



Published in final edited form as:

Neuropsychologia. 2009 March ; 47(4): 1096–1106. doi:10.1016/j.neuropsychologia.2008.12.033.

The Role of Spectral and Durational Properties on Hemispheric Asymmetries in Vowel Perception

Brendan Britton, Sheila E. Blumstein, Emily B. Myers, and Christopher Grindrod¹
Brown University

Abstract

The aim of the current study is to investigate potential hemispheric asymmetries in the perception of vowels and the influence of different time scales on such asymmetries. Activation patterns for naturally produced vowels were examined at three durations encompassing a short (75 ms), medium (150 ms), and long (300 ms) integration time window in a discrimination task. A set of 5 corresponding non-speech sine wave tones were created with frequencies matching the second formant of each vowel. Consistent with earlier hypotheses, there was a right hemisphere preference in the superior temporal gyrus for the processing of spectral information for both vowel and tone stimuli. However, observed laterality differences for vowels and tones were a function of heightened right hemisphere sensitivity to long integration windows, whereas the left hemisphere showed sensitivity to both long and short integration windows. Although there were a number of similarities in the processing of vowels and tones, differences also emerged suggesting that even fairly early in the processing stream at the level of the STG, different mechanisms are recruited for processing vowels and tones.

INTRODUCTION

Understanding the neural basis of speech perception is essential for mapping out the neural systems underlying language processing. An important outstanding area of this research focuses on the functional role the two hemispheres play in decoding the speech signal. Recent neuroimaging experiments suggest a hierarchical organization of the phonetic processing stream with early auditory analysis of the speech signal occurring bilaterally in Heschl's gyri and the superior temporal lobes and later stages of phonetic processing occurring in the left middle and anterior STG and STS of the left, dominant language hemisphere (Scott et al. 2000; Liebenthal et al. 2005).

Speech sounds themselves, however, are not indissoluble wholes; they are comprised of a set of acoustic properties or phonetic features. For example, the perception of place of articulation in stop consonants requires the extraction of rapid spectral changes in the 10s of ms at the release of the consonant. The perception of voicing in stop consonants requires the extraction of a number of acoustic properties (Lisker, 1978), among them, voice-onset time,

Correspondence address: Prof. Sheila Blumstein, Department of Cognitive and Linguistic Sciences, Brown University, 190 Thayer Street, Box 1978, Providence, RI 02912, Sheila_Blumstein@brown.edu, Tel. 401-863-2849, Fax 401-96[0-9]x[0-9]255.

Reprint requests and correspondence should be sent to Sheila Blumstein, Department of Cognitive and Linguistic Sciences, Brown University, 190 Thayer St., Box 1978, Providence, RI 02912; Sheila_Blumstein@brown.edu.

¹now at University of Illinois, Urbana Champaign

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.

corresponding to the timing between the release of the stop consonant and the onset of vocal cord vibration. The perception of vowel quality requires the extraction of quasi-steady-state spectral properties associated with the resonant properties of the vocal tract. What is less clear are the neural substrates underlying the mapping of the different acoustic properties or features that give rise to these speech sounds.

Several studies have explored potential differences in lateralization for spectral and temporal properties using synthetic nonspeech stimuli with spectral and temporal properties similar to speech (Hall et al. 2002; Jamison et al. 2005; Boemio et al. 2005) and have found differential effects of both parameters on hemispheric processing. Results suggest that at early stages of auditory processing, the functional role of the two hemispheres may differ (Ivry and Robertson 1998; Poeppel 2001; Zatorre et al. 2002). In particular, although both hemispheres process both spectral and temporal information, there is differential sensitivity to this information. It has been shown in fMRI studies (Zatorre and Belin 2001; Boemio et al. 2005) as well as studies of aphasia (Van Lancker and Sidtis, 1992) that the right hemisphere has a preference for encoding pitch or spectral change information and it does so most efficiently over long integration time windows, whereas the left hemisphere has a preference for encoding spectral information and particularly for integrating rapid spectral changes. Moreover, it has been shown that this preference may be modulated by task demands (Brechmann, 2005). Evidence from intracerebral evoked potentials further suggests that processing of fine-grained durational properties of both speech and non-speech stimuli is localized to the left Heschl's Gyrus (HG) and planum temporale (Liegeois-Chauvel, 1999). Other evidence suggests that this division may not be so clear-cut. For instance, in an fMRI study of non-speech stimuli, bilateral HG responded to spectral variation, whereas in the STG temporal information was left-lateralized, and spectral properties were right-lateralized (Schonwiesner, 2005).

Poeppel's Asymmetrical Sampling in Time theory (AST) (2003) hypothesizes that laterality effects in processing auditory stimuli arise due to different temporal specializations of the two hemispheres. In particular, the AST proposes that the two hemispheres have different temporal integration windows with the left hemisphere preferentially extracting information from short (25–50 ms) windows and the right hemisphere preferentially extracting information from long (150–300 ms) windows. To test this hypothesis, Poeppel and colleagues (Boemio et al. 2005) designed an fMRI study in which subjects listened to narrow-band noise segments with temporal windows ranging from 25ms to 300ms. They found that relative to a constant frequency tone, the temporal noise segments produced increasing right hemisphere activation as they increased both the overall duration and the number of different pitch variations. However, their results only partially support the AST model. While an increase in right hemisphere activity was observed with increasing duration, no corresponding increase in left hemisphere activity was found for short durations.

Two questions are raised by these studies. The first is whether the acoustic properties of speech are processed by a domain-general spectro-temporal mechanism (for discussion see Zatorre and Gandour, 2007). If this is the case, then a speech property that favored spectral analysis over a relatively long time domain should recruit right hemisphere mechanisms. The second is whether manipulating the time window over which spectral information is integrated would affect hemispheric preferences.

The previous studies described above demonstrated right hemisphere preference using non-speech stimuli varying in their long-term spectral properties. What is less clear is whether similar right hemisphere preferences would emerge for vowels which require extraction and integration of spectral properties over a particular time window and the potential effects of the size of that window on hemispheric laterality. Given the hypothesized differences in the computational processes of the two hemispheres, there are two properties of vowels that are

likely to recruit right hemisphere processing mechanisms. The perception of vowel quality requires the extraction of spectral properties over a relatively long steady-state. For example, differences in vowel quality such as [i] as in *beet* and [u] as in *boot* are determined by the first two formant frequencies. Moreover, the spectral properties of vowels tend to be fairly stable over 150 msec or more, and should therefore have a 'long' integration window. Thus, presumably at early stages of processing, there should be a right hemisphere preference for the processing of vowels.

Nonetheless, the hypotheses that there are functional differences in the processing of speech sounds in the two hemispheres are largely based on the results of neuroimaging studies using nonspeech stimuli (Zatorre and Belin 2001; Boemio et al. 2005; Jamison et al. 2005). While it is assumed that at early stages of processing both nonspeech and speech engage similar computational mechanisms (Uppenkamp et al. 2006), the question remains as to what effect the presence of a linguistic stimulus may have on hemispheric laterality (Narain et al. 2003).

Indeed, early literature exploring the hemispheric processing of auditory stimuli using the dichotic listening technique showed hemispheric differences as a function of the type of stimulus. Linguistic stimuli such as numbers, words, nonsense syllables and even consonants showed a right ear/left hemisphere advantage, whereas non-linguistic stimuli such as music and sound effects showed a left ear/right hemisphere advantage (Kimura, 1961, 1964; Spellacy and Blumstein, 1970; Studdert-Kennedy and Shankweiler, 1970). Similar dichotomies in the hemispheric processing of linguistic and nonlinguistic stimuli have also been shown in patients with left and right temporal lobectomies, with deficits in processing speech sounds with left temporal lobectomies and deficits in processing pure tones and other aspects of musical processing with right temporal lobectomies (Milner, 1962). Interestingly, results for vowel stimuli were mixed; a small right hemisphere advantage was shown in some experiments (Shankweiler and Studdert-Kennedy, 1967), a left hemisphere advantage in others (Godfrey, 1974; Weiss and House, 1973); and no ear advantage in still others (Spellacy and Blumstein, 1970).

A few functional neuroimaging studies have been conducted using real and synthetic vowel stimuli, but none have analyzed directly potential hemispheric asymmetries for the processing of vowels. One study investigated vowel processing and showed bilateral activation for both vowels and nonspeech control stimuli of equal complexity with greater activation for vowels than for the nonspeech stimuli (Uppenkamp et al. 2006). Obleser et al. (2006) also showed bilateral activation for vowels in a study investigating the topographical mapping of vowel features associated with differences in formant frequencies, and hence spectral, patterns of the stimuli. Their results suggested stronger right hemisphere activation for the processing of these vowel features, although laterality differences were not tested statistically (cf. also Guenther et al., 2004).

The aim of the current study is to investigate potential hemispheric asymmetries in the perception of vowel quality and the influence of different time scales on such asymmetries. To this end, activation patterns for naturally produced vowels, which have a quasi-steady-state, constant spectral formant pattern will be examined at three time scales or durations, encompassing a short (75 ms), medium (150 ms), and long (300 ms) integration time window. If, as discussed above, there is a right hemisphere domain-general mechanism for extracting relatively steady-state spectral properties of auditory stimuli, then a right hemisphere preference for the processing of vowels will emerge. However, the magnitude of the asymmetry should be influenced by vowel duration with an increased right hemisphere preference for long vowels. Models such as Poeppel's AST also predict an increased left hemisphere preference for short vowels. Nonetheless, given the results of Boemio et al. (2003), it is not clear whether there will be increased left hemisphere activation for vowels at shorter (75ms) timescales. The

locus of this asymmetry for vowels should emerge in the STG and STS, and, in particular, in the anterior STG and STS reflecting the recruitment of the auditory ‘what’ processing stream relating to the perception of ‘auditory objects’ or speech sounds (Obleser et al. 2006; Hickock and Poeppel 2000).

A discrimination task with a short (50 ms) interstimulus interval (ISI) will be utilized. There are two reasons for using this paradigm. First, discrimination judgments will be based on the perception of differences in the spectral properties (i.e. the formant frequencies) of the vowel stimuli. Second, it is generally assumed that a discrimination task taps early stages of phonetic processing since it focuses attention on potential differences in the acoustic properties of the stimuli, rather than on the phonetic category membership of the stimuli (Lieberman et al. 1957).

A tone discrimination task will be used as a control task. A tone discrimination task has been used as a control condition in other studies investigating the perception of speech (Burton et al., 2000; Jancke et al., 2002; Sebastian and Yasin, 2008). Although the discrimination of tones reflects perception of differences in pitch and the discrimination of vowels reflects perception of differences in vowel quality, both share the acoustic property of periodicity with tones being fully periodic and vowels being quasi-periodic. Importantly, tone discrimination has shown right hemisphere lateralization (Binder et al., 1997). The tone stimuli will be single frequency sine wave tones which share the second formant frequency and the duration parameters (75, 150, 300 ms) of the vowel stimuli. Given the hypotheses described above that there is a right hemisphere domain-general mechanism for extracting steady-state properties of auditory stimuli, simple sine wave tones should show similar lateralization patterns to those of vowels. In general, there should be right hemisphere lateralization for the discrimination of pitch contrasts between the tone stimuli (cf. also Milner, 1962; Binder et al. 1997), and the patterns of asymmetry as a function of duration should mirror those for vowels with an increased right hemisphere preference for long duration tones. When contrasted with the tone stimuli, however, there may be less right hemisphere activation due to the linguistically relevant properties of the vowel.

MATERIALS & METHODS

Subjects

Fifteen healthy volunteers (11 females, 4 males), ages 18–54 (mean = 23±9 yrs), participated in the study. All subjects were native English speakers and right-handed according to the Edinburgh Handedness Inventory (Oldfield 1971). Participants gave written consent prior to participation in accordance with guidelines established by the Human Subjects Committee of Brown University and Memorial Hospital. Each participant received moderate monetary compensation for their time. In addition to the fMRI experiment, subjects also participated in a behavioral pretest to ensure that subjects could accurately perform the experimental task.

Stimuli

Stimuli consisted of pairs of vowels and tones, selected from a 5 vowel or 5 tone set. The five-vowel stimulus set included [i], [e], [a], [o], and [u], and was produced by an adult male speaker native of English. The formant frequencies of the vowel stimuli are listed in Table 1. The durations of the produced vowels varied between 284–336 ms. They were then all adjusted to 300 ms by means of a frequency-independent time stretching program (Adobe Audition), and then edited to 75ms and 150ms in duration using the BLISS audio software package (Mertus 1989). Stimulus onsets and offsets were tapered with 10ms Hamming windows to prevent acoustic transients.

A set of 5 corresponding sine wave tones were created using the NCH Tone Generator (<http://www.nch.com.au/tonegen/>) with frequencies matching the second formant of each vowel. Similar to the vowel stimuli, the tone stimuli were of three durations (75, 150, and 300 ms) and were tapered at stimulus onset and offset using a 10ms Hamming window. The resulting stimulus set consisted of 15 speech and 15 corresponding non-speech stimuli, with 5 vowels and 5 tones in each of the 3 duration conditions.

Discrimination pairs matched in duration were created for the vowel and tone test stimuli. The stimuli within each discrimination pair were separated by an interstimulus interval (ISI) of 50 ms. Within each of the three (75, 150, 300 ms) duration conditions, there were 30 same trials in which each vowel or tone stimulus was paired with itself and 30 different trials. For the different trials, all possible combinations of stimuli occurred twice within each condition, with ten of the stimulus pairs occurring a second time in reverse order (i.e. [i-e] and [e-i]). In total, there were 360 discrimination pairs, two runs of 90 vowel pairs and two runs of 90 tone pairs.

Procedure

The experiment consisted of two runs for the vowel discrimination task and two runs for the tone discrimination task. Participants received the four runs in a fixed order (Vowel, Tone, Vowel, Tone). Each run consisted of 15 discrimination pairs in each of the stimulus conditions (15 pairs \times 3 duration conditions \times 2 response conditions, totaling 90 stimuli per run. (See *Imaging* for details of stimulus timing). Stimuli within each run occurred in a fixed, pseudo-randomized order.

Stimuli were presented over MR-compatible headphones (Resonance Technology, Inc., Northridge, California) and the sound level was adjusted to ensure a comfortable listening level. Participants performed an AX discrimination task in which they were required to determine whether the two stimuli in the pair were the same or different by pressing one of two buttons on an MR-compatible button box using their right hand. The button mapping was counterbalanced across subjects. Responses were scored for both accuracy and reaction time (RT), with RT latencies measured from the onset of the second stimulus. Presentation of stimuli and collection of response data were controlled by a laptop running the BLISS software suite (Mertus 1989).

Imaging

Scanning was performed on a 1.5T Symphony Magnetom MR system (Siemens Medical Systems; Erlangen, Germany) at Memorial Hospital in Pawtucket, RI. Subjects were aligned to magnetic field center and instructed to remain motionless with eyes closed during the duration of the experiment.

For each subject, an anatomical T1-weighted rapid acquisition gradient-echo (MPRAGE) sequence (TR=1900ms, TE=4.15ms, TI=1100ms, 1mm³ isotropic voxel size, 256 \times 256 matrix) reconstructed into 160 slices was acquired for co-registration with functional data. The functional scans consisted of interleaved multislice echo planar (EPI) sequences of 15 axial slices centered on the corpus callosum. Each slice was 5mm in thickness, with 3 mm² FOV, flip angle=90°, TE=38ms, TR=2000ms. The imaging data were acquired so that the auditory stimuli were presented during silent gaps (Belin et al. 1999; Hall et al. 1999). In particular, 15 slices were acquired in a 1200 ms interval (80 ms per slice) followed by a 800 ms silent gap in which a discrimination pair was presented. This yielded an effective volume repetition time of 2000 ms (see Figure 1). Both vowel and tone runs contained 274 volumes per run.

Four runs of alternating vowel and tone pairs were presented in an event-related design. Stimuli were jittered to whole-multiple intervals of the 2000 ms repetition time and presented at equally

probable combinations of trial onset asynchronies (TOA=2,4,6,8,10s). To reduce T1 saturation of functional images, four functional dummy volumes were collected before the presentation of the first stimulus pair in each run, and these volumes were excluded from subsequent analysis.

Analysis

Behavioral Data—The behavioral data were scored for both performance and reaction-time (RT). Incorrect responses and responses that were greater than 2 standard deviations from a subject's mean for that condition were excluded from the analysis. A 3-way repeated measures analysis of variance (ANOVA) was conducted comparing Duration (75ms/150ms/300ms), Condition (Vowel/Tone), and Response (Same/Different).

Because the focus of this study was on sensitivity to the perception of *changes* in vowel quality or *changes* in pitch in tones as a function of duration, only the Different responses were considered in the functional imaging analysis. Moreover, as discussed below, there was no difference in overall performance between vowels and tones in the Different responses, whereas a behavioral difference emerged for vowels and tones for Same responses. Thus, focusing on Different responses assured that variations in activation patterns between vowels and tones would reflect differences in perceptual processing and not in difficulty of processing.

MR Data—Image processing and statistical analysis of MR imaging data was performed using the AFNI software platform (Cox and Hyde 1997). Slice timing correction was applied to the EPI images to correct for interleaved slice acquisition. All functional volumes were aligned to the fourth collected volume by means of a rigid body transform to correct for head motion. Next, functional voxels were resampled into a 3mm³ (originally 3×3×5mm) isotropic voxel space, co-registered to the anatomical MPRAGE images and then transformed into Talairach and Tournoux (1998) coordinate space. Finally, a 6mm FWHM Gaussian blur was finally applied to the functional voxels for spatial smoothing, and subject masks were created which excluded voxels outside of subject brain space.

Deconvolution analyses were performed on individual subject EPI data to model the hemodynamic activation as a function of the twelve experimental conditions (3 length conditions × Same/Different × Vowel/Tone). Canonical gamma functions were convolved with stimuli onset times to create ideal time-series waveforms used as covariates in the deconvolution. The six output parameters from the motion correction algorithm (x, y, and z translation, roll, pitch and yaw) were also included as covariates of no interest. The resulting correlation fit coefficients showed the condition-dependent activation in each voxel during the course of the experiment. The raw fit coefficients were finally converted to percent signal change by dividing each voxel by its estimated mean baseline across the entire experiment. Statistical maps were clipped to exclude any voxel which was not imaged in each of the 15 subjects.

A mixed-effects 2-way ANOVA was computed on the results of the deconvolution analyses with stimulus condition as a fixed effect and subject as a random effect. In addition to condition means, a number of contrasts were calculated including Vowels vs. Tones across the three durations, Vowels vs. Tones at each duration condition (75, 150, 300ms), and within each of the Vowels and Tone conditions comparisons of duration (e.g. Vowel 75 vs. Vowel 300, Vowel 75 vs. Vowel 150, Vowel 150 vs. Vowel 300). The resulting group statistical map showed the activation of each condition and contrast for the entire experiment.

To correct for Type-I multiple comparison errors, Monte Carlo simulations were run to determine a minimum reliable cluster size at a voxel threshold of $p < 0.05$. Using a cluster

threshold of $p < 0.05$, the minimum cluster size was determined to be 131 contiguous voxels of 3mm^3 .

In addition to the statistical comparisons described above, Region of Interest (ROI) analyses were performed on the Superior Temporal Gyrus (STG) where significant clusters emerged. All areas were subdivided into left and right hemisphere components to investigate laterality effects. The anatomical boundaries for these regions were determined by the AFNI Talairach Daemon (Lancaster et al. 2000) and visually verified using a standard brain atlas (Duvernoy 1999), and included both the STG as well as the dorsal bank of the superior temporal sulcus (STS). Previous studies have suggested different roles for anterior and posterior regions of the STG in auditory and speech processing shown in a number of studies (Hickok and Poeppel 2004; Scott and Wise 2004; Liebenthal et al. 2005; Obleser et al. 2006). In keeping with these findings, the STG was divided into anterior and posterior regions along the y-axis to perform ROI analyses within these regions independently. The y-axis was divided at the STG midpoint by a vertical (x-z) plane at $y = -18$. The resulting anterior and posterior STG regions in the left hemisphere contained 258 voxels and 555 voxels, respectively, and in the right hemisphere 233 voxels (anterior) and 458 voxels (posterior) (see Figure 5, inset). The imbalance in voxel numbers between anterior and posterior regions was a result of anterior temporal lobe signal dropout in the imaging procedure, rather than a result of the choice of dividing boundary.

Finally, repeated measures ANOVAs were performed on condition and subject means for each ROI, with condition categories of left/right, vowel/tone, and duration (75ms 150ms 300ms).

RESULTS

Behavioral

Figure 2 shows the performance and Figure 3 shows the RT results for the behavioral data. As Figure 2 shows, overall performance was above 90% for both vowels and tones. “Different” and “same” performance responses were submitted to separate 2-way ANOVAs with factors of duration and stimulus type (see Farrell, 1985 for discussion about potential differences in processing mechanisms for same and different responses). For “different” responses, only a main effect of duration ($F(2, 28) = 7.905, p < .002$) was observed, showing an increase of performance with increasing duration. The “same” responses showed an effect of condition ($F(1, 14) = 5.094, p < .041$) only, with an overall better performance for vowels than for tones. Thus, subjects did better in discriminating the vowel stimuli than the tone stimuli for “same” pairs, but for “different” pairs the performance was the same for the two types of stimuli. For the reaction-time data, two 2-way ANOVAs (Duration \times Stimulus Type) were separately conducted for the different and same responses. For the different discrimination pairs, there was a main effect for duration ($F(2, 28) = 30.840, p < .001$) with decreasing RT latencies as stimulus duration increased. There was neither a main effect for stimulus type (vowels vs. tones) nor was there a significant interaction. Analysis of the behavioral data for the same discrimination pairs revealed a main effect for duration ($F(2, 28) = 23.869, p < .001$) (similar to the different discrimination pairs, RT latencies decreased as stimulus duration increased), a main effect for stimulus type ($F(1, 14) = 5.580, p < .033$) (vowels had faster RT latencies than did tones), and no interactions.

fMRI

Activated clusters ($p < 0.05$, corrected) for statistical contrasts of vowel and tone stimuli are shown in Table 2. Maximum intensity locations of each cluster are shown in Talairach and Tournoux coordinates; however, several clusters extended into multiple regions.

Vowels versus Tones—Comparison of activation for vowels vs. tones overall yielded greater activation for vowels versus tones in the temporal lobes bilaterally. Maximum intensity foci occurred in anterior regions of the left and right Superior Temporal Gyrus (STG), with the left hemisphere showing the largest activation. In addition, as Table 1 shows, cluster size in the STG was larger on the left (425 voxels) than on the right (292 voxels).

Nonetheless, the pattern of activation for vowels compared to tones differed as a function of duration (see Table 2 and Fig. 4). As Fig. 4 shows, there was extensive temporal lobe activation with greater activation for vowels compared to tones only in the left hemisphere for short durations (75 ms) and bilateral activation at the longer durations (150 ms and 300 ms). In particular, at 75 ms durations, contrasts of vowels versus tone stimuli showed one activated cluster in the L STG (see Table 2). At intermediate durations (150ms), clusters appeared in the L STG, and R Middle Temporal Gyrus (MTG) extending into the R STG, and the R IPL. At the longest duration (300ms), contrasts showed greater activation for vowels bilaterally in the STG. Although there was bilateral activation at the intermediate and long durations, it is worth noting that in the comparison of vowels and tones the cluster size was larger on the left than on the right.

The cluster analyses showed, as expected, extensive temporal lobe activation in the discrimination of the spectral patterns of vowels and tones. Since the goal of this paper was to examine potential hemispheric asymmetries and the effects of duration on those asymmetries, the remainder of our analyses will focus on the patterns of activation in the STG. Further cluster analyses showing the results of pair-wise comparisons within the vowel stimuli and within the tone stimuli as a function of duration are shown in the Appendix.

Region of Interest Analysis—In order to examine in more detail potential laterality differences as well as the influences of duration on laterality, a series of region of interest (ROI) analyses were conducted focusing on the anterior and posterior portions of the STG bilaterally. To this end, an anatomical Region of Interest (ROI) analysis was computed on the percent signal change for four regions in the left and right anterior and posterior STG for both vowels and tones over the 3 stimulus durations for each of the 15 subjects. Factors included hemisphere (L/R), site (anterior/posterior), condition (vowel/tone), and duration (75ms/150ms/300ms). The results were submitted to a 4-way ANOVA which resulted in main effects of hemisphere ($F(1, 14)=9.43$, $p<0.009$), condition ($F(1, 14)=13.1$, $p<0.003$), and duration ($F(2, 28)=19.4$, $p<0.001$), and a hemisphere \times duration interaction ($F(2, 28)=5.93$, $p<0.007$). In addition, two further interactions approached significance: hemisphere \times site ($F(1, 14)=4.00$, $p<0.065$) and a 4-way hemisphere \times site \times condition \times duration ($F(2, 28)=2.82$, $p<0.078$) interaction.¹ As discussed earlier, results of recent neuroimaging experiments suggest different functional roles for the aSTG and pSTG (Scott and Wise 2004; Hickock and Poeppel 2004). The fact that the interaction effects approached significance by site is consistent with these findings. For these reasons, the remaining analyses considered the activation patterns for the aSTG and pSTG separately.

The results of the ROI analyses are shown in Fig. 5 for the anterior Superior Temporal Gyrus (aSTG, top panel) and the posterior Superior Temporal Gyrus (pSTG, bottom panel). Several observations can be made across these two sites. First, there is greater activation in the R STG than the L STG for both vowels and tones, although the magnitude of the laterality difference appears to be greater for vowels than for tones, particularly in the aSTG. Second, there is

¹In order to better localize voxels which show an interaction between effects of duration (75 msec, 150 msec, and 300 msec) and stimulus type (vowel versus tone), a 2-way ANOVA was performed on the percent signal change values for different stimulus pairs. A single cluster emerged in the precuneus and posterior cingulate. No clusters emerged which showed a duration by type interaction in the STG.

increased activation for vowels as duration of the stimuli increases. This effect appears to be less robust for tones.

To examine the effects statistically, two 3-way ANOVAs were conducted, one for the aSTG and the other for the pSTG with factors of hemisphere (left/right), condition (vowel/tone), and duration (75ms, 150ms, 300ms). For the aSTG, there were main effects of hemisphere ($F(1, 14)=10.8, p<0.005$) (the RH showed more activation than the LH), condition ($F(1, 14)=9.71, p<0.008$) (there was greater activation for vowels than tones), and duration ($F(2, 28)=23.0, p<0.001$) (activation increased as the stimuli were longer). There was also a hemisphere \times condition interaction ($F(1, 14)=5.13, p<0.04$) and a hemisphere \times duration ($F(1, 14)=6.78, p<0.004$) interaction. The hemisphere \times condition interaction was due to a greater magnitude in the RH preference for vowels compared to tones. The hemisphere \times duration interaction showed that the RH lateralization increased as stimulus duration increased. A three-way hemisphere \times condition \times duration ($F(2, 28)=2.62, p<0.09$) interaction approached significance. This interaction emerged because the right hemisphere was more sensitive to changes in duration for vowels than for tones in anterior sites. This effect can be seen in Fig. 6 which shows the magnitude of the laterality effect (subtracting the percent signal change in the LH from the RH) as a function of condition and duration. In anterior sites (shown in black), the right hemisphere shows a greater increase in activation with duration for vowels, whereas for tones the difference between right and left activation remains essentially constant across duration. To confirm these results statistically, two 2-way ANOVAs were performed on the vowel and tone conditions separately with factors of hemisphere and duration. Results yielded main effects of hemisphere (vowels: $F(1, 14)=12.0, p<0.004$; tones: $F(1, 14)=8.34, p<0.002$) and duration (vowels: $F(2, 28)=11.4, p<0.001$; tones: $F(2, 28)=9.32, p<0.001$) for both vowels and tones, with greater activation in right versus left hemisphere and activation increasing with duration. Of importance, only the vowel condition showed a hemisphere \times duration interaction ($F(2, 13)=10.8, p<0.001$). This interaction was due to increased right hemisphere activation as a function of duration for vowels. No interaction emerged for the tones.

Posterior regions show a different pattern of results. While main effects emerged for condition ($F(1, 14)=11.8, p<0.004$) and duration ($F(2, 28)=12.8, p<0.001$), showing similar patterns to those for the aSTG, there was no main effect of hemisphere nor were there any interactions. This pattern of results can also be seen in Fig 6, where the magnitude of the laterality effects in pSTG (as shown by the gray bars) is similar for vowels and tones across the three durations.

Discussion

The results of this study provide some further insight into the computational mechanisms underlying the processing of both speech and nonspeech. Although previous studies have examined hemispheric processing of spectral and temporal information in nonspeech stimuli, these parameters have typically covaried (cf. Zatorre and Belin 2001; Hall et al. 2002) making it difficult to determine the role of each of these parameters separately and to assess potential interactions between them. In the current study, the processing of components of vowel and tone stimuli were studied in two ways: potential differences in hemispheric processing of spectral information for vowels and pitch information for tones were assessed and compared both within and between duration conditions, and comparisons were made within vowel and tone stimulus conditions as a function of their duration. Consistent with earlier hypotheses that at early stages of processing a common mechanism is used for extracting the acoustic properties of both speech and non-speech (Poeppel, 2003; Zatorre et al., 2002b), there is a right hemisphere preference for the processing of relatively steady-state spectral properties of vowels and pitch properties of tones. Of importance, this hemispheric difference emerges for natural speech vowel stimuli as well as for nonspeech tone stimuli. Thus, even though the vowel stimuli are perceived as speech, the language dominant LH is not preferentially recruited

in a discrimination task for ‘specialized’ processing to extract the steady-state spectral structure signaling differences in vowel quality.

Nonetheless, the findings of this study challenge the strong form of the hypothesis that there are hemispheric differences in the integration window or time scale for processing spectral information (Hickock and Poeppel 2004). The strong form of this hypothesis posits that short duration stimuli should show left-lateralized activation. Differences in acoustic complexity notwithstanding, both vowels and tones have steady-state quasi-periodic and periodic properties respectively and thus the lateralization of the signal should correlate with duration if a hypothesis that favors a purely temporal (sub-phonetic) timescale is correct. In the present study, the ROI analyses revealed that neither vowel nor tone stimuli show left-lateralized activation at shorter timescales, nor at any timescale. Moreover, comparisons across the three duration conditions for the vowel stimuli in the ROI analyses showed bilateral activation with increased activation in the left as well as the right hemisphere as the stimuli were longer. These findings emerged in both the aSTG and pSTG. Similarly, the results of the ROI analysis for the tone stimuli showed increased activation bilaterally as duration increased. Thus again, increase in activation as a function of duration was not restricted to the RH. Confirmatory evidence of the effects of duration on the activation patterns in temporal areas for both vowels and tones also emerged in the cluster analyses shown in Appendix 1. In particular, contrasts of 300 ms to 75 ms in both vowels and tones revealed clusters in temporal areas with greater activation for the longer duration stimuli.

Taken together, increase in activation as a function of duration occurs bilaterally for both vowels and for tones. These findings are consistent with recent results of Boemio et al. (2005) who showed increased activation for frequency modulated noise and tone stimuli in both left and right hemispheres as the longer stimuli contained increasing pitch variation. However, similar to the vowel stimuli in the current study, slowly modulated changes in pitch variation in their stimuli preferentially increased activation in the right hemisphere, and failed to show left hemisphere activation at short stimulus timescales. In general, it seems to be the case that observed laterality differences for vowels and tones are a function of a heightened right hemisphere sensitivity to long integration windows, whereas the left hemisphere shows sensitivity to both long and short integration windows (cf. Belin et al., 1998).

The results of the current study show that although there are a number of similarities in the processing of vowels and tones, differences also emerged suggesting that even fairly early in the processing stream at the level of the STG. Results of the cluster analysis revealed greater activation and larger cluster sizes for vowels than for tones overall and at all duration conditions. Moreover, comparison of vowel-tone contrasts across the duration conditions revealed lateralization at short time-scales and bilateral activation at longer timescales. In particular, the comparison of vowels vs. tones at 75 ms revealed a single left hemisphere cluster with peak intensity in the aSTG. At 150ms, contrasts between vowels and tones showed a cluster in the LSTG cluster as well as in the RMTG extending into the RSTG, and at 300ms bilateral activation bilaterally in the STG.

The greater activation for vowels compared to tones overall and at each of the three duration conditions could be due to stimulus complexity. Nonetheless, Uppenkamp et al. (2006) showed greater activation for vowels than for control stimuli of equal complexity, making it unlikely that just complexity of the stimuli gave rise to greater activation in the current experiment. One possibility for the increased activation overall for vowels compared to tones is that the vowel stimuli were not only produced by a human vocal tract but listeners have more experience with them since they are part of the speech inventory. The behavioral data are consistent with this view in that overall performance was better for vowels than for tones. Importantly, however,

there were no reaction-time differences between vowels and tones suggesting that the discrimination task was equally easy/difficult for the subjects across the two stimulus types.

The differences in asymmetry revealed in the comparison between vowels and tones at the three duration conditions are also likely not due to stimulus complexity. Similar patterns of asymmetry should have emerged across the three duration conditions if stimulus complexity were the basis for asymmetry of processing since the difference in stimulus complexity between vowels and tones was in principle the same at each of the duration conditions. And yet, the results revealed a left hemisphere asymmetry at the 75 ms condition and bilateral activation at the longer, 150 and 300 ms, conditions.

The left hemisphere activation in general may reflect the linguistic relevance and in this case, the phonetic relevance, of stimuli. Such findings would support a number of studies suggesting that hemispheric lateralization of stimuli reflects the functional role they play, with left hemisphere lateralization for linguistic stimuli and right hemisphere lateralization for non-linguistic stimuli (VanLancker and Sidtis, 1992; Kimura 1961, 1964). This left hemisphere activation is present for all stimulus durations. However, the right hemisphere is also recruited for longer (150 and 300 ms durations). That the right hemisphere is also recruited at the longer intervals is likely due to the fact that vowels of 150 and 300 ms are not the canonical form in which vowels are found in natural speech, which are typically shorter in duration and involve considerable spectral change due to coarticulatory effects with adjacent segments and vowel diphthongization (Liberman et al. 1967). Moreover, vowels presented in isolation have been shown to be perceived as less linguistic than vowels produced in context (Rakerd 1984). Thus, at the longer intervals, the vowel stimuli may be more nonspeech-like resulting in right hemisphere in addition to left hemisphere activation.

Further differences in the processing of vowels and tone emerged from the ROI analyses when considering the magnitude of the laterality effect across duration conditions. In contrast to the vowel stimuli, the magnitude of the laterality effect for the tone stimuli remained constant at both temporal lobe sites. In contrast, the vowel stimuli in the aSTG showed an increase in the magnitude of the laterality effect as stimulus duration increased. That this interaction effect emerges only in the aSTG and only for vowels provides further evidence for a functional distinction between the aSTG and pSTG in auditory processing and supports the view that the aSTG is sensitive to spectral complexity and to the phonetic content of the stimuli (Binder et al. 2000; Scott et al. 2000; Davis and Johnsrude 2003; Giraud et al. 2004).

The results of the current study raise questions about how the acoustic properties giving rise to the speech sounds of language are integrated online as listeners receive speech input. If it is the case that there is a right hemisphere preference for vowels and a left hemisphere preference for those consonants which require the extraction of rapid spectral changes such as stop consonants, how does this information get integrated across the hemispheres into syllable-sized units or into words that may range from 100 ms to 500 ms in duration? In other words, is it the case that asymmetries for different speech sound means that this information is separately extracted by the 'preferred' hemisphere and then integrated by the left hemisphere in the phonetic processing stream? Even more problematic, the sounds of language have very different time scales – from 20–40 ms for stop consonants to hundreds of ms for fricative consonants, and the duration of these sound segments may vary as a function of a number of factors including speaking rate, stress and linguistic context (cf. Scott and Wise 2004). For example, the average duration of vowels is about 300 ms in citation form, 126 ms in reading a script (Crystal and House 1990), and these durations may be considerably shorter in conversational speech. Are there hemispheric differences in the processing of these same vowel sounds based on their different time scales?

We think not. We suggest that integration of auditory information occurs early in the phonetic processing stream, and we would expect most likely in the left hemisphere. As shown in the current study, the left hemisphere does process vowel stimuli across all duration conditions, and even shows left lateralization compared to tones at the short 75 ms duration condition. Ultimately the question of how the neural system integrates information across time scales is a critical question for our understanding of speech processing.

Acknowledgments

This research was supported in part by the Dana Foundation, NIH Grant DC006220 to Brown University and the Ittelson Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

References

- Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y. Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience* 1998;10:536–540. [PubMed: 9712682]
- 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, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience* 1997;17:353–362. [PubMed: 8987760]
- Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience* 2005;8:389–395.
- Brechmann A, Scheich H. Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex* 2005;15:578–587. [PubMed: 15319313]
- Burton MW, Blumstein SE, Small SL. The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience* 2000;12:679–690. [PubMed: 10936919]
- Cox RW, Hyde JS. Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine* 1997;10:171–178. [PubMed: 9430344]
- Crystal TH, House AS. Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustic Society of America* 1990;88:101–112.
- Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience* 2003;23:3423–3431. [PubMed: 12716950]
- Duvernoy, HM. *The human brain: surface, blood supply, and three-dimensional sectional anatomy*. New York, NY: SpringerWien; 1999.
- Farrell B. “Same”-“Different” judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin* 1985;98:415–456.
- Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex* 2004;14:247–255. [PubMed: 14754865]
- Godfrey JJ. Perceptual difficulty and the right ear advantage for vowels. *Brain and Language* 1974;1:323–336.
- Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA. Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research* 2004;47:46–57.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. “Sparse” temporal sampling in auditory cortex. *Human Brain Mapping* 1999;7:213–223. [PubMed: 10194620]
- Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield Q. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* 2002;12:140–149. [PubMed: 11739262]
- Hewson-Stoate N, Schonwiesner M, Krumbholz K. Vowel processing evokes a large sustained response anterior to primary auditory cortex. *European Journal of Neuroscience* 2006;24:2661–2667. [PubMed: 17100854]

- Hickock G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Science* 2000;4:131–138.
- Hickock 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]
- Ivry, RB.; Robertson, LC. *The two sides of perception*. Cambridge, MA: MIT Press; 1998.
- Jamison HL, Watkins KE, Bishop DVM, Matthews PM. Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex* 2005;16:1266–1275. [PubMed: 16280465]
- Jancke L, Wu, Wustenberg T, Scheich H, Heinze HJ. Phonetic perception and the temporal cortex. *Neuroimage* 2002;15:733–746. [PubMed: 11906217]
- Kimura D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology* 1961;15:166–171.
- Kimura D. Left-right difference in the perception of melodies. *Quarterly Journal of Experimental Psychology* 1964;16:355–358.
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kockunov PV, Nickerson D, Mikiten SA, Fox PT. Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping* 2000;10:120–131. [PubMed: 10912591]
- 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, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychology Review* 1967;74:431–461.
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. *Cerebral Cortex* 2005;15:1621–1631. [PubMed: 15703256]
- Liegeois-Chauvel C, deGraaf JB, Laguitton V, Chauvel P. Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex* 1999;9:484–496. [PubMed: 10450893]
- Lisker L. *Rapid vs. rabid*: A catalogue of acoustic features that may cue the distinction. Haskins Laboratories Status Report of Speech Research 1978;SR-54:127–132.
- Mertus, J. Bliss Audio Software Package. 1989. www.mertus.org
- Milner, B. Laterality effects in audition. In: Mountcastle, VB., editor. *Interhemispheric relations and cerebral dominance*. Baltimore: Johns Hopkins Press; 1962.
- Narain C, Scott SK, Wise RS, Rose S. Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex* 2003;13:1362–1368. [PubMed: 14615301]
- Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping* 2006;27:562–571. [PubMed: 16281283]
- Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 1971;9:97–113. [PubMed: 5146491]
- Poeppel D. Pure word deafness and the bilateral processing of the speech code. *Cognitive Science* 2001;25:679–693.
- Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication* 2003;41:245–255.
- Rakerd B. Vowels in consonantal context are perceived more linguistically than are isolated vowels. *Perception and Psychophysics* 1984;35:123–136. [PubMed: 6718209]
- Schonwiesner M, Rubsamen R, von Cramon DY. Spectral and temporal processing in the human auditory cortex—revisited. *Annals of the New York Academy of Science* 2005;1060:89–92.
- Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. *Trends in Neuroscience* 2003;106:100–107.
- Scott SK, Wise RJS. The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 2004;92:13–45. [PubMed: 15037125]
- Scott S, Blank C, Rosen S, Wise RJS. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 2000;123:2400–2406. [PubMed: 11099443]

- Sebastian C, Yasin I. Speech versus tone processing in compensated dyslexia: Discrimination and lateralization with a dichotic mismatch negativity (MMN) paradigm. *International Journal of Psychophysiology*. 2008 epub ahead of print.
- 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]
- Spellacy F, Blumstein S. The influence of language set on ear preference in phoneme recognition. *Cortex* 1970;6:430–439. [PubMed: 5514515]
- Studdert-Kennedy M, Shankweiler D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America* 1970;48:579–594. [PubMed: 5470503]
- Talairach, J.; Tournoux, P. A co-planar stereotaxic atlas of a human brain. Stuttgart, Germany: Thieme; 1988.
- Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD. Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage* 2006;31:1284–1296. [PubMed: 16504540]
- Van Lancker, Sidtis J. The identification of affective-prosodic stimuli by left-and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research* 1992;35:963–970. [PubMed: 1447930]
- Weiss MS, House AS. Perception of dichotically presented vowels. *Journal of the Acoustical Society of America* 1973;53:51–58. [PubMed: 4687671]
- Zatorre R, Belin P. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* 2001;11:946–953. [PubMed: 11549617]
- Zatorre RJ, Bouffard M, Ahad P, Belin P. Where is ‘where’ in the human auditory cortex? *Nature Neuroscience* 2002a;5:905–909.
- Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. *Trends in Cognitive Science* 2002b;6:37–46.
- Zatorre RJ, Gandour JT. Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B* 2007:1–18.



Figure 1.
Schematic for scanning protocol used for stimulus presentation.

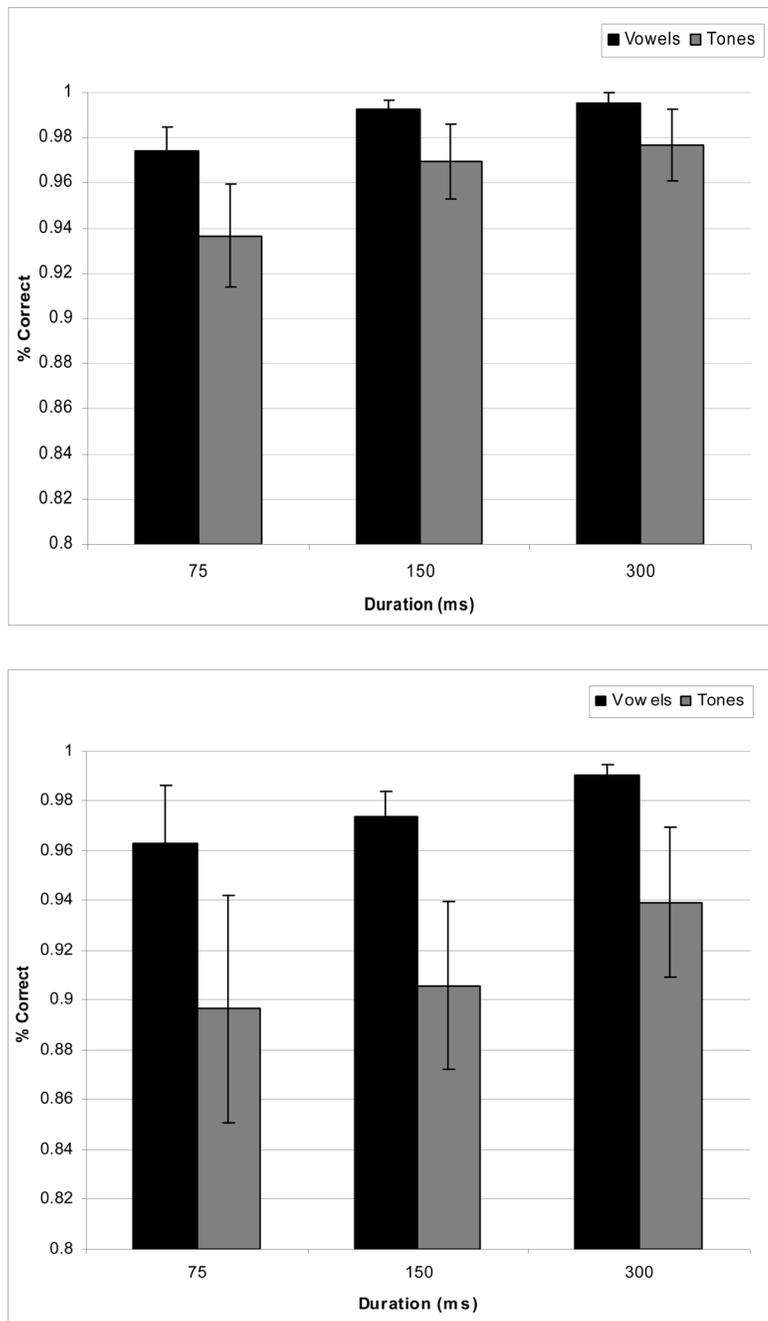


Figure 2. Mean performance for different responses (top panel) and the same responses (bottom panel) in the discrimination of vowel and tone stimuli.

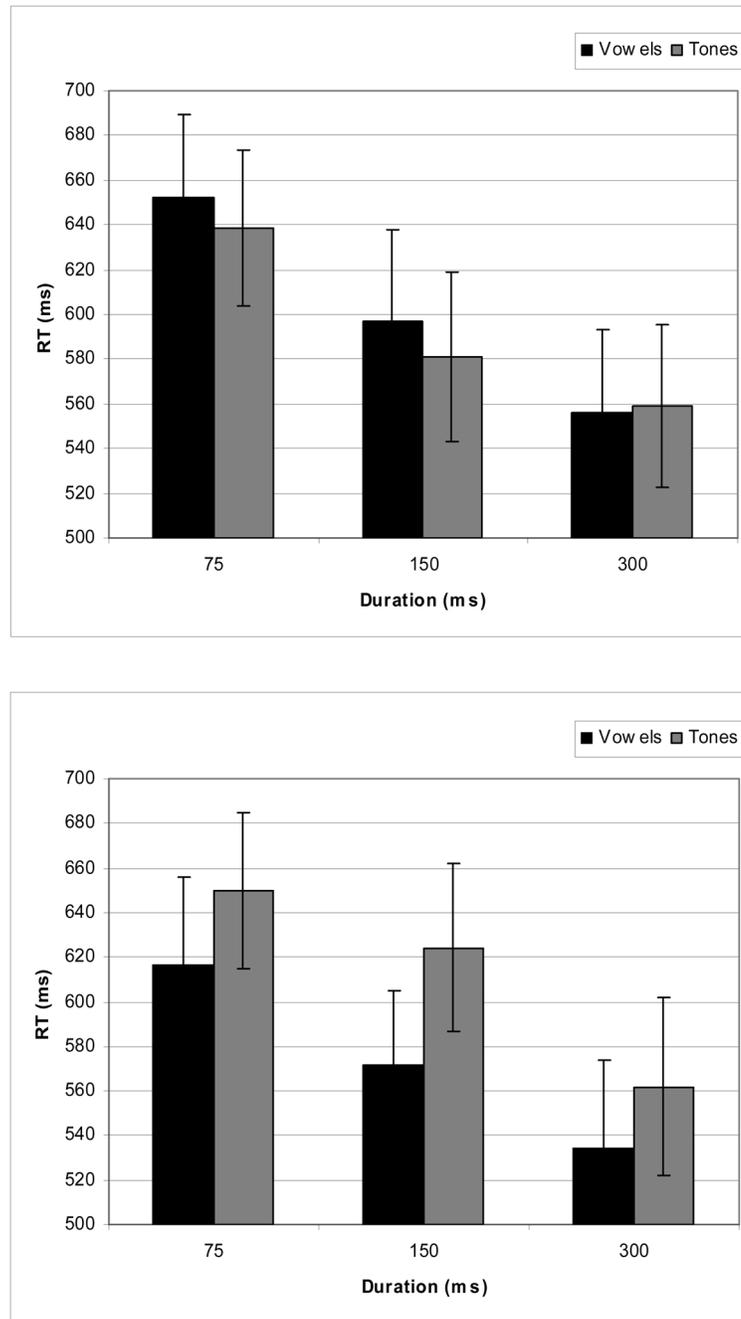


Figure 3. Mean reaction-time latencies for different responses (top panel) and the same responses (bottom panel) in the discrimination of vowel and tone stimuli.

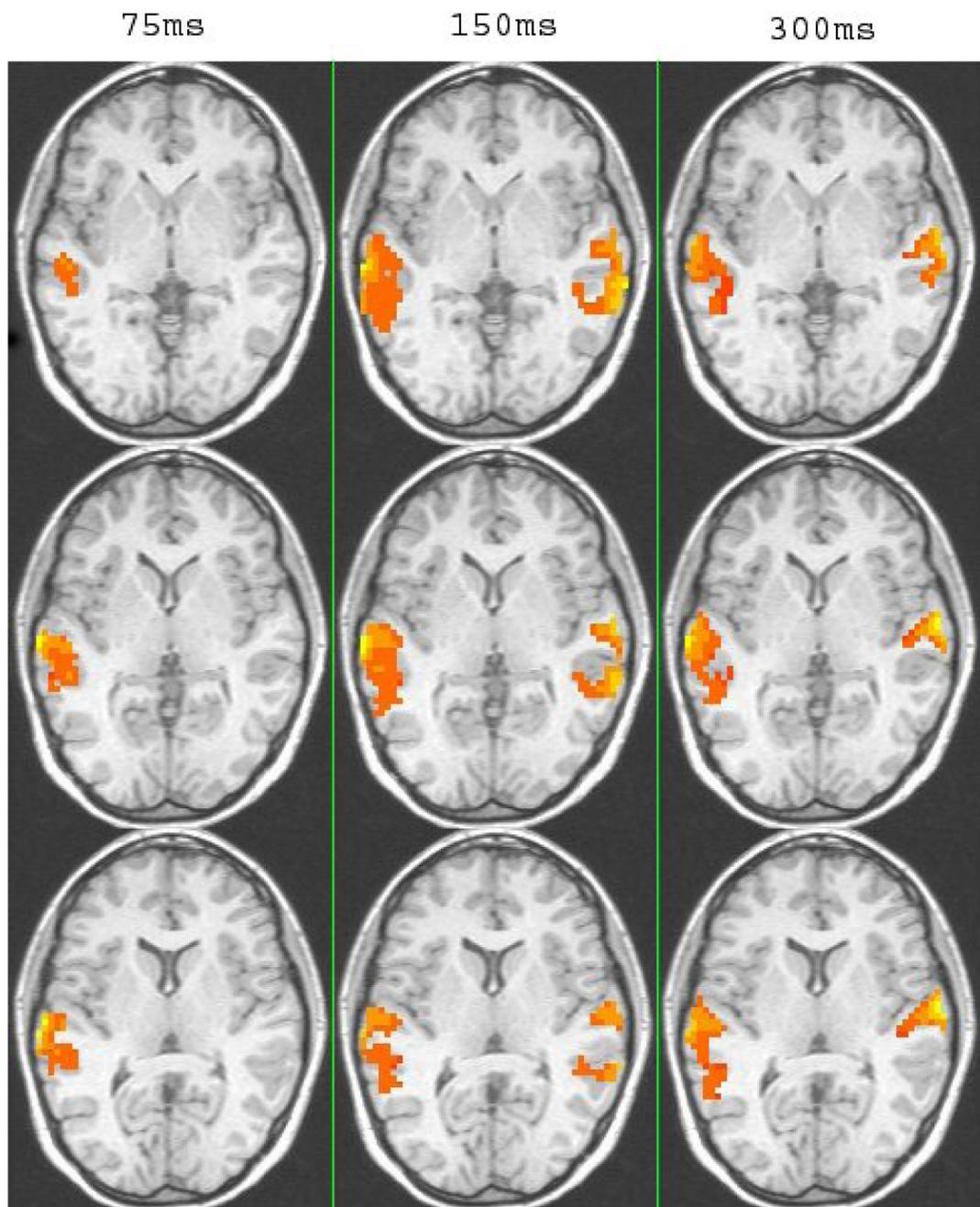


Figure 4. Montage of clusters showing significantly greater activation for vowels than for tones at the three duration conditions ($p < .05$ corrected). Axial views are shown at $z = 67$ for the top panel and $z = 73$ for the bottom panel.

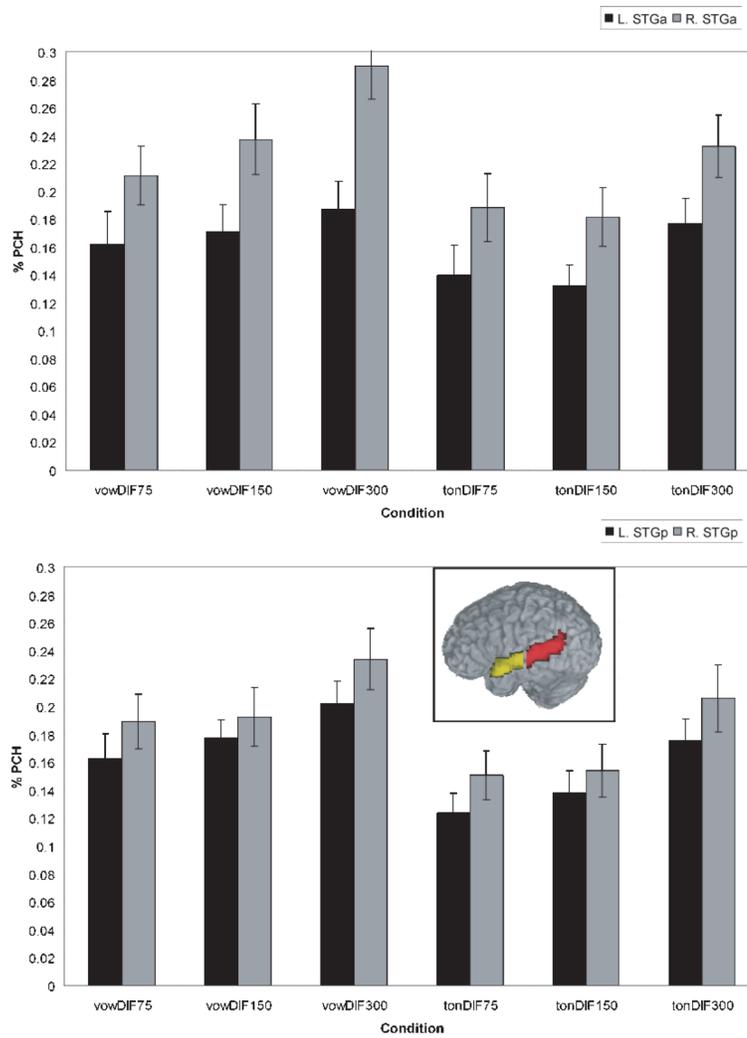


Figure 5. Percent signal change for the right and left anterior STG (top panel) and posterior STG (bottom panel) across the vowel and tone duration conditions. Inset shows anterior (yellow) and posterior (red) STG region of interest mask.

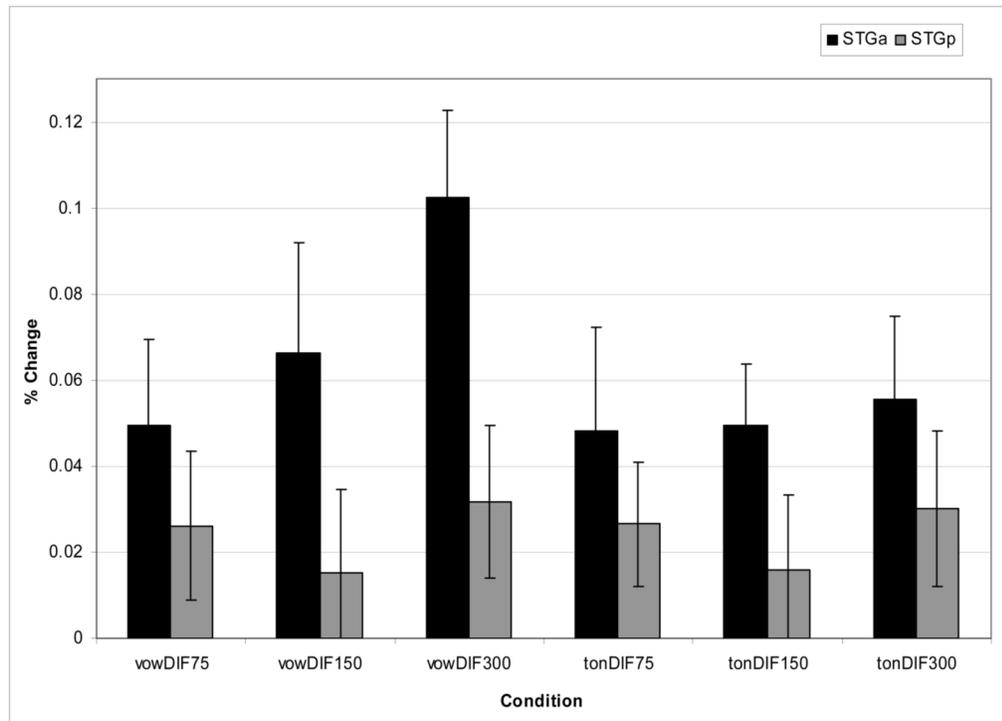


Figure 6. Percent signal change differences in the right versus left hemispheric activation in the anterior and posterior STG (see text).

Table 1

Formant values (in Hz) for the vowel stimuli

	[i]	[e]	[a]	[o]	[u]
F1	352	545	873	410	250
F2	2221	1914	1301	1129	1239

Table 2

Regions Exhibiting Significant Differences between Conditions

Regions Exhibiting Significant Differences between Conditions (Cluster Threshold $p < 0.05$, Voxel Level Threshold $p < 0.05$)						
Condition	Region	x	y	z	# of Voxels	Max T
<i>Vowels > Tones</i>						
	L. Superior Temporal Gyrus	-64	-16	5	425	5.296
	R. Superior Temporal Gyrus	59	-7	5	292	4.495
<i>Vow 75 > Ton 75</i>						
	L. Superior Temporal Gyrus	-65	-16	5	125	3.429
<i>Vow 150 > Ton 150</i>						
	L. Posterior Cingulate	-1	-46	20	844	3.473
	L. Superior Temporal Gyrus	-64	-19	5	490	7.622
	R. Middle Temporal Gyrus	65	-31	2	242	3.154
	R. Inferior Parietal Lobule	38	-61	44	145	2.735
<i>Vow 300 > Ton 300</i>						
	L. Superior Temporal Gyrus	-64	-16	5	266	4.282
	R. Superior Temporal Gyrus	59	-7	8	176	5.821
<i>Ton 300 > Vow 300</i>	L. Inferior Parietal Lobule	-55	-49	38	141	3.191

Appendix

Regions Exhibiting Significant Differences between Vowel Conditions and Tone Conditions.

Regions Exhibiting Significant Differences between Vowel Conditions and Tone Conditions (Cluster Threshold $p < 0.05$, Voxel Level Threshold $p < 0.05$)						
Condition	Region	x	y	z	# of Voxels	Max T
<i>Vow 75 > Vow 300</i>						
	R. Inferior Frontal Gyrus	47	20	-1	895	3.747
	R. Inferior Parietal Lobule	53	-37	44	393	4.172
	L. Medial Frontal Gyrus	-1	17	44	292	3.144
<i>Vow 300 > Vow 75</i>						
	R. Superior Temporal Gyrus	56	-7	8	867	6.476
	L. Posterior Cingulate	-1	-52	20	558	2.148
	L. Transverse Temporal Gyrus	-55	-19	11	555	5.272
<i>Vow 75 > Vow 150</i>						
	R. Inferior Frontal Gyrus	47	20	-1	251	2.63
	L. Medial Frontal Gyrus	-1	17	44	184	3.018
<i>Vow 150 > Vow 75</i>						
	L. Posterior Cingulate	-1	-46	8	543	4.631
	R. Parahippocampal Gyrus	11	-31	-4	292	2.359
	R. Precentral Gyrus	59	-1	11	146	2.434
	L. Anterior Cingulate	-1	44	-4	140	4.023
<i>Vow 150 > Vow 300</i>						
	R. Precuneus	2	-73	41	207	2.43
	L. Parahippocampal Gyrus	-1	-34	5	199	3.562
<i>Vow 300 > Vow 150</i>						
	R. Superior Temporal Gyrus	59	-7	8	390	3.321
	L. Transverse Temporal Gyrus	-61	-19	11	272	2.705
<i>Tone 300 > Tone 75</i>						
	R. Postcentral Gyrus	62	-19	17	971	2.814

Regions Exhibiting Significant Differences between Vowel Conditions and Tone Conditions (Cluster Threshold $p < 0.05$, Voxel Level Threshold $p < 0.05$)						
Condition	Region	x	y	z	# of Voxels	Max T
	L. Superior Temporal Gyrus	-64	-31	11	613	4.503
	L. Posterior Cingulate	-1	-61	11	272	2.233
	L. Cuneus	-4	-91	17	117	2.214
<i>Tone 75 > Tone 150</i>	R. Hypothalamus	2	-4	-7	276	2.74
	L. Inferior Frontal Gyrus	-43	5	26	153	5.644
<i>Tone 300 > Tone 150</i>	L. Precuneus	-19	-79	41	1752	3.405
	R. Poscentral Gyrus	62	-22	17	521	3.213
	L. Superior Temporal Gyrus	-64	-31	14	366	3.873