Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using fMRI

Corianne Rogalsky[1], **Feng Rong**[2], **Kourosh Saberi**[2], and **Gregory Hickok**[2,*]

[1] Dana and David Dornsife Cognitive Neuroscience Imaging Center & Brain and Creativity Institute, University of Southern California

[2] Department of Cognitive Sciences, University of California, Irvine

Abstract

Language and music exhibit similar acoustic and structural properties and both appear to be uniquely human. Several recent studies suggest that speech and music perception recruit shared computational systems, and a common substrate in Broca's area for hierarchical processing has recently been proposed. However, this claim has not been tested by directly comparing the spatial distribution of activations to speech and music processing within-subjects. In the present study, participants listened to sentences, scrambled sentences, and novel melodies. As expected, large swaths of activation for both sentences and melodies were found bilaterally in the superior temporal lobe, overlapping in portions of auditory cortex. However, substantial nonoverlap was also found: sentences elicited more ventrolateral activation, whereas the melodies elicited a more dorsomedial pattern, extending into the parietal lobe. Multitvariate pattern classification analyses indicate that even within the regions of BOLD response overlap, speech and music elicit distinguishable patterns of activation. Regions involved in processing hierarchical related aspects of sentence perception were identified by contrasting sentences with scrambled sentences revealing a bilateral temporal lobe network. Music perception showed no overlap whatsoever with this network. Broca's area was not robustly activated by any stimulus type. Overall, these findings suggest that basic hierarchical processing for music and speech recruit distinct cortical networks, neither of which involve Broca's area. We suggest that previous claims are based on data from tasks that tap higher-order cognitive processes, such as working memory and/or cognitive control, which can operate in both speech and music domains.

Keywords

Introduction

Language and music share a number of interesting properties spanning acoustic, structural, and even possibly evolutionary domains. Both systems involve the perception of sequences of acoustic events that unfold over time with both rhythmic and tonal features; both systems involve a hierarchical structuring of the individual elements to derive a higher-order combinatorial representation; and both appear to be uniquely human biological capacities (Patel, 2008; Lerdahl & Jackendoff, 1983; McDermott & Hauser, 2005). As such, investigating the relation between neural systems supporting language and music processing

[*]corresponding author: Gregory Hickok, Department of Cognitive Sciences, University of California, Irvine, 3151 Social Sciences Plaza, Irvine, CA 92697, greg.hickok@uci.edu.

can shed light on the underlying mechanisms involved in acoustic sequence processing and higher-order structural processing, which in turn may inform theories of the evolution of language and music capacity.

Recent electrophysiological, functional imaging, and behavioral work has suggested that structural processing in language and music draw on shared neural resources (Patel, 2007; Fadiga et al., 2009). For example, presentation of syntactic or musical "violations" (e.g., ungrammatical sentences or out-of-key chords) result in a similar modulation of the P600 event-related potential (ERP) (Patel, et al. 1998). Functional MRI studies have reported activation of the inferior frontal gyrus during the processing of aspects of musical structure (Levitin and Menon, 2003; Tillmann et al., 2003), which is a region that is also active during the processing of aspects of sentence structure (Just et al., 1996; Stromswold et al., 1996; Caplan et al., 2000; Friederici et al., 2003; Santi and Grodzinsky, 2007), although there is much debate regarding the role of this region in structural processing (Hagoort, 2005; Novick et al., 2005; Grodzinsky and Santi, 2008; Rogalsky et al., 2008; Rogalsky and Hickok, 2009). Behaviorally, it has been shown that processing complex linguistic structures interacts with processing complex musical structures: accuracy in comprehending complex sentences in sung speech is degraded if the carrier melody contains an out-of-key note at a critical juncture (Fedorenko, et al. 2009).

Although this body of work suggests that there are similarities in the way language and music structure are processed under some circumstances, a direct within-subject investigation of the brain regions involved in language and music perception has not, to our knowledge, been reported. As noted by Peretz and Zatorre (Peretz and Zatorre, 2005), such a direct comparison is critical before drawing conclusions about possible relations in the neural circuits activated by speech and music. This is the aim of the present fMRI experiment. We assessed the relation between music and language processing in the brain using a variety of analysis approaches including whole brain conjunction/disjunction, region of interest (ROI), and multivariate pattern classification analyses (MVPA). We also implemented a rate manipulation in which linguistic and melodic stimuli were presented 30% faster or slower than their normal rate. This parametric temporal manipulation served as a means to assess the effects of temporal envelope modulation rates, an stimulus feature that appears to a major role in speech perception (Shannon et al., 1995; Luo and Poeppel, 2007), and provided a novel method for assessing the domain-specificity of processing load effects by allowing us to map regions that were modulated by periodicity manipulations within and across stimulus types.

Methods

Participants

Twenty right-handed native English speakers (9 male, 11 female; mean age= 22.6 years, range 18–31) participated in this study. Twelve participants had some formal musical training (mean years of training = 3.5, range 0–8). All participants were free of neurological disease (self report) and gave informed consent under a protocol approved by the Institutional Review Board of the University of California, Irvine (UCI).

Experimental Design

Our mixed block and event-related experiment consisted of the subject listening to blocks of meaningless "jabberwocky" sentences, scrambled jabberwocky sentences, and simple novel piano melodies. We used novel melodies and meaningless sentences to emphasize structural processing over semantic analysis. Within each block, the stimuli were presented at three different rates, with the mid-range rate being that of natural speech/piano playing (i.e., the

rate at which a sentence was read or a composition played without giving the reader/player explicit rate instructions). The presentation rates of the stimuli in each block were randomized, with the restriction that each block had the same total amount of time in which stimuli were playing. Each trial consisted of a stimulus block (27 sec) followed by a rest period (jittered around 12 sec). Within each stimulus block, the interval between stimulus onsets was 4.5 seconds. Subjects listened to 10 blocks of each stimulus type, in a randomized order, over eight scanning runs.

Stimuli

As mentioned above, three types of stimuli were presented: jabberwocky sentences, scrambled jabberwocky sentences, and simple novel melodies. The jabberwocky sentences were generated by replacing the content words of normal, correct sentences with pseudowords (e.g. "It was the glandar in my nederop"); these sentences had a mean length of 9.7 syllables (range = 8–13). The scrambled jabberwocky sentences were generated by randomly rearranging the word order of the previously described jabberwocky sentences; the resulting sequences, containing words and pseudowords, were recorded and presented as concatenated lists. As recorded, these sentences and scrambled sentences had a mean duration of 3 seconds (range = 2.5–3.5). The sentences and scrambled sentences then were digitally edited using sound-editing software to generate the two other rates of presentation: each stimulus' tempo was increased by 30%, and also decreased by 30%. The original (normal presentation rate) sentences and scrambled sentences averaged 3.25 syllables/ second; the stimuli generated by altering the tempo of the original recordings averaged 4.23 syllables/second (30% slower tempo than normal), and 2.29 syllables/second (30% faster tempo than normal), respectively.

The melody stimuli were composed and played by a trained pianist. The composition of each melody outlined a common major or minor chord in the system of tonal Western harmony, such as C major, F major, G major, or D minor. Most pitches of the melodies were in the fourth register (octave) of the piano; pitches ranged from G3 to D5. Durations in the sequence were chosen to sound relatively rhythmic according to typical Western rhythmic patterns. There are 5 to 17 pitches in each melodic sequence (on average, ~8 pitches per melody). Pitch durations range from 106 ms to 1067 ms (mean = 369 ms, SD = 208 ms). The melodies were played on a Yamaha Clavinova CLP-840 digital piano and were recorded through the MIDI interface, using MIDI sequencing software. The melodies, as played by the trained pianist, averaged 3s in length. The melodies (like the sentence and scrambled sentences) were then digitally edited to generate versions of each melody that were 30% slower, and 30% faster, respectively, than the original melodies.

fMRI Data Acquisition & Processing

Data were collected on the 3T Phillips Achieva MR scanner at the UCI Research Imaging Center. A high-resolution anatomical image was acquired, in the axial plane, with a 3D SPGR pulse sequence for each subject (FOV = 250mm, TR =13ms, flip angle = 20 deg., voxel size = 1mm × 1mm × 1mm). Functional MRI data was collected using single-shot echo-planar imaging (FOV = 250 mm, TR = 2 sec, TE = 40ms, flip angle = 90 deg., voxel size = 1.95 mm × 1.95 mm × 5 mm). MRIcro (Rorden & Brett, 2000) was used to reconstruct the high-resolution structural image, and an in-house Matlab program was used to reconstruct the echo-planar images. Functional volumes were aligned to the sixth volume in the series using a 6-parameter rigid-body model to correct for subject motion (Cox & Jesmanowicz, 1999). Each volume then was spatially filtered (FWHM = 8mm) to better accommodate group analysis.

Data Analysis Overview

Analysis of Functional NeuroImaging (AFNI) software (http://afni.nimh.nih.gov/afni) was used to perform analyses on the time course of each voxel's BOLD response for each subject (Cox & Hyde, 1997). Initially, a voxel-wise multiple regression analysis was conducted, with regressors for each stimulus type (passive listening to jabberwocky sentences, scrambled jabberwocky sentences, and simple piano melodies), at each presentation rate. These regressors (in addition to motion correction parameters and the grand mean) were convolved with a hemodynamic response function to create predictor variables for analysis. An F-statistic was calculated for each voxel, and activation maps were created for each subject to identify regions that were more active while listening to each type of stimulus at each presentation rate compared to baseline scanner noise. The functional maps for each subject were transformed into standardized space and resampled into $1 \times 1 \times 1$mm voxels (Talairach & Tournoux, 1988) to facilitate group analyses. Voxel-wise repeated-measures t-tests were performed to identify active voxels in various contrasts. We used a relatively liberal threshold of $p = 0.005$ to ensure that potential non-overlap between music and speech activated voxels was not due to overly strict thresholding. In these contrasts, we sought to identify regions that were (i) active for sentences compared to baseline (rest), (ii) active for melodies compared to baseline, (iii) equally active for sentences and melodies (i.e., the conjunction (i) & (ii)), (iv) more active for sentences than music, (v) more active for music than sentences, (vi) more active for sentences than scrambles sentences (to identify regions selective for sentence structure over unstructured speech), (vii) more active for sentences than either scrambled sentences and melodies (to determine whether there are regions selective for sentence structure compared to unstructured speech *and* melodic structure). In addition, several of the above contrasts were carried out with a rate covariate (see immediately below).

Analysis of temporal envelope spectra

Sentences, scrambled sentences, and melodies differ not only in the type of information conveyed, but also potentially in the amount of acoustic information being presented per unit time. Speech, music, and other band-limited waveforms have a quasiperiodic envelope structure that carries significant temporal-rate information. The envelope spectrum provides important information about the dominant periodicity rate in the temporal-envelope structure of a waveform. Thus, additional analyses explored how the patterns of activation were modulated by a measure of information presentation rate. To this end, we performed a Fourier transform on the amplitude envelopes extracted from the Hilbert transform of the waveforms, and determined the largest frequency component of each stimulus' envelope spectrum using MATLAB (MathWorks, Inc.). The mean peak periodicity rate of each stimulus type's acoustic envelope varies between the three stimuli types at each presentation rate (Fig 1, Table 1). Of particular importance to the interpretation of the results reported below is that, at the normal presentation rate, the dominant frequency components of the sentences' temporal envelope ($M = 1.26$, $sd = 0.78$) are significantly different than those of the scrambled sentences ($M = 2.13$, $sd = 0.42$): $t(82) = 14.1$, $p < .00001$, however, at the normal presentation rate, the sentences' principal frequency components are not significantly different than those of the melodies ($M = 1.19$, $sd = 0.82$): $t(82) = 0.34$, $p = 0.73$. To account for temporal envelope modulation rate effects, we conducted a multiple regression analysis, including the principal frequency component of each stimulus as a covariate regressor (R.W. Cox, Regression with Event-dependent Amplitudes, http://afni.nimh.nih.gov/afni). This allowed us to assess the distribution of brain activation to the various stimulus types controlled for differences in peak envelope modulation rate. In addition, we assessed the contribution of envelope modulation rate directly by correlating the BOLD response with envelope modulation rate in the various stimulus categories.

Regions of Interest Analyses

Voxel clusters identified by the conjunction analyses described above were further investigated in two ways: (i) mean peak amplitudes for these clusters across subjects were calculated using AFNI's 3dMaskDump program and MATLAB and (ii) a multivariate pattern analysis (MVPA) was conducted on the overlap regions. MVPA provides an assessment of the *pattern* of activity within a region and is thought to be sensitive to weak tuning preferences of individual voxels driven by non-uniform distributions of the underlying cell populations. (Haxby et al., 2001; Kamitani and Tong, 2005; Norman et al., 2006; Serences and Boynton, 2007).

The MVPA was performed to determine if the voxels of overlap in the conjunction analyses were in fact exhibiting the same response pattern to both sentences and melodies. This analysis proceeded as follows: regions of interest (ROIs) for the MVPA were defined for each individual subject by identifying, in the four odd-numbered scanning runs, the voxels identified by both (i.e., the conjunction of) the sentences > rest and the melodies > rest comparisons ($p < .005$) across all presentation rates. In addition, a second set of ROIs was defined in a similar manner, but also included the envelope modulation rate for each stimulus as a covariate. All subjects yielded voxels meeting these criteria in both hemispheres. The average ROI sizes for the left and right hemispheres were: 93.4 (range 22–279) and 37.2 (range 6–94) voxels without the covariate, respectively, and 199.6 (range 15–683) and 263.7 voxels (range 26–658) with the envelope modulation rate covariate.

MVPA was then applied in this sentence-melody overlap ROI in each subject to the even-numbered scanning runs (i.e. the runs not used in the ROI identification process). In each ROI, one pairwise classification was performed to explore the spatial distribution of activation that varies as a function of the presentation of sentence versus melody stimuli, across all presentation rates. MVPA was conducted using a support vector machine (SVM) (Matlab Bioinformatics Toolbox v3.1, The MathWorks, Inc., Natick, MA) as a pattern classification method. This process is grounded in the logic that if the SVM is able to successfully distinguish (i.e. classify) the activation pattern in the ROI for sentences from that for melodies, then the ROI must contain information that distinguishes between sentences and melodies.

Before classification, the EPI data were motion-corrected (described above). The motion-corrected data were then normalized so that in each scanning run a z-score was calculated for each voxel's BOLD response at each time point. In addition, to ensure that overall amplitude differences between the conditions were not contributing to significant classification, the mean activation level across the voxels within each trial was removed prior to classification. We then performed MVPA on the preprocessed data set using a leave-one-out cross validation approach (Vapnik, 1995). In each iteration, data from all but one even-numbered session was used to train the SVM classifier and then the trained SVM was used to classify the data from the remaining session. The SVM-estimated labels of sentences and melody conditions were then compared with the actual stimuli types to compute a classification accuracy score. Classification accuracy for each subject was derived by averaging across the accuracy scores across all leave one-out sessions. An overall accuracy score was then computed by averaging across all subjects.

We then statistically evaluated the classification accuracy scores using nonparametric bootstrap methods (Lunneborg, 2000). Similar classification procedures were repeated 10,000 times for the pairwise classification within each individual subject's data set. The only difference in this bootstrapping method from the classification method described above is that in the bootstrapping, the labels of "sentences" and "melodies" were randomly reshuffled on each repetition. This process generates a random distribution of the bootstrap

classification accuracy scores ranging from 0 to 1 for each subject for the pairwise classification, where the ideal mean of this distribution is at the accuracy value of 0.5. We then tested the null hypotheses that the original classification accuracy score equals the mean of the distribution via a one-tailed accumulated percentile of the original classification accuracy score in the distribution. If the accumulated $p > 0.95$, the null hypothesis was rejected and it would be concluded that, for that subject, signal from the corresponding ROI can classify the sentence and melody conditions. In addition, a bootstrap-T approach was used to assess the significance of the classification accuracy on the group level. For each bootstrap repetition, a t-test accuracy score across all subjects against the ideal accuracy score (0.5) was calculated. The *t*-score from the original classification procedures across all subjects was then statistically tested against the mean value of the distributed bootstrap t-scores. An accumulated $p > 0.95$ was the criterion for rejecting the null hypothesis and concluding that the accuracy score is significantly greater than chance (this is the criterion as used for the individual subject testing).

Results

Activations for sentences and melodies compared with rest

Extensive activations to both sentence and melodic stimuli were found bilaterally in the superior temporal lobe. The distribution of activity to the two classes of stimuli is far from identical: a gradient of activation is apparent with more dorsomedial temporal lobe regions responding preferentially to melodic stimuli and more ventrolateral regions responding preferentially to sentence stimuli, with a region of overlap in between, bilaterally. The melody-selective response in the dorsal temporal lobe extended into the inferior parietal lobe. Sentences activated some frontal regions (left premotor, right prefrontal), although not Broca's area (Fig 2A).

Consistent with the claim that speech and music share neural resources, our conjunction analysis showed substantial overlap in the response to sentence and music perception in both hemispheres. However, activation overlap does not necessarily imply computational overlap or even the involvement of the same neural systems at a finer-grained level of analysis. Thus we used MVPA to determine if the activation within the overlap region may be driven by non-identical cell assemblies. Classification accuracy for sentence versus melody conditions in the ROI representing the region of overlap was found to be significantly above chance in both hemispheres: left hemisphere, classification accuracy (proportion correct, PC) = 0.79, t = 12.75, p = .0015, classification accuracy (d′) = 2.90, t = 10.87, p = .0012; right hemisphere, classification accuracy (PC) = 0.79, t = 11.34, p = .0003, classification accuracy (d′) = 3.19, t = 10.61, p = .0006. This result indicates that even within regions that activate for both speech and music perception, the two types of stimuli activate non-identical neural ensembles, or activate them to different degrees.

Analysis of the conjunction between speech and music (both versus rest) using the envelope modulation rate covariate had two primary effects on the activation pattern: (i) the speech-related frontal activation disappeared and (ii) the speech stimuli reached threshold in the medial posterior superior temporal areas leaving the anterior region bilaterally as the dominant location for music-selective activations, although some small posterior foci remained (Fig 2B). MVPA was run on the redefined region of overlap between music and speech. The results were the same in that the ROI significantly classified the two conditions in both hemispheres: left hemisphere, classification accuracy (proportion correct, PC) = 0.70, t = 6.49, p = .0026, classification accuracy (d′) = 1.88, t = 3.73, p = .0027; right hemisphere, classification accuracy (PC) = 0.75, t = 8.66, p = .0008, classification accuracy (d′) = 2.49, t = 6.50, p = .0003.

Because the envelope modulation rate covariate controls for a lower-level acoustic feature and tended to increase the degree of overlap between music and speech stimuli indicating greater sensitivity to possibly shared neural resources, all subsequent analyses use the rate covariate.

Contrasts of Speech & Music

A direct contrast between speech and music conditions using the envelope modulation rate covariate was carried out to identify regions that were selective for one stimulus condition relative to the other. Figure 3 shows the distribution of activations revealed by this contrast. Speech stimuli selectively activated more lateral regions in the superior temporal lobe bilaterally while music stimuli selectively activated more medial anterior regions on the supratemporal plane and extending into the insula, primarily in the right hemisphere. It is important not to conclude from this apparently lateralized pattern for music that the right hemisphere preferentially processes music stimuli as is often assumed. As is clear from the previous analysis (Fig 2), music activates both hemispheres rather symmetrically; the lateralization effect is in the *relative* activation patterns to music versus speech.

Previous work has suggested the existence of sentence-selective regions in the anterior and posterior temporal lobe (bilaterally). Contrasts were performed to identify these sentence selective regions as they have been defined previously, by comparing structured sentences with unstructured lists of words, and then to determine if this selectively holds up in comparison to musical stimuli, which share hierarchical structure with sentences.

The contrast between listening to sentences compared to scrambled sentences identified bilateral ATL regions that were more active for sentence than scrambled sentence stimuli (left −53 −2 −4, right 58 −2 −4) (Fig 4, Table 2). No IFG activation was observed in this contrast and no regions were more active for scrambled than non-scrambled sentences. Figure 4 shows the relation between sentence selective activations (sentences > scrambled sentences) and activations to music perception (music > rest). No overlap was observed between the two conditions.

As a stronger test of sentence-selective responses, we carried out an additional analysis that looked for voxels that were more active for sentences than scrambled sentences *and* more active for sentences than music. This analysis confirmed that sentences yield significantly greater activity than scrambled sentences and melodies in the left anterior (−53 7 −4; −55 −5 0), mid (−57 −21 4) and posterior temporal lobe (−57 −35 2); fewer voxels survived this contrast in the right hemisphere however, which were limited to the anterior temporal region (53 −25 −2; 57 −3 2) (Fig. 5, Table 2).

Envelope modulation rate effects

In addition to using envelope modulation rate as a covariate to control for differences in the rate of information presentation in the different stimulus conditions, we carried out an exploratory analysis to map brain regions that were correlated by envelope modulation rate directly and whether these regions varied by stimulus condition. The logic is that the rate manipulation functions as a kind of parametric load manipulation in which faster rates induce more processing load. If a computation is unique to a given stimulus class (e.g., syntactic analysis for sentences) then the region involved in this computation should be modulated only for that stimulus class. If a computation is shared by more than one stimulus type (e.g., pitch perception) then the regions involved in that computation should be modulated by rate similarly across those stimulus types.

Envelope modulation rate of sentences positively correlated with activity in the anterior and mid portions of the superior temporal lobes bilaterally as well as a small focus in the right

posterior temporal lobe (45 −39 10) (Fig 6A, Table 3). Positive correlations with the envelope modulation rate of scrambled sentences had a slightly more posterior activation distribution which overlapped that for sentences in the mid portion of the superior temporal lobes bilaterally (center of overlap: LH:−55 −11 4, RH: 59 −15 2). There was no negative correlation with envelope modulation rate for either sentences or scrambled sentences, even at a very liberal threshold (p = .01).

Envelope modulation rate of the music stimuli was positively correlated with a very different activation distribution that included the posterior inferior parietal lobe bilaterally, more medial portions of the dorsal temporal lobe, which likely include auditory areas on the supratemporal plane, the insula, and prefrontal regions. There was no overlap in the distribution of positive envelope modulation rate correlations for music and either of the two speech conditions. A negative correlation with envelope modulation rate of the music stimuli was observed primarily in the lateral left temporal lobe (see Supplemental Figure 2).

Direct contrasts of envelope modulation rate effects between stimulus conditions revealed a large bilateral focus in the mid portion of the superior temporal lobe with extension into anterior regions that was more strongly correlated with sentence than melody envelope modulation rate (LH: −53 −27 2; RH: 53 −21 −4; Figure 6B). Anterior to this region were smaller foci that were correlated more strongly with sentence than scrambled sentence envelope modulation rate and/or with sentence than melody envelope modulation rate. Thus the envelope modulation rate analysis was successful at identifying stimulus specific effects and confirms the specificity of more anterior temporal regions in processing sentence stimuli.

Sentences and Scrambled Sentences at Normal Rates

The contrast between sentences and scrambled sentences at the normal rate identified voxels that activated more to sentences (*p* = .005) in two left ATL clusters (−51 15 −13, −53 −6 1), in a left posterior MTG cluster (−43 −51 −1), in the left inferior frontal gyrus (IFG) (-46 26 8), as well as in the right ATL (55 2 −8), right IFG (32 26 9), and right postcentral gyrus (49 −9 18) (Supp. Fig. 1). No other regions were more active for sentences than for scrambled sentences at this threshold (see Supp. Table for peak *t* values), and no regions were found to be more active for the scrambled sentences than sentences. Signal amplitude plots for a left ATL (Supp. Fig. 1, middle graph) and the left frontal ROI (Supp. Fig 1, right graph) are presented for each condition for descriptive purposes. It is relevant that in both ROIs the music stimuli activation is low (particularly in the ATL) and in fact do not appear to generate any more activation than the scrambled sentence stimuli.

Discussion

The aim of the present study was to assess the distribution of activation associated with processing speech (sentences) and music (simple melodies) information. Previous work has suggested common computational operations in processing aspects of speech and music, particularly in terms of hierarchical structure (Patel, 2007; Fadiga et al., 2009), which predicts significant overlap in the distribution of activation for the two stimulus types. The present study indeed found some overlap in the activation patterns for speech and music, although this was restricted to relatively early stages of acoustic analysis, and not in regions, such as the anterior temporal cortex or Broca's area, that have previously been associated higher-level hierarchical analysis. Specifically, only speech stimuli activated anterior temporal regions that have been implicated in hierarchical processes (defined by the sentence vs. scrambled sentence contrast) and Broca's area did not reliably activate to either stimulus type once lower-level stimulus features were factored out. Furthermore, even within the region of overlap in auditory cortex, multivariate pattern classification analysis

showed that the two classes of stimuli yielded distinguishable patterns of activity, likely reflecting the different acoustic features present in speech and music (Zatorre et al., 2002). Overall, these findings seriously question the view that hierarchical processing in speech and music rely on the same neural computation systems.

If the neural systems involved in processing higher-order aspects of language and music are largely distinct, why have previous studies suggested common mechanisms? One possibility is that similar computational mechanisms are implemented in distinct neural systems. This view, however, is still inconsistent with the broader claim that language and music share computational resources (e.g. (Fedorenko et al., 2009)). Another possibility is that the previously documented similarities derive not from fundamental perceptual processes involved in language and music processing but from computational similarities in the tasks employed to assess these functions. Previous reports of common mechanisms employ structural violations in sentences and melodic stimuli and it is in the response to *violations* that similarities in the neural response have been observed (Patel et al., 1996; Levitin and Menon, 2003). The underlying assumption of this body of work is that a violation response reflects the additional structural load involved in trying to integrate an unexpected continuation of the existing structure and therefore index systems involved in more natural structural processing. However, this is not the only possibility. A structural violation could trigger working memory processes (Rogalsky et al., 2008) or so-called "cognitive control" processes (Novick et al., 2005) each of which have implicated regions, Broca's area in particular, that are thought to be involved in structural processing. The fact that Broca's area was not part of the network implicated in the processing of structurally coherent language and musical stimuli in the present study argues in favor of the view that the commonalities found in previous studies indeed reflect a task-specific process that is not normally invoked under more natural circumstances.

The present study employed a temporal rate manipulation that, we suggest, may function as a kind of parametric load manipulation that can isolate stimulus-specific processing networks. If this reasoning is correct, such a manipulation may be a better alternative to the use of structural violations to induce load. We quantified our rate manipulation by calculating the envelope modulation rate of our stimuli and used these values as a predictor of brain activity. Envelope modulation rate has emerged in recent years as a potentially important feature in the analysis of acoustic input in that it may be important in driving endogenous cortical oscillations (Luo and Poeppel, 2007). We reasoned that correlations with envelope modulation rate would tend to isolate higher-level, integrative, aspects of processing on the assumption that such levels will be more affected by rate modulation (although this is clearly an empirical question). Indeed, based on the pattern of correlated activity, this appears to be the case. Regions that correlated with the modulation rate of sentences included anterior and mid portions of the superior temporal lobes bilaterally as well as a small focus in the right posterior temporal lobe (Fig 6A, Table 3). No core auditory areas showed a correlation with modulation rate of sentences, nor did Broca's region show a correlation with modulation rate. When structural information is largely absent, as in scrambled sentences, modulation rate does not correlate with more anterior temporal regions, but instead has a more mid- to posterior distribution. In the absence of structural information, the load on lexical-phonological level processes may be amplified explaining the involvement of more posterior superior temporal regions which have been implicated in phonological level processes (for review see (Hickok and Poeppel, 2007)).

Regions that correlated with the modulation rate of melodies were completely non-overlapping with regions correlated with the modulation rate of sentences. The modulation rate of the melodies correlated with dorsomedial regions of the anterior temporal lobe, likely including portions of auditory cortex, inferior parietal-temporal regions, and prefrontal

cortex. A previous study which manipulated prosodic information in speech, which is similar to a melodic contour, also reported activation of anterior dorsomedial auditory regions (Humphries et al., 2005) and other studies have implicated the posterior supratemporal plane in aspects of tonal perception (Binder et al., 1996; Griffiths and Warren, 2002; Hickok et al., 2003). Lesion studies of music perception, similarly have implicated both anterior and posterior temporal regions (Stewart et al., 2006), consistent with our findings.

The rate modulation analysis should be viewed as preliminary and requires further empirical exploration regarding the present assumptions. Although the pattern of activity revealed by these analyses is consistent with our reasoning that rate modulation correlations isolate higher-order aspects of processing, this has not been demonstrated directly. Another aspect of this approach that warrants investigation is the possibility of a non-monotonic function relating rate modulation and neural load. For example, one might expect that both very slow or very fast rates induce additional processing loads, perhaps in different ways. In addition, it is important to note that the lack of a response in core auditory regions to the temporal modulation rate does not imply that auditory cortex is insensitive to modulation rate generally. In fact, previous studies have demonstrated that auditory cortex is sensitive to the temporal modulation rate of various types of noise stimuli (Giraud et al., 2000; Schonwiesner & Zatorre, 2009; Wang et al., 2003). However, the modulation rates in our stimuli were below the modulation rates used in these previous studies and the range of rates in our stimuli was relatively small. It is therefore difficult to compare the present study with previous work on temporal modulate rate.

Nonetheless, in the present study, both the traditional conjunction/disjunction and rate modulation analyses produced convergent results, suggesting that higher-order aspects of language and music structure are processed largely within distinct cortical networks, with music structure being processed in more dorsomedial temporal lobe regions and language structure being processed in more ventrolateral structures. This division of labor likely reflects the different computational building blocks that go into constructing linguistic- compared to musical structure. Language structure is largely dependent on the grammatical (syntactic) features of words, whereas musical structure is determined by pitch and rhythmic contours (i.e., acoustic features). Given that computation of hierarchical structure requires the integration of these lower-level units of analysis to derive structural relations, given that the representational features of the language and melodic units are so different, and given that the "end-game" of the two processes are different (derivation of a combinatorial semantic representation vs. acoustic recognition or perhaps an emotional modulation) it seems unlikely that a single computational mechanism could suffice for both.

In terms of language organization, available evidence points to a role of the posterior-lateral superior temporal lobe in analyzing phonemic and lexical features of speech (Binder et al., 2000; Vouloumanos et al., 2001; Liebenthal et al., 2005; Obleser et al., 2006; Okada and Hickok, 2006; Hickok and Poeppel, 2007; Vaden et al., 2009) and more anterior regions playing some role in higher-order sentence-level processing (Mazoyer et al., 1993; Friederici et al., 2000; Humphries et al., 2001; Vandenberghe et al., 2002; Humphries et al., 2005; Humphries et al., 2006; Rogalsky and Hickok, 2009). This is consistent with the present finding from the rate modulation analysis of an anterior distribution for sentences and a more posterior distribution for speech with reduced sentence structure (scrambled sentences). Perhaps music perception has a similar posterior-to-anterior, lower-to-higher- level gradient, although this remains speculative.

One aspect where there may be some intersection is with prosodic aspects of language, which, like music rely on pitch and rhythm contours, and which certainly inform syntactic

structure (e.g., see (Cutler et al., 1997; Eckstein and Friederici, 2006)). As noted above, prosodic manipulations in sentences activates a dorsomedial region in the anterior temporal lobe (Humphries et al., 2005) that is similar to the region responsive to musical stimuli in the present study. This point of contact warrants further investigation on a within-subject basis.

Summary and conclusions

Despite increasing enthusiasm for the idea that music and speech share important computational mechanisms involved in hierarchical processing, the present direct within-subject comparison failed to find compelling evidence for this view. Music and speech stimuli activated largely distinct neural networks except for in an around core auditory regions, and even in these overlapping regions, distinguishable *patterns* of activation were found. Many previous studies hinting at shared processing systems may have induced higher-order cognitive mechanisms, such as working memory or cognitive control systems, that may be the basis of the apparent process similarity rather than a common computation system for hierarchical processing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. Cerebral Cortex. 2000; 10:512–528. [PubMed: 10847601]

Binder JT, Frost JA, Hammeke TA, Rao SM, Cox RW. Function of the left planum temporale in auditory and linguistic processing. Brain. 1996; 119:1239–1247. [PubMed: 8813286]

Caplan D, Alpert N, Waters G, Olivieri A. Activation of Broca's area by syntactic processing under conditions of concurrent articulation. Hum Brain Mapp. 2000; 9:65–71. [PubMed: 10680763]

Cutler A, Dahan D, van Donselaar W. Prosody in the comprehension of spoken language: a literature review. Lang Speech. 1997; 40 (Pt 2):141–201. [PubMed: 9509577]

Eckstein K, Friederici AD. It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. J Cogn Neurosci. 2006; 18:1696–1711. [PubMed: 17014374]

Fadiga L, Craighero L, D'Ausilio A. Broca's area in language, action, and music. Ann N Y Acad Sci. 2009; 1169:448–458. [PubMed: 19673823]

Fedorenko E, Patel A, Casasanto D, Winawer J, Gibson E. Structural integration in language and music: evidence for a shared system. Mem Cognit. 2009; 37:1–9.

Friederici AD, Meyer M, von Cramon DY. Auditory langue comprehension: An event-related fMRI study on the processing of syntactic and lexical information. Brain and Language. 2000; 74:289–300. [PubMed: 10950920]

Friederici AD, Ruschemeyer SA, Hahne A, Fiebach CJ. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb Cortex. 2003; 13:170–177. [PubMed: 12507948]

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A. Representation of the temporal envelope of sounds in the human brain. J Neurophysiol. 2000; 84(3):1588–1598. [PubMed: 10980029]

Griffiths TD, Warren JD. The planum temporale as a computational hub. Trends Neurosci. 2002; 25:348–353. [PubMed: 12079762]

Grodzinsky Y, Santi A. The battle for Broca's region. Trends Cogn Sci. 2008; 12:474–480. [PubMed: 18930695]

Hagoort P. On Broca, brain, and binding: a new framework. Trends Cogn Sci. 2005; 9:416–423. [PubMed: 16054419]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001; 293:2425–2430. [PubMed: 11577229]

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8:393–402. [PubMed: 17431404]

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. Journal of Cognitive Neuroscience. 2003; 15:673–682. [PubMed: 12965041]

Humphries C, Willard K, Buchsbaum B, Hickok G. Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. Neuroreport. 2001; 12:1749–1752. [PubMed: 11409752]

Humphries C, Love T, Swinney D, Hickok G. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Human Brain Mapping. 2005; 26:128–138. [PubMed: 15895428]

Humphries C, Binder JR, Medler DA, Liebenthal E. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. J Cogn Neurosci. 2006; 18:665–679. [PubMed: 16768368]

Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR. Brain activation modulated by sentence comprehension. Science. 1996; 274:114–116. [PubMed: 8810246]

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 2005; 8:679–685. [PubMed: 15852014]

Levitin DJ, Menon V. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. Neuroimage. 2003; 20:2142–2152. [PubMed: 14683718]

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cereb Cortex. 2005; 15:1621–1631. [PubMed: 15703256]

Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007; 54:1001–1010. [PubMed: 17582338]

Mazoyer BM, Tzourio N, Frak V, Syrota A, Murayama N, Levrier O, Salamon G, Dehaene S, Cohen L, Mehler J. The cortical representation of speech. Journal of Cognitive Neuroscience. 1993; 5:467–479.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 2006; 10:424–430. [PubMed: 16899397]

Novick JM, Trueswell JC, Thompson-Schill SL. Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. Cogn Affect Behav Neurosci. 2005; 5:263–281. [PubMed: 16396089]

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP. Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI. Cereb Cortex. 2006

Okada K, Hickok G. Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. Neuroreport. 2006; 17:1293–1296. [PubMed: 16951572]

Patel, AD. Music, Language, and the Brain. Oxford: Oxford University Press; 2007.

Patel, AD.; Gibson, E.; Ratner, J.; Besson, M.; Holcomb, P. Examining the specificity of the P600 ERP component to syntactic incongruity in language: A comparative event-related potential study of structural incongruity in language and music. Talk presented at the 9th annual conference on human sentence processing; City University of New York; 1996.

Peretz I, Zatorre RJ. Brain organization for music processing. Annu Rev Psychol. 2005; 56:89–114. [PubMed: 15709930]

Rogalsky C, Hickok G. Selective Attention to Semantic and Syntactic Features Modulates Sentence Processing Networks in Anterior Temporal Cortex. Cereb Cortex. 2009; 19:786–796. [PubMed: 18669589]

Rogalsky C, Matchin W, Hickok G. Broca's Area, Sentence Comprehension, and Working Memory: An fMRI Study. Front Hum Neurosci. 2008; 2:14. [PubMed: 18958214]

Santi A, Grodzinsky Y. Working memory and syntax interact in Broca's area. Neuroimage. 2007; 37:8–17. [PubMed: 17560794]

Schonwiesner M, Zatorre RJ. Spectro-temporal modulation transfer function of single voxels in human auditory cortex measured with high-resolution fMRI. Proc Natl Acad Sci USA. 2009; 106(34): 14611–14616. [PubMed: 19667199]

Serences JT, Boynton GM. Feature-based attentional modulations in the absence of direct visual stimulation. Neuron. 2007; 55:301–312. [PubMed: 17640530]

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270:303–304. [PubMed: 7569981]

Stewart L, von Kriegstein K, Warren JD, Griffiths TD. Music and the brain: disorders of musical listening. Brain. 2006; 129:2533–2553. [PubMed: 16845129]

Stromswold K, Caplan D, Alpert N, Rauch S. Localization of syntactic comprehension by positron emission tomography. Brain and Language. 1996; 52:452–473. [PubMed: 8653390]

Tillmann B, Janata P, Bharucha JJ. Activation of the inferior frontal cortex in musical priming. Brain Res Cogn Brain Res. 2003; 16:145–161. [PubMed: 12668222]

Vaden KI Jr, Muftuler LT, Hickok G. Phonological repetition-suppression in bilateral superior temporal sulci. Neuroimage. 2009

Vandenberghe R, Nobre AC, Price CJ. The response of left temporal cortex to sentences. J Cogn Neurosci. 2002; 14:550–560. [PubMed: 12126497]

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci. 2001; 13:994–1005. [PubMed: 11595101]

Wang X, Lu T, Liang L. Cortical processing of temporal modulations. Speech Commun. 2003; 41:107–121.

Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. Trends in Cognitive Sciences. 2002; 6:37–46. [PubMed: 11849614]
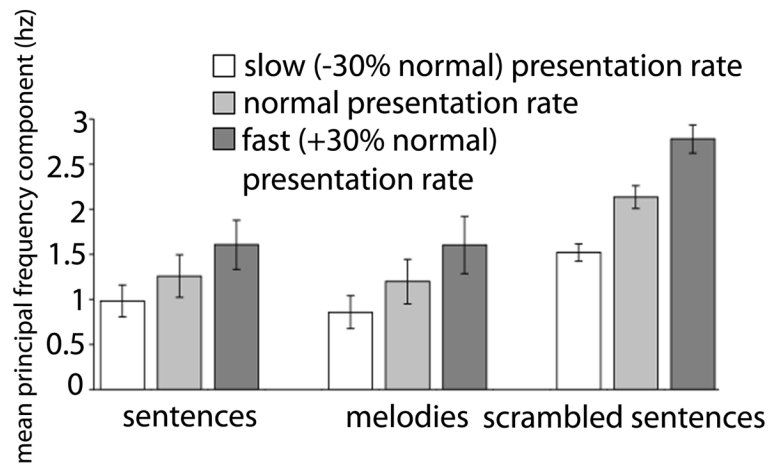
Figure 1.
Mean principal frequency component (i.e. modulation rate) of each stimulus type's acoustic envelope, within each presentation rate. Error bars represent 95% confidence intervals.
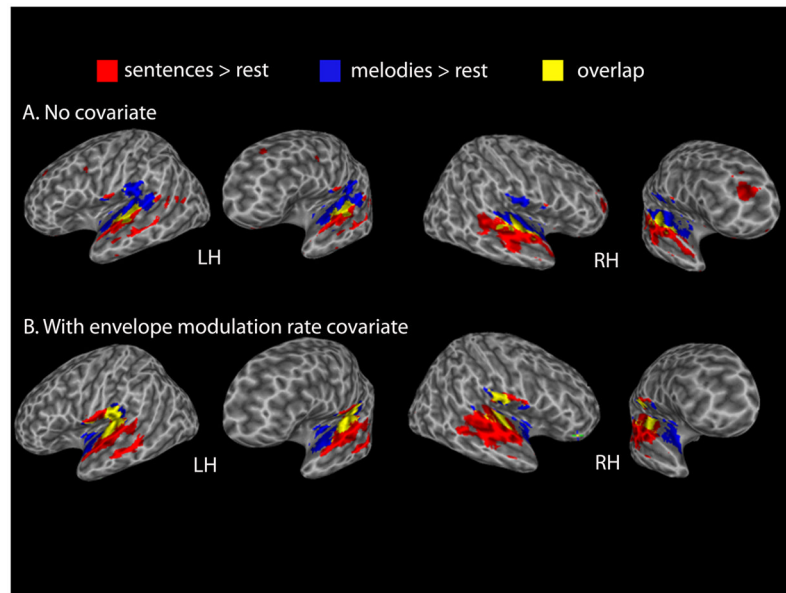
Figure 2.
Voxels more active (p < .005) for sentences versus rest (red), melodies versus rest (blue), and overlap between the two (yellow): *A*. across all presentation rates, *B*. also across all presentation rates, but with the envelope modulation rate information included as a covariate in the regression analysis.
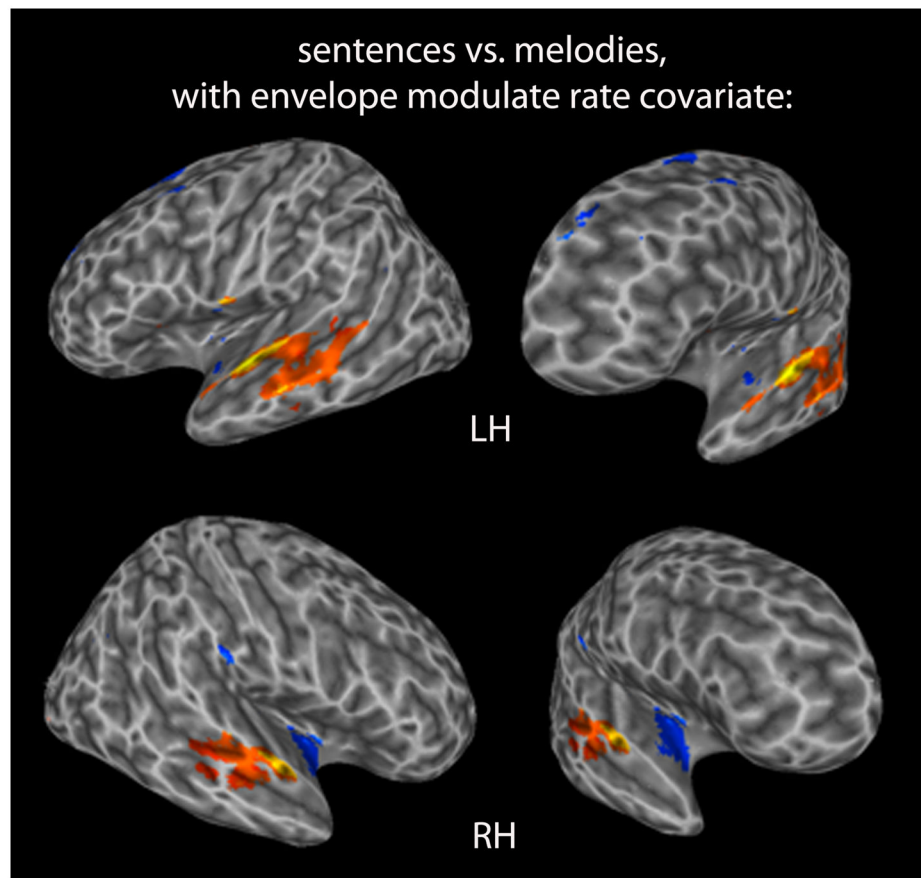
Figure 3.
Voxels more active for sentences than melodies (warm colors), and more active for melodies than sentences (cool colors) with the envelope modulation rate covariate, p < .005.
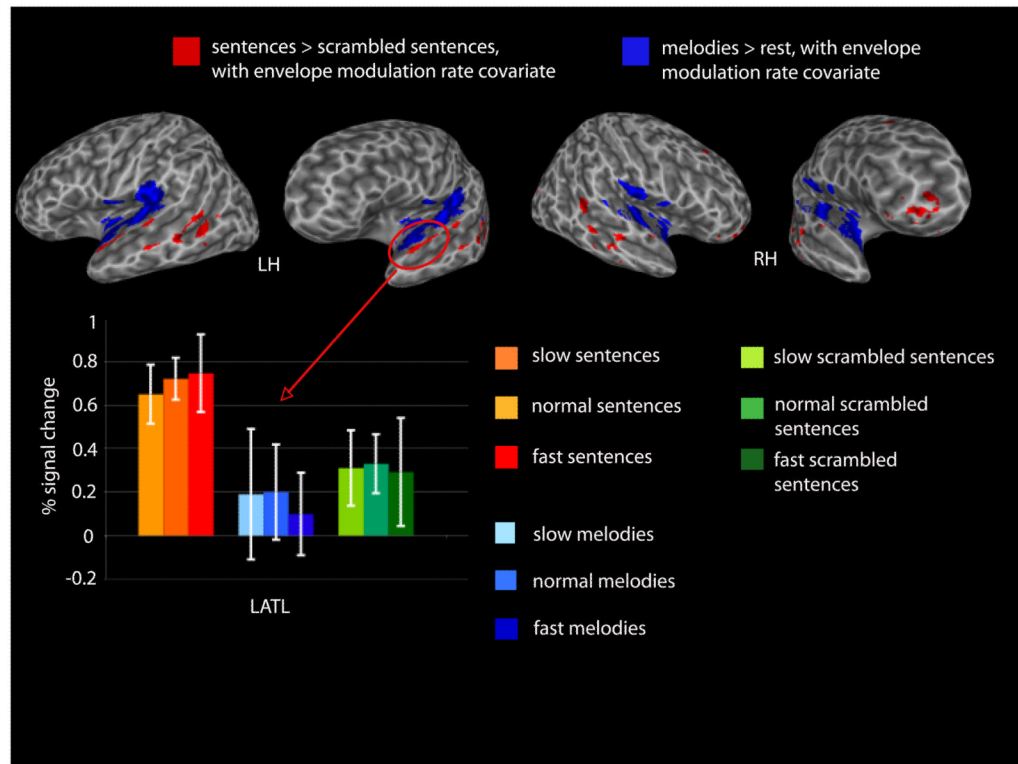
Figure 4.
Top. Voxels more active for sentences than scrambled sentences (red), and more active for melodies than rest (blue), with the envelope modulation rate covariate, p < .005. *Bottom.* Mean peak amplitudes for each stimulus type, at each presentation rate for the left anterior temporal ROI more active for active for sentences than scrambled sentences. Error bars represent 95% confidence intervals.

overlap of sentences > scrambled sentences & sentences > melodies, with envelope modulation rate covariate:

LH   RH
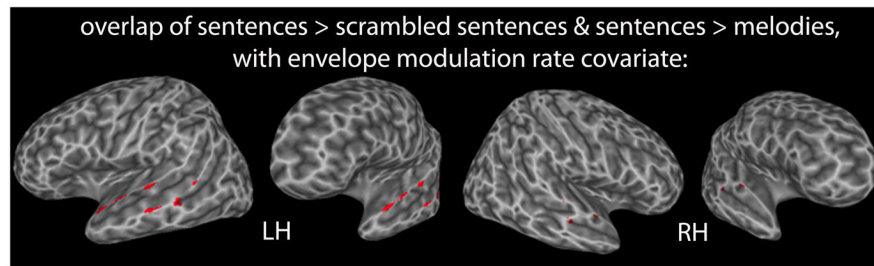
Figure 5.
Overlap (red) of voxels passing threshold for both the sentences > scrambled sentences and for the melodies > rest repeated-measures t-tests, with the envelope modulation rate covariate, p < .005.
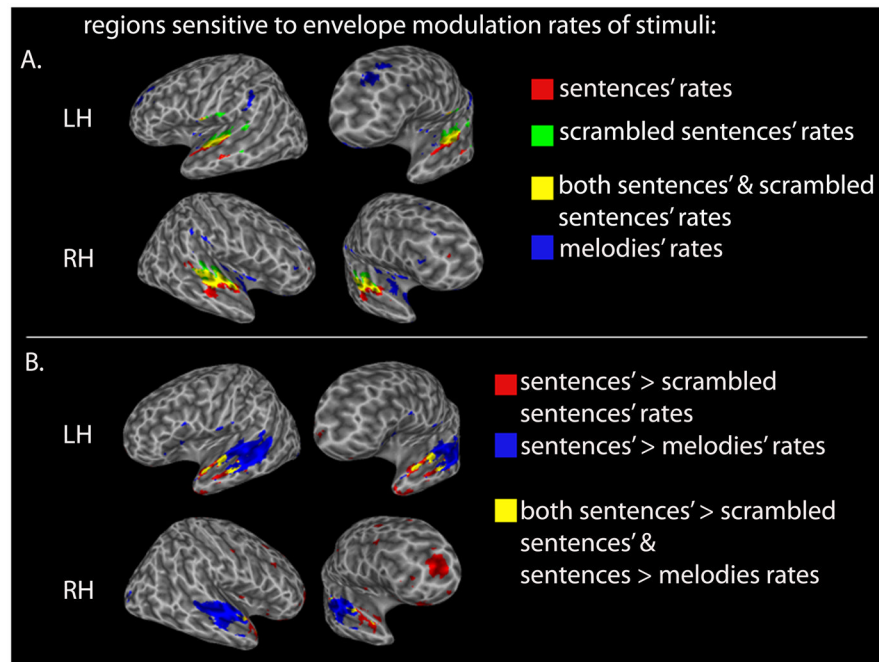
Figure 6.
Voxels whose activation is sensitive to the envelope modulation rates of sentences, scrambled sentences, and melodies, respectively, p < .005. Voxels sensitive to both sentences' and scrambled sentences' rates are also displayed (yellow). *B*. Voxels whose activation is more sensitive to the envelope modulation rate of the sentences than the scrambled sentences (red), more sensitive to the envelope modulation rate of sentences than melodies (blue), respectively, p < .005. Regions of overlap for the two contrasts are also displayed (yellow).

Table 1

Summary of descriptive statistics and independent-samples t-tests (df = 82) comparing the mean peak modulation rate of each stimulus type's acoustic envelopes.

slow (−30%) presentation rate:	*M/sd*
sentences	0.98/0.59
melodies	0.86/0.60
scrambled sentences	1.52/0.32

	t/p
sentences vs scrambled sentences	11.6/< .00001
sentences vs melodies	3.35/.0012

normal rate:	*M/sd*
sentences	1.26/0.78
melodies	1.20/0.82
scrambled sentences	2.13/0.42

	t/p
sentences vs scrambled sentences	14.1/< .00001
sentences vs melodies	0.34/0.73

fast (+30% normal) rate:	*M/sd*
sentences	1.60/0.90
melodies	1.60/1.05
scrambled sentences	2.79/0.52

	t/p
sentences vs scrambled sentences	12.1/< .00001
sentences vs melodies	0.01/0.99

Table 2

Talairach coordinates for voxel clusters that passed threshold (p < .005) for the listed contrasts, averaged across trials and subjects. The t values and Talairach coordinates of each cluster's peak voxel also are listed.

region	Brodmann area(s)	center of mass			peak t-score in ROI for contrast			
		x	y	z	t score	x	y	z
w/no frequency covariate:								
sentences > rest								
L STG/MTG/TTG	21/22/41/42	−55	−11	3	9.33	−60	−9	5
L STG/MTG/TTG	21/22	−53	−44	9	4.77	−57	−36	9
L PCG	6	−31	0	30	3.78	−31	2	27
L MFG	9	−25	32	25	4.03	−21	35	23
R STG/MTG	21/22/41/42	58	−14	0	7.65	59	−6	−4
R MFG	9	18	32	26	3.72	29	36	13
melodies > rest								
L STG/TTG/insula/PCG	13/22/40/41/42	−48	−19	9	10.62	−46	−20	7
R STG/TTG/insula/PCG	13/22/40/41/42	49	−14	5	12.13	51	−4	2
overlap of sentences > rest & melodies > rest								
L STG/TTG	22/41	−51	−17	6				
L STG	22	−55	−38	11				
R STG/TTG	22/41	56	−16	5				
w/frequency covariate:								
sentences > rest								
L STG/MTG	22	−53	−13	5	8.7	−63	−19	4
R STG/MTG	22	58	−15	1	16.5	60	−15	0
melodies > rest								
L insula/STG/TTG	13/22/41	−43	−15	7	9.92	−43	−3	−1
R insula/TTG/STG	13/22/41	46	−7	6	6.8	45	−1	−4
overlap of sentences > rest & melodies > rest								
L STG/TTG	22/41	−43	−21	9				
R STG	22	49	−14	8				
sentences > melodies								
L STG/MTG	22	−56	−16	0	5.7	−61	−13	1
L STG	38/22	−52	7	−3	4.45	−52	9	−5
L PHG	20	−36	−9	−30	3.6	−35	−11	−30
R STG/MTG	21/22	57	−12	−2	4.88	63	−5	−2
R Caudate	--	19	−23	21	3.75	19	−25	23

region	Brodmann area(s)	center of mass			peak t-score in ROI for contrast			
		x	y	z	t score	x	y	z
melodies > sentences								
L insula	13	−44	−3	1	4.34	−43	−3	1
L thalamus	--	−11	−3	5	4.32	−7	−3	3
L SFG	9	−21	53	29	5.05	−21	59	29
L MFG/SFG	6	−21	16	55	4.86	−17	16	54
R insula	13	42	−2	−4	5.88	43	−2	−6
R IPL/PCG	40	63	−26	22	4.39	64	−25	21
sentences > scrambled sentences								
L STG	22	−53	−2	−4	23.05	−54	9	−5
L STG/MTG	21/22	−59	−21	2	20.59	−62	−17	5
L STG/MTG	21/22/37	−55	−47	1	12.94	−54	−48	−3
R STG	21/22	46	−44	11	8.12	39	−44	13
R MTG	21	56	−31	−2	8.88	56	−33	−3
R STG/MTG	21/22	58	−2	−4	21.06	65	−1	2
R MFG	10	34	47	−4	8.85	29	47	−7
overlap of sentences > scramb. sent. & sentences >melodies								
L STG	22/38	−53	7	−4				
L STG	22	−55	−5	0				
L STG/MTG	21/22	−57	−21	4				
L MTG	21	−57	−35	2				
R STG/MTG	21/22	53	−25	−2				
R STG	22	57	−3	2				
R Caudate	--	19	−24	24				

STG = superior temporal gyrus, MTG = middle temporal gyrus, TTG = transverse temporal gyrus, PCG = precentral gyrus, MFG = middle frontal gyrus, PHG = parahippocampal gyrus, SFG = superior frontal gyrus, IPL = intraparietal lobule, IFG = inferior frontal gyrus

Table 3

Talairach coordinates for voxel clusters that passed threshold (p < .005) for being significantly modulated by the envelope modulation rate of a given stimulus type, or for a contrast of modulation rate between two stimulus types.

region	Brodmann area(s)	center of mass			peak t-score in ROI for contrast			
		x	y	z	t score	x	y	z
sensitivity to frequency envelope of:								
sentences								
L STG/MTG	22	−57	−11	4	9.33	−61	−9	4
R STG/MTG	21/22	57	−15	−2	7.65	59	−7	−2
R MFG	10	33	45	10	3.63	33	43	10
melodies								
L insula/STG	13/22	−43	−5	2	6.54	−45	−3	2
L IPL	40	−53	−33	32	4.92	−51	−33	30
L MFG	9	−25	35	26	4.76	−27	33	38
L STG/MTG	21/22	−60	−14	−1	−6.74	−65	−14	1
L STG/MTG	21/22	−61	−40	2	−4.89	−62	−32	−10
L MFG	10	−37	47	−6	−4.31	−37	47	−10
L SFG	9	−13	57	32	−4.05	−15	59	32
R insula/STG	13/22	44	2	−5	5.28	45	4	−5
R STG	38	35	6	−19	4.99	36	6	−19
R STG/MTG	21/22	52	−13	−3	−4.75	50	−20	−2
R MFG	10	35	56	−3	−4.3	35	56	0
R SFG	10	7	69	9	−5.14	12	69	8
scrambled sentences								
L STG/TTG	22/42	−57	−8	9	8.23	−56	−9	5
L STG/MTG	22	−59	−36	9	5.1	−59	−34	11
R STG/MTG	21/22	58	−15	1	6.44	59	−15	−1
overlap of sentences & scramb. sent.								
L STG/MTG	22	−55	−11	4				
R STG/MTG	21/22	59	−15	2				
sentences > melodies								
L STG/MTG	21/22	−53	−27	2	9.23	−59	−13	2
L STG	22	−53	7	−10	5.32	−55	10	−10
L STG	38	−45	21	−16	4.38	−42	21	−16

region	Brodmann area(s)	center of mass			peak t-score in ROI for contrast			
		x	y	z	t score	x	y	z
L IFG	45/44/9	−55	13	22	3.78	−56	14	23
L IFG	47	−42	24	0	3.55	−43	25	−1
R STG/MTG	21/22	53	−21	−4	7.29	59	−7	−1
R STG	38	47	13	−16	3.75	47	15	−18
melodies > sentences								
L insula	13	−41	−4	−2	6.78	−41	−3	−2
R insula	13	43	0	−2	7.32	43	−1	−4
sentences > scramb. sent.								
L STG/MTG	1	−57	−6	−3	6.01	−58	1	−4
L STG/MTG	38	−34	15	−31	5	−34	4	−29
L SFG/MeFG	10	−14	54	4	6.74	−19	54	4
R STG/MTG	21/38	45	11	−18	4.93	44	11	−17
R MFG/SFG	10	18	49	4	6.36	21	49	2
R MFG	8	30	18	44	5.31	31	13	42
R PCG	6	43	−12	29	3.86	43	−12	30
R IFG	47	27	27	−4	8.99	26	31	−9
overlap of sentences > scramb. sent. & sentences >melodies								
L STG/MTG	21/22/38	−52	8	−9				
L STG/MTG	21/22	−53	−10	−6				
R STG/MTG	21/22	57	−2	−5				
R STG	38	46	15	−18				