



Published in final edited form as:

Neuroimage. 2008 March 1; 40(1): 342–352. doi:10.1016/j.neuroimage.2007.10.064.

An Event-Related fMRI Investigation of Voice-Onset Time Discrimination

Emmette R. Hutchison, Sheila E. Blumstein, and Emily B. Myers

Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI

Abstract

The discrimination of voice-onset time, an acoustic-phonetic cue to voicing in stop consonants, was investigated to explore the neural systems underlying the perception of a rapid temporal speech parameter. Pairs of synthetic stimuli taken from a [da] to [ta] continuum varying in voice-onset time (VOT) were presented for discrimination judgments. Participants exhibited categorical perception, discriminating 15 ms and 30 ms between-category comparisons and failing to discriminate 15 ms within-category comparisons. Contrastive analysis with a tone discrimination task demonstrated left superior temporal gyrus activation in all three VOT conditions with recruitment of additional regions, particularly the right inferior frontal gyrus and middle frontal gyrus for the 15 ms between-category stimuli. Hemispheric differences using anatomically-defined regions of interest showed two distinct patterns with anterior regions showing more activation in the right hemisphere relative to the left hemisphere and temporal regions demonstrating greater activation in the left hemisphere relative to the right hemisphere. Activation in the temporal regions appears to reflect initial acoustic-perceptual analysis of VOT. Greater activation in the right hemisphere anterior regions may reflect increased processing demands, suggesting involvement of the right hemisphere when the acoustic distance between the stimuli are reduced and when the discrimination judgment becomes more difficult.

Introduction

The perception of speech and the mapping of sound structure to higher levels of language is a fundamental property of the language processing system, yet it is still a poorly understood phenomenon. Similar to other language functions, speech perception has been traditionally viewed as left hemisphere dominant. Patients with left hemisphere lesions involving either frontal structures or temporo-parietal structures display impairments in speech perception (Blumstein, 2000). In addition, behavioral data from dichotic listening tasks with unimpaired populations support a right ear (left hemisphere) advantage for the perception of consonants as well as for phonetic features such as voicing and place of articulation (Shankweiler & Studdert-Kennedy, 1967; Spellacy & Blumstein, 1970; Studdert-Kennedy & Shankweiler, 1970).

Nonetheless, there is evidence that challenges the view that the left hemisphere is the exclusive domain for the processing of speech. Boatman et al. (1998) found that the only receptive language ability that was spared in a seizure patient after disabling the left hemisphere with sodium amytal was the discrimination of CV syllables distinguished by voicing or place of

Correspondence address: Sheila E. Blumstein, Brown University, Department of Cognitive and Linguistic Sciences, Box 1978, 190 Thayer Street, Providence, RI 02912, Sheila_Blumstein@brown.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

articulation, suggesting that the right hemisphere may have a role in discrimination of these types of phonetic contrasts. Additionally, a converging body of evidence from neuroimaging studies has shown involvement of both left and right hemisphere structures in various speech perception tasks (Hickok & Poeppel, 2000; Binder & Price, 2001; Scott & Johnsrude, 2003).

Consistent with these findings are several hypotheses which propose bilateral involvement in the early stages of speech perception (Poeppel, 2001; Zatorre et al., 2002). In this case, early stages of processing refer to the extraction of the spectral and temporal properties of the stimuli which ultimately provide the basic parameters for perceiving the sounds of speech. Despite differences in their details, these hypotheses share two assumptions. First, they propose that the temporal lobe structures of both hemispheres provide the substrate for constructing sound-based representations (Binder & Price, 2001; Hickok & Poeppel, 2000, 2004). Second, they hypothesize that the computational capacities of the two hemispheres differ and as such preferentially process different aspects of speech as a function of their intrinsic acoustic properties. For example, fine spectral detail over a longer time window that characterizes formant patterns and serves as a cue to vowel quality should be preferentially processed by the right hemisphere. In contrast, temporal parameters of short duration, such as the rapid spectral changes that distinguish place of articulation in stop consonants, or voice-onset time (VOT), a short duration (0-40 ms) parameter that distinguishes voiced and voiceless stop consonants, should be preferentially processed by the left hemisphere.

The findings from several recent event-related fMRI studies investigating the perception of voicing in stop consonants (Burton et al., 2000; Blumstein et al., 2005; Myers, 2007) are consistent with the view that there are both bilateral (Binder & Price, 2001; Scott & Johnsrude, 2003) and left-lateralized (cf. Scott & Wise, 2004) components to the processing stream for speech. Burton et al. showed bilateral STG activation for the discrimination of natural speech stimuli differing in the voicing of the initial stop consonant, e.g. *dip* vs. *tip*. When the stimuli were such that the subjects had to segment out the initial stop consonant in order to discriminate the voicing of the initial stop consonant, e.g. *dip* vs. *ten*, there was additional unilateral activation in the left inferior frontal gyrus (IFG). Similar findings were shown by Blumstein et al. (2005) in a study investigating the perception of voice-onset time (VOT), an acoustic cue distinguishing voiced and voiceless stop consonants. Subjects were required to make phonetic category decisions on a synthetic [da] to [ta] continuum varying in VOT. Overall bilateral temporal areas were implicated in initial processing of VOT, showing responsiveness as a function of the “goodness of fit” of tokens to their phonetic category (cf. also Myers, 2007), whereas left IFG areas showed graded activation as a function of phonetic category membership as the speech signal was mapped onto a linguistic representation.

Other studies focusing on the perception of VOT have suggested a bilateral component to its processing. Using auditory evoked responses, Molfese and collaborators found both right and left hemisphere sensitivity to stimuli varying in voice-onset time (see Simos et al., 1997 for review.) Of interest, the RH temporal component observed in these studies emerged for stimulus contrasts that crossed the phonetic boundary. In addition, Papanicolaou et al. (2003), using MEG to examine responses to stimuli varying in VOT and tone-onset time (TOT) in temporal areas of both hemispheres, showed that early responses (60 to 130ms) in the auditory cortex were bilateral, whereas middle (130ms to 800ms) and late responses (800ms to 1400ms) showed a greater magnitude of response in the left hemisphere than in the right hemisphere.

These results suggest that early stages of auditory processing in which the temporal parameters associated with VOT are extracted may recruit both hemispheres bilaterally and that left hemisphere lateralization emerges later in the phonetic processing stream when these temporal parameters are mapped to the voiced and voiceless phonetic categories. Thus the right

hemisphere may be recruited in addition to the left hemisphere in extracting the auditory parameters of the stimuli, whereas left hemisphere processes may be recruited in later stages of processing such as phonetic segmentation or phonetic categorization.

It is the goal of this study to explore the neural systems underlying the processing of voice-onset time and potential laterality differences in such processing. There are two questions of interest. First, although the studies reviewed above have shown bilateral STG activation for the processing of VOT, none have directly compared the activation patterns of the left and right STG to determine whether there is a lateral preference for the processing of VOT. Based on the hypotheses suggesting different computational properties of the two hemispheres, it is expected that there will be greater left than right STG activation for the processing of VOT, which requires resolving temporal order changes ranging from 0 to 40 ms.

The second question concerns the role of anterior brain structures in the processing of VOT. The results of Burton et al. (2000) suggest that the IFG is recruited only in tasks that require the mapping of the spectral-temporal patterns of speech on to a linguistic phonetic category (cf. also Burton, 2001). However, the stimuli used in that study were natural speech exemplars which contained multiple acoustic cues distinguishing voicing in stop consonants including voice-onset time, aspiration noise, and fundamental frequency differences at consonantal release. The question is whether frontal areas will be recruited using the same task as Burton et al. (2000), a same-different discrimination task with a short (50 ms) ISI when subjects are required to attend to fine acoustic differences in VOT and presumably are not attending to phonetic category membership (Liberman et al, 1957).

VOT is typically perceived in a categorical-like manner. Listeners categorize stimuli varying parametrically in voice-onset time into two distinct phonetic categories, showing a sharp discontinuity between the two categories, typically at a particular voice-onset time 'boundary' stimulus. However, they are only able to discriminate reliably those stimuli that they have categorized as members of different phonetic categories ([da] vs. [ta]), and are at chance in discriminating stimuli that lie within a phonetic category (Liberman, Delattre, & Cooper, 1958). Nonetheless, recent studies have shown that listeners are in fact sensitive to the acoustic differences within a phonetic category (Miller, 1994; Iverson & Kuhl, 1995). They show slower reaction-time latencies to different members of the same phonetic category than to two acoustically identical stimuli, indicating some perceptual sensitivity to within-category variation in VOT differences (Pisoni & Tash, 1974). Thus, examining the processing of VOT in a same-different discrimination task allows for the investigation of the role of phonetic factors, on the one hand, by comparing between and within-category differences while equating for acoustic distance, and acoustic factors, on the other, by comparing activation patterns for between phonetic category differences that differ in their acoustic distance.

We hypothesize that although there will be both right and left hemisphere activation for the discrimination of tokens differing only in voice-onset time, the processing of VOT will be left-lateralized, particularly in the STG. In the event that differences emerge in the activation patterns of the STG as a function of the different VOT stimulus conditions, we will examine whether similar differences emerge even earlier in the processing stream in the transverse temporal gyrus (TTG). We also expect activation of frontal structures such as the IFG, particularly as the discrimination judgments become more difficult. In particular, increased frontal activation should emerge for 15 ms within phonetic category discriminations compared to 15 ms between phonetic category discriminations because the discrimination judgment is more difficult in the former case as it is based solely on acoustic differences in the stimulus pairs whereas in the latter case the discrimination judgment it is based on both acoustic and phonetic differences in the stimuli. Moreover, increased frontal activation should also emerge for the discrimination of between-category stimulus pairs and within-category stimulus pairs

that differ by 15 ms compared to between-category stimuli that differ by 30 ms, reflecting the increased difficulty of the discrimination judgment.

A simple tone discrimination task was utilized to serve as a non-speech control. Since the goal of the study was to examine the perception of a fine-scale temporal parameter, we did not want to factor out this parameter in the non-speech stimuli. Hence, the non-speech stimuli were pure tones which differed from each other only in frequency, but were otherwise matched with the speech stimuli in both duration and amplitude. In this way, the tone condition was not designed to equate for task difficulty but rather to serve as a control for low-level auditory processing as well as for task-related components such as planning and executing a button press. Of import, the experimental hypotheses explored in this study depend on differences in activation between syllable stimuli varying in voice-onset time, having subtracted out from each syllable condition low-level auditory processing via the tone control condition. Such simple tone stimuli have been used as a control for speech stimuli when the goal of the study was to identify differences in activation patterns associated with the processing of phonetic category structure (cf. Blumstein et al., 2005).

Materials and methods

Participants

Seventeen healthy adult native English speakers were recruited from the Brown University community. The educational level of the participants ranged from current undergraduate students to Ph.D.s. Participants gave written informed consent according to guidelines established and approved by the Human Subjects Committees of both Brown University and the Memorial Hospital of Rhode Island. All participants were screened for lack of bodily ferromagnetic objects. The Edinburgh Handedness Inventory (Oldfield, 1971) was administered to confirm that all participants were strongly right handed (the mean handedness index for the participants was 88 with a range of 74 to 100). Modest monetary compensation was given for participation in the experiment. Of the original 17 participants, the fMRI data from 14 were analyzed: one participant failed to complete the scanning session and the behavioral data from two others indicated that they could not reliably discriminate any of the stimulus conditions. The remaining 14 participants included 5 men and 9 women from 21-60 years of age (mean 30, SD 3.1).

Stimuli

Stimuli consisted of six speech stimuli taken from a synthetic [da]-[ta] continuum created by a parallel resonance synthesizer made available at Haskins Laboratories (see Figure 1). The six stimuli had VOT values of 0, 10, 15, 25, 30 and 40 ms. The selection of the VOT values for these stimuli was based on an earlier fMRI study of VOT categorization which varied VOT in 10 ms step from 0 ms ([da]) to 40 ms ([ta]) and showed that the phonetic boundary between the two phonetic categories was at 20 ms (Blumstein et al., 2005). Thus, VOT values of 0, 10, and 15 corresponded to voiced stimuli ([da]) and VOT values of 25, 30, and 40 corresponded to voiceless stimuli ([ta]). VOT was manipulated by replacing the periodic source with an aperiodic source in 10, 15, 25, 30, or 40 msec segments, starting from the onset of the syllable. All stimuli consisted of 5 formant patterns with onset values for F1-F5 being 200 Hz, 1350 Hz, 3100 Hz, 3600 Hz and 4500 Hz respectively. F1-F3 had formant transitions that lasted 40 ms followed by steady-state values of 720 Hz, 1250 Hz, and 2500 Hz. All of the stimuli were 230 ms in length.

In addition to the speech stimuli, there were two tone stimuli created for use in the tone discrimination task. These stimuli were the same as those used in Blumstein et al. (2005) investigating the perception of voice-onset time using a phonetic categorization task and

consisted of a high tone (a sine wave at 1320 Hz) and a low tone (a sine wave at 910 Hz). The pitch of these tones was selected to lie within the range of the vowel steady-state formant frequencies of the speech stimuli. The duration of the tone stimuli was the same as the VOT stimuli and the amplitude of the tone stimuli was matched to that of the steady-state portion of the vowels of the VOT stimuli.

Tasks

Participants performed a same-different discrimination task in which each trial consisted of either pairs of syllable or tone stimuli separated by a 50 ms ISI. The experiment consisted of two syllable runs and two tone runs, with the order of syllable runs counterbalanced across participants. Participants were required to listen to each pair of stimuli and to press one of two buttons in determining whether the stimulus pairs were the same or different; the button mapping for the run was counterbalanced across participants. Subjects were asked to respond as quickly and as accurately as possible. Each subject received practice trials for the discrimination task on both syllable and tone pairs during the anatomical scan prior to the functional runs. To reduce eye movement artifacts and activation of visual areas, scan room lights were dimmed and participants were asked to keep their eyes closed during functional runs. Stimuli were presented in a fixed, pseudo-randomized order within each run.

The syllable stimuli were grouped into same and different conditions. The different discrimination condition consisted of three trial types: a 30ms between category comparison (0 vs 30 ms VOT and 10 vs 40 ms VOT), a 15ms between category comparison (10 vs 25 ms VOT and 15 vs 30 ms VOT), and a 15ms within category comparison (0 vs 15 ms VOT and 25 vs 40 ms VOT). The order of the different stimulus pairs was counterbalanced across trial types (e.g. 0 vs. 15 ms VOT and 15ms VOT vs. 0 ms VOT). The Same discrimination condition paired each VOT stimulus with itself. In each syllable run, 60 different (20 of each condition) and 60 same discrimination pairs were presented.

In the tone runs there were two conditions with 20 presentations each: different tone discriminations (high vs. low, low vs. high) and same tone discriminations (high vs. high, low vs. low) resulting in 40 trials for each of the two tone runs. Similar to the syllable stimuli, the order of the different stimulus pairs was counterbalanced.

Stimuli were presented over sound attenuating air conduction headphones and the sound level was adjusted to ensure a comfortable listening level. Responses were collected with an MR-compatible button box and scored for both accuracy and reaction time (RT), with RT latencies measured from the onset of the stimulus. Presentation of stimuli and collection of response data were controlled by a laptop running the BLISS software suite (Mertus, 1989).

Image Collection

All images were acquired on a 1.5 Tesla Symphony Magnetom MR system (Siemens Medical Systems, Erlangen, Germany). A high-resolution anatomical dataset was acquired for each participant (TR=1900ms, TE=4.15ms, TI=1100ms, 1mm³ voxels and a FoV of 256mm), prior to acquiring functional data. A total of 968 (362 images for each syllable run and 122 images for each tone run) echo-planar images (EPIs) were acquired in a transverse plane using blood-oxygenation-level-dependent (BOLD) imaging. Each EPI image consisted of 15 slices covering a 75mm area and was aligned to each participant's corpus callosum which allowed for the collection of functional data from bilateral perisylvian cortex (TR= 2000ms with 1200ms of image acquisition and 800ms of a silent interval, TE= 38ms, a FoV of 192mm with a 64mm² matrix, a slice thickness of 5mm and an in-plane resolution of 3mm² resulting in a voxel volume of 45mm³). The imaging data were acquired so that the auditory stimuli were presented during silent gaps in between functional scans (Belin et al. 1999; Hall et al. 1999;

Edmister et al., 1999). In particular, the 15 slices were acquired in a 1200 ms interval followed by a 800 ms silent gap in which a discrimination pair was presented. To reduce T1 saturation effects, four dummy volumes were collected before the presentation of the first stimulus pair in each run, and these volumes were excluded from subsequent analysis.

An event-related design was utilized in which stimuli were jittered according to a uniform distribution of six trial onset asynchronies (TOA values ranging from 2 to 10s in 2 second steps, resulting in an average TOA of 6 seconds).

Data Analysis

Images acquired during the experiment were processed and analyzed on a 12 node Linux cluster using AFNI (Cox, 1996; Cox & Hyde, 1997; <http://afni.nimh.nih.gov/afni/>). Functional and anatomical datasets for each participant were co-registered using positioning information from the MR. Functional datasets were motion corrected using a six-parameter rigid body transformation (Cox & Jesmanowicz, 1999). Anatomical datasets were normalized to the Talairach and Tournoux template (1988) as implemented in AFNI and motion-corrected functional datasets were registered to the normalized anatomical dataset. Normalized functional datasets were resampled to 3mm^3 and smoothed using a 6mm^3 full-width at half-minimum Gaussian kernel. A brain mask was created using each participant's normalized anatomical dataset which encompassed only brain tissue. These masks were merged to form a representative brain mask that included all voxels defined as brain tissue in a minimum of 12 participants.

Single-subject analysis was conducted on voxels defined in the group brain mask. Vectors containing stimulus onset times for all 6 conditions (15 msec Within-Category, 15 msec Between-Category, 30 msec Between-Category, Syllable Same, Tone Different and Tone Same) were created for each participant and convolved with a gamma-variate function to create reference functions for each condition. The smoothed functional data and reference functions were submitted to deconvolution to estimate the hemodynamic response during performance of each condition on a voxel by voxel basis. The six parameters from the motion correction algorithm were included in the deconvolution analysis as additional reference waveforms. Estimated coefficients for each condition were normalized to the participant's experiment-wise average BOLD to convert intensity into percent signal change.

Group analysis was accomplished by submitting percent change datasets to a two-factor ANOVA using stimulus condition (fixed effect) and participant (random effect) as independent variables. Planned comparisons were conducted between the different stimulus pairs for the speech and tone stimuli. Contrasts included comparisons of the speech and tone stimuli (Syllables vs. Tones) and the three syllable different conditions compared to the tone different condition (30ms Between-Category vs. Tone Different, 15ms Between-Category vs. Tone Different, 15ms Within-Category vs. Tone Different). In addition, planned comparisons were conducted for the different syllable discrimination stimuli. Pair-wise comparisons were conducted between the three syllable different conditions (15 ms Within-Category vs. 15 ms Between-Category, 15 ms Between-Category vs. 30 Between-Category ms, and 15 ms Within-Category and 30 ms Between-Category). Monte-Carlo simulations were run to obtain cluster level thresholds. A voxel level threshold of $p < 0.01$ and a cluster level threshold of $p < 0.01$ (44 contiguous voxels) was used for all comparisons. To investigate whether activation observed in these comparisons was attributable to differences in difficulty as measured by reaction time, a regression analysis was performed which identified areas in which activation correlated with reaction time.

To assess activation patterns in specific anatomical regions as well as potential differences in laterality, a region-of-interest (ROI) analysis was conducted. ROIs were selected based upon

the hypotheses described earlier and defined by the Talairach and Tournoux atlas (1988), as implemented by AFNI. Left and right hemisphere ROIs were created for each of the following areas of interest: Transverse Temporal Gyrus (TTG), Superior Temporal Gyrus (STG), and the Inferior Frontal Gyrus (IFG). For each of these regions, mean percent signal change was extracted for every participant and submitted to a two factor (Condition x Hemisphere) analysis of variance.

RESULTS

Behavioral Results

Mean behavioral results for both performance and reaction-time for all 14 participants are displayed in Figure 2 for all syllable different conditions. Figure 2A shows the performance results. As expected, performance was better for the 30 ms between category discrimination pairs (91%) than for the 15 ms between category discrimination pairs (77%), and participants discriminated stimuli in the 15ms Within-Category condition at below-chance levels (29%). This behavioral pattern reflects what is typically seen in discrimination tasks involving categorically perceived stimuli (Liberman et al., 1957). A one-way repeated measures analysis of variance (ANOVA) confirmed these findings showing a significant main effect of pair type ($F(2, 26) = 108.514, p < 0.001$). Pairwise post-hoc comparisons revealed that all three conditions were significantly different from each other.

As Figure 2B shows, reaction-time (RT) latencies were inversely related to accuracy. The 30ms Between-Category condition showed the fastest latencies (648 ms), followed by the 15ms Between-Category condition (701 ms) and the 15ms Within-Category condition (798 ms). A one-way repeated measures ANOVA for RT was significant ($F(2, 26) = 30.473, p < 0.001$). Pair-wise post-hoc comparisons revealed significant differences between the 15ms and 30 ms Between-Category conditions compared to the 15ms Within-Category condition. However, the two between category different conditions were not different from each other.

For the tone discrimination task, performance results showed 99% accuracy ($SD=0.023$) for the different discrimination judgments and 99% accuracy ($SD=0.013$) for the same discrimination judgments. Reaction times for the two tone conditions were 536ms ($SD = 82$) for the tone different and condition 539ms ($SD=63$) for the tone same condition. Thus, overall and as expected, the tone discrimination task was easier than the VOT discrimination task.

fMRI Results

Tables 1-3 show the results of the planned comparisons between the four different stimulus types.

Syllables versus Tones—A number of significant clusters emerged in both hemispheres for the comparison between syllables and tones (see Table 1). Greater activation for syllables was observed in a number of clusters including the right inferior frontal gyrus (R IFG), the left superior temporal gyrus (L STG), and in medial structures in both the right and left hemispheres. Greater activation for tones emerged in several clusters including the right superior temporal gyrus (R STG), right inferior parietal lobule (R IPL), right middle temporal gyrus (R MTG), the left supramarginal gyrus (L SMG) and the left insula.

Comparison of Syllable Different Conditions to Tone Different—To further explore differences between the three syllable different conditions, comparisons were made between each syllable different condition and the Tone Different condition (see Table 2 and Figure 3). Comparison of the 30 ms Between-Category condition to the Tone Different condition (Figure 3A) revealed greater activation for the Between-Category condition in the L STG and greater

activation for the Tone condition in the R IFG and the IPL bilaterally. Similarly, comparison of the 15 ms Between-Category condition to the Tone condition (Figure 3B) revealed greater activation for the Between-Category condition in the L STG. Unlike the 30 ms Between-Category comparison, however, there was also greater activation for the 15 ms Between-Category condition in the R IFG and R MFG, the L insula, and medial structures including the cingulate. Regions observed as more active for tones in this comparison included the R IPL, bilateral insula and the L SMG.

Comparison of the activation patterns for the 15 ms Within-Category condition and the Tone condition (Figure 3C) revealed significantly greater activation for the Within-Category condition in the L STG as well as in medial structures including the cingulate gyrus and medial frontal gyrus. There was greater activation for tones in the R. IPL and insula as well as in left hemisphere structures including the SMG and insula.

In sum, when all three syllable category conditions were contrasted with a control (Tone Different) condition, the only area of activation that consistently emerged was the L STG. Additional areas were recruited as a function of the particular syllable category discrimination. More medial structures including the cingulate gyrus were observed for both 15ms conditions. The 15ms Between-Category comparison, the most difficult discrimination that participants were able to perform, selectively recruited both right and left anterior regions including the R IFG and R MFG and the L insula.

Pairwise comparison of the Syllable Different Conditions—Pairwise comparisons were conducted between the three syllable different conditions (15 ms Between-Category vs. 15 ms Within-Category, 15 ms Between-Category vs. 30 ms Between-Category, and 15 ms Within-Category and 30 ms Between-Category). As Table 3 shows, significant clusters emerged in only one of the three comparisons. In particular, activation was observed for the 15ms Between-Category condition compared to the 30 ms Between-Category condition in the R IFG and R IPL.

Regression Analysis—Results of the regression analysis looking at those clusters where there was a significant relation between reaction-time and degree of activation are shown in Table 4. As the Table shows significant clusters emerged in the R Cingulate Gyrus, the L Precuneus, and the L Anterior Cingulate. The absence of significant effects in the IFG or STG suggests that the significant activation patterns in the L STG and R IFG shown in the previous cluster analyses are not due to the difficulty of the discrimination judgments (as measured by RT).

ROI Analysis—A Hemisphere x Syllable condition ANOVA was performed for each ROI. Two patterns of interest were observed as shown in Figure 4. Left hemisphere lateralization was observed in posterior regions including the TTG and the STG, an effect which was significant in the former, and approached significance in the latter ($F(1,13)=12.995$, $p < 0.003$ and $F(1,3)=4.086$, $p < 0.064$ respectively). Conversely, a pattern of significant right hemisphere lateralization was observed for the IFG ($F(1,13)=8.352$, $p < 0.013$).

For the IFG, results of the ANOVA also showed that overall there was greater activation for the 15 ms Between-Category discrimination than either the 15 ms Within-Category or 30 ms Between-Category discrimination, as revealed by the significant main effect for Condition ($F(2,26)=4.415$, $p < 0.022$). In addition, there was a Condition X Hemisphere interaction which approached significance ($F(2,26)=2.505$, $p < 0.101$), indicating that the magnitude of the laterality effects in the IFG varied as a function of condition. The magnitude of the size of this effect showed a partial eta-squared value of .162 and a power analysis revealed a medium effect size of .46 (Cohen, 1988). Nonetheless, analysis of simple effects revealed that both the 15 ms

Between-Category and 15 ms Within-Category discriminations had a greater degree of RH lateralization than did the 30 ms Between-Category discrimination.

In addition to the main effect for laterality for the STG owing to left hemisphere lateralization across the discrimination conditions, results of the ANOVA also revealed a Condition X Hemisphere interaction that was due to a smaller laterality difference for the 15 ms Within-Category discrimination condition compared to both the 15 ms and 30 ms Between-Category discrimination conditions ($F(2,26)=4.341$, $p<0.024$). Post-hoc tests revealed that there was greater left hemisphere activation in the STG for the 30 and 15 ms Between-Category discriminations, and no laterality difference for the 15 ms Within-Category discrimination.

DISCUSSION

The current study investigated the neural systems underlying early stages of phonetic processing by utilizing a sub-lexical discrimination task with a short 50 ms ISI designed to focus listeners' attention on the acoustic-phonetic properties inherent in the stimuli (cf. Liberman et al., 1957). While the relatively poor temporal resolution of the fMRI paradigm does not allow for determination of the time course of processing, the discrimination task was designed to tap processing without requiring the subject to make a decision about the phonetic content of the stimuli. Results showed that the perception of voice-onset time, a rapid temporal cue to voicing in stop consonants, recruits an extensive bilateral network with left hemisphere lateralization for the acoustic processing of voice-onset time and right hemisphere lateralization for decision processes. Even relatively early in the phonetic processing stream, i.e. in the STG, differences in activation patterns emerged as a function of the phonetic category status of the stimuli, i.e. between versus within category discriminations, and as a function of the acoustic distance between stimuli.

Phonetic Category Structure and Acoustic Distance

Perception of Between-Category and Within-Category Discrimination—The behavioral results showed sensitivity to phonetic category structure and the acoustic distance between stimuli. As expected, participants showed more accurate discrimination and faster reaction-time latencies for stimulus pairs which belonged to different phonetic categories (15 msec Between-Category, 30 msec Between-Category) as well as more accurate performance for a larger acoustic step (30 ms Between-Category) than a smaller acoustic step (15 ms Between-Category). Additionally, better discrimination of 15 ms between phonetic category pairs than 15 within phonetic category pairs indicated that despite the fact that the acoustic distance between the discrimination pairs was the same, there was a processing advantage for those discrimination pairs in which stimuli belonged to different phonetic categories, even when the acoustic distance was equated.

The neuroimaging results also showed sensitivity to both phonetic category structure and acoustic distance. In particular, pairwise comparison of the VOT discrimination conditions showed greater activation for the 15ms Between-Category discrimination compared to the 30ms Between-Category discrimination, particularly in frontal regions which are typically associated with tasks requiring increased computational resources (Binder et al., 2004; Scott and Wise, 2004). In addition, the near-significant laterality X hemisphere interaction for the IFG in the ROI analysis showed greater right hemisphere lateralization for both the 15 ms Between-Category and 15 ms Within-Category discriminations compared to the 30 ms Between-Category discriminations. This activation pattern is consistent with the interpretation that increased processing resources are needed as the acoustic distance between stimuli is reduced and the phonetic discrimination is made more difficult.

What is less clear is what the nature of these processing resources may be. Increased bilateral frontal activation has been shown for both language and nonlanguage stimuli across a range of tasks having different cognitive demands including response conflict, stimulus novelty, working memory, perceptual difficulty, and attention (Duncan and Owen, 2000; Fuster, 2000). That similar bilateral frontal areas are activated across these disparate cognitive demands makes it difficult to ascribe a common functional role to them. Nonetheless, in the case of the current experiment, it appears as though increased frontal activation occurs as a function of the difficulty of the discrimination judgment as measured by both performance and reaction-time latency.

Consistent with these findings are the results of a recent study showing that the strength of the MMN was modulated by the acoustic distance between stimuli drawn from two phonetic categories (Joanisse et al., 2007). Such sensitivity suggests that these effects do not rely on overt attention or memory processes and have a strong sensory component to them. Nonetheless, the results of the current experiment suggest that the increased activation of the R IFG is not due merely to the acoustic distance between the stimuli and the resultant difficulty of the discrimination, but that phonetic category structure also plays a role. In particular, the R IFG showed sensitivity to acoustic distance but only for those stimuli that belonged to different phonetic categories. As described above, the pairwise comparisons of the 15ms Between-Category and the 30 ms Between-Category conditions showed increased frontal activation, whereas the comparison of the 15ms Within-Category and the 30 ms Between-Category conditions failed to show any difference. In both cases, the magnitude of the acoustic distance between the pair types was the same (i.e. a 15 ms discrimination vs. a 30 ms discrimination). Moreover, the ROI analysis revealed that the 15 ms Between-Category discrimination showed significantly more activation in both right and left hemispheres than the 15 ms Within-Category.

Comparison of the three VOT conditions to the tone control condition revealed a similar pattern of results as those shown for the direct comparison between the VOT conditions and the ROI analysis. Namely, a significant cluster emerged in R IFG only for the 15 ms Between-Category discrimination. No frontal clusters emerged for either the 15 ms Within-Category discrimination or for the 30 ms Between-Category discrimination. The lack of a significant IFG cluster for the 30 ms Between-Category discrimination most likely reflects the fact that the discrimination is relatively easy and requires few resources. What is surprising is that none of the analyses (comparisons of conditions, ROI analysis, and comparison of the three VOT conditions with the tone condition) showed increased activation in the R IFG cluster for the 15 ms Within-Category discrimination relative to the 15 ms Between-Category discrimination. The question is why?

It was originally hypothesized that the 15ms Within-Category discrimination would show the poorest performance, the slowest reaction-time latencies, and the greatest neural activation compared to the other VOT discriminations. As predicted, participants failed to discriminate the 15 ms Within-Category pairs as indicated by a mean accuracy score of 29%. However, despite the obvious difficulty of the discrimination, there was no difference in RT latencies between the 15 ms Between-Category and 15 ms Within-Category discriminations. Moreover, the regression analysis showed no correlation in the IFG between degree of activation and reaction-time latencies. Thus, the failure to show increased R IFG activation in the 15 ms Within-Category discrimination cannot be attributed to the difficulty of the discrimination per se. These results suggest that the RH is recruited under conditions of stimulus difficulty and uncertainty, but that there is increased activation only when the difference between stimuli is perceptible. When subjects fail to discriminate stimuli, the RH shows a reduction in activation, presumably reflecting reduced processing on these stimuli.

A study by Poldrack et al. (2001) also showed changes in the activation patterns in the IFG as a function of the extent of temporal compression of speech and its effects on subsequent ‘comprehensibility’ of the stimuli. Using a sentence verification task, they showed a convex pattern of activation in frontal areas as a function of the extent of speech compression. Increased activation was observed bilaterally in the IFG as compression increased (and the stimuli became less comprehensible as measured by task performance). However, when speech compression resulted in chance performance on the sentence verification task, IFG activation decreased. In the present study, the 15ms Between-Category condition was a more difficult discrimination than the 30 ms Between-Category discrimination, but participants were still able to reliably phonetically discriminate this condition as “different”, recruiting similar frontal regions to those seen in the Poldrack et al. study, although only in the right hemisphere. In contrast, when discrimination performance was at chance in the Within-Category condition, the expected frontal activation failed to emerge.

Laterality Effects

Two general patterns of interest were observed. First, activation in anterior areas was right hemisphere lateralized and activation in posterior areas was left hemisphere lateralized. Second, the magnitude of the laterality effect in both the IFG and STG varied as a function of the type of VOT discrimination.

Posterior Structures—Although bilateral activation was observed in posterior regions in the discrimination of VOT, an overall left-lateralization was observed across syllable conditions as early as the transverse temporal gyrus (TTG). Of interest, the pattern of lateralization observed in the ROI analysis was the same across the three VOT discrimination conditions. That lateral differences emerged as early as the TTG for acoustic properties of speech has not been reported previously in fMRI experiments, but lateral differences have been shown in the processing of voiced and voiceless stop consonants using intracerebral evoked potentials (Liegeois-Chauvel et al., 1999). The failure to show these effects in fMRI studies may be due, in part, to the methods used in analyzing the data. As an example, Liebenthal et al. (2005) investigated the discrimination of formant transitions cueing place of articulation in stop consonant and found increased LH activation only in the ventral STG/STS but not in the TTG. In their study, Liebenthal et al. compared the discrimination of speech stimuli to non-speech controls closely matched to the speech sounds in terms of their acoustic properties. By comparing the activation patterns between speech and non-speech, Liebenthal et al. subtracted out potential laterality differences for the processing of rapid spectral changes in formant transitions. Thus, it is possible that there were laterality differences in the processing of spectral changes for either speech or non-speech stimuli as early as the TTG. In the current study, differences in the processing of voice-onset time were compared directly and showed left TTG lateralization. It would be of interest to determine whether similar laterality effects would emerge for the processing of rapid temporal cues in non-speech stimuli patterned after voice-onset time such as tone-onset time. Of interest, both Zatorre and Belin (2001) and Jamison et al. (2006) showed lateral differences as early as Heschl’s gyrus in the processing of temporal properties of non-speech stimuli. Such findings suggest that hemispheric asymmetries occur in primary auditory areas for acoustic attributes of both speech and nonspeech and challenge the view that at the earliest stages of the phonetic processing stream both hemispheres equally provide the substrate for constructing sound based representations (Hickok & Poeppel, 2000; Binder et al., 2004).

Consistent with current models of the neural systems underlying the processing of speech, the L STG was activated for all three VOT conditions (cf. Hickok & Poeppel, 2000,2004;Poeppel & Hickock, 2004;Zatorre et al., 2002). These results emerged in the pairwise comparisons of

the VOT conditions, in the comparison of the three VOT conditions to the simple non-speech tone discrimination task, and in the ROI analysis.

The magnitude of LH activation differed as a function of the syllable conditions, with significantly greater LH lateralization in both the 30 ms and 15 ms Between-Category conditions compared to the 15 ms Within-Category discriminations. Of interest, there was not a concomitant significant increase in the magnitude of activation in either right or left hemisphere posterior structures for the 15ms Within-Category condition, the condition in which participants failed to reliably discriminate the stimuli. What distinguishes the 15 ms Between and Within-Category conditions is whether or not the stimulus pairs fall within a phonetic category or belong to two different phonetic categories, suggesting that the perceptual system is 'aware' of phonetic category structure in those anatomical areas considered to be early in the phonetic processing stream. If it is assumed that the phonetic processing system is both a feedforward (TTG to STG to frontal areas) and feedback network, then it would not be surprising to find differential neural responses early in the neural processing stream to stimulus pairs that map on to a common phonetic category compared to ones that map on to a different phonetic category.

The finding that the left posterior STG is sensitive to the phonetic properties of stimulus pairs, even when controlling for acoustic distance, is consistent with those of a recent study by Joanisse and colleagues (2006). In this study, greater release from adaptation of the fMR signal was observed in the left posterior STG for four-syllable trains in which the last stimulus of the train was from a different phonetic category than the first three (i.e. *ga ga ga da*), than for those for those in which the last syllable was different by an equivalent acoustic step, but belonged to the same phonetic category (i.e. *gal gal gal ga2*).

Anterior Structures—A pattern of right hemisphere lateralization was observed for the IFG in all analyses. Additionally, as described earlier, the 15 ms Between-Category discrimination showed increased frontal activation presumably because of increased response selection demands owing to the difficulty of the discrimination. Sharp et al. (2004) showed activation in the right dorsolateral prefrontal cortex in a semantic decision task when the stimuli were acoustically degraded, and a similar but smaller effect in a syllable decision task under similar conditions of acoustic degradation. Similar to the results of the Poldrack et al. study discussed above, this system appeared to be recruited only when the stimuli were sufficiently 'degraded' such that listeners were unable to accurately discriminate them. Nonetheless, the current study showed RH lateralization in the IFG for the discrimination of VOT across all syllable conditions, suggesting that this system may be active in phonetic processing even in the absence of acoustic degradation. What appears to modulate activity in this right hemisphere system are two factors: the sensitivity of the system to the acoustic parameters of the stimulus and the difficulty of the phonetic discrimination.

Of importance, not all speech discrimination tasks appear to recruit this right pre-frontal system. Burton et al. (2000) failed to show increased frontal activation (in either the RH or LH) in a voicing discrimination task using clearly produced natural speech tokens. In the current study, voicing discriminations were also required but the stimuli were synthetic varying along a single acoustic parameter, and the magnitude of the acoustic differences between the stimuli was small at either 15 ms or 30 ms. The difficulty of the between phonetic category discrimination increased the computational resources required to perform the task, thus recruiting the right hemisphere prefrontal network.

In contrast to the current study, Joanisse et al. (2003) showed increased activation in the L IFG and the STG bilaterally in the discrimination of rapid temporal and spectral cues to both speech and nonspeech continua. They proposed that the increased activation in the L IFG was due to

difficulty of the discrimination task. However, unlike the current study, they collapsed the 'easy' and 'difficult' discriminations within the particular stimulus continuum. Thus, it is possible that they too would have found increased activation in right frontal structures had they compared the easy and difficult discriminations within the particular stimulus continuum.

That anterior structures are involved in executive decisions is not surprising (cf. Binder et al., 2004; Scott and Wise, 2004). What was unexpected is that the activation emerged in right frontal structures and not in left frontal structures as well. In a study investigating the discrimination of place of articulation in two synthetically produced syllables in the presence of a parametrically varied noise mask, Binder et al. (2004) found a bilaterally symmetric system in which the IFG showed increased activation as a function of increased masking of the speech stimuli by noise. Myers (2007) found similar bilateral increases in activation in a phonetic categorization task for stimuli varying in VOT as those stimuli approached the phonetic category boundary. Although the current study did not show significant clusters in the L IFG for the more difficult 15 ms Between-Category discrimination, post-hoc tests of changes in lateralization in the LH in the ROI analysis (see Figure 4) revealed a significant increase in activation in the L IFG for the 15 ms Between Category condition compared to the 30 ms Between Category comparison. Thus, the increase in activation was bilateral, but not bilaterally symmetric as found by Binder et al. It may be that the perceptual judgments required in the current study were more difficult than the task demands of the Binder et al. and Myers studies. As a consequence, it is possible that the left hemisphere executive system was operating at ceiling even for the easiest 30 ms Between-Category condition and the right hemisphere executive system was recruited under the more difficult 15 ms Between-Category condition. Another possibility is that in contrast to the current study the task demands in both the Binder et al. and Myers study required subjects to phonetically identify the stimuli rather than focus solely on the acoustic properties of the stimuli. It is likely that under such conditions there would be increased left hemisphere activation.

Acknowledgements

This research was supported in part by NIH grant DC006220 to SEB, a Dana Foundation Grant to SEB, and the Ittleson Foundation .

References

- Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B. Event-related fMRI of the auditory cortex. *Neuroimage* 1999;10:417–429. [PubMed: 10493900]
- Binder, JR.; Price, C. Functional neuroimaging of language. In: Cabeza, R.; Kingstone, A., editors. *Handbook of Functional Neuroimaging of Cognition*. MIT press; Cambridge: 2001. p. 187-251.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience* 2004;7:295–301.
- Blumstein, SE. Deficits of speech production and speech perception in aphasia. In: Berndt, R., editor. *Handbook of Neuropsychology*. 2nd edition. 2. Elsevier Science; the Netherlands: 2000. p. 95-113.
- Blumstein SE, Myers EB, Rissman J. The perception of voice-onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience* 2005;17:1353–1366. [PubMed: 16197689]
- Boatman D, Hart J Jr. Lesser RP, Honeycutt N, Anderson NB, Miglioretti D, Gordon B. Right hemisphere speech perception revealed by amobarbital injection and electrical interference. *Neurology* 1998;51:458–64. [PubMed: 9710019]
- Burton MW, Small SL, Blumstein SE. The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience* 2000;12:679–690. [PubMed: 10936919]
- Burton MW. The role of inferior frontal cortex in phonological processing. *Cognitive Science* 2001;25:695–709.
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum; New Jersey: 1988.

- Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research* 1996;29
- Cox RW, Hyde JS. Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine* 1997;10:171–8. [PubMed: 9430344]
- Cox RW, Jesmanowicz A. Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine* 1999;42:1014–8. [PubMed: 10571921]
- Duncan J, Owen AM. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neuroscience* 2000;23:475–483.
- Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM. Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping* 1999;7:90–97.
- Fuster JM. Executive frontal functions. *Experimental Brain Research* 2000;133:66–70.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. Sparse temporal sampling in auditory fMRI. *Human Brain Mapping* 1999;7:213–223. [PubMed: 10194620]
- Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Science* 2000;4:131–138.
- Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 2004;92:67–99. [PubMed: 15037127]
- Iverson P, Kuhl PK. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustic Society of America* 1995;97:553–562.
- Jamison HL, Watkins E, Bishop DVM, Matthews PM. Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex* 2006;16:1266–1275. [PubMed: 16280465]
- Joanisse MF, Gati JS. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage* 2003;19:64–79. [PubMed: 12781727]
- Joanisse MF, Robertson EK, Newman RL. Mismatch negativity reflects sensory and phonetic speech processing. *NeuroReport* 2007;18:901–905. [PubMed: 17515798]
- Joanisse MF, Zevin JD, McCandliss BD. Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*. 2006(publication ahead of print)
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. *Cerebral Cortex* 2005;15:1621–1631. [PubMed: 15703256]
- Lieberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 1957;54:358–368. [PubMed: 13481283]
- Lieberman AM, Delattre PC, Cooper FS. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech* 1958;1:153–167.
- Liegeois-Chauvel C, de Graaf JB, Laguitton V, Chauvel P. Specialization of left auditory cortex for speech perception in man depends on temporal encoding. *Cerebral Cortex* 1999;9:484–496. [PubMed: 10450893]
- Mertus, J. BLISS user's manual. Providence, Brown University; 1989.
<http://www.cog.brown.edu/localSites/mertus/BlissHome.htm>
- Miller JL. On the internal structure of phonetic categories: A progress report. *Cognition* 1994;50:271–285. [PubMed: 8039364]
- Myers EB. Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: an fMRI investigation. *Neuropsychologia* 2007;45:1463–1473. [PubMed: 17178420]
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97–113. [PubMed: 5146491]
- Papanicolaou AC, Castillo E, Breier JI, Davis RN, Simos PG, Diehl RL. Differential brain activation patterns during perception of voice and tone onset time series: a MEG study. *Neuroimage* 2003;18:448–59. [PubMed: 12595198]
- Pisoni D, Tash J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 1974;15:289–290.

- Poeppel D. Pure word deafness and the bilateral processing of the speech code. *Cognitive Science* 2001;25:679–693.
- Poeppel D, Hickok G. Towards a new functional anatomy of language. *Cognition* 2004;92:1–12. [PubMed: 15037124]
- Poldrack RA, Temple E, Protopapas A, Nagarajan S, Tallal P, Merzenich M, Gabrieli JD. Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *Journal of Cognitive Neuroscience* 2001;13:687–97. [PubMed: 11506664]
- Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences* 2003;26:100–107. [PubMed: 12536133]
- Scott SK, Wise RJ. The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 2004;92:13–45. [PubMed: 15037125]
- Shankweiler D, Studdert-Kennedy M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology* 1967;19:59–63. [PubMed: 6041684]
- Sharp DJ, Scott SK, Wise RJ. Monitoring and the controlled processing of meaning: distinct prefrontal systems. *Cerebral Cortex* 2004;14:1–10. [PubMed: 14654452]
- Simos PG, Molfese DL, Brenden RA. Behavioral and electrophysiological indices of voicing-cue discrimination: laterality patterns and development. *Brain and Language* 1997;57:122–50. [PubMed: 9126410]
- Spellacy F, Blumstein S. The influence of language set on ear preference in phoneme recognition. *Cortex* 1970;6:430–9. [PubMed: 5514515]
- Studdert-Kennedy M, Shankweiler D. Hemispheric specialization for speech perception. *Journal of the Acoustic Society of America* 1970;48:579–94.
- Talairach, J.; Tournoux, P. A co-planar stereotaxic atlas of a human brain. Thieme; Stuttgart: 1988.
- Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* 2001;11:46–953.
- Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Neuroscience* 2002;6:37–47.

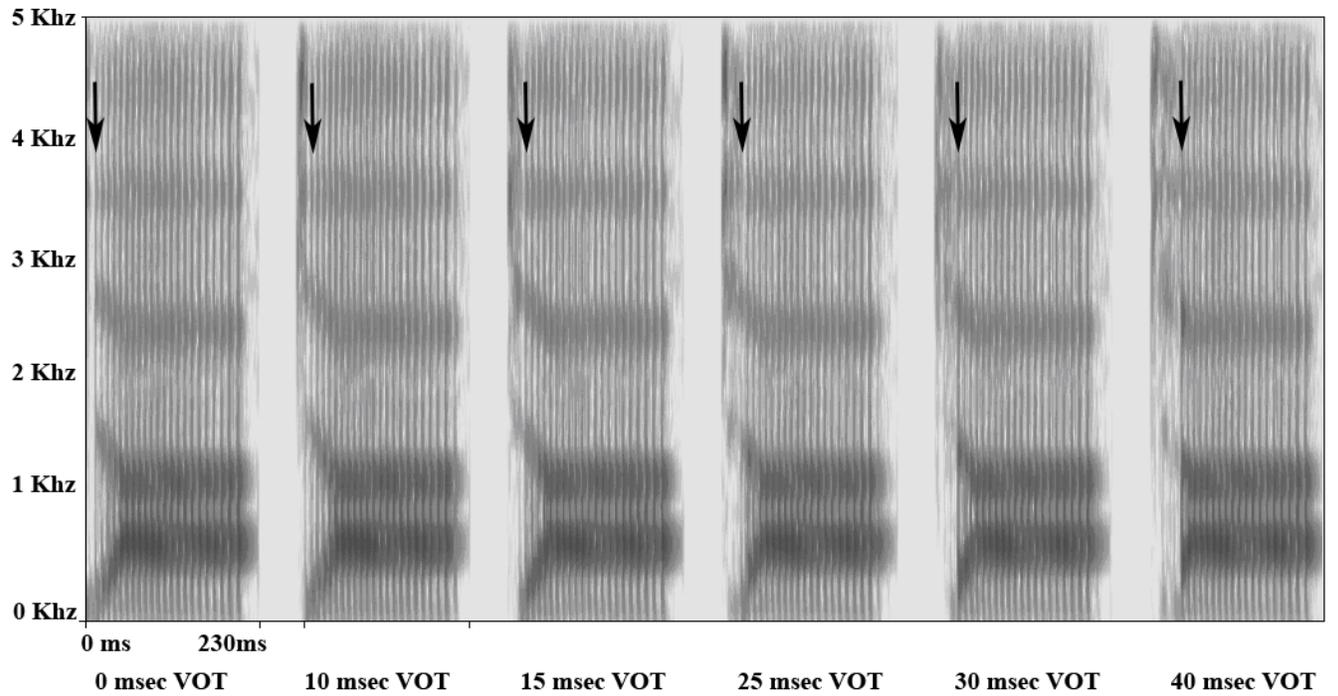
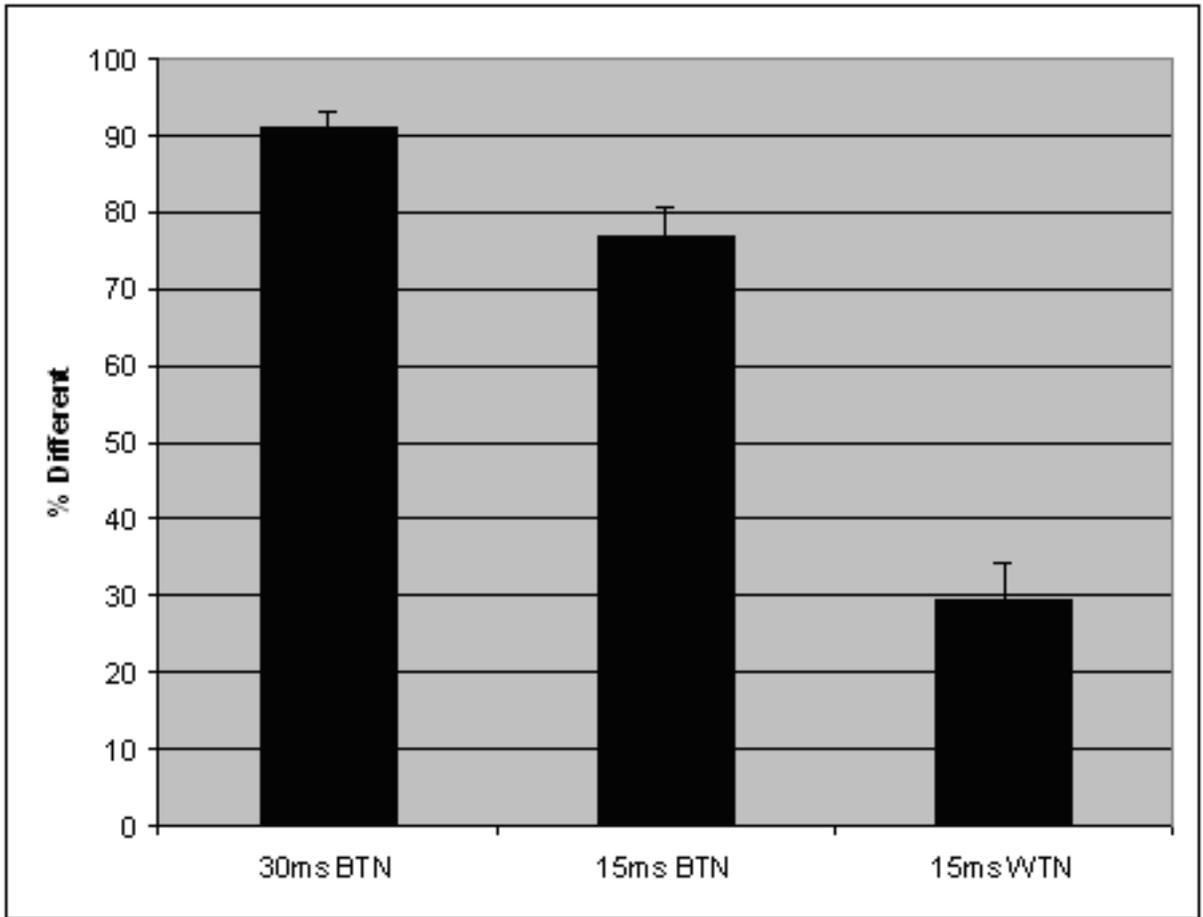


Figure 1. Spectrograms of the six synthetic syllable stimuli, ranging from a VOT of 0ms (left) to 40ms (right). VOT values used were 0ms, 10ms, 15ms, 25ms, 30ms and 40ms.



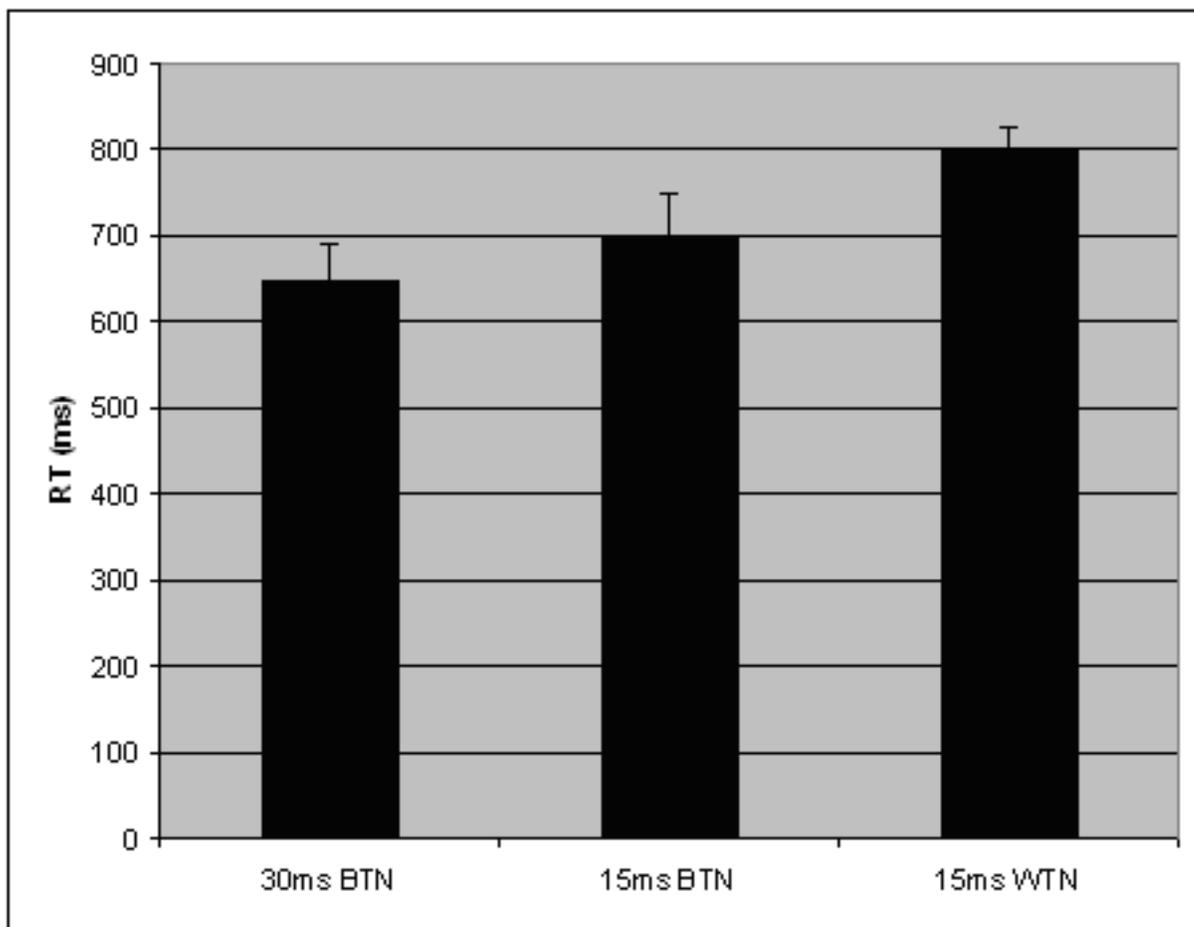


Figure II.

The top panel shows the mean % correct ('different') responses on the three syllable conditions, 30 ms Between Phonetic Category, 15 ms Between Phonetic Category, and 15 ms Within Phonetic Category, for all 14 participants. Error bars represent standard error. The bottom panel shows the mean RT (in ms) for correct different responses for all 14 participants. Error bars represent standard error.

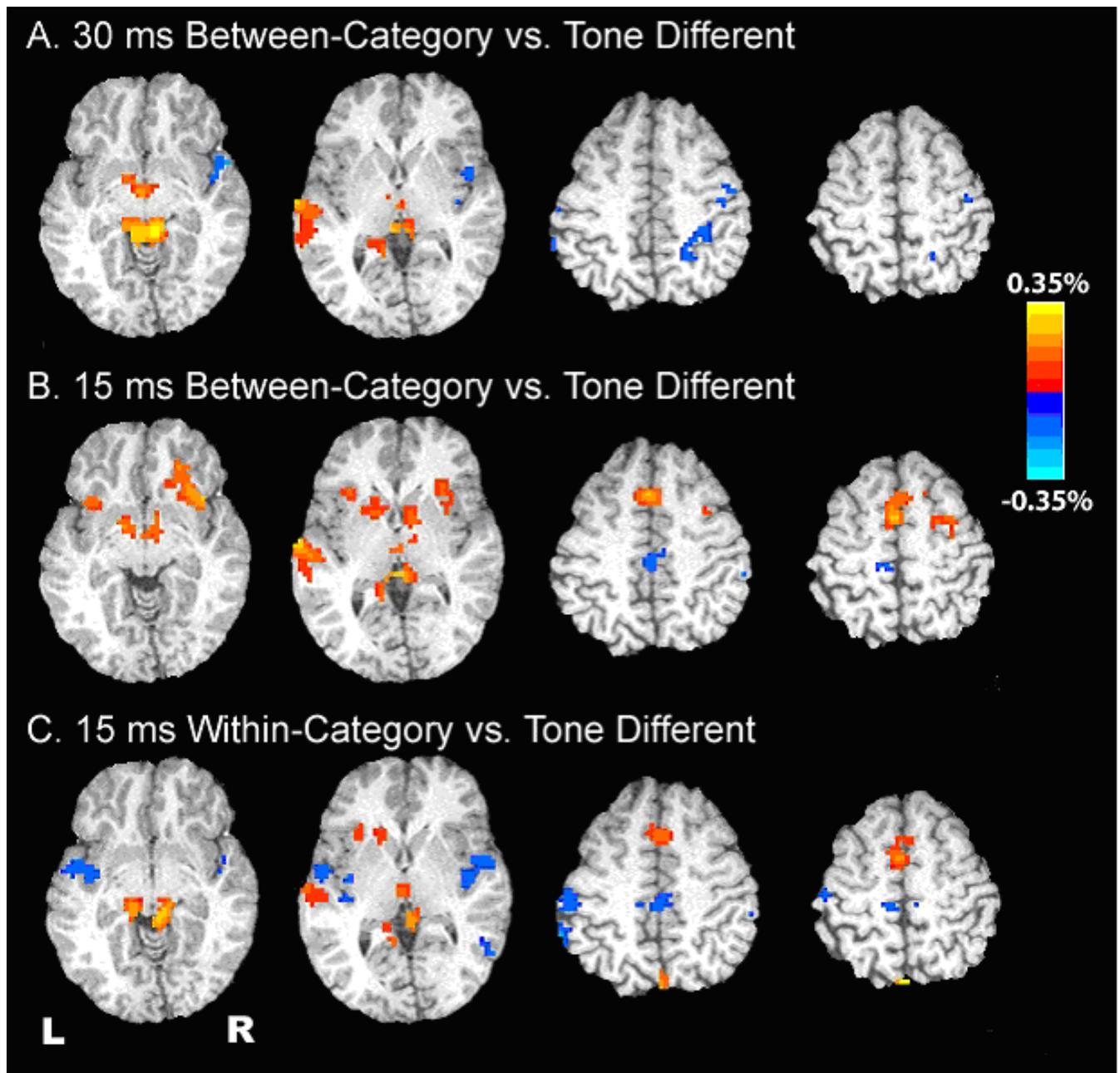
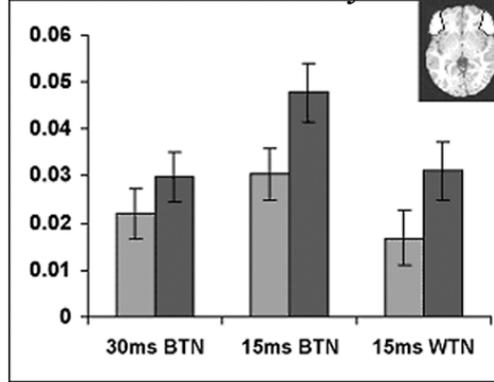


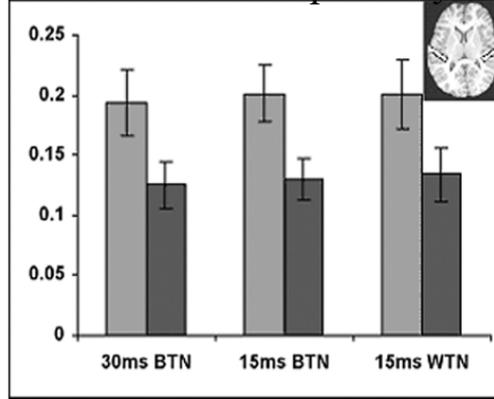
Figure III.

A and B. Regions showing significant differences in activation level at a voxel level threshold of $p < 0.01$ and a corrected threshold of $p < 0.01$ (44 contiguous voxels) for the three VOT discrimination conditions compared to the tone different condition. Axial views at $z=60$, $z=69$, $z=107$ and $z=114$ (from left to right) are shown. The top panel shows the 30ms Between-Category vs. Tone Different, the middle panel the 15ms Between-Category vs. Tone Different and the bottom panel the 15ms Within-Category vs. Tone Different. Images are scaled from a percent change range of -0.35 to 0.35.

A. Inferior Frontal Gyrus



B. Transverse Temporal Gyrus



C. Superior Temporal Gyrus

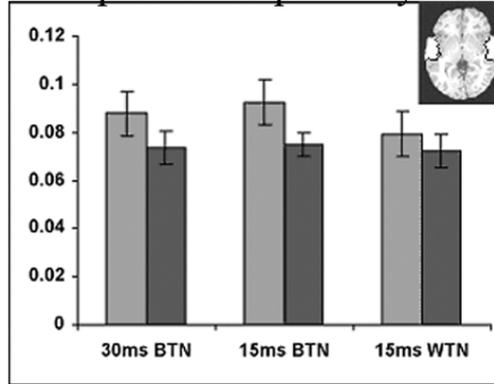


Figure IV.

Mean percent signal change for the three syllable different conditions is displayed for anatomically defined ROIs in the left and right hemisphere where significant differences emerged. The anatomically defined ROI from AFNI's implementation of the Talairach and Tournoux atlas is displayed in the upper right hand corner of each graph.

Table 1

Syllables versus Tones

Areas of significant activation in the comparison between all syllable stimuli and all tone stimuli at a voxel-level threshold of $p < 0.01$ and a corrected threshold of $p < 0.01$. Coordinates displayed are in the standardized Talairach and Tournoux template and indicate the center of mass for each cluster

Region (CM)	CM x	CM y	CM z	Volume (µL)	Max % Change
Syllables > Tones					
L. Caudate	-5.9	0.6	17	47250	0.1879
L. STG	-58.2	-19.6	6.6	4887	0.2144
R. IFG	28.5	22	-1.8	4428	0.1908
R. Parahippocampal Gyrus	29.7	-45.1	5.3	2511	0.0729
R. Precuneus	4.7	-78	37.8	1485	0.26
Tones > Syllables					
R. Inferior Parietal Lobule, R. BA40	51	-27.2	23.6	22491	0.1935
L. Supramarginal Gyrus, L. BA40	-56	-39.4	30.5	7884	0.1297
L. Precuneus	-0.3	-31.7	43.5	3105	0.0821
L. Insula	-40.4	-4.4	5.4	2808	0.0868
R. Middle Temporal Gyrus	40.5	-57.2	13.4	1593	0.0962
L. Postcentral Gyrus	-52.5	-10.2	23.9	1242	0.0704

Table II

VOT Category conditions versus Tones

Areas of significant activation in the comparisons between the three syllable different conditions and the tone different condition. A voxel-level threshold of $p < 0.01$ and a corrected threshold of $p < 0.01$ were used. Coordinates displayed are in the standardized Talairach and Tournoux template and indicate the center of mass for each cluster

Region (CM)	CM x	CM y	CM z	Volume (μL)	Max % Change
30ms Between > Tone					
N/A (subcortical)	-1.1	-33.1	2.5	5184	0.2772
L. STG	-59.1	-21.7	5.4	2268	0.2246
L. Thalamus	-3.2	-6.8	-1.7	1458	0.1661
Tones > 30ms Between					
R. Inferior Parietal Lobule	56.6	-38.3	21.8	7803	0.193
R. IFG	48.4	-1.3	20.1	5346	0.1526
L. Inferior Parietal Lobule, L. BA40	57.5	-42.5	27.5	3132	0.1702
R. Precuneus	27.4	-41.7	41.3	2025	0.0908
L. Postcentral Gyrus	-52.8	-14.8	29	1701	0.1127
15ms Between > Tones					
L. Caudate & L. Caudate Head	-1.2	2.4	4.1	5859	0.1781
L. Medial Frontal Gyrus, L. BA6	-0.9	14	45.4	4833	0.1892
R. IFG	28.7	22.8	-1.7	4077	0.2549
L. Insula, L. BA13	-34.9	18.1	6.5	2943	0.2618
L. Thalamus	-0.4	-29.8	6	2727	0.1971
L. STG	-58.6	-18.9	6.5	2646	0.2529
R. Middle Frontal Gyrus, R. BA6	31.8	4	49.1	1269	0.1074
Tones > 15ms Between					
R. Inferior Parietal Lobule, R. BA40	56.6	-33.3	24	5130	0.2429
R. Middle Temporal Gyrus, R. BA22	40.4	-56.9	13.9	1620	0.1327
L. Paracentral Lobule, L. BA31	-2.5	-23.6	44.1	1188	0.1138
15ms Within > Tones					
L. Cingulate Gyrus, L. BA6	0.4	14.8	41.8	5481	0.1838
R. Thalamus	4.6	-34.5	5.1	3402	0.2537
L. Thalamus	-1.9	-3.8	11.5	2646	0.1481
L. Claustrum	-21.6	22.4	8.5	1512	0.0845
L. Parahippocampal Gyrus	-11.4	-34.3	0.4	1458	0.2164
L. STG	-58.7	-18.4	6.9	1431	0.2571
R. Precuneus, R. BA7	2.6	-75.1	40.7	1215	0.3347
Tones > 15ms Within					
R. Inferior Parietal Lobule	55.4	-39	22	10233	0.2794
L. Supramarginal Gyrus, L. BA40	-57.3	-37.6	31.5	8127	0.1944
L. Postcentral Gyrus	-49.1	-12.5	21.5	2457	0.1088
L. Insula	-44.7	-4.7	0.9	2376	0.1404
R. Insula	45.6	-2.6	4.8	2025	0.1504
L. Paracentral Lobule, L. BA31	-1.8	-25.4	44.9	1539	0.1038

Table III

Syllable Condition Comparisons

Areas of significant activation in pair-wise comparisons between the three syllable different conditions at a voxel-level threshold of $p < 0.01$ and a corrected threshold of $p < 0.01$. Coordinates displayed are in the standardized Talairach and Tournoux template and indicate the center of mass for each cluster

Region (CM)	CM x	CM y	CM z	Volume (μ L)	Max % Change
15ms Between > 30 ms Between					
R. Inferior Frontal Gyrus	45.9	12.5	23	2160	0.1382
R. Inferior Parietal Lobule	42.5	-48.2	40	1566	0.1331

Table IV

Results of Condition vs. RT regression analysis. Results in the table are using a voxel level threshold of $p < 0.01$ and a corrected threshold of $p < 0.01$ (44 contiguous voxels)

Region (CM)	CM x	CM y	CM z	Volume
R. Cingulate Gyrus	2.3	-20.1	39.9	1296
L. Precuneus, L. BA31	-3.1	-46.4	30.3	1242
L. Anterior Cingulate, L. BA24	-6.1	38	5.3	1215