

# The Perception of Voice Onset Time: An fMRI Investigation of Phonetic Category Structure

Sheila E. Blumstein, Emily B. Myers, and Jesse Rissman\*

## Abstract

■ This study explored the neural systems underlying the perception of phonetic category structure by investigating the perception of a voice onset time (VOT) continuum in a phonetic categorization task. Stimuli consisted of five synthetic speech stimuli which ranged in VOT from 0 msec ([da]) to 40 msec ([ta]). Results from 12 subjects showed that the neural system is sensitive to VOT differences of 10 msec and that details of phonetic category structure are retained throughout the phonetic processing stream. Both the left inferior frontal gyrus (IFG) and cingulate showed graded activation as a function of category membership with increasing activation as stimuli approached the phonetic category boundary. These results are consistent with the view that the left IFG is involved in phonetic decision processes,

with the extent of activation influenced by increased resources devoted to resolving phonetic category membership and/or selecting between competing phonetic categories. Activation patterns in the cingulate suggest that it is sensitive to stimulus difficulty and resolving response conflict. In contrast, activation in the posterior left middle temporal gyrus and the left angular gyrus showed modulation of activation only to the “best fit” of the phonetic category, suggesting that these areas are involved in mapping sound structure to its phonetic representation. The superior temporal gyrus (STG) bilaterally showed weaker sensitivity to the differences in phonetic category structure, providing further evidence that the STG is involved in the early analysis of the sensory properties of speech. ■

## INTRODUCTION

Perceptual systems typically require the mapping of information from the relevant sensory system to higher levels of processing. As this information is conveyed upstream, it is generally assumed to go through multiple levels of processing such that the information is transformed into increasingly more abstract representations. An example of such a perceptual system is speech.

Most models of speech perception assume multiple levels of processing between the auditory input and the mapping of sound structure to meaning. In this view, the auditory input from the peripheral auditory system is transformed from simple acoustic features of frequency and amplitude to more generalized auditory patterns or properties. These spectral-temporal patterns are in turn converted to a more abstract category representation corresponding to the phonetic categories of speech. It is generally assumed that lexical entries are represented in terms of these categories. Consistent with the speech literature, we refer to this type of representation as a phonetic category. The term phonological category is often used to refer to a unit of sound structure upon which phonological rules operate. This article does not attempt to distinguish between phonetic and phonolog-

ical categories. For purposes of simplicity, we will use the term phonetic rather than phonological throughout the article whether referring to sound categories (phonetic categories) or sound contrasts (phonetic contrasts).

Much recent work has focused on understanding the neural systems underlying speech perception processes. Given that the primary auditory area, Heschl's gyrus, surfaces in the temporal lobe, it has traditionally been assumed that speech perception recruits posterior and, in particular, temporal lobe structures. Early evidence from the aphasias supports this view. Wernicke's aphasics with damage in left temporal structures displayed severe auditory comprehension deficits (Geschwind, 1965). Nonetheless, there is considerable evidence from behavioral studies of aphasic patients, as well as neuroimaging studies with normal subjects, suggesting that speech perception involves a distributed neural system involving anterior as well as posterior brain structures and the right as well as the left hemisphere. In particular, both Broca's and Wernicke's aphasics have shown deficits in discriminating phonetic contrasts such as “pill” versus “bill” (Blumstein, Baker, & Goodglass, 1977), and also perceiving acoustic-phonetic properties distinguishing phonetic categories such as voice onset time (VOT) or formant transitions (Blumstein, Tartter, Nigro, & Statlender, 1984; Blumstein, Cooper, Zurif, & Caramazza, 1977). In fact, pure word-deaf patients are

---

Brown University

\*Now at University of California, Berkeley

the only aphasic group for which impairments appear to be limited to speech perception (Poeppel, 2001). These patients typically have either bilateral temporal lobe lesions or lesions that involve the left auditory cortex and extend deep, cutting off right hemisphere auditory input. Neuroimaging studies support these findings from the aphasias. They have shown bilateral temporal lobe activation for speech as well as for complex auditory nonspeech stimuli (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Binder, Frost, Hammeke, Bellgowan, Springer, et al. 2000), and activation of anterior as well as posterior structures of the left hemisphere (Zatorre, Belin, & Penhune, 2002; Burton et al., 2001; Zatorre, Meyer, Gjedde, & Evans, 1996).

Recent neuroimaging research has attempted to resolve the relative contributions of posterior and anterior peri-sylvian areas to phonetic processing (see Binder & Price, 2001; Burton, 2001 for review). Differences have emerged as a function of task demands and presumably the different cognitive mechanisms and levels of speech processing required by such tasks (Poeppel, 1996). In particular, the primary auditory cortex and auditory association areas (superior temporal gyrus [STG]) bilaterally have been implicated in early stages of acoustic-phonetic processing (Belin, Zatorre, Hoge, Evans, & Pike, 1999; Zatorre, Meyer, et al., 1996; Price et al., 1992). Tasks that require short-term phonological store have shown activation of left inferior parietal areas (Jonides et al., 1998; Awh et al., 1996; Paulesu, Frith, & Frackowiak, 1993), and tasks that involve overt segmentation of the speech signal activate frontal regions, particularly the inferior frontal gyrus (IFG) and, to a lesser extent, the middle frontal gyrus (MFG) (Burton, Small, & Blumstein, 2000; Zatorre, Meyer, et al., 1996).

Most of the neuroimaging experiments exploring phonetic categorization and discrimination have used stimuli that are good exemplars of their phonetic category. However, speech perception studies have shown that there is a structure to phonetic categories. That is, all members of a phonetic category are not “equal.” Listeners are sensitive to within-category, subphonetic, distinctions, and although they identify stimuli along an acoustic-phonetic continuum, such as VOT, in a categorical-like fashion, some members of the phonetic category are better exemplars than others. Thus, although listeners judge both good exemplar voiceless stop consonants and voiceless stop consonants with reduced VOTs as belonging to the same phonetic category, they show longer reaction times (RTs) in discrimination tasks to such stimulus pairs than to acoustically identical pairs (Pisoni & Tash, 1974). They also show longer response latencies for stimuli with reduced VOTs in a phonetic categorization task (Andruski, Blumstein, & Burton, 1994), and they rate these stimuli as poorer exemplars of the voiceless phonetic category (Miller & Volaitis, 1989). These findings indicate that the perception of phonetic categories is graded. Moreover, short-

ening the VOT of a voiceless stop consonant not only makes it a poorer exemplar of the voiceless phonetic category, but also renders it closer to its contrasting voiced counterpart, resulting in increased competition between the voiced and voiceless phonetic categories.

That gradients of a phonetic category are all classified as members of the same category makes good ecological sense. Speech occurs in noisy environments and speakers do not have sufficient articulatory control to produce an utterance in exactly the same way every time. Thus, the listener needs to be able to recover a stable percept despite the various sources of variability that affect the speech input. The fact that listeners can perceive subtle differences within categories may also be advantageous in that it allows for sensitivity to acoustic variants occurring in different phonetic contexts which may ultimately aid in the on-line processing of the sound structure of language.

These two aspects of the phonetic categorization process—categorization of different acoustic exemplars as members of the same category and perceptual sensitivity to graded category membership—should have neural consequences. It is the goal of the current study to investigate this issue by examining the neural systems underlying the perception of the phonetic categories of speech. To this end, we investigated the perception of VOT, a temporal cue that distinguishes initial voiced (e.g., [d]), from voiceless consonants (e.g., [t]), in English using a phonetic categorization task. A synthetic VOT continuum was used ranging from 0 to 40 msec VOT in 10 msec steps, and subjects were asked to categorize the stimuli as either [d] or [t]. We examine potential differences in neural activation patterns in the perception of a speech continuum containing stimuli which are good exemplars of a phonetic category (end-point stimuli, VOT 0 and 40), stimuli that are perceived as members of the phonetic category but are near to the acoustic-phonetic boundary (within-category stimuli, VOT 10 and 30), and a stimulus that is inconsistently perceived as a member of either category (boundary-value stimulus, VOT 20). In addition, there was a tone control task in which subjects were required to categorize the stimuli as either high tone or low tone.

We hypothesized that the STG should show extensive bilateral activation for the processing of a VOT continuum. What was less clear is whether this region is responsive to differences among the different acoustic exemplars of the continuum. It is possible that the role of the STG is to do an early acoustic analysis of the sensory information inherent in the stimuli, and to pass this information to other brain regions for higher levels of processing (cf. Scott & Wise, 2004; Binder, Frost, Hammeke, Bellgowan, Springer, et al., 2000). In such a case, although there would be extensive activation of the STG bilaterally, there would be no modulation of activation as a function of the stimulus type. Alternatively, it is possible that the STG will show sensitivity to the

acoustic–phonetic differences among the stimuli, and hence, will show modulation of activation. Several MEG and EEG studies using the mismatch negativity (MMN) paradigm have shown modulation of activation for within- and between-phonetic category stimulus pairs for VOT, suggesting that sensitivity to such acoustic differences occurs early in the processing stream (Sharma & Dorman, 1999; cf. Phillips, 2001 for a review).

Because the phonetic categorization task is a linguistic one, it is expected that the activation patterns of areas other than the STG will be left-lateralized. In particular, there should be activation of both left frontal and temporo-parietal areas. Nonetheless, it is hypothesized that the patterns of activation in these areas will reflect different aspects of the phonetic categorization process. In particular, because frontal areas appear to be involved in decision processes and are sensitive to the difficulty of the decision, it is hypothesized that areas such as the left IFG and cingulate should show modulation of activation across the VOT continuum as a function of the goodness of fit (GF) of the stimulus to its phonetic category. Thus, increasingly greater activation should occur for the within-category stimuli and the boundary-value stimulus compared with the endpoint stimulus, reflecting the increasing resources needed to make a phonetic category decision for these tokens.

It is also the case that the within-category stimuli are nearer to the phonetic boundary and closer in acoustic space to the contrasting phonetic category. As such, these stimuli are not only more difficult to map on to a phonetic category, but there is also greater competition between the voiced and voiceless phonetic categories. Such increased competition may also potentially recruit frontal areas such as the IFG (cf. Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997).

As described earlier, the phonetic categorization process requires that listeners categorize different exemplars as members of the same phonetic category. There are several processes involved in determining such phonetic category membership. These include mapping

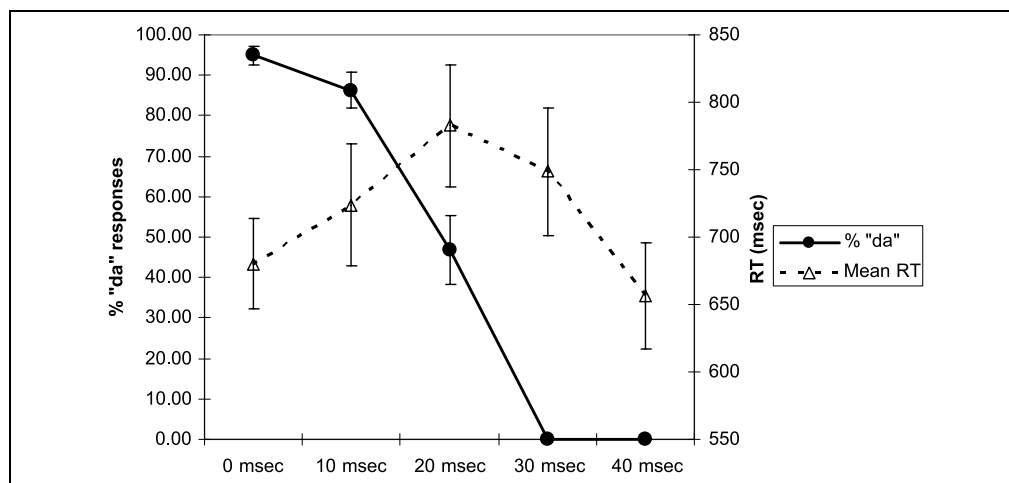
of the sound structure to its phonetic representation, and categorizing the stimulus by its name, processes which are linked to temporo-parietal areas such as the left middle temporal gyrus (MTG) (Hickok & Poeppel, 2004; Scott & Wise, 2004) and the angular gyrus (AG) (Geschwind, 1965). These areas, in particular, may show the greatest activation to those tokens that correspond to the “best fit” to that phonetic category.

## RESULTS

### Behavioral Results

The behavioral results for both RT and categorization data for VOT stimuli are shown in Figure 1, and mean and standard deviation values for the tone stimuli are shown in Table 1. Subjects exhibited the classical categorical perception identification function for VOT. A one-way repeated-measures ANOVA using the percentage of “da” categorizations for each subject for each stimulus on the continuum was significant [ $F(4,44) = 133.7, p < .001$ ]. Post hoc Newman–Keuls tests revealed significant differences in percent “da” responses between the 10- and 20-msec VOT stimuli and between the 20- and 30-msec VOT categories at  $p < .05$ . There was no significant difference between the 0- and 10-msec stimuli, nor was there a significant difference between the 30- and 40-msec stimuli, indicating that within-category differences of VOT have little effect on phonetic categorization decisions. In contrast, the RT results indicate sensitivity to within-category VOT differences, with RTs increasing as the stimuli approach the phonetic boundary. A repeated-measures ANOVA indicated a main effect of VOT on RT [ $F(4,44) = 6.503, p < .001$ ]. Post hoc Newman–Keuls tests indicated no significant differences between any individual VOT stimulus and its VOT neighbors. However, when stimuli were grouped into endpoint (0- and 40-msec VOT), within category (10- and 30-msec VOT), and boundary (20-msec VOT) categories as was done in the functional magnetic

**Figure 1.** Mean RT in msec (triangles) and mean % “da” responses (circles) for the 12 MR participants. Bars represent standard error of the mean.



**Table 1.** Behavioral Data from 12 Participants in the MRI System for the Tone Stimuli

<i>Stimulus</i>	<i>% Correct</i>	<i>RT (msec)</i>
High Tone	96 (1.20)	646 (29)
Low Tone	98 (1.07)	633 (35)

Standard error of the mean for both percentage scores and RT is indicated in parentheses.

resonance imaging (fMRI) analysis, post hoc tests indicated significant differences at  $p < .05$  between all stimulus categories.

## Imaging Results

### Group Activation Maps

To qualitatively compare the extent of activation across conditions, mean activation maps were generated for all stimulus categories. Figure 2 shows a representative stimulus from the endpoint (0 msec), within category (10 msec), and tones (high tone) categories together with the boundary stimulus (20 msec). In general, all syllable conditions yielded extensive bilateral activation centering in peri-sylvian and midline regions. Activation tended to be of greater intensity and extent in left hemisphere regions than in the homologous right hemisphere regions. In addition, a greater amount of activation was observed for the boundary-value stimuli than the within-category stimuli, and for the within-category stimuli than for the endpoint stimuli. The tone stimuli

activated regions similar to those activated in the syllable condition, but both activation intensity and extent of activation was less lateralized.

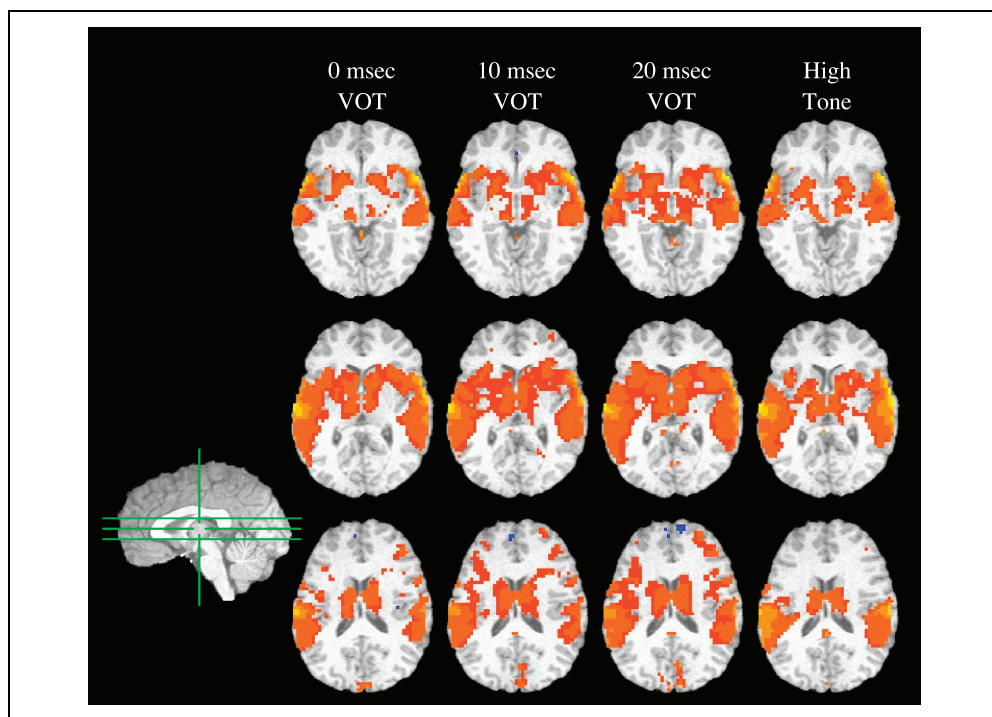
### Planned Comparisons

A summary of all clusters activated in the planned comparisons, which were significant at a corrected  $p < .01$ , is shown in Table 2. The following discussion will focus on the planned comparisons for the endpoint versus tones stimuli, the boundary versus endpoint stimuli, the endpoint versus within-category stimuli, and the boundary versus within-category stimuli.

### Endpoint versus Tones

The comparison of endpoint VOT stimuli with tone stimuli yielded one cluster which was more active for endpoint stimuli than tones. This cluster was located in the anterior left insula, and extended deep into the lentiform nucleus, the caudate, and the putamen. In addition, three clusters were more active for tones than for endpoint stimuli. The largest of these (608 voxels) was centered in the right insula, but extended primarily along the STG and Heschl's gyrus, with some small extension into the right inferior parietal lobule (IPL) and precentral gyrus. A smaller cluster on the left side (255 voxels) mirrored the larger cluster on the right, and was centered in the left IPL and left insula, with activation extending along Heschl's gyrus and the STG. Finally, a cluster in the left and right cingulate was also more active for tones than endpoint stimuli.

**Figure 2.** Group activation maps, thresholded at a voxelwise  $p < 4.0 \times 10^{-5}$ , uncorrected. Axial slices at  $z = -2$ ,  $z = 8$ , and  $z = 18$ , bottom to top.



**Table 2.** Areas of Activation Significant in Planned Comparisons, Thresholded at Voxel Level  $p < .025$ , Cluster Level  $p < .01$  ( $\geq 81$  Contiguous Activated Voxels)

Anatomical Region	Maximum Intensity			No. of Activated Voxels	Local Maximum (t value)
	x	y	z		
Boundary > Endpoint					
Left IFG	-50	26	24	259	3.581
Left Cingulate	-5	17	45	212	4.370
Endpoint > Within-Category					
Left Anterior Cingulate	-2	47	3	156	3.541
Left MTG	-44	-80	30	148	3.242
Left Precuneus	-2	-59	21	138	3.092
Boundary > Within-Category					
Left Cingulate	-5	10	39	163	3.770
Left Subcortical	-17	8	6	136	3.402
Left IPL	-62	-20	30	88	3.160
Endpoint > Tones					
Left Insula	-41	11	-7	274	2.490
Tones > Endpoint					
Right Insula	62	-14	12	608	3.031
Left IPL	-65	-32	15	255	3.534
Left Cingulate	2	-20	45	215	5.543

Coordinates indicate the maximum intensity voxel for that cluster. All coordinates are in Talairach and Tournoux space.

### Boundary versus Endpoint

A comparison of endpoint stimuli and boundary stimuli yielded two significant clusters, both of which were more active for the boundary condition than the endpoint condition (Figure 3). The larger cluster was centered in the left IFG, with activation extending dorsally into the MFG, and ventrally into the insula. The extracted hemodynamic time course for that cluster as shown in Figure 3 revealed modulation of activation, with boundary > within-category > endpoint stimuli. This observation was confirmed in an analysis of the time-series data, using the mean clusterwise percent signal change as the dependent variable, and time point and condition as the independent variables in a repeated-measures, within-subject ANOVA. Data from time points at 2, 4, and 6 sec were used in the analysis, as these times best captured the peak of the hemodynamic response in this area. This analysis showed a significant

main effect of condition [ $F(2,22) = 13.305$ ,  $p < .001$ ]. Post hoc Newman-Keuls tests revealed significant simple effects between all three stimulus types ( $p < .05$ ).

A second cluster centered in the left cingulate and extending into the right cingulate was also activated. As in the inferior frontal cluster, the time-series data (see Figure 3) revealed a graded activation pattern, with boundary stimuli showing greater activation than within-category, which in turn showed greater activation than endpoint stimuli. A repeated-measures ANOVA yielded a significant main effect of condition [ $F(2,22) = 4.323$ ,  $p < .026$ ]. Post hoc tests revealed a significant difference between the endpoint and boundary conditions ( $p < .05$ ), but no other significant differences.

### Endpoint versus Within Category

Three clusters showed significant differences between the endpoint and within-category stimuli, with more activation for the endpoint than within-category stimuli in all clusters (see Figure 4). Activation emerged in a midline area centered in the left anterior cingulate, extending into the right anterior cingulate and the left medial frontal gyrus, in a posterior middle temporal area that extended into the AG, and in the precuneus bilaterally, with activation extending into the left and right posterior cingulate. However, an inspection of the time-series data revealed that in each cluster, all conditions were deactivated, reaching peak deactivation at 6–8 sec poststimuli, and in each case, endpoint stimuli were less deactivated than within-category stimuli (Figure 4).

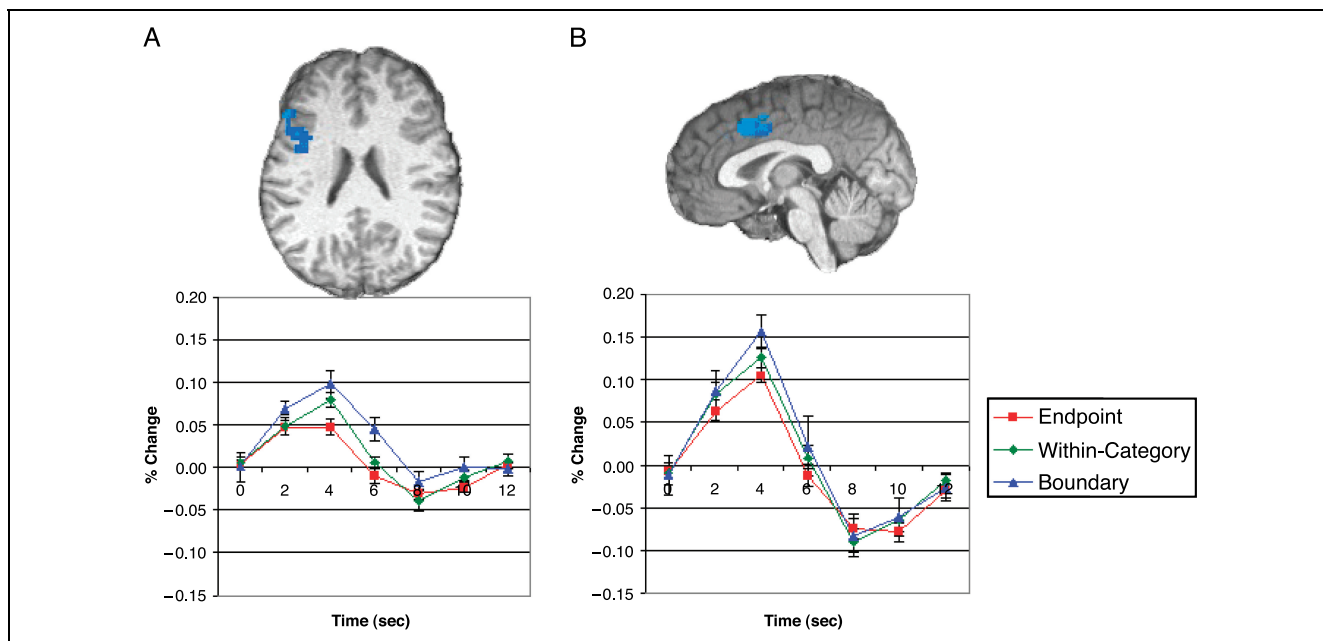
Given that the left MTG and AG have both been implicated in speech processing, time series from this cluster was examined in more detail. Mean clusterwise percent signal change values were submitted to an ANOVA as described above, the only difference being that data from time points at 4, 6, and 8 sec were used. There was a main effect of condition [ $F(2,22) = 7.192$ ,  $p < .004$ ], and post hoc tests showed significant differences between the endpoint category and both the within-category and boundary groups ( $p < .01$ ), but no difference between within-category and boundary stimuli.

### Boundary versus Within Category

Three clusters were activated in this analysis. All showed greater activation for the boundary-value stimuli than the within-category stimuli. These areas included the left and right cingulate, the left IPL extending into the left STG, and a left subcortical area centered in the lentiform nucleus.

### Region-of-Interest Analysis

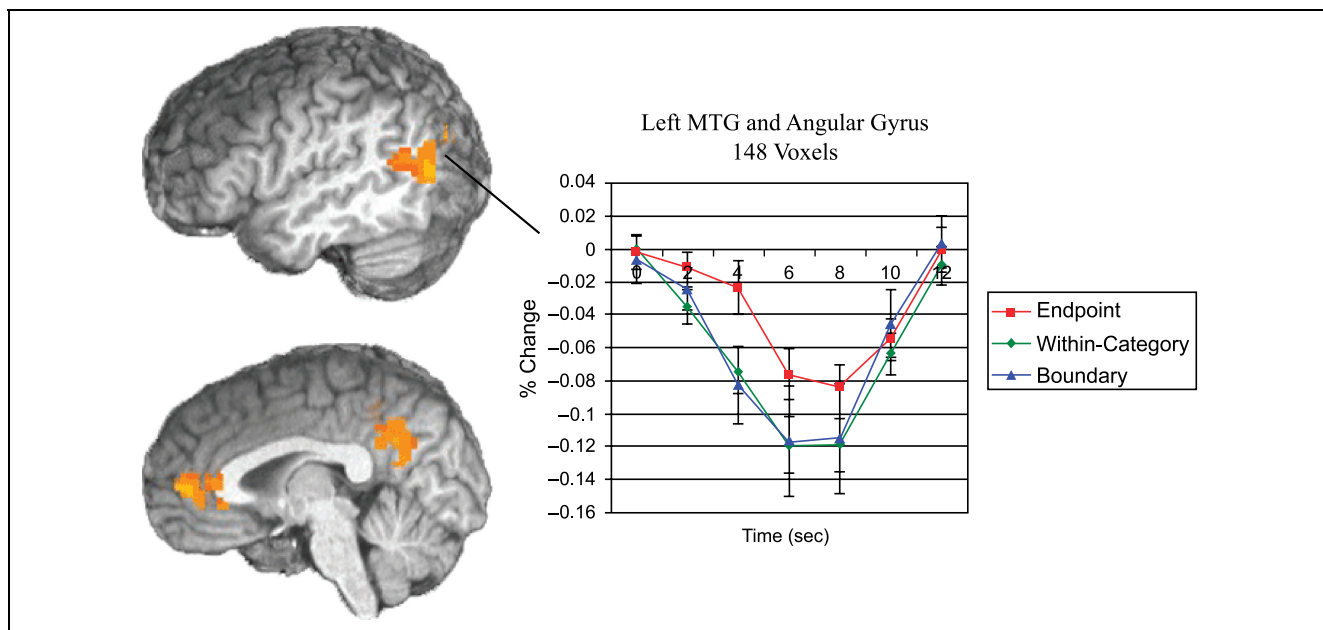
Under the hypothesis that auditory and auditory association areas such as Heschl's gyrus and the STG are



**Figure 3.** Clusters with significant activation differences in the endpoint versus boundary comparison ( $p < .01$ , corrected). Both clusters are more active for boundary stimuli than for endpoint stimuli. (A) The location of the left IFG cluster in an axial slice ( $z = 24$ ) and time series of activation for that cluster. (B) The same for the cingulate cluster in a sagittal view ( $x = -2$ ).

sensitive to phonetic category structure, one would expect to find activation differences in these areas as a function of VOT category. No clusters centered in auditory areas were found in any of the VOT comparisons (although see the comparison between endpoints and tones, above). In order to further explore this

hypothesis, a region-of-interest (ROI) analysis was carried out individually on the left and right STG and on Heschl's gyri bilaterally. Results showed no effect of VOT condition within the left or right Heschl's gyrus [left:  $F(2,22) = 0.891$ ; right:  $F(2,22) = 1.987$ ]. However, there was a significant main effect of VOT condition in



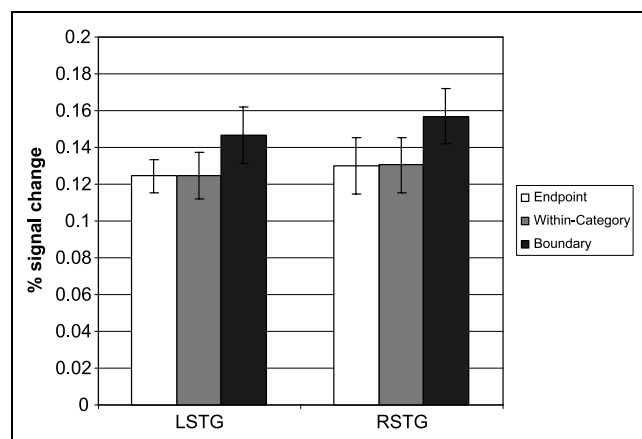
**Figure 4.** Clusters with significant activation differences for the endpoint versus within-category comparison ( $p < .01$ , corrected). All clusters showed deactivation for all stimulus types, with greater deactivation for within-category stimuli than for endpoint stimuli. Sagittal views are shown at  $x = -50$  (top) and  $x = -4$  (bottom). The plot on the right shows the time series of activation extracted from the left MTG cluster.

the right STG [ $F(2,22) = 5.779, p < .01$ ], and in the left STG, the effect approached significance [ $F(2,22) = 3.235, p < .059$ ] (Figure 5). In the right STG, post hoc tests revealed that the main effect was driven by a difference in activation for boundary stimuli versus the endpoint and within VOT categories, ( $F > 4.102; p < .05$  for both effects). The pattern of results was similar in the left STG, although the only comparison that reached significance was the boundary versus within-category comparison ( $F = 3.026, p < .05$ ).

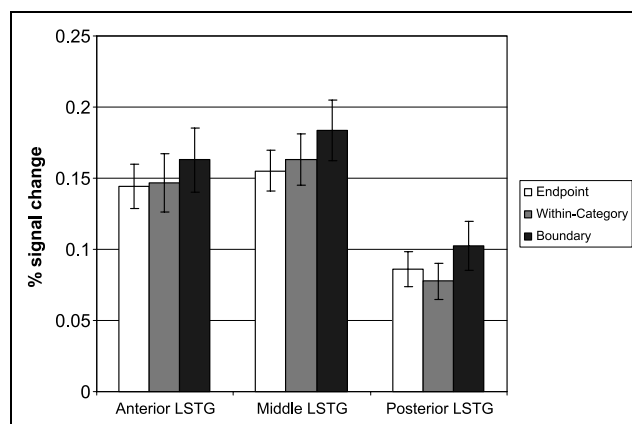
A subdivision of the left STG into thirds along the anterior to posterior extent yielded the same general pattern of activation for all portions of the STG, with little difference in activation between endpoint and within-category stimuli, and more activation for boundary stimuli (Figure 6). Only in the posterior STG region were there any significant differences among the speech stimuli [ $F(2,22) = 3.790, p < .039$ ]. Post hoc Newman-Keuls tests revealed no significant differences between endpoint, within-category, or boundary stimuli within this area.

#### *“Goodness of Fit” versus Reaction Time Regression*

The modulation of activation that emerged in a number of areas showing increased activation as stimuli approached the phonetic boundary raises the question of whether this modulation of activation reflects the phonetic categorization processes invoked by the task, or the increasing difficulty of the task as stimuli become more difficult to categorize. To investigate this issue, a regression analysis was carried out to decouple the effects of “goodness-of-fit” (GF), which assumed a linear relationship between activation and proximity to the phonetic boundary, and difficulty of processing, which is assumed to be related to increases in RT as



**Figure 5.** Mean percent signal change values for endpoint, within-category, and boundary-value stimuli from an ROI analysis of the left STG and right STG. Error bars indicate standard error of the mean.



**Figure 6.** Mean percent signal change values for endpoint, within-category, and boundary stimuli in the anterior, middle, and posterior portions of the left STG. The anterior portion extended forward from a plane at  $y = -8$ , the middle portion extended between planes at  $y = -8$  and  $y = -35$ , and the posterior portion extended posteriorly from a plane at  $y = -36$ . Error bars indicate standard error of the mean. The only area within which significant differences between stimulus types were observed was the posterior portion of the left STG.

stimuli approached the phonetic boundary. To this end, we examined the portion of the variance in any voxel that could be absorbed by a GF regressor which was not already absorbed by linear, quadratic, and logarithmic RT regressors (see Methods). Although there is likely to be some correlation between GF and RT, this analysis was designed to factor out effects of stimulus difficulty on activation levels.

Results revealed a number of areas in which there was a significant correlation between the signal and the GF regressor, over and beyond any correlation with RT (Table 3). Among these activations was a cluster in the left temporo-parietal junction. The maximum intensity voxel of this cluster fell in the left IPL, although the cluster fell primarily in the posterior left STG, with some activation in the left MTG. A similar area in the right

**Table 3.** Areas Active in the GF Analysis, Thresholded at Voxel Level  $p < .025$ , Cluster Level  $p < .01$  ( $\geq 81$  Contiguous Activated Voxels)

Anatomical Region	Maximum Intensity			No. of Activated Voxels	Local Maximum (t value)
	x	y	z		
Bilateral Occipital, Right STG	20	-95	12	1023	3.849
Right Superior Parietal Lobule	17	-71	54	89	2.726
Left IPL	-53	-38	26	82	3.840

Coordinates indicate the maximum intensity voxel for that cluster. All coordinates are in Talairach and Tournoux space.

posterior STG was also activated, which extended into the right MTG as well as the right Heschl's gyrus. This area was connected by a one-voxel bridge to the large bilateral occipital cluster noted in the cluster table, and as such, is technically considered part of that cluster. Finally, a cluster in the right superior parietal lobule and right precuneus was also activated. Neither the cingulate nor the left IFG showed any significant activation attributable to GF that was not already accounted for by the RT regressors.

## DISCUSSION

The results of this study show that there is a distributed neural network involved in the processing of phonetic category structure. These areas include the STG bilaterally, the left IFG, the left MTG extending to the AG, the left IPL, and the right and left cingulate. Different patterns of activation emerged in these areas as a function of the VOT properties of the stimulus and its consequent phonetic category status as an endpoint, within-category, or boundary stimulus, presumably reflecting different functional roles played by these areas in processing the phonetic categories of speech. In particular, frontal areas, including the IFG and cingulate, showed modulation of activation as a function of graded category membership (boundary > within-category > endpoint), and temporo-parietal areas, including the posterior MTG and AG, showed modulation of activation as a function of "best fit" to the phonetic category (endpoint vs. boundary and within-category). Of interest, these results emerged using a VOT continuum that varied in 10-msec steps, indicating the sensitivity of the neural systems to acoustic fine structure along the phonetic processing stream.

### Behavioral Results

The behavioral results showed the typical categorical identification function for the perception of a VOT continuum. Participants showed consistent identification of stimuli in the voiced and voiceless phonetic categories and chance performance to the boundary stimulus. The RT data showed increased RT latencies as a function of the "goodness" of the stimuli as members of a phonetic category. The endpoint stimuli were responded to fastest, followed by the within-category stimuli, and the slowest RT latencies were to the boundary stimulus.

### Patterns of Modulation of Activation

#### *Frontal Structures and Graded Category Membership*

Frontal structures have been implicated in many aspects of language processing including phonetic processing (Burton, 2001). Additionally, these structures have been

implicated in executive decisions with increased activation as those decisions become more difficult (Wagner, Pare-Blagoev, Clark, & Poldrack, 2001). In this study, the extent of phonetic processing necessary to map a token to a phonetic category largely covaried with the difficulty of the executive decision necessary to perform the task. That is, within-category and boundary-value stimuli are not only more difficult to map to a phonetic category, and thus, require increased processing to resolve phonetic category membership, but the increased RT latencies as the stimuli approach the phonetic boundary indicate that the executive decision becomes more difficult as well. As such, either the hypothesis that frontal structures have a role in phonetic processing or the hypothesis that frontal structures have a role in executive decisions would predict that modulation of activation should emerge as a function of the stimulus category, with the greatest activation for boundary-value stimuli, less activation for within-phonetic category stimuli, and the least activation for endpoint stimuli. The results of the analysis comparing the endpoint and boundary stimuli are consistent with this prediction. There was greater activation for boundary stimuli versus endpoint stimuli in the left IFG extending dorsally into the MFG and ventrally into the insula and, as well, for the left cingulate extending into the right cingulate. The extracted hemodynamic time course for these clusters showed graded activation for phonetic category membership, with boundary > within-category > endpoint stimuli.

The question remains whether this pattern of graded activation reflects properties of phonetic category structure or reflects the difficulty of the decision. In a study aimed at distinguishing sensory and decision processes in phonetic perception, Binder, Liebenthal, et al. (2004) found that activation in a portion of the bilateral opercula and insulae correlated with RT but not with accuracy. They propose that accuracy reflects sensory components and RT reflects decision components, hence, attributing an executive role to medial portions of the operculum. What is not clear is whether the left frontal activation in the current study is due to stimulus difficulty, to phonetic processing, or to some combination of the two. The regression analysis provided a possible means of distinguishing these possibilities.

In particular, the results of the regression analysis showed no significant activation in the left IFG over and above that accounted for by RT, suggesting that activation in the left IFG is closely related to stimulus difficulty, and presumably, the consequent response conflict (Gehring & Knight, 2000; MacDonald, Cohen, Stenger, & Carter, 2000). Likewise, all of the cingulate activation could be accounted for solely by RT. In contrast, the regression analysis revealed significant activation in the bilateral posterior STG correlated with the "GF" of the stimulus to its phonetic category, but not with difficulty. This result supports the view that



both the cingulate and the left IFG play a role in executive processes, whereas the sensory component of phonetic processing is subserved by the bilateral posterior STG.

There is another aspect to the processing of phonetic category structure which could account for the modulation of activation observed in the left IFG. Not only are the non-endpoint tokens more difficult to map onto a phonetic category, they also are processed under conditions of increased competition. A within-category stimulus is closer in acoustic space to the contrasting phonetic category, and a boundary-value stimulus is equivocal in its category membership. Thus, the increased frontal activation could also reflect phonetic competition (cf. Thompson-Schill et al., 1997). Such increases in phonetic competition would result in increases in decision difficulty, and thus, RT, and would therefore be accounted for in the RT regression analysis.

#### *Temporo-Parietal Areas and Best Fit to the Phonetic Category*

Several significant clusters, among them one in the posterior tip of the left MTG extending into the AG, emerged showing significantly greater activation for endpoint stimuli versus within-category stimuli. Thus, in contrast to the modulation of activation described above for the other clusters, these clusters appeared to show the *greatest* activation for the stimulus on the continuum that was the “best fit” to its phonetic category. However, time-series analysis of the MTG/AG cluster showed deactivation for all stimuli, with the greatest deactivation for both the within-category and boundary-value stimuli compared with the endpoint stimuli. Thus, this area shows differential sensitivity to the “best fit” stimuli versus all other speech stimuli, with activation levels for all stimuli remaining below baseline.

Deactivation in parietal areas including the AG has been shown in a number of studies. There has been considerable discussion about the role that deactivation plays. Some have proposed that it reflects either reallocation of processing resources from deactivated areas to those areas involved in the task (McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003), with the extent of deactivation a function of the processing resources required to do the task, or a “gating” role, that is, controlling input to areas not involved in the task (Gusnard & Raichle, 2001; Shulman et al., 1997). Binder, Frost, Hammeke, Bellgowan, Rao, et al. (1999) suggest that these areas are involved in ongoing processes during resting states, processes which involve access to conceptual knowledge, hence, these areas become deactivated when exogenously presented input requiring attention and response are presented.

Although the proposals about the role of deactivation vary, they all share the view that deactivation indicates

that these areas are not actively involved in the processing of the stimuli in the task. However, the results of the current study suggest that indeed the MTG/AG are actively involved in phonetic processing and appear to be “tuned” to the best match of the acoustic input to the phonetic category or to the label for that category. If this area were to reflect either reallocation of resources or “gating” input with the extent of activation a function of the degree of resources required, then graded activation should have emerged for the stimulus categories, as was found in the IFG and the cingulate. However, the activation patterns in the MTG/AG showed no differences in the patterns of deactivation for the poorer exemplars (the within-category and boundary-value stimuli) (see Figure 4), and significantly less deactivation for the best exemplars (the endpoint stimuli). Thus, the MTG/AG did not show increased deactivation as a function of stimulus difficulty, but rather showed sensitivity only to the “best” fit of the phonetic category, suggesting that this area is actively involved in processing the stimuli.

Evidence from lesion studies suggests that the inferior parietal area including the AG is actively involved in language processing. Classic pure anomic patients have damage to the left temporo-parietal junction including the AG and show selective deficits in naming (Goodglass, 1993; Geschwind, 1965). These patients understand words and show intact conceptual knowledge. However, they appear to be unable to access the sound shape or “name” of the word. Moreover, recent neuroimaging results showed increased activation in the left lateral temporo-parietal region under conditions of increased lexical competition, consistent with the view that this region is involved with mapping sound structure to lexical form and accessing the sound shape of the word from the lexicon (Prabhakaran, Blumstein, Myers, & Hutchison, submitted).

Taken together, the results suggest that the MTG/AG is involved in the mapping of the acoustic input to the representation of the phonetic category. As a consequence, there is the least deactivation for the best fit to the category. The results are also compatible with the hypothesis that the temporo-parietal junction is involved in access to lexical representations. The phonetic categorization task is essentially a naming task. Thus, it is also possible that this area is recruited for implicit naming of the stimulus as a “d” or a “t,” with the least deactivation for the stimulus that is the best exemplar of the category.

#### *The Role of the Superior Temporal Gyrus*

Although the mean activation maps showed extensive activation in the STG bilaterally, analyses comparing the endpoint, within-category, and boundary-value stimuli failed to reveal any areas of significant difference centered in either the left or right STG. These results

suggest that the STG is involved in the early analysis of the acoustic properties of speech, but sends this information to other areas for higher levels of processing relating to phonetic category structure. Nonetheless, an ROI analysis investigating the percent signal change across the VOT categories in Heschl's gyrus and the STG bilaterally suggests that there is some modulation of activation early in the processing stream. In particular, although Heschl's gyrus showed no modulation of activation as a function of VOT, greater activation was found for the boundary-value stimuli compared to the within-category and endpoint stimuli for both the right and left STG, indicating sensitivity to VOT differences. Consistent with these results, the regression analysis showed that the bilateral posterior STG were sensitive to the GF of a stimulus to its phonetic category, but were insensitive to the difficulty of the stimulus, as measured by RT. This evidence further supports a role for the bilateral STG, particularly the posterior portions, in the early phonetic processing of speech stimuli.

That modulation of activation emerged between the boundary stimuli versus the within-category and endpoint stimuli in this analysis is consistent with electrophysiological findings using the MMN paradigm. As described earlier, Sharma and Dorman (1999) showed a larger MMN for between-phonetic category stimulus pairs than for within-phonetic category stimulus pairs in a VOT continuum. Thus, it appears that auditory cortex is involved in the early acoustic analysis of the sensory information inherent in the stimuli. Nonetheless, although these results support the hypothesis that the STG bilaterally is sensitive to phonetic category structure, such a conclusion needs to be made with caution. Although MMN sensitivity has been observed for VOT and also for a number of vowel contrasts, it has not emerged consistently for another phonetic contrast, place of articulation in stop consonants (cf. Phillips, 2001 for review). Thus, it is not clear whether the modulation of the MMN reflects phonetic category structure per se or natural acoustic discontinuities in these continua.

Of interest is the failure in this study to find differences among syllable stimuli in the left anterior temporal lobe. The left temporal lobe lateral and anterior to Heschl's gyrus has been implicated in analysis of spectro-temporal cues in speech, with eventual mapping of these cues to the lexicon (Scott & Johnsrude, 2003), and activation in a similar area bilaterally has been shown to correlate with sound identification accuracy (Binder, Liebenthal, et al., 2004). An informal post hoc analysis of activation in the left STG showed that although anterior, middle, and posterior thirds of the left STG showed similar patterns of activation, with boundary stimuli more active than either endpoint or within-category stimuli, this pattern only reached significance in posterior portions of the left STG. One possible explanation for the lack of a significant difference in

anterior areas is that the greatest signal dropout occurred in the lateral anterior temporal lobe, and thus, differences did not emerge in this area. A second possibility is that the arbitrarily defined thirds of the STG include more than one functionally distinct area. Indeed, some have suggested that a gradient of functionality exists along the medial to lateral surface of the superior temporal plane, as well as in the anterior to posterior direction (Scott & Johnsrude, 2003). A third possibility is that although the anterior temporal lobes may be involved in spectro-temporal analysis (Scott & Johnsrude, 2003; Wise et al., 2001), the differences in the spectro-temporal properties of speech stimuli in this study are so minimal so as to engage the analysis process equally. All VOT stimuli in this study shared the same spectral properties, and the variation in the temporal property of VOT, although phonetically meaningful, is on the order of tens of msec. The fact that the posterior STG shows sensitivity to the phonetic parameters of the stimuli with significantly different activation for boundary stimuli versus other speech stimuli, taken together with the observation that an adjacent area in the left posterior MTG shows preferential sensitivity to the stimuli which are the "best fit" to a phonetic category, suggests that a posterior stream of phonetic processing may be engaged in mapping acoustic input onto the abstract representation which constitutes the phonetic category.

### **The Role of Phonetic Category Structure in Language Processing**

The results of this study suggest that the neural system is not only sensitive to phonetic category structure but the details of phonetic category membership are also retained throughout the phonetic processing stream. One issue is whether the sensitivity to this information is a function of the demands of the phonetic categorization task itself or whether sensitivity to phonetic category structure has consequences at higher levels of processing. Evidence from behavioral studies suggests that phonetic category structure influences not only phonetic categorization processes but also influences lexical access and ultimately access to the lexical-semantic network. In particular, it has been shown that not only the magnitude of form priming in a lexical decision task is reduced if the prime stimulus is a poorer exemplar of the phonetic category, for example, *c\*at-cat* shows less priming than *cat-cat* (Ju, 2004), but so is the magnitude of semantic priming, for example, *c\*at-dog* shows less semantic priming than *cat-dog* (Andruski et al., 1994). These results suggest that phonetic category structure plays a role in higher levels of language processing, and would presumably influence the neural activation patterns during such processing. Further research is needed to investigate this issue. However, taken together with the results of the current study, it suggests that

sensory information is retained and used beyond the sensory system that is responsible for its analysis.

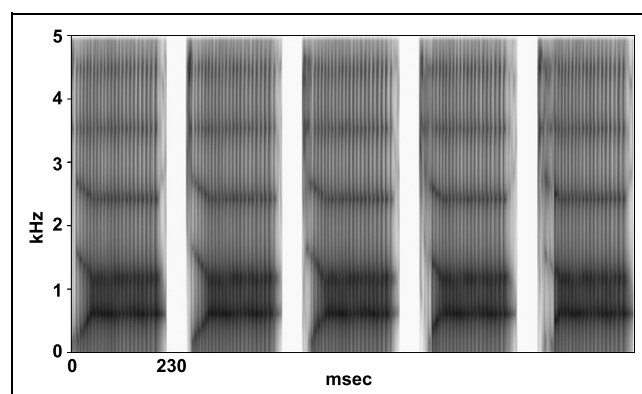
## METHODS

### Participants

Thirteen adult native speakers of English (11 women) between the ages of 20 and 59 (mean age =  $25.5 \pm 10.5$  years) participated in the study, each screened for a history of neurological disease or injury. All reported normal hearing and were right-handed as determined by Oldfield's Handedness Inventory (Oldfield, 1971). Participants gave written informed consent according to guidelines established and approved by the Human Subjects Committees of Brown University and Memorial Hospital of Rhode Island, and were screened for magnetic resonance (MR) safety before entering the scanner. Participants received modest monetary compensation for their time. One subject was excluded from both behavioral and functional analyses due to improper image acquisition parameters.

### Materials

Stimuli consisted of five synthetic speech stimuli taken from a larger continuum ranging from [da] to [ta] synthesized at Haskins Laboratory using a parallel synthesizer. The five stimuli ranged in VOT from 0 msec ([da]) to 40 msec ([ta]) in 10-msec steps (see Figure 7). VOT was manipulated by replacing the periodic source with an aperiodic source in 10-msec increments starting from the stimulus onset. Two sinewave tones with frequencies of 910 Hz ("low tone") and 1320 Hz ("high tone") were also generated for use in the tone task. All stimuli were 230 msec in length.



**Figure 7.** Wide-band spectrograms of speech stimuli used in the study, ranging in VOT from 0 to 40 msec (left to right). Each stimulus consisted of a five-formant pattern. The onset frequencies were 200 Hz (F1), 1350 Hz (F2), and 3100 Hz (F3). Formant transitions into the vowel were 40 msec to a steady-state vowel with formant frequencies of 720, 1250, and 2500 Hz, respectively. F4 and F5 remained steady throughout at 3600 and 4500 Hz.

### Behavioral Procedure

The experiment consisted of two runs in each of two tasks, a phonetic categorization task (PC), and a tone categorization task (TC). Participants received the four runs in a fixed order (PC, TC, PC, TC). Each PC run consisted of 20 repetitions of each of the five synthesized syllables, and each TC run consisted of 20 repetitions of each of the two tone stimuli (see *MR Imaging* for details of stimulus timing). Stimuli within each run were presented in a fixed, pseudorandomized order. During scanning, participants listened to stimuli through MR-compatible headphones (Resonance Technology, Northridge, CA) set at the maximum comfortable volume, and indicated responses via two buttons of an MR-compatible button box placed by the right hand (Resonance Technology). Stimuli were presented using an IBM ThinkPad running the AVRunner program, designed for accurate timing of auditory stimuli presentation and collection of RT data (Mertus, 1989).

For the PC task, participants were instructed to listen to each syllable and to decide whether the syllable was [da] or [ta], by pressing a corresponding button as quickly and accurately as possible. For the TC task, subjects were told to listen to each tone stimulus, and to decide whether it was the "high" tone (1320 Hz) or the "low" tone (910 Hz), by pressing the corresponding button. The mapping of the button (left or right) to the response type ("ta" or "da"; "high" or "low") was counterbalanced across subjects. Participants were given five trials of each task for practice during acquisition of the anatomical dataset. RT and categorization data were collected for all subjects; RTs were measured from the onset of the stimulus.

### MR Imaging

Whole-brain MRI was performed with a 1.5-T Symphony Magnetom MR system (Siemens Medical Systems, Erlangen, Germany) equipped with echo-planar imaging (EPI) capabilities. Each participant's head was aligned to the magnetic field center. Participants were instructed to refrain from moving the head during MR imaging, and were reminded to keep their eyes closed.

For anatomical co-registration, we acquired a 3-D T1-weighted magnetization prepared rapid acquisition gradient-echo (MPRAGE) sequence (TR = 1900 msec, TE = 4.15 msec, TI = 1100 msec, 1 mm<sup>3</sup> isotropic voxel size, 256 × 256 matrix) reconstructed into 160 slices. A multislice, ascending, interleaved (EPI) sequence with 15 axial slices constituted the functional scans (5 mm thickness, 3 mm<sup>2</sup> axial in-plane resolution, 64 × 64 matrix, 192 mm<sup>2</sup> FOV, FA = 90°, TE = 38 msec, TR = 2000 msec). Before EPI images were acquired, the center of the imaged slab was aligned to each participant's corpus callosum using a sagittal localizer image, which allowed for the collection of functional

data from bilateral peri-sylvian cortex. A functional acquisition sequence was used, in which auditory stimuli were presented during silent gaps between volume acquisitions (Belin et al., 1999; Hall et al., 1999). Each EPI volume acquisition was obtained in 1200 msec (80 msec per slice) followed by 800 msec of silence, during which the auditory stimulus was presented (Figure 8), yielding an effective volume repetition time of 2000 msec.

We used an event-related design presenting stimuli at different equally probable trial onset asynchronies (TOA = 2, 4, 6, 8, 10 sec). Four volumes were acquired prior to the onset of the first stimulus to avoid contamination of functional data by T1 saturation. These four volumes were eliminated from further analysis. Each of the two PC runs consisted of 304 echo-planar volumes, and each TC run consisted of 124 volumes, for a total of 856 EPI volumes.

## Data Analysis

### Behavior

RT and categorization data were collected for all participants. RT means were computed for each VOT stimulus for each subject after eliminating responses having RT > 2 sec or > 2 SD from the mean for a particular stimulus for that subject. The percentage of “da” responses was also calculated for each subject for each VOT value. Repeated-measures ANOVAs were performed separately on the RT means and on the percent “da” values for each subject. In addition, RT means and percentage correct scores were calculated for the tone stimuli.

## Magnetic Resonance

### Image Preprocessing

Analysis of imaging data was carried out using AFNI (Cox & Hyde, 1997). Functional MR images were corrected for head motion by aligning all volumes to the fourth collected volume using a six-parameter rigid-body transform (Cox & Jesmanowicz, 1999) and then resampled to 3-mm isotropic voxels and warped to Talairach & Tournoux (1988) space using AFNI tools. The images were

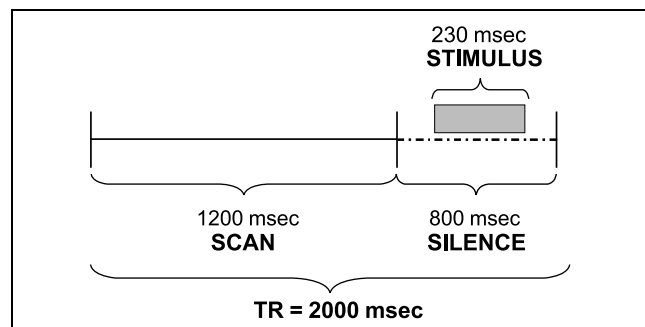
then spatially smoothed with a 6-mm full width at half maximum Gaussian kernel. These preprocessed datasets became the input to a deconvolution analysis.

### Statistical Analysis

Each subject’s EPI data were submitted to deconvolution analysis to estimate the individual hemodynamic response during each stimulus condition. A reference time series was created for each of the seven trial types (five VOT conditions, two tone conditions) by convolving the stimulus presentation times with a stereotypic gamma-variate hemodynamic response curve provided by AFNI (Cohen, 1997). The resulting seven reference waveforms were used as covariates in the deconvolution analysis. The six output parameters of the motion correction analysis ( $x$ ,  $y$ , and  $z$  translations, roll, pitch, and yaw) were also included as covariates in order to remove residual motion artifacts from the EPI time series. These 13 reference functions were then submitted to AFNI’s 3dDeconvolve program, which output raw fit coefficients for each subject for each condition on a voxel-by-voxel basis. The fit coefficients were then converted to percent change by dividing the coefficient from each voxel by the baseline for that voxel, which was calculated by computing the mean for that voxel across the entire experiment.

The percent change values for each condition and each participant were entered into a two-way, mixed-factor ANOVA using stimulus condition as the fixed factor and participant as the random factor. A group statistical map was created for each stimulus condition. To reduce the number of statistical comparisons, the five VOT conditions were grouped into three stimulus categories: the 0-msec and 40-msec tokens constituted the “endpoint” category, the 10- and 30-msec tokens constituted the “within-category” stimuli, and the 20-msec VOT stimulus was referred to as the “boundary” stimulus. Four planned comparisons were carried out: endpoint versus tone stimuli, endpoint versus within-category, endpoint versus boundary, and within-category versus boundary. Using standard AFNI methods, we used a voxel-level threshold of  $p < .025$  and a cluster threshold of 81 contiguous 3 mm<sup>3</sup> voxels in Talairach space which, on the basis of Monte Carlo simulations, yielded a corrected threshold of  $p < .01$ .

A second analysis was performed to investigate the possibility that some portion of the activation differences revealed in the first ANOVA might be attributable to differences in the difficulty of the task across stimuli (as estimated by RT) rather than differences in the processing of the stimuli per se. To test this hypothesis, we used only the VOT stimuli in a linear regression analysis having one “GF” regressor which assumed a linear increase in activation as stimuli approached the phonetic boundary (endpoint stimuli: 0 msec and 40 msec VOT = 1; within-category stimuli: 10 msec and 30 msec



**Figure 8.** Scanning schematic for stimulus presentation.

VOT = 2; boundary stimuli: 20 msec VOT = 3), and one regressor for each subject. Mean by-condition RT for each subject, the squares of these means, and the log of those means served as the final three regressors, which were respectively meant to factor out any linear, quadratic, or logarithmic relationship between activation and RT. In this analysis, each VOT stimulus type was entered into the regression separately to assess the contribution of RT to individual variation among the five VOT stimulus types. The employed model tested what portion of the variance could be absorbed by the “GF” regressor, which was not already accounted for by RT and subject variables. The resultant statistical maps were thresholded as above. Tone stimuli were not included in the analysis, as there was no significant difference in RT (one-tailed  $t = 1.058$ ,  $p < .158$ ) or in accuracy (one-tailed  $t = 0.919$ ,  $p < .189$ ) between high and low tones, as indicated by paired  $t$  tests.

In order to examine the time course of the hemodynamic response within those regions identified in the previous analyses, a second deconvolution analysis was performed in which no a priori assumptions were made about the shape or temporal properties of the hemodynamic function. For this analysis, a binary stimulus function was created for each condition, with ones at time points when stimuli were presented and zeroes elsewhere. Deconvolution was carried out for each subject as described above, with the binary stimulus functions substituting for the convolved stimulus functions. Voxelwise scaling coefficients were output at 2-sec intervals ranging from 0 to 20 sec poststimulus, and coefficients were converted to percent signal change values as described in the gamma-based deconvolution analysis. Time-series graphs were generated for each subject by taking the means of all activated voxels within a cluster at a range of poststimulus time lags, and these were then averaged to create a group time-series graph for the cluster.

ROI analyses were carried out in Heschl’s gyrus and the STG bilaterally because of a priori hypotheses predicting activation differences in the primary auditory and auditory association areas as a function of stimulus type. These areas were defined on the basis of rough anatomical localization masks provided by AFNI (Lancaster et al., 2000). Additionally, to investigate the hypothesis that functional divisions exist along the anterior to posterior extent of the left STG, the left STG mask was divided into approximate thirds, with the anterior third bounded posteriorly by a plane at  $y = -8$ , the middle third extending between planes at  $y = -8$  and  $y = -35$ , and the posterior third bounded anteriorly by a plane at  $y = -36$ . The mean percent signal change value from the gamma-based deconvolution analysis was taken of all voxels for each subject and condition within an area defined by the anatomical masks provided by AFNI. Only voxels which were imaged in all subjects and also

fell within the anatomical area of interest contributed to the mean. By-condition and by-subject means were entered into separate within-subject, repeated-measures ANOVAs for each anatomical area, with stimulus condition (endpoint, within-category, or boundary) as the variable of interest.

## Acknowledgments

This research was supported in part by the Dana Foundation, NIH Grant DC006220 to Brown University, NIH Grant F31006520 to Brown University, and the Ittleson Foundation. We thank James Eliassen, Jerome Sanes, and Emmette Hutchison for their helpful comments and advice.

Reprint requests should be sent to Sheila E. Blumstein, Department of Cognitive and Linguistic Sciences, Brown University, 190 Thayer Street, Box 1978, Providence, RI 02912, or via e-mail: Sheila\_Blumstein@brown.edu.

The data reported in this experiment have been deposited with The fMRI Data Center archive ([www.fmridc.org](http://www.fmridc.org)). The accession number is 2-2005-118RP.

## REFERENCES

- Andruski, J., Blumstein, S. E., & Burton, M. (1994). The effects of subphonetic differences on lexical access. *Cognition*, 52, 163–187.
- Awh, E., Jonides, J., Smith, E., Schumacher, W., Koeppel, R., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from PET. *Psychological Science*, 7, 25–31.
- Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, 10, 417–429.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience*, 11, 80–93.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512–528.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7, 295–301.
- Binder, J. R., & Price, C. (2001). Functional neuroimaging of language. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 187–251). Cambridge: MIT Press.
- Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, 15, 19–30.
- Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, 15, 371–383.
- Blumstein, S. E., Tartter, V. C., Nigro, G., & Statlender, S. (1984). Acoustic cues for the perception of place of articulation in aphasia. *Brain and Language*, 22, 128–149.
- Burton, M. W. (2001). The role of inferior frontal cortex in phonological processing. *Cognitive Science*, 25, 695–709.
- Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*, 12, 679–690.

- Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *Neuroimage*, 6, 93–103.
- Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10, 171–178.
- Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine*, 42, 1014–1018.
- Gehring, W. J., & Knight, R. T. (2000). Prefrontal–cingulate interactions in action monitoring. *Nature Neuroscience*, 3, 516–520.
- Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain*, 88, 237–294, 585–644.
- Goodglass, H. (1993). *Understanding aphasia*. New York: Academic Press.
- Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews: Neuroscience*, 2, 685–694.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., & Bowtell, R. W. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping*, 7, 213–223.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99.
- Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, O. A., Marshuetz, C., & Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience*, 18, 5026–5034.
- Ju, M. K. (2004). Representational specificity of within-category phonetic variation in the mental lexicon. Unpublished doctoral dissertation. State University of New York at Buffalo.
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A., & Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, 10, 120–131.
- MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating of the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288, 1835–1838.
- McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of Cognitive Neuroscience*, 15, 394–408.
- Mertus, J. (1989). *BLISS user's manual*. Providence: Brown University. Available at: [www.cog.brown.edu/localSites/mertus/BlissHome.htm](http://www.cog.brown.edu/localSites/mertus/BlissHome.htm).
- Miller, J., & Volaitis, L. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46, 505–512.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Paulesu, E., Frith, C. D., & Frackowiak, R. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342–345.
- Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science*, 25, 711–731.
- Pisoni, D., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15, 289–290.
- Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain and Language*, 55, 317–351.
- Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25, 679–693.
- Prabhakaran, R., Blumstein, S. E., Myers, E. B., & Hutchison, E. (submitted). *An event-related fMRI investigation of phonological-lexical competition*.
- Price, C., Wise, R., Ramsey, S., Friston, K., Howard, D., & Patterson, K. (1992). Regional response differences within the human auditory cortex when listening to words. *Neuroscience Letters*, 146, 179–182.
- Shulman, G. L., Corbetta, M., Buckner, R., Fiez, J. A., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: I. Increases in subcortical structures and cerebellum but not in nonvisual cortex. *Journal of Cognitive Neuroscience*, 9, 624–647.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100–107.
- Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, 92, 13–45.
- Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America*, 106, 1078–1083.
- Talairach, J., & Tournoux, P. (1988) *A co-planar stereotaxic atlas of a human brain*. Stuttgart: Thieme.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, U.S.A.*, 94, 14792–14797.
- Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, 31, 329–338.
- Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within ‘Wernicke’s area’. *Brain*, 124, 83–95.
- Zatorre, R., Meyer, E., Gjedde, A., & Evans, A. (1996). PET studies of phonetic processing of speech: Review, replication, and reanalysis. *Cerebral Cortex*, 6, 21–30.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Neurosciences*, 6, 37–47.

**This article has been cited by:**

1. David W. Gow, Bruna B. Olson. 2016. Using effective connectivity analyses to understand processing architecture: response to commentaries by Samuel, Spivey and McQueen, Eisner and Norris. *Language, Cognition and Neuroscience* **31**:7, 869-875. [[CrossRef](#)]
2. Yakov Kronrod, Emily Coppess, Naomi H. Feldman. 2016. A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin & Review* . [[CrossRef](#)]
3. Samuel Evans, Carolyn McGettigan, Zarinah K. Agnew, Stuart Rosen, Sophie K. Scott. 2016. Getting the Cocktail Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience* **28**:3, 483-500. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]
4. Matthew H. Davis The Neurobiology of Lexical Access 541-555. [[CrossRef](#)]
5. Sheila E. Blumstein Psycholinguistic Approaches to the Study of Syndromes and Symptoms of Aphasia 923-933. [[CrossRef](#)]
6. Katya Krieger-Redwood, Catarina Teige, James Davey, Mark Hymers, Elizabeth Jefferies. 2015. Conceptual control across modalities: graded specialisation for pictures and words in inferior frontal and posterior temporal cortex. *Neuropsychologia* **76**, 92-107. [[CrossRef](#)]
7. D. Poeppel Speech Perception 429-434. [[CrossRef](#)]
8. Emily B. Myers, Laura M. Mesite. 2014. Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language* **76**, 80-93. [[CrossRef](#)]
9. Karsten Specht, Florian Baumgartner, J  rg Stadler, Kenneth Hugdahl, Stefan Pollmann. 2014. Functional asymmetry and effective connectivity of the auditory system during speech perception is modulated by the place of articulation of the consonant-A 7T fMRI study. *Frontiers in Psychology* **5** . [[CrossRef](#)]
10. Mathias Scharinger, Molly J. Henry, Julia Erb, Lars Meyer, Jonas Obleser. 2014. Thalamic and parietal brain morphology predicts auditory category learning. *Neuropsychologia* **53**, 75-83. [[CrossRef](#)]
11. Laurel Lawyer, David Corina. 2014. An investigation of place and voice features using fMRI-adaptation. *Journal of Neurolinguistics* **27**:1, 18-30. [[CrossRef](#)]
12. Jessica Junger, Katharina Pauly, Sabine Br  hr, Peter Birkholz, Christiane Neusch  fer-Rube, Christian Kohler, Frank Schneider, Birgit Derntl, Ute Habel. 2013. Sex matters: Neural correlates of voice gender perception. *NeuroImage* **79**, 275-287. [[CrossRef](#)]
13. Hannah E. Thompson, Elizabeth Jefferies. 2013. Semantic control and modality: An input processing deficit in aphasia leading to deregulated semantic cognition in a single modality. *Neuropsychologia* **51**:10, 1998-2015. [[CrossRef](#)]
14. Domenic Minicucci, Sara Guediche, Sheila E. Blumstein. 2013. An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical-semantic network. *Neuropsychologia* **51**:10, 1980-1988. [[CrossRef](#)]
15. Sara Guediche, Caden Salvata, Sheila E. Blumstein. 2013. Temporal Cortex Reflects Effects of Sentence Context on Phonetic Processing. *Journal of Cognitive Neuroscience* **25**:5, 706-718. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]
16. Mirko Grimaldi. 2012. Toward a neural theory of language: Old issues and new perspectives. *Journal of Neurolinguistics* **25**:5, 304-327. [[CrossRef](#)]
17. Emily B. Myers, Kristen Swan. 2012. Effects of Category Learning on Neural Sensitivity to Non-native Phonetic Categories. *Journal of Cognitive Neuroscience* **24**:8, 1695-1708. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]
18. David W. Gow. 2012. The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language* **121**:3, 273-288. [[CrossRef](#)]
19. L.A. Petitto, M.S. Berens, I. Kovelman, M.H. Dubins, K. Jasinska, M. Shalinsky. 2012. The "Perceptual Wedge Hypothesis" as the basis for bilingual babies' phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language* **121**:2, 130-143. [[CrossRef](#)]
20. Jianfeng Yang, Xiaojuan Wang, Hua Shu, Jason D. Zevin. 2012. Task by stimulus interactions in brain responses during Chinese character processing. *NeuroImage* **60**:2, 979-990. [[CrossRef](#)]
21. References 261-314. [[CrossRef](#)]
22. Evelina Fedorenko, Alfonso Nieto-Casta  n, Nancy Kanwisher. 2012. Syntactic processing in the human brain: What we know, what we don't know, and a suggestion for how to proceed. *Brain and Language* **120**:2, 187-207. [[CrossRef](#)]
23. Caden Salvata, Sheila E. Blumstein, Emily B. Myers. 2012. Speaker invariance for phonetic information: An fMRI investigation. *Language and Cognitive Processes* **27**:2, 210-230. [[CrossRef](#)]
24. Patti Adank, Matthijs L. Noordzij, Peter Hagoort. 2012. The role of planum temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping* **33**:2, 360-372. [[CrossRef](#)]



25. Ji Young Kim. 2011. L1-L2 Phonetic interference in the production of Spanish heritage speakers in the US. *The Korean Journal of Hispanic Studies* 4:null, 1-28. [[CrossRef](#)]
26. Stefan Elmer, Martin Meyer, Lucas Marrama, Lutz Jäncke. 2011. Intensive language training and attention modulate the involvement of fronto-parietal regions during a non-verbal auditory discrimination task. *European Journal of Neuroscience* 34:1, 165-175. [[CrossRef](#)]
27. Daniel Mirman, Eiling Yee, Sheila E. Blumstein, James S. Magnuson. 2011. Theories of spoken word recognition deficits in Aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language* 117:2, 53-68. [[CrossRef](#)]
28. Michael Wolmetz, David Poeppel, Brenda Rapp. 2011. What Does the Right Hemisphere Know about Phoneme Categories?. *Journal of Cognitive Neuroscience* 23:3, 552-569. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]
29. Evelina Fedorenko, Nancy Kanwisher. 2011. Functionally Localizing Language-Sensitive Regions in Individual Subjects With fMRI: A Reply to Grodzinsky's Critique of Fedorenko and Kanwisher (2009). *Language and Linguistics Compass* 5:2, 78-94. [[CrossRef](#)]
30. E. Liebenthal, R. Desai, M. M. Ellingson, B. Ramachandran, A. Desai, J. R. Binder. 2010. Specialization along the Left Superior Temporal Sulcus for Auditory Categorization. *Cerebral Cortex* 20:12, 2958-2970. [[CrossRef](#)]
31. Edward F Chang, Jochem W Rieger, Keith Johnson, Mitchel S Berger, Nicholas M Barbaro, Robert T Knight. 2010. Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience* 13:11, 1428-1432. [[CrossRef](#)]
32. Peter E Turkeltaub, H. Branch Coslett. 2010. Localization of sublexical speech perception components. *Brain and Language* 114:1, 1-15. [[CrossRef](#)]
33. David January, John C. Trueswell, Sharon L. Thompson-Schill. 2009. Co-localization of Stroop and Syntactic Ambiguity Resolution in Broca's Area: Implications for the Neural Basis of Sentence Processing. *Journal of Cognitive Neuroscience* 21:12, 2434-2444. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]
34. Christopher W. Bishop, Lee M. Miller. 2009. A Multisensory Cortical Network for Understanding Speech in Noise. *Journal of Cognitive Neuroscience* 21:9, 1790-1804. [[Abstract](#)] [[Full Text](#)] [[PDF](#)] [[PDF Plus](#)]
35. Jared M. Novick, Irene P. Kan, John C. Trueswell, Sharon L. Thompson-Schill. 2009. A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology* 26:6, 527-567. [[CrossRef](#)]
36. Jie XI, Wei JIANG, Lin-Jun ZHANG, Hua SHU. 2009. Categorical Perception of VOT and Lexical Tones in Chinese and the Developmental Course. *Acta Psychologica Sinica* 41:7, 572-579. [[CrossRef](#)]
37. Sheila E. Blumstein. 2009. Auditory Word Recognition: Evidence from Aphasia and Functional Neuroimaging. *Language and Linguistics Compass* 3:4, 824-838. [[CrossRef](#)]
38. Emily B. Myers, Sheila E. Blumstein, Edward Walsh, James Eliassen. 2009. Inferior Frontal Regions Underlie the Perception of Phonetic Category Invariance. *Psychological Science* 20:7, 895-903. [[CrossRef](#)]
39. Daniel Voyer, Cheryl Techentin. 2009. Dichotic listening with consonant-vowel pairs: The role of place of articulation and stimulus dominance. *Journal of Phonetics* 37:2, 162-172. [[CrossRef](#)]
40. David W. Gow, Jennifer A. Segawa. 2009. Articulatory mediation of speech perception: A causal analysis of multi-modal imaging data. *Cognition* 110:2, 222-236. [[CrossRef](#)]
41. Bob McMurray, Michael K. Tanenhaus, Richard N. Aslin. 2009. Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language* 60:1, 65-91. [[CrossRef](#)]
42. David W. Gow, Jennifer A. Segawa, Seppo P. Ahlfors, Fa-Hsuan Lin. 2008. Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage* 43:3, 614-623. [[CrossRef](#)]
43. M. Bedny, M. McGill, S. L. Thompson-Schill. 2008. Semantic Adaptation and Competition during Word Comprehension. *Cerebral Cortex* 18:11, 2574-2585. [[CrossRef](#)]
44. Meghan Clayards, Michael K. Tanenhaus, Richard N. Aslin, Robert A. Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108:3, 804-809. [[CrossRef](#)]
45. Alfredo Brancucci, Anita D'Anselmo, Federica Martello, Luca Tommasi. 2008. Left hemisphere specialization for duration discrimination of musical and speech sounds. *Neuropsychologia* 46:7, 2013-2019. [[CrossRef](#)]
46. Richard E. Frye, Janet McGraw Fisher, Thomas Witzel, Seppo P. Ahlfors, Paul Swank, Jacqueline Liederman, Eric Halgren. 2008. Objective phonological and subjective perceptual characteristics of syllables modulate spatiotemporal patterns of superior temporal gyrus activity. *NeuroImage* 40:4, 1888-1901. [[CrossRef](#)]
47. E. B. Myers, S. E. Blumstein. 2008. The Neural Bases of the Lexical Effect: An fMRI Investigation. *Cerebral Cortex* 18:2, 278-288. [[CrossRef](#)]



48. Uri Hasson, Jeremy I. Skipper, Howard C. Nusbaum, Steven L. Small. 2007. Abstract Coding of Audiovisual Speech: Beyond Sensory Representation. *Neuron* **56**:6, 1116-1126. [[CrossRef](#)]
49. U. Hasson, H. C. Nusbaum, S. L. Small. 2007. Brain Networks Subserving the Extraction of Sentence Information and Its Encoding to Memory. *Cerebral Cortex* **17**:12, 2899-2913. [[CrossRef](#)]
50. Rajeev D.S. Raizada, Russell A. Poldrack. 2007. Selective Amplification of Stimulus Differences during Categorical Processing of Speech. *Neuron* **56**:4, 726-740. [[CrossRef](#)]
51. M. F. Joanisse, J. D. Zevin, B. D. McCandliss. 2007. Brain Mechanisms Implicated in the Preattentive Categorization of Speech Sounds Revealed Using fMRI and a Short-Interval Habituation Trial Paradigm. *Cerebral Cortex* **17**:9, 2084-2093. [[CrossRef](#)]
52. Marina Bedny, Justin C. Hulbert, Sharon L. Thompson-Schill. 2007. Understanding words in context: The role of Broca's area in word comprehension. *Brain Research* **1146**, 101-114. [[CrossRef](#)]
53. Alexander L. Francis, Courtney Driscoll. 2006. Training to use voice onset time as a cue to talker identification induces a left-ear/right-hemisphere processing advantage. *Brain and Language* **98**:3, 310-318. [[CrossRef](#)]
54. Fatima T. Husain, Christopher M. McKinney, Barry Horwitz. 2006. Frontal cortex functional connectivity changes during sound categorization. *NeuroReport* **17**:6, 617-621. [[CrossRef](#)]
55. David Poeppel, William Idsardi. Recognizing words from speech: The perception-action-memory loop . [[CrossRef](#)]