimaging studies have studied categorical perception ability using overt identification and discrimination tasks. The typical experimental logic has been to identify brain regions that are preferentially active while participants make overt judgments about speech sounds compared with non-speech signals (Demonet et al. 1992; Burton et al. 2000; Vouloumanos et al. 2001; Joanisse and Gati 2003; Zatorre et al. 2004). Such studies have identified a number of brain regions that appear to be selectively activated during speech perception, including left STG/STS and IFG. The present study took a different approach. First, a passive-listening paradigm was used that did not require participants to generate overt responses or attend to any specific aspect of the stimuli. This had the benefit



of minimizing activation that might be due to attention and decision processes not usually engaged during everyday speech perception. Second, the short-interval habituation paradigm took advantage of neuronal adaptation effects to identify regions that are sensitive to acoustic changes in stimulus properties that specifically signal a change in phonetic category. Third, this study did not rely on a nonspeech baseline condition to help identify regions that are selective for phonetic processing. Finding an appropriate type of signal to serve as a comparison measure has proven to be difficult, because speech involves many acoustic characteristics that make it unique, and humans have much more extensive experience with speech sounds compared with other types of stimuli. The present study sought to overcome these issues by using speech sounds for both a phonetic condition (between category or BW) and a non-phonetic condition that involve an acoustically similar stimulus shift (within category or WN).

The results replicate and extend previous studies of phonetic categorization. In particular, by comparing responses on trials containing a change in phonetic category with trials containing an acoustic change, we were able to identify 2 regions of the left hemisphere that selectively respond to phonetic change even when participants are not performing a conscious perceptual task or generating an overt response.

#### ***General Responses to Speech Stimuli Throughout STG Bilaterally***

The SPCH > SIL contrast identified brain regions that are engaged during passive auditory presentation of speech sounds. As anticipated, this analysis identified large bilateral areas of activation extending across STG, STS, and some parts of SMG. All these regions are known to be implicated in basic auditory perception (Belin et al. 2000). The current study found significant activation for speech throughout STG including some portions of primary auditory cortex (A1, Penhune et al. 1996; Rademacher et al. 2001). This indicates that the current paradigm can reliably identify stimulus-correlated activation in A1, allowing us to address questions about the role of auditory cortex in phonetic processing. For instance, it validates the use of subsequent contrasts that investigate whether phonetic processing occurs within primary auditory cortex.

Activation in A1 was not observed in the earlier Zevin and McCandliss study (Zevin and McCandliss 2005). There are several explanations for this. First, the present study scanned participants at a higher field strength (4 rather than 3 T) and used gradient coils specifically optimized for 3D functional imaging. Both factors could have provided a moderate increase in signal-to-noise ratio, increasing our ability to identify activation in this region. Second, acoustic noise was less of a factor in this experiment due to a shorter volume acquisition time (1200 vs. 1500 ms), the use of a custom sound dampening bore liner (Mechefske et al. 2002), and a generally quieter scan sequence (i.e., spiral-in/out sequences like the one used by Zevin and McCandliss creates a greater amount of acoustic noise than the spiral-in sequence used here, Preston et al. 2004; Gaab et al. forthcoming). The reduced masking effects of scanner noise on the auditory stimuli likely improved our ability to detect stimulus-dependent BOLD signal increases in A1. Finally, the stimuli used in this study may have yielded stronger A1 activation because they required finer grained acoustic processing. The stop consonants /g/ and /d/ are signaled by relatively brief 30-ms frequency sweeps; in contrast, the earlier study used

/r/ and /l/ liquids as stimuli, which are signaled by temporally coarser grained acoustic cues (80–100 ms). These stimulus differences might also explain why we observed a less sustained event-related BOLD response in the present study (Fig. 4B) than in the previous one (Fig. 4C).

#### ***Responses to Phonetic Change in Left Posterior Regions***

The critical conditions in this study were the 2 “deviant” conditions designed to identify regions that showed a dishabituation response to a change in a repeated speech sound. We suggest that a neural habituation response occurs when a sound is presented repeatedly and that dishabituation occurs when an oddball stimulus is played, marked by increased fMRI signal levels (Grill-Spector and Malach 2001). We hypothesized that different types of acoustic shifts can lead to different patterns of neural activation; specifically, the BW condition involved an acoustic shift that crossed a phonetic boundary, whereas the WN condition involved the same magnitude of acoustic shift but did not result in a phonetic change. We found that the BW condition did in fact yield a reliable increase in fMRI activation in perisylvian regions when contrasted with the WN condition. We suggest that these regions of activation reflect neural mechanisms forming the basis of the categorization and subsequent discrimination of phonetic information in speech.

This effect appeared to be restricted to the left hemisphere, a finding that contrasts with what was found for SPCH > SIL, in which we observed a strongly bilateral effect in STG. It indicates that, even though bilateral auditory processing mechanisms are engaged for passive perception of speech sounds, specifically phonetic mechanisms are more narrowly localized in the left hemisphere. This finding is also consistent with prior studies indicating stronger left hemisphere preference for speech sounds versus nonspeech (Vouloumanos et al. 2001; Liebenthal et al. 2005).

Also of interest was whether the A1 region of STG would show phonetic habituation effects. Although this region was strongly activated for the SPCH > SIL contrast, we did not observe activation in this region for the BW > WN contrast. This result suggests that, even though the imaging and stimulus presentation paradigm was sufficient to identify A1 activation in response to speech, this region does not appear to play a more specific role in phonetic discrimination. Instead, regions that were selective for between-category changes in consonants were located lateral and inferior to A1. This seems consistent with prior imaging studies suggesting that belt regions surrounding A1 are selective for spectrally and temporally complex sounds (Thivard et al. 2000; Wessinger et al. 2001). On the other hand, it suggests that phonetic processing occurs at a somewhat later stage of the auditory stream.

Regions selective for a phonetic contrast included a narrowly proscribed region along the lateral plane of the temporal lobe spanning STS and MTG. This region is consistent with previous fMRI studies comparing phonetic and nonphonetic perception (Celsis et al. 1999; Jacquemot et al. 2003; Callan et al. 2004). In addition, we also observed activation in a posterior parietal region extending between SMG and AG, which has also been reported in some earlier studies (Callan et al. 2004; Dehaene-Lambertz et al. 2005). One question that arises is why an earlier study by Dehaene-Lambertz et al. (2005) failed to reveal a significant contrast for a similar between- versus within-category contrast in the posterior temporal regions. One reason could be that their study used sine wave speech, which tends to

be perceived as quite unnatural and requires some experience to be perceived reliably. Their study also used an overt discrimination task, which might also have modulated how temporal regions responded to the stimuli.

We also noted that our findings with respect to the BW > WN contrast were restricted to posterior brain regions. Some previous studies have also observed left IFG activation during speech processing (e.g., Poldrack et al. 1999; Joanisse and Gati 2003; Blumstein et al. 2005). However, it is notable that these studies involved active listening tasks. The present results seem to support the hypothesis that left IFG is engaged by executive processing demands that arise during active processing, rather than sensory components of phonetic processing (Binder et al. 2004; Blumstein et al. 2005). On this explanation, the lack of frontal activation in the face of robust posterior activation likely reflects the minimization of executive processing demands in the current paradigm. This pattern is also highly convergent across the present study and the earlier study by Zevin and McCandliss (2005).

The greater activation for the BW condition relative to the other 2 conditions in posterior temporal cortex is interpreted as resulting from a neural dishabituation response to a phonetic change. This is well supported in the superior/middle temporal region, where activation for the WN condition was similar to what was observed in the REP condition. However, the same cannot be said for the inferior parietal region, where we observed greater activation for both the REP and BW conditions compared with the WN condition. This seems contrary to a dishabituation explanation because the REP condition did not consist of any type of change, either phonetic or acoustic. The explanation could instead be that the inferior parietal region showed selective activation to a phonetic stimulus change due to a suppression of activation when a nonphonetic acoustic stimulus change occurs (i.e., the WN condition). This suppression did not occur in the REP condition because no change occurred, phonetic or otherwise. Although this conclusion is speculative, it does underline an interesting distinction between the apparently phonetic responses observed in temporal versus inferior parietal regions.

A potential limitation of the present study has to do with the stimuli that were used. It is argued that the key difference between the BW and WN conditions is that the BW condition involves a change to a different phonetic category. However, there is another difference between the 2 conditions that should be pointed out. As illustrated in Figure 1, the BW condition involved a change from a falling to a rising  $F_3$  transition, whereas the WN condition did not. This difference is strictly coincidental because the direction of change in formant transitions does not represent a useful cue to discriminating phonetic categories under most circumstances. That is, no single acoustic cue such as this one will reliably correspond to a phonetic distinction, a phenomenon that has come to be known as the “lack of invariance problem” in speech perception. Nevertheless, it is an open question whether this acoustic cue contributed to the habituation effects observed here.

A second potential limitation concerns the relationship between the imaging data and the behavioral data. Whereas the activation in left STG demonstrated strong sensitivity to the BW > WN contrast, the behavioral data demonstrated less dramatic sensitivity to this difference. Note, however, that the behavioral data are consistent with several other studies that investigated discrimination of small acoustic changes in synthetic speech stimuli in ideal listening conditions (e.g., Sussman

1991) and in the fMRI environment (Joanisse and Gati 2003). More importantly, participants were more likely to identify the BW condition as “different” than the WN and REP conditions, consistent with the classical definition of categorical perception, that listeners show poorer within-category discrimination than between-category discrimination. As such, they indicate the BW and WN conditions used in the fMRI experiment do reflect the desired manipulation. Finally, the strong differences in fMRI responses to the BW and WN conditions indicate that any difficulties participants had with actively discriminating these stimuli were due to task demands imposed by the overt discrimination task. Indeed, the disjunction in the imaging and behavioral data suggests that overt and passive speech processing do engage markedly different processing mechanisms. The fMRI activation pattern demonstrates that the BW stimuli were highly discriminable by the human nervous system. However, unique aspects of overt discrimination tasks serve to lower performance on these same stimuli.

### ***Lack of Within-Category Habituation Effects***

The WN condition failed to reveal regions of significant activation compared with the REP condition, notwithstanding a small subcortical region discussed above. We did not find a habituation response to this condition in classical auditory cortex even though the acoustic form of the stimulus was nevertheless being manipulated in the WN condition. The fact that neither A1 nor secondary auditory cortical regions showed such a response suggests that dishabituation does not occur for such complex acoustic stimuli unless the stimulus change also signals a change in phonetic category. An alternative explanation is that there are regions of auditory cortex that are sensitive to nonphonetic acoustic changes in speech but that our ability to adequately image these regions was impeded by acoustic noise produced by the spiral scanning sequence, which is itself spectrally complex and highly frequency modulated. Even though we used a clustered volume acquisition paradigm to minimize the effects of acoustic gradient noise and we were able to demonstrate modulation of A1 in SPCH > SIL, it remains possible that subtle shifts in auditory cortex activation are still being masked by influences of scanner noise.

The present finding does not preclude the possibility that other acoustically simpler sounds could show auditory habituation effects; rather, it suggests that speech stimuli contain complex temporal and spectral information that do not readily lend themselves to such effects. Habituation effects for speech thus seem to occur as a result of the formation of a higher level abstract template for such sounds (Näätänen 2001).

### **Conclusions**

Using a passive-listening paradigm, we have identified regions of left temporal and parietal cortices that appear to comprise neural populations specialized for the categorization of speech sounds. Importantly, these regions are specifically sensitive to an acoustic change that results in a shift of a sound’s phonetic category versus an acoustic change of the same distance on the continuum but that does not coincide with a phonetic shift. Two specific regions show a phonetic-specific effect during passive discrimination: lateral regions of left temporal cortex adjacent to A1 and a left lateralized inferior parietal region subtending AG and SMG. These posterior brain regions appear to be consistent with what has been observed in previous studies using active discrimination (e.g., Jacquemot et al. 2003,



Callan et al. 2004). Moreover, we were able to observe this activation without requiring an overt judgment task. This represents an important step forward from a technical point of view because this passive and unattended paradigm lends itself well to studying speech perception in special populations. For instance, there is significant interest in understanding the brain mechanisms supporting auditory speech perception in second language learners (Best et al. 1988) and in individuals with reading and language impairments (Werker and Tees 1987; Joanisse et al. 2000). The advantage of passive discrimination in these individuals is obvious: overt tasks place demands on attention and short-term memory that can make them unnecessarily challenging and that might lead to groupwise activation differences that are due to these attendant factors.

Results such as the present one are consistent with a division between a “ventral” stream for recovering the semantic form of speech and a “dorsal” stream for recovering phonetic information (Hickok and Poeppel 2000). This is similar to what is proposed in the visual domain (Milner and Goodale 1993), in terms of having both “what” and “where/how” pathways. The present study was primarily concerned with putative dorsal auditory processing stream mechanisms supporting phonetic processing, especially as they relate to the categorical perception of consonants. One interpretation is that the left temporal and inferior parietal regions we identified in this study are critical to speech perception because they represent the locus of abstract phonetic representations that mediate the basic auditory perception mechanisms located in STG and the motor, executive, and short-term memory systems situated in IFG (Hickok and Poeppel 2000). However, the findings run counter to the suggestion that such representations are also located in primary auditory cortex itself. Finally, our study leaves open the question of whether the mechanisms encoded within this putative dorsal stream are speech specific; for instance, these brain regions might also respect preattentive categorization of other types of auditory stimuli, something that was not tested here, but which merits future study given the present results.

## Notes

This research was supported by an Operating Grant and New Investigator Award from the Canadian Institutes for Health Research to MFJ; National Institutes of Health F32-DC006352 to JDZ; and NSF-REC-0337715 and J.F.Merck Scholars Award to BDM. *Conflict of Interest:* None declared.

Funding to pay the Open Access publication charges for this article was provided by NSF grant REC-0337715 to BDM.

Address correspondence to email: marcj@uwo.ca.

## Appendix

Klatt synthesis parameters used to create the consonant-vowel syllables

Time	AV	F <sub>0</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	A1	A2	A3	A4	AB	B1	B2	B3	B4	BGP
0	60	103	220	—	—	3600	0	56	52	48	36	50	70	110	250	110
10	60					3600	60	56	52	48	36	50	70	110	250	110
35	60	125	720	1240	2500	3600	60	56	52	48	36	50	70	110	250	110
55	60	120	720	1240	2500	3600	60	56	52	48	36	50	70	110	250	110
140	60	80	720	1240	2500	3600	60	56	52	48	36	50	70	110	250	110
150	0	80	720	1240	2500	3600	60	56	52	48	36	50	70	110	250	110

A continuum between /ga/ and /da/ was created by manipulating the F<sub>2</sub> and F<sub>3</sub> parameters at time = 0 (F<sub>2</sub>: 1640–1703 Hz in 7 Hz steps; F<sub>3</sub>: 2100–2802 Hz in 78 Hz steps). Three stimuli were selected from the continuum: ga<sub>1</sub> (F<sub>2</sub>: 1640 Hz, F<sub>3</sub>: 2100 Hz start frequencies); ga<sub>4</sub> (F<sub>2</sub>: 1661 Hz, F<sub>3</sub>: 2334 Hz); and da<sub>7</sub> (F<sub>2</sub>: 1682 Hz, F<sub>3</sub>: 2568 Hz start frequencies). Aside from F<sub>2</sub> to F<sub>3</sub>, all parameters were held constant across items. (Note: F<sub>0</sub>–F<sub>3</sub> frequencies at time = 10 ms have been left blank as these values are interpolated from 0 to 35 ms.)

## References

- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. *Nature*. 403:309–312.
- Best CT, McRoberts GW, Sithole NM. 1988. Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *J Exp Psychol Hum Percept Perform*. 14:345–360.
- Binder J, Frost JA, Hammeke TA, Bellowgan P, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex*. 10:512–528.
- Binder J, Liebenthal E, Possing E, Medler D, Ward B. 2004. Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci*. 7:295–301.
- Blumstein S, Myers EB, Rissman J. 2005. The perception of voice onset time: an fMRI investigation of phonetic category structure. *J Cogn Neurosci*. 17(9):1353–1366.
- Burton MW. 2001. The role of inferior frontal cortex in phonological processing. *Cogn Sci*. 25:695–709.
- Burton MW, Small SL, Blumstein SE. 2000. The role of segmentation in phonological processing: an fMRI investigation. *J Cogn Neurosci*. 12:679–690.
- Callan DE, Jones JA, Callan AM, Akahane-Yamada R. 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage*. 22:1182–1194.
- Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, Chollet FL. 1999. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage*. 9: 135–144.
- Dehaene-Lambertz G. 1997. Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*. 8:914–924.
- Dehaene-Lambertz G, Baillet S. 1998. A phonological representation in the infant brain. *Neuroreport*. 9:1885–1888.
- Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S. 2005. Neural correlates of switching from auditory to speech perception. *Neuroimage*. 24:21–33.
- Demonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulous JL, Wise R, Rascol A, Frackowiak R. 1992. The anatomy of phonological and semantic processing in normal subjects. *Brain*. 115:1753–1768.
- 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(5):636–647.
- Fry D, Abramson A, Eimas P, Liberman A. 1962. The identification and discrimination of synthetic vowels. *Lang Speech*. 5:171–179.
- Gaab N, Gabrieli J, Glover G. Forthcoming. Assessing the influence of scanner background noise on auditory processing—II: an fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse temporal sampling design. *Hum Brain Mapp*.
- Godfrey JJ, Syrdal-Lasky AK, Millay K, Knox CM. 1981. Performance of dyslexic children on speech perception tests. *J Exp Child Psychol*. 32:401–424.
- Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol*. 107:293–321.
- Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci*. 4:131–138.
- Jacquemot C, Pallier C, Lebihan D, Dehaene S, Dupoux E. 2003. Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J Neurosci*. 23:9541–9546.
- Jäncke L, Wurstenberg T, Scheich H, Heinze H-J. 2002. Phonetic perception and the temporal cortex. *Neuroimage*. 15:733–746.
- Joanisse MF, Gati JS. 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage*. 19:64–79.
- Joanisse MF, Manis FR, Keating P, Seidenberg MS. 2000. Language deficits in dyslexic children: speech perception, phonology, and morphology. *J Exp Child Psychol*. 77:30–60.

- Jusczyk P. 1997. *The discovery of spoken language*. Cambridge (MA): MIT press.
- Klassen LM, Menon RS. 2004. Robust automated shimming technique using arbitrary mapping acquisition parameters (RASTAMAP). *Magn Reson Med*. 51(5):881-887.
- Klatt DH. 1980. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am*. 67:971-995.
- Kluender KR, Lotto AJ. 1999. Virtues and perils of an empiricist approach to speech perception. *J Acoust Soc Am*. 105:503-511.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 255:606-608.
- Liberman AM, Delattre PC, Cooper FS. 1958. Some cues for the distinction between voiced and voiceless stops in initial position. *Lang Speech*. 1:153-167.
- Liberman AM, Mattingly IG. 1985. The motor theory of speech perception revised. *Cognition*. 21:1-36.
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. *Cereb Cortex*. 15:1621-1631.
- Massaro D. 1997. Categorical partition: a fuzzy-logical model of categorization behavior. In: Harnad S, editor. *Categorical perception: the groundwork of cognition*. Cambridge (UK): Cambridge University Press. p. 254-283.
- McCandliss BD, Fiez JA, Protopapas A, Conway M, McClelland JL. 2002. Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn Affect Behav Neurosci*. 2:89-108.
- Mechefske C, Geris R, Gati J, Rutt B. 2002. Acoustic noise reduction in a 4T MRI scanner. *Magn Reson Mater*. 13:172-176.
- Milner AD, Goodale MA. 1993. Visual pathways to perception and action. *Prog Brain Res*. 95:317-337.
- Näätänen R. 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*. 38:1-21.
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, et al. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*. 385:432-434.
- Oyama S. 1978. The sensitive period and comprehension of speech. *Work Pap Biling*. 6:1-17.
- Penhune VB, Zatorre RJ, MacDonald JD, Evans AC. 1996. Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex*. 6:661-672.
- Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T. 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. *J Cogn Neurosci*. 12:1038-1055.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JDE. 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*. 10(1):15-35.
- Preston AR, Thomason ME, Ochsner KN, Cooper JC, Glover GH. 2004. Comparison of spiral-in/out and spiral-out BOLD fMRI at 1.5T and 3T. *Neuroimage*. 21:291-301.
- Price C, Thierry G, Griffiths T. 2005. Speech-specific auditory processing: where is it? *Trends Cogn Sci*. 9:271-276.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund H, Zilles K. 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage*. 13:669-683.
- Scott SK, Blank CC, Rosen S, Wise RJS. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*. 123:2400-2406.
- Serniclaes W, Ventura P, Morais J, Kolinsky R. 2005. Categorical perception of speech sounds in illiterate adults. *Cognition*. 98:B35-B44.
- Sharma A, Dorman MF. 1999. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am*. 106:1078-1083.
- Sittiprapaporn W, Tervaniemi M, Chindaduangratn C, Kotchabhakdi N. 2005. Pre-attentive discrimination of across-category and within-category change in consonant-vowel syllable. *Neuroreport*. 16(13):1513-1518.
- Sussman J. 1991. Stimulus ratio effects on speech discrimination by children and adults. *J Speech Hear Res*. 34:671-678.
- Talairach J, Tournoux P. 1988. *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.
- Talavage T, Edmister W, Ledden P, Weisskoff R. 1999. Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Hum Brain Mapp*. 7:79-88.
- Thivard L, Belin P, Zilbovicius M, Poline JB, Samson Y. 2000. A cortical region sensitive to auditory spectral motion. *Neuroreport*. 11:2969-2972.
- Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cogn Neurosci*. 13:994-1005.
- Werker JF, Tees R. 1987. Speech perception in severely disabled and average reading children. *Can J Psychol*. 41(1):48-61.
- Werker JF, Tees RC. 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav Dev*. 7:49-63.
- Wessinger C, Vanmeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP. 2001. Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci*. 13:1-7.
- Zatorre RJ, Bouffard M, Belin P. 2004. Sensitivity to auditory object features in human temporal neocortex. *J Neurosci*. 24:3637-3642.
- Zatorre RJ, Meyer E, Gjedde A, Evans AC. 1996. PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb Cortex*. 6:21-30.
- Zevin JD, McCandliss BD. 2005. Dishabituation of the BOLD response to speech sounds. *Behav Brain Funct*. 1:4.