A possible functional localizer for identifying brain regions sensitive to sentence-level prosody

Evelina Fedorenko[1], **Po-Jang Hsieh**[2], and **Zuzanna Balewski**[1]

[1]Brain and Cognitive Sciences Department, MIT

[2]Neuroscience and Behavioral Disorders Program, Duke-NUS Graduate Medical School

Abstract

Investigations of how we produce and perceive prosodic patterns are not only interesting in their own right but can inform fundamental questions in language research. We here argue that functional magnetic resonance imaging (fMRI) in general – and the functional localization approach in particular (e.g., Kanwisher et al., 1997; Saxe et al., 2006; Fedorenko et al., 2010; Nieto-Castañon & Fedorenko, 2012) – has the potential to help address open research questions in prosody research and at the intersection of prosody and other domains. Critically, this approach can go beyond questions like "where in the brain does mental process *x* produce activation" and toward questions that probe the nature of the representations and computations that subserve different mental abilities. We describe one way to functionally define regions sensitive to sentence-level prosody in individual subjects. This or similar "localizer" contrasts can be used in future studies to test hypotheses about the precise contributions of prosody-sensitive brain regions to prosodic processing and cognition more broadly.

Keywords

fMRI prosody

Introduction

Patterns of pitch and loudness in speech, as well as the ways in which words are grouped temporally, provide an important source of information in both language acquisition (e.g., Gleitman & Wanner, 1982; Jusczyk et al., 1993; Jusczyk, 1997; Mattys et al., 1999), and language processing (e.g., Marslen-Wilson et al., 1992; Bader, 1998). Investigating how people produce and/or perceive certain aspects of prosody is not only interesting in its own right, but can also inform broader issues in the architecture of human language. For example, investigations of prosody have been used to ask about the basic meaning units of language (e.g., Selkirk, 1984; cf. Watson & Gibson, 2004), or about whether we produce language with a comprehender in mind (e.g., Albritton et al., 1996; Snedeker & Trueswell, 2003; Breen et al., 2010; cf. Schafer, 2000; Kraljic & Brennan, 2005). Although we have learned a tremendous amount over the last several decades, some key questions about the

Send correspondence to: Evelina Fedorenko, MIT, 43 Vassar Street, 46-3037G, Cambridge, MA 02139, evelina9@mit.edu.

mechanisms that support prosodic processing remain unanswered (or at least are still actively debated). For example, are the same or distinct mechanisms used for processing prominence patterns vs. temporal grouping information in the speech signal? Does extracting prosodic information from speech rely on specialized cognitive and neural machinery, or is it instead supported by some of the mechanisms that are engaged by other mental processes, like musical processing or social cognition? The answers to these questions would importantly constrain the possibilities for the kinds of representations that mediate prosodic processing, which might in turn allow us to tackle even more challenging questions. For example, how do prosodic representations interact with syntactic/semantic representations in both constructing utterances in the course of production and extracting meaning from the linguistic signal in the course of comprehension? In the current paper we will i) argue that functional magnetic resonance imaging (fMRI) is a powerful – and currently under-used in the domain of prosody – tool that can help address some of these open questions; and ii) describe an fMRI approach that we think is promising for doing so.

The rest of the paper is organized as follows: We begin with a brief summary of previous investigations of the brain basis of prosodic processing. We then outline an fMRI approach, which has been successful in other domains but has not yet been applied in the domain of prosody. In particular, we argue for the importance of defining regions of interest functionally in individual subjects (e.g., Saxe et al., 2006; Fedorenko et al., 2010). We then motivate and present the current study, which provides one possible way to identify brain regions sensitive to sentence-level prosody. We conclude with a discussion of how this (or similar) "functional localizers" can be used in future work to tackle theoretically important questions in the domain of prosody and beyond.

Previous investigations of the brain basis of prosodic processing

As in other cognitive domains, two primary sources of evidence have informed our understanding of how prosodic processing is instantiated in the brain: a) studies of patients with brain damage, and b) neuroimaging (PET, fMRI) investigations. Both kinds of studies have implicated a large number of cortical regions in the temporal, frontal and parietal lobes, as well as some subcortical structures (for reviews see e.g., Baum & Pell, 1999; Van Lancker & Breitenstein, 2000; Wong, 2002; Friderici & Alter, 2004, Kotz et al., 2006; Van Lancker Sidtis et al., 2006; Wildgruber et al., 2006).

A question that has perhaps received the most attention in the literature is that of lateralization of prosodic processing. Some early patient findings have suggested that deficits in prosody perception and production – especially affective prosody – typically arise after damage to the right hemisphere (e.g., Heilman et al., 1975; Tucker et al., 1977; Ross & Mesulam, 1979; Ross, 1981; Weintraub et al., 1981; Bowers et al., 1987; Bryan, 1989; Bradvik et al., 1991; Darby, 1993; Starkstein et al., 1994; Dykstra et al., 1995; Schmitt et al., 1997; Ross et al., 1997; see also Blumstein & Cooper, 1974, Ley & Bryden, 1982, and Herrero & Hillix, 1990, for behavioral evidence of right hemisphere superiority for prosodic processing from healthy individuals). The right hemisphere superiority for processing affective prosody fits nicely with work on emotional processing in other domains (e.g., Sackheim et al., 1978; Strauss & Moscovitch, 1981; cf. Caltagrione et al., 1989; Kowner,

1995). However, other studies have shown that the left hemisphere also plays an important role, especially for linguistic prosody, leading to the argument that both hemispheres may be important (e.g., Blumstein & Goodglass, 1972; Zurif & Meldelsohn, 1972; Schlanger et al., 1976; Goodglass & Calderon, 1977; Van Lancker, 1980; Baum et al., 1982; Seron et al., 1982; Heilman et al., 1984; Speedie et al., 1984; Shapiro & Danly, 1985; Emmorey, 1987; Behrens, 1988; Van Lancker & Sidtis, 1992; Pell & Baum, 1997; Wertz et al., 1998). In a recent review and meta-analysis of neuropsychological investigations, Witterman et al. (2011; see also Kotz et al., 2006) conclude that both hemispheres are necessary for both emotional and linguistic prosodic perception, but the right hemisphere plays a relatively greater role in processing emotional prosody. In particular, right hemisphere damage leads to a) greater deficits in emotional, compared to linguistic, prosody, and b) greater deficits in emotional prosody, compared to left hemisphere damage. Consistent with this review, neuroimaging studies often find bilateral cortical activations for prosodic (including affective prosodic) manipulations, as well as some additional subcortical structures (e.g., George et al., 1996; Imaizumi et al., 1997; Phillips et al., 1998; Kotz et al., 2003; Grandjean et al., 2005).

In summary, although many cortical and subcortical brain regions in both hemispheres have been implicated in prosodic processing and a number of proposals have been advanced for the functions of these regions (e.g., Friederici & Alter, 2004; Ethofer et al., 2006b; Van Lancker Sidtis et al., 2006; Wildgruber et al., 2008; Kotz & Schwartze, 2010), most researchers agree that more work is needed in order to understand the precise contributions of these different regions to perceiving and producing prosody.

Functional localization as an alternative to the traditional fMRI approach

Many previous neuroimaging studies have focused on the question of *which hemisphere* or *which particular brain region* is engaged by some aspect(s) of prosodic processing. This kind of a question is a necessary starting point, but the ultimate goal of cognitive science and cognitive neuroscience is to understand the function(s) of each relevant component of the mind/brain. In particular, for any given brain region, we would like to know what kinds of knowledge representations it stores and works with, and/or what computations it performs on particular stimuli. To be able to answer – or at least begin to answer – these questions, multiple hypotheses need to be evaluated about each key brain region. As a result, no single study will be sufficient. In order to accumulate knowledge across studies and labs, it is important to be able to refer to the "same" region from one brain to the next. We have recently been arguing that the traditional fMRI approach is not well suited for comparing results across studies as needed for accumulating knowledge (e.g., Fedorenko & Kanwisher, 2009; Fedorenko et al., 2010, 2012a). In particular, in the traditional group-based approach, brains are aligned in the common stereotaxic space and activation overlap is examined across individual brains. However, because of anatomical variability (e.g., Brodmann, 1909; Geschwind and Levitsky, 1968; Ono et al., 1990; Zilles, 1997; Amunts et al., 1999; Tomaiuolo et al., 1999; Miller et al., 2002; Wohlschlager et al., 2005; Juch et al., 2005), individual activations do not line up well across brains, especially in the frontal and temporal lobes (e.g., Frost & Goebel, 2011; Tahmasebi et al., 2011). Consequently,

locations (e.g., sets of {x,y,z} coordinates) in the stereotaxic space are not optimally suited for comparing results across individuals and studies.

For example, imagine a scenario where one study reports activation in or around some anatomical location (e.g., superior temporal gyrus, STG) for a manipulation of affective prosody, and another study reports a nearby location (also within the STG) for a manipulation of linguistic prosody. Based on this pattern, one could arrive at two opposite conclusions about the relationship between affective and linguistic prosody. On the one hand, it could be argued that the two locations are close enough to each other (falling within the same broad anatomical region) to count as the "same" region, which would imply that affective and linguistic prosody rely on the same mechanisms. On the other hand, it could be argued that because the activations do not fall in exactly the same coordinates in the stereotaxic space, they are two nearby but distinct regions, which would imply that affective and linguistic prosody are supported by different mechanisms. We know that in many parts of the brain small but functionally distinct regions lie side by side (e.g., the fusiform face area and the fusiform body area - Schwarzlose et al., 2005; or different regions within the left inferior frontal gyrus – Fedorenko et al., 2012b). Consequently, without comparing the two manipulations to each other in the same individual, it is impossible to determine which interpretation is correct.

An approach that has been proposed as an alternative to the traditional fMRI approach involves i) identifying regions of interest *functionally* in each individual brain (i.e., regions that exhibit a particular functional signature), and then ii) probing the functional profiles of those regions in additional studies in an effort to narrow down the range of possible hypotheses about their function(s). For example, using a contrast between faces and objects, Kanwisher et al. (1997) identified a region in the fusiform gyrus that responds more strongly during the processing of faces than during the processing of objects. This region can be robustly found in any individual brain in just a few minutes of scanning. Then, across many subsequent studies, the responses of this region were examined to many new stimuli and tasks to try to understand what drives the stronger response to faces (e.g., Kanwisher et al., 1998; Downing et al., 2006). Because the same "localizer" task (the faces > objects contrast in this example) is used across studies, the results can be straightforwardly compared. This "functional localization" approach is the standard approach in the field of vision research, and it has recently been successfully extended to other domains (e.g., social cognition – Saxe & Kanwisher, 2003; speech perception – Belin et al., 2000; Hickok et al., 2009; language – Pinel et al., 2007; January et al., 2009; Fedorenko et al., 2010). In addition to facilitating knowledge accumulation, the functional localization approach yields higher sensitivity and functional resolution, i.e., it is more likely to detect an effect when it is present, and it is better at distinguishing between nearby functionally different regions, which is especially important for addressing questions of functional specificity (e.g., Nieto-Castañón & Fedorenko, 2012).

In summary, the functional localization approach is more conducive to asking deep questions about the nature of the representations and computations underlying a particular mental process, compared to the traditional fMRI approach[1]. We therefore advocate the adoption of this approach for investigating prosodic processing. A prerequisite for this

approach is a "localizer" task that can identify domain- or process-relevant regions at the level of individual subjects. We here propose one possible "localizer" for brain regions sensitive to sentence-level prosody.

Experiment

It is not obvious what experimental contrast(s) are best suited for discovering brain regions sensitive to prosodic processing. Previous studies have used several approaches: i) stimulus manipulations where different kinds of prosodic contours are compared to one another (e.g., Kotz et al., 2003; Doherty et al., 2004; Grandjean et al., 2005); ii) stimulus manipulations where a prosodic contour is compared to some control condition(s) where some aspects of prosody are degraded (e.g., Humphries et al., 2005; Wiethoff et al., 2008; Zhao et al., 2008; Newman et al., 2010); and iii) task manipulations where participants perform either a task that draws attention to prosody (e.g., classifying the emotion that the intonation is conveying) vs. some control task on the same stimuli (e.g., George et al., 1996; Buchanan et al., 2000; Gandour et al., 2003; Mitchell et al., 2003; Ethofer et al., 2006a,b). The first approach only makes sense if we assume that the neural representations/processes of different prosodic contours (either different affective contours, like happy vs. sad prosody, or different linguistic prosodic contours, like statements vs. questions) are spatially distinct. It is not clear that such an assumption is warranted. And if the same patch of cortex processes different kinds of contours, then any contrast of this sort would not reveal those regions. For example, the fusiform face area (FFA; Kanwisher et al., 1997) responds to a wide range of face stimuli, and contrasting e.g., sad vs. neutral faces or male vs. female faces would miss this region.

The second and third approaches seem more promising. Any brain region engaged in prosodic processing should respond more when the signal contains a prosodic contour, compared to when some features of this contour are present to a lesser extent or absent. Similarly, brain regions engaged in prosodic processing may be expected to respond more when the task requires paying attention to the prosodic features of the stimulus compared to some other features, although this approach may not work well if perceiving prosody is highly automatic. In that case, prosody-sensitive regions would be engaged to a similar extent regardless of the specific task and may thus be subtracted out in a task-based contrast. Consistent with this notion, some studies that have used task manipulations report activations in what appear to be the highly domain-general regions of the fronto-parietal network (e.g., Buchanan et al., 2000), which respond across a wide range of cognitive demands and which are generally sensitive to salient stimuli (e.g., Duncan & Owen, 2000; Fedorenko et al., 2013; see Duncan, 2010, for a recent review of this brain system). As a result of these potential concerns with the stimulus manipulations that compare different prosodic contours, and task manipulations, in the current study we chose to use a stimulus manipulation that compares stimuli that have sentence prosody to those that do not.

---

[1]In principle, studies that ask questions like "does a particular manipulation activate brain region *x*?" could also inform deep issues in cognitive science, but this is only possible in cases where region *x* is characterized sufficiently well to serve as a neural "marker" of a particular mental process. With a few exceptions, most brain regions lack such detailed functional characterization and thus are not suitable for use as markers of particular mental processes (see e.g., Poldrack, 2006).

The design used here has not been previously used to the best of our knowledge. We used "structured" linguistic stimuli (sentences and Jabberwocky sentences, which obey the rules of English syntax but use pseudowords instead of the content words) and "unstructured" linguistic stimuli (lists of words and lists of pseudowords; see (2) below for sample stimuli)[2]. These linguistic materials were presented visually and auditorily. The contrast between structured and unstructured linguistic stimuli has some features that are the same regardless of the modality of presentation. For example, structured stimuli contain syntactic information and compositional semantic information, and that holds for both visually and auditorily presented materials. Importantly, however, auditorily presented structured stimuli – read naturally (cf. Humphries et al., 2005) – involve sentence-level prosodic contours, which are not present, or present to a lesser degree, in the unstructured stimuli (see Methods), as shown schematically in (1). Thus, subtracting the unstructured conditions from the structured conditions in the auditory materials should activate brain regions sensitive to sentence-level prosodic contours. But this subtraction also includes whatever makes the structured materials structured (i.e., syntax and compositional semantics). In order to isolate the prosody-relevant component of auditory structured stimuli, we contrasted the "structured > unstructured" comparison for the auditorily presented stimuli with the same comparison for the visually presented stimuli.

(1) A schematic illustration of the logic of the experimental design (Structured stimuli: sentences, Jabberwocky sentences; Unstructured stimuli: word lists, pseudoword lists):

	Structured > Unstructured VISUAL presentation	Structured > Unstructured AUDITORY presentation
Syntax	+	+
Compositional semantics	+	+
Sentence-level prosody	–	+

Although silent reading of sentences has been argued to activate prosodic representations (Fodor, 1998; see e.g., Bader, 1998; Hirose, 2003; Swets et al., 2007, for experimental evidence), previous work on visual imagery has established that to the extent that mental simulations of a particular sensory experience activate the corresponding sensory cortices, these activations are not nearly as robust as those elicited by actual sensory stimulation (e.g., O'Craven & Kanwisher, 2000). With respect to our design, this finding suggests that the sentence-level prosodic response to visually presented sentences and Jabberwocky sentences will be much weaker than to the same materials presented auditorily. Consequently, we reasoned that we can target brain regions that are sensitive to sentence-level prosodic contours with a conjunction of two contrasts: i) a greater response to structured than unstructured auditory stimuli, and ii) no difference between structured and unstructured visual stimuli[3]. In other words, this conjunction of contrasts is aimed at brain regions that *selectively* respond to the presence of structure in the auditory stimuli.

---

[2]Note that this experiment was not originally designed to study prosodic processing. Hence the inclusion of both meaningful (Sentences, Word lists) and meaningless (Jabberwocky, Pseudoword lists) conditions may seem unmotivated. However, we think it ends up being a *strength* of this experiment to be able to generalize across the presence of meaning in the stimuli: as we will show in the Results, similar patterns hold for meaningful and meaningless conditions when examined separately.

To discover these prosody-sensitive[4] regions we use a method that was recently developed for investigating high-level language regions (Fedorenko et al., 2010) and subsequently validated on the well-known ventral visual stream regions (Julian et al., 2012; see also Fedorenko et al., 2012c, for the application of this method to musical processing). This method – the group-constrained subject-specific (GSS) method – is an alternative to the traditional random-effects analysis, where individual activation maps are aligned in the common space and a *t*-test is performed in every voxel. The GSS analysis discovers patterns that are spatially systematic across subjects without requiring voxel-level overlap, thus accommodating inter-subject variability in the anatomical location of functional regions. This method thus improves sensitivity and functional resolution in cases where the effects are anatomically variable (Nieto-Castañón & Fedorenko, 2012).

Methods

Participants

12 participants (7 females) between the ages of 18 and 30 – students at MIT and members of the surrounding community – were paid for their participation[5]. Participants were right-handed native speakers of English. All participants had normal or corrected-to-normal vision and were naïve to the purposes of the study. All participants gave informed consent in accordance with the requirements of the Internal Review Board at MIT.

Design, materials and procedure

Each participant performed several runs of the visual version of the experiment and several runs of the auditory version of the experiment. The entire scanning session lasted between 1.5 and 2 hours.

Materials—There were four types of stimuli: sentences, word lists, Jabberwocky sentences, and pseudoword lists. 160 items were constructed for each condition, and each item was 8 words/pseudowords long. Sample items for each type of stimulus are shown in (2) below. For details of how the materials were created, see Fedorenko et al. (2010; Experiments 2-3). All the materials are available from: http://web.mit.edu/evelina9/www/funcloc.html. After each stimulus, a word (for Sentences and Word-lists conditions) or a pseudoword (for Jabberwocky and Pseudoword-lists conditions) appeared and participants were asked to decide whether this word/pseudoword appeared in the immediately preceding stimulus. Participants were instructed to press one of two buttons to respond. (In previous work, we have established that activations in the language-sensitive brain regions are similar

[3]One important caveat to keep in mind is that individuals may differ with respect to how strongly they activate prosodic representations during silent reading. In the extreme case, if an individual activated prosodic representations during silent reading to the same degree as during auditory linguistic processing, then the contrast proposed here would fail to identify any prosody-sensitive regions in that individual. As will be shown below, the proposed contrast successfully identifies regions with the specified functional properties in the majority of individuals, suggesting that a substantial proportion of individuals have brain regions that respond more to the presence of structure in the auditory stimuli than in the visual stimuli (which we argue plausibly reflects sensitivity to sentence-level prosody). Once these prosody-sensitive brain regions are established as robust to irrelevant differences in the materials, task, etc., investigating individual differences in their response profiles and relating them to behavior will be a fruitful avenue for future research.

[4]Although in the remainder of the paper we will use the term "prosody-sensitive", the reader should keep in mind that we are referring to brain regions that are sensitive to *sentence-level prosodic contours*.

[5]The dataset used for the current study is the same dataset as that used in Experiment 3 in Fedorenko et al. (2010).

regardless of whether this memory probe task is included or whether participants are simply reading the materials with no task (Fedorenko et al., 2010; Fedorenko, submitted).)

(2) Sample items:

Sentences:	THE DOG CHASED THE CAT ALL DAY LONG
	A RUSTY LOCK WAS FOUND IN THE DRAWER
Word lists:	BECKY STOP HE THE LEAVES BED LIVE MAXIME'S
	FOR THE JUICE UP AROUND GARDEN LILY TRIES
Jabberwocky:	THE GOU TWUPED THE VAG ALL LUS RALL
	A CLISY NYTH WAS SMASP IN THE VIGREE
Pseudoword lists:	BOKER DESH HE THE DRILES LER CICE FRISTY'S
	FOR THE GRART UP AROUND MEENEL LALY SMEBS

128 of the 160 items were used for the auditory versions of the materials (fewer materials were needed because four, not five, items constituted a block of the same length as the visual presentation, because visual presentation is typically faster than the average speaking rate). The materials were recorded by a native speaker using the Audacity software, freely available at http://audacity.sourceforge.net/. The speaker was instructed to produce the sentences and the Jabberwocky sentences with a somewhat exaggerated prosody. These two conditions were recorded in parallel in order to make the prosodic contours across each pair of a regular sentence and a Jabberwocky sentence as similar as possible. The speaker was further instructed to produce the Word-lists and Pseudoword-lists conditions not with a list intonation, where each word/pseudoword is a separate intonational phrase (e.g., Schubiger, 1958; Couper-Kuhlen, 1986), but in a way that would make them sound more like a continuous stream of speech. This was done to make the low-level acoustic properties (e.g., frequency of boundaries in a stimulus) more similar between the structured and the unstructured stimuli (see Discussion and Appendix C for the results of the acoustic analyses of the materials, and for sample pitch tracks). We reasoned that with this recording strategy the auditory structured stimuli would only differ from the auditory unstructured stimuli in the presence of intonation patterns typical of English sentences (present in the structured stimuli vs. absent – or present to a lesser extent – in the unstructured stimuli).

Visual presentation—Words/pseudowords were presented in the center of the screen one at a time in all capital letters. No punctuation was included in the Sentences and Jabberwocky conditions, in order to minimize differences between the Sentences and Jabberwocky conditions on the one hand, and the Word-lists and Pseudoword-lists conditions on the other. Each trial lasted 4800 ms, which included (a) a string of 8 words/ pseudowords each presented for 350 ms, (b) a 300 ms fixation, (c) a memory probe appearing on the screen for 350 ms, (d) a period of 1000 ms during which participants were instructed to press one of two buttons, and (e) a 350 ms fixation. Participants could respond any time after the memory probe appeared on the screen. There were five trials in each block.

Auditory presentation—Stimuli were presented over scanner-safe earphones. Each trial lasted 6000 ms, which included (a) the stimulus (whose total duration varied between 3300

ms and 4300 ms), (b) a 100 ms beep tone indicating the end of the sentence, (c) a memory probe presented auditorily (maximum duration 1000 ms), (d) a period (lasting until the end of the trial) during which participants were instructed to press one of two buttons. Participants could respond any time after the onset of the memory probe. There were four trials in each block.

Each run lasted 464 sec (7 min 44 sec) and consisted of 16 24-second blocks, grouped into four sets of 4 blocks with 16-second fixation periods at the beginning of the run and after each set of blocks. Condition order was counterbalanced across runs and participants, and auditory and visual runs were alternated. Each item was presented in either visual or auditory modality, but not both; which items were presented in which modality varied across participants. Each participant, except for one, completed four visual and four auditory runs. The remaining participant completed 2 visual and 7 auditory runs.

fMRI data acquisition

Structural and functional data were collected on the whole-body 3 Tesla Siemens Trio scanner with a 12-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted structural images were collected in 128 axial slices with 1.33 mm isotropic voxels (TR = 2000 ms, TE = 3.39 ms). Functional, blood oxygenation level dependent (BOLD), data were acquired using an EPI sequence (with a 90 degree flip angle), with the following acquisition parameters: thirty-two 4 mm thick near-axial slices acquired in the interleaved order (with 10% distance factor), 3.1 mm × 3.1 mm in-plane resolution, FoV in the phase encoding (A >> P) direction 200 mm and matrix size 96 mm × 96 mm, TR=2000 ms and TE=30 ms. The first eight seconds of each run were excluded to allow for steady state magnetization.

fMRI data analyses

MRI data were analyzed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm) and custom matlab scripts (available from http://web.mit.edu/evelina9/www/funcloc.html). Each subject's data were motion corrected and then normalized into a common brain space (the Montreal Neurological Institute, MNI template) and resampled into 2mm isotropic voxels. Data were smoothed using a 4 mm Gaussian filter, and high-pass filtered (at 200 seconds).

To identify brain regions sensitive to sentence-level prosodic contours, we performed a group-constrained subject-specific (GSS) analysis (Fedorenko et al., 2010; Julian et al., 2012). To do so, we first created – for each individual subject – a map containing voxels that satisfy the following two criteria: (i) a significant effect of structure, i.e., Sentences +Jabberwocky > Word-lists+Pseudoword-lists (at p<.01 uncorrected level) for the auditory materials; and ii) no significant effect of structure (at p<.1) for the visual materials. We then overlaid these maps to create a probabilistic overlap map, and then divided this map into "parcels" using the watershed image parcellation algorithm. This algorithm discovers key topographical features (i.e., the main "hills") of the activation landscape (see Fedorenko et al., 2010, for details). We then identified parcels which – when intersected with the individual maps – contained responses in at least 9/12 individual subjects (i.e., ~ 75%). 12 parcels satisfied this criterion. Finally, because there is currently less agreement about how

to interpret deactivations in fMRI, we selected a subset of these 12 parcels that responded to each of the i) auditory Sentences, and ii) auditory Jabberwocky sentences conditions reliably above the fixation baseline. Four parcels remained. [See Appendix A for figures showing all 12 parcels and their responses to the eight experimental conditions. Note that some of these parcels – not discussed below – look spatially similar to regions implicated in some previous studies of prosodic processing (e.g., medial frontal regions; e.g., Heilman et al., 2004; Alba-Ferrara et al., 2011) and may thus require further investigation.]

For each of the four key parcels, we estimated the response magnitude to each condition in individual subjects using an n-fold cross-validation procedure, so that the data used to define the functional regions of interest (fROIs) and to estimate the responses were independent (e.g., Vul & Kanwisher, 2010; Kriegeskorte et al., 2009). Each parcel from the GSS analysis (Fig. 1) was intersected with each subject's activation map containing voxels that satisfy the criteria described above (i.e., an effect of structure for auditory materials, and a lack of such an effect for visual materials) for all but one run of the data. All the voxels that fell within the boundaries of the parcel were taken as that subject's fROI (see Appendix B for figures showing sample individual fROIs). This procedure was iterated across all possible partitions of the data, and the responses were then averaged across the left-out runs to derive a single response magnitude per subject per region per condition (see Appendix A for responses to individual conditions). Statistical tests were performed on these values. Two contrasts were examined to test for sensitivity to structure in the visual and auditory conditions, respectively: a) Sentences+Jabberwocky > Word-lists+Pseudoword-lists for visual materials; and b) Sentences+Jabberwocky > Word-lists+Pseudoword-lists for auditory materials.

Results

We discovered four regions in which the majority of subjects showed a selective response to structure in the auditory materials (i.e., a greater response to sentences and Jabberwocky sentences than to word and pseudoword lists for auditory, but not for visual, materials). These regions (parcels) are shown in Figure 1 (see Table 1 for a summary of key properties) and include bilateral regions in the superior temporal poles, and bilateral regions in the posterior inferior temporal lobes. (Note that the activations of any individual subject within a parcel typically constitute only a small fraction of the parcel; see Table 1 for average sizes of the individual fROIs; see also Appendix B.) Each of the four regions was present in at least 9/12 subjects; the left temporal pole region was present in 10/12 subjects[6].

---

[6]Not being able to define a fROI in every single subject using the fixed-threshold approach – i.e., when parcels are intersected with thresholded individual activation maps – is not uncommon. For example, when developing a localizer for high-level language regions, Fedorenko et al. (2010) considered a region meaningful if it could be defined in 80% or more of individual participants (see also Julian et al., 2012, where 60% or more of individual participants is used as a criterion for selecting meaningful high-level visual regions). An alternative that would enable one to define a fROI in every single subject would be to move away from the fixed-threshold approach. In particular, once a region has "established itself" (i.e., once we know that it emerges consistently across people, is stable within individuals, and is robust to various properties of the localizer contrast), we can simply take the top – with respect to the *t*-values for the relevant functional contrast – 5 or 10% of voxels within some spatial constraint (defined anatomically or with the use of functional parcels obtained from a GSS analysis). This approach ensures that i) a fROI is defined in every individual (and thus the results are generalizable to the whole population, as opposed to the proportion of the population for whom the fROIs could be defined), and ii) fROIs are of the same size across individuals (see Nieto-Castañon & Fedorenko, 2012, for a discussion). The reason that we used the fixed-threshold approach in the current paper is that it is mathematically not trivial to use the top-n-voxels approach for the conjunction of multiple contrasts, which is what we use in the current study.

In Figure 2 we present the responses of these prosody-sensitive regions to the presence of structure in the visual and auditory conditions (estimated using cross-validation, as described in Methods). The sensitivity to structure in the auditory conditions is highly robust in each of these regions (ps<.001). However, none of the regions show an effect of structure in the visual conditions (ts<1.24; see Table 2 for the details of the statistics).

Discussion

Four brain regions in the temporal cortices were found to be sensitive to sentence-level prosodic contours, as evidenced by a stronger response to structured (sentences, Jabberwocky sentences) compared to unstructured (word lists, pseudoword lists) conditions for the auditory, but not visual, presentation. These regions include bilateral superior temporal pole regions, and bilateral regions in the posterior inferior temporal lobes. We now discuss a few theoretical and methodological points that the current study raises.

The prosody-sensitive regions discovered in the current study

What can we currently say about the four regions discovered in the current experiment? We begin by ruling out a few possibilities. *First*, these regions are not likely engaged in low-level auditory processing given that i) both structured and unstructured auditory stimuli require basic auditory analysis; and ii) they are not located in or around primary auditory cortex (e.g., Morosan et al., 2001).

Second, by design, we know that these four regions are not part of the "language network" (e.g., Binder et al., 1997; Fedorenko et al., 2010), whose regions respond to linguistic stimuli similarly across modalities (see also Braze et al., 2011). For example, in Figure 3 below we show sensitivity to structure in the visual vs. auditory conditions in the language regions (defined by a greater response to sentences than to lists of pseudowords, as described in Fedorenko et al., 2010). As can be clearly seen, all of the regions of the "language network" show highly robust sensitivity to structure in both modalities (all ps<.001). Note that this is in spite of the fact that language activations encompass extended portions of the temporal cortices, including anterior temporal regions, especially in the left hemisphere. It appears then that language and prosody-sensitive regions are located near each other but are nevertheless functionally distinct. This result is consistent with the findings from the neuropsychological literature that at least some aphasic individuals have intact prosodic abilities (e.g., Hughlings-Jackson, 1931; Schmitt et al., 1997).

It is worth noting, however, that one possibility that needs to be tested in future work is that the regions identified in the current study are, in fact, part of the language network, but are less sensitive to structure (i.e., syntax and combinatorial semantics) than the "core" language regions (e.g., Fig. 3 below). Sentence-level prosodic contours (present in the auditory materials) may make the structural information more salient thus recruiting these possibly less sensitive-to-structure brain regions. To evaluate this interpretation, the response in the regions reported here needs to be tested to stimuli that contain sentence-level prosodic contours but do not contain any linguistic information (e.g., using degraded auditory materials like those used in Meyer et al., 2002). A robust response to such stimuli would

suggest sensitivity to prosodic contours rather than reinforcement of the structural information in linguistic stimuli.

And *third*, we know that these regions are not part of the domain-general fronto-parietal "multiple-demand" network (e.g., Duncan, 2010) whose regions respond across a wide range of demanding cognitive tasks: regions of the multiple-demand network respond more to unstructured compared to structured stimuli (Fedorenko et al., 2012b, 2013). Again, this is in spite of the fact that demanding cognitive tasks frequently activate posterior inferior temporal cortices (e.g., Duncan, 2006).

So what do we think the prosody-sensitive regions discovered here might be doing? We hypothesize that at least some of the regions we report here may store typical prosodic contours. Their response may thus be driven by how well the prosodic contour of an incoming stimulus matches these stored prosodic "templates". This proposal is similar to Peretz et al.'s proposal for some of the music-sensitive brain regions, which are proposed to store pitch / rhythm patterns characteristic of the music that the person has been exposed to (e.g., Peretz & Coltheart, 2003; also Fedorenko et al., 2012c).

In order to better understand the acoustic features that may be associated with sentence-level prosodic contours, we performed a series of acoustic analyses on the auditory materials used in the current study. In particular, we identified word/pseudoword boundaries in each audio file using the Prosodylab-Aligner tool (Gorman et al., 2011), extracted a set of acoustic features from each word/pseudoword using Praat (Boersma & Weenink, 2005), and then analyzed those features statistically. The procedure and the results are described in detail in Appendix C, but we here highlight the key results.

Although, as intended in creating these materials, the structured and unstructured stimuli turned out to be quite well matched on a number of acoustic dimensions (e.g., mean pitch, a falling pitch and intensity pattern across the stimulus, etc.), we did observe some differences. First, the maximum pitch was higher, the minimum pitch was lower, and the power was lower in the structured compared to the unstructured conditions. And second, there was greater variability across the stimulus in the structured conditions than in the unstructured conditions with respect to duration, maximum pitch, and center pitch, and lower variability with respect to power.

These observed acoustic differences between structured and unstructured conditions provide some preliminary hints about the relevant features of the sentence-level prosodic contours that may be contributing to or driving the fMRI effects. As discussed above, in order to understand how exactly these regions contribute to prosodic processing, many studies are needed that would examine the responses of these functionally defined regions to a variety of new stimuli and tasks. For example, by manipulating different acoustic features above separately in a controlled fashion we would be able to narrow in on the ones that contribute the most to the observed effects.

In terms of relating the current findings to previous work on prosody, the superior temporal pole regions look similar to regions that have been reported in previous neuroimaging studies (e.g., Mayer et al., 2002; Humphries et al., 2005). For example, Humphries et al.

(2005) reported some regions in the vicinity of superior temporal pole that respond more to stimuli with sentence-like prosody (compared to those with list prosody) even in cases where the stimuli were not linguistically meaningful. That would be consistent with the possible interpretation above, that these regions store typical prosodic "templates" and respond when there is a match between a stimulus and these stored representations. The regions in the posterior inferior temporal lobes have not been previously reported in investigations of prosody to the best of our knowledge, except for one unpublished study (Davis et al., 2010).

A number of previous studies have reported activations in and around the regions we observed in the current study for linguistic contrasts that do not appear to have anything to do with prosodic processing. For example, activations in or near the superior temporal pole regions have been reported for violations of syntactic structure (e.g., Friederici et al., 2003), and effortful speech perception (e.g., Adank, 2012). And parts of the inferior posterior temporal cortex have been implicated in visual word recognition (e.g., Petersen et al., 1990; Polk & Farah, 1998; Cohen et al., 2000; Baker et al., 2007; for a review see e.g., McCandliss et al., 2003), recalling word-forms in non-alphabetical languages (e.g., Kawahata et al., 1988; Soma et al., 1989; Nakamura et al., 2000), and spelling (e.g., Rapcsak & Beeson, 2004). In examining these studies, we should keep in mind that there is no way to determine with certainty whether or not activated regions are the "same" as the regions we report here, as discussed in the Introduction. To find that out, one would need to examine activations for the different contrasts within the same individuals. Nevertheless, examining activations observed in and around the regions discussed here may be important in generating hypotheses to be evaluated in future work.

How good is the current "localizer" as a localizer for prosody-sensitive regions?

We want to stress that the "localizer" task proposed here is certainly not the only way to find prosody-sensitive brain regions, and we are not even arguing that this is necessarily the *best* contrast to use. The goal of the current study is to provide a proof of concept. In particular, we show that it is possible to define prosody-sensitive regions (with stable functional profiles, as indicated by the replicability of the effects with across-runs cross-validation) in individual subjects. This result suggests that the functional localization approach is feasible in prosody research.

As with any localizer, before investigating the responses of the functional ROIs reported here to new conditions, it is important to first establish the robustness of this localizer contrast. In particular, a good localizer should not be sensitive to changes in e.g., the specific materials used, in the particular speaker used, or in the specific task. Furthermore, only a subset of the conditions from the current study may suffice for fROI definition in future investigations. For example, in Figure 4 below, we show that qualitatively and quantitatively similar profiles obtain when only four of the eight conditions are used (i.e., only meaningful materials, i.e., sentences and word lists visually and auditorily presented – Fig. 4a, or only meaningless materials, i.e., Jabberwocky sentences and pseudoword lists visually and auditorily presented – Fig. 4b).

Other localizer contrasts might work just as well or better. There are different approaches to developing a functional localizer for a particular mental process / set of mental processes. One can imagine initially "casting a wide net" with a broad functional contrast. In the extreme, you can imagine starting with something like "listening to sentences > fixation". This contrast activates a wide range of brain regions including those in the primary and secondary auditory cortices, the whole language network (e.g., Fedorenko et al., 2010) and some of the regions in the "multiple-demand" network (e.g., Duncan, 2010). Across many studies, one could then try to narrow in on the parts of this extended set of brain regions that may be more specifically engaged in dealing with prosody by examining the responses of these various regions to new conditions. In the current study we chose a narrower starting point. As discussed in the beginning of the paper, this contrast may not (and almost certainly does not) include all of the brain regions important for prosodic processing, which may or may not a problem, depending on the goals of a particular research study / program. For example, brain regions that respond to implicit prosody would not be included in the current set given the design we chose.

As with any experimental approach, many possibilities are perfectly valid in using the functional localization approach, as long as the researcher is i) clear about what mental process(es) the localizer contrast targets; and ii) careful in interpreting the results in line with all the possibilities for what the regions could be responding to (see Saxe et al., 2006; Fedorenko et al., 2010, for discussions). For example, Wiethoff et al. (2008) have reported stronger neural responses in the right middle superior temporal gyrus (STG) for emotional compared to neutral prosodic stimuli. However, they then demonstrated that this greater response can be explained by a combination of arousal and acoustic parameters like mean intensity, mean fundamental frequency and variability of fundamental frequency. As a result, the region in the STG is engaged during the processing of emotional prosody but only by virtue of the acoustic characteristics of those stimuli. For any candidate prosody-sensitive region – including those reported here – it is important to consider all of the possible alternatives for what could be driving the response to prosody. The functional localization approach is perfectly suited for doing so, allowing for easy comparisons across studies and labs.

Some general guidelines for creating powerful yet quick "localizers" include the following (for general advice on fMRI designs, see e.g., Huettel et al., 2008):

a. Use a blocked, not event-related, design (blocked designs are much more powerful due to the additive nature of the BOLD signal; e.g., Friston et al., 1999; Birn et al., 2002).

b. Use as few conditions as possible (from that perspective, the localizer used in the current experiment is not ideal; this is because this study was originally designed to address different research questions than the one asked here).

c. Given that the recommended block length is between ~10 and ~40 seconds (e.g., Friston et al., 1999), use blocks that are as short as possible within this range. However, this choice also depends on how long individual trials are, because it is

also advisable to include as many trials in a block as possible. Typical blocks are between 16 and 24 seconds in duration.

d. Unless the manipulation you are examining is very subtle (which is probably not a good idea for a localizer contrast anyway), 10-12 blocks per condition is generally sufficient to obtain a robust effect.

e. It is generally advisable to distribute the blocks across two or more "runs" so that data can be easily split up for cross-validation (i.e., using some portion of the data to define the regions of interest, and the other portion to estimate the responses; see also Coutanche & Thompson-Schill, 2012).

Keeping in mind i) the guidelines above, and ii) the discussion at the beginning of the Experiment section, the best contrast for identifying prosody-responsive brain regions in future work may be a two-condition contrast between prosodically "intact" and prosodically degraded stimuli (e.g., Humphries et al., 2005; Wiethoff et al., 2008; Zhao et al., 2008; Newman et al., 2010). Our recommendation is also to remove lexical content and syntactic structure from the contrast by using either speech in an unfamiliar language or speech that has been filtered so as to remove linguistic information.

Do we really need individual functional ROIs?

To illustrate the importance of using individual fROIs, consider Figure 5 where we show the response obtained with individually defined fROIs for the region in the left superior temporal pole (same data as in Figure 2) vs. an anatomical region in a similar location where we simply extract the response from all the voxels in that ROI in each subject (see Saxe et al., 2006; Fedorenko et al., 2010, 2012a, for similar demonstrations). As can be clearly seen, although the general patterns of response are similar, the effects are much larger and more statistically robust for individually defined fROIs. Given that individual fROIs constitute only a small portion of the relevant parcel (see Table 1), this is not surprising: in the case of the subject-independent anatomical ROI many voxels that are included in the analysis do not have the right functional properties and thus "dilute" the effects (see Nieto-Castañon & Fedorenko, 2012).

Future research

The ability to define prosody-sensitive regions functionally in individual subjects opens the door to a research program investigating the functional profiles of these (or other functionally defined prosody-sensitive) regions in an effort to understand their contributions to prosodic processing. For example, it is important to discover the necessary and sufficient features that a stimulus must possess in order to elicit a response in these regions. This question applies both to linguistic stimuli (e.g., different types of prosodic contours) and non-linguistic stimuli. For example, if some of these regions indeed store prosodic "templates" (i.e., commonly encountered prosodic patterns), we can probe the nature of these representations. We can ask how long these prosodic chunks have to be to elicit a response in these regions, by presenting foreign or low-pass filtered speech split up into prosodic patterns of various durations. Or we can ask how abstract these representations are,

by examining adaptation effects in these regions to the same/similar prosodic patterns produced by speakers with different voices or presented at different speeds.

Cognitive tasks across several domains have been shown to activate regions in and around temporal poles (see e.g., Olson et al., 2007 for a review), including music (e.g., Peretz & Zatorre, 2005; Fedorenko et al., 2012c), social cognition (e.g., Fletcher et al., 1995) and abstract semantics (e.g., Patterson et al., 2007). It is possible to construct multiple hypotheses about possibly shared computations between prosody and each of these other domains, especially music and social cognition (see some of the reviews cited in the introduction for discussions of some ideas along these lines). Examining the response of the regions reported here to musical stimuli and to non-prosodic socially-salient stimuli is necessary for evaluating such hypotheses.

Once we make progress in functionally characterizing these regions, we can start investigating the relationship between these regions and other regions / networks in the brain, by examining anatomical connectivity (e.g., using DTI) and functional resting-state correlations (e.g., Fox & Raichle, 2007). In particular, we can examine the relationship between prosody-sensitive regions and primary auditory regions (e.g., Morosan et al., 2001), regions that have been implicated in pitch processing (Patterson et al., 2002), language regions (Fedorenko et al., 2010), multiple-demand regions (Duncan, 2010), and regions that support social cognition (Saxe & Kanwisher, 2003), among others.

Multi-voxel pattern analyses (e.g., Norman et al., 2006) might also prove valuable in studying prosody-sensitive regions. In particular, as discussed in the introduction, there is no reason to necessarily expect differences in the mean BOLD response for different types of prosodic contours. However, a region that is important in prosodic processing may be able to distinguish among these various prosodic contours in its fine-grained pattern of spatial activity, in spite of showing similar BOLD responses. This method can thus be used to ask which distinctions are represented in each of these regions.

Furthermore, once we have sufficiently narrowed down the range of hypotheses about the functions of these regions, we can use these regions as targets for transcranial magnetic stimulation (TMS) to investigate their causal role in various computations.

Conclusions

We have here argued that fMRI in general, and the functional localization approach in particular, holds promise for asking and answering theoretically important questions in the domain of prosody. We have presented one possible functional localizer for identifying prosody-sensitive brain regions in individual subjects, demonstrating the feasibility of this method for investigating prosodic processing. We hope that this approach can help shift the field from asking questions about *where* something happens in the brain to *how* it happens.
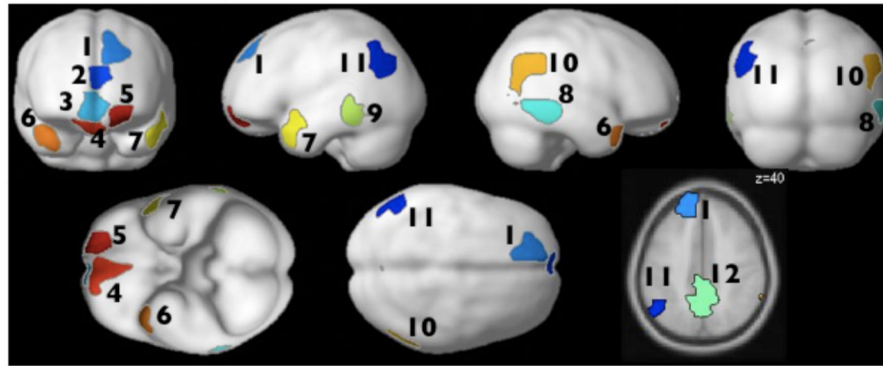
Acknowledgments

Appendix A

Figure A1.
Twelve parcels that satisfied the functional criteria (i.e., a selective response to structure in the auditory stimuli; see Methods for details) and were present in at least 9/12 subjects. (Regions #6-9 are the regions discussed in the main body of the paper and shown in Figures 1 and 2.)
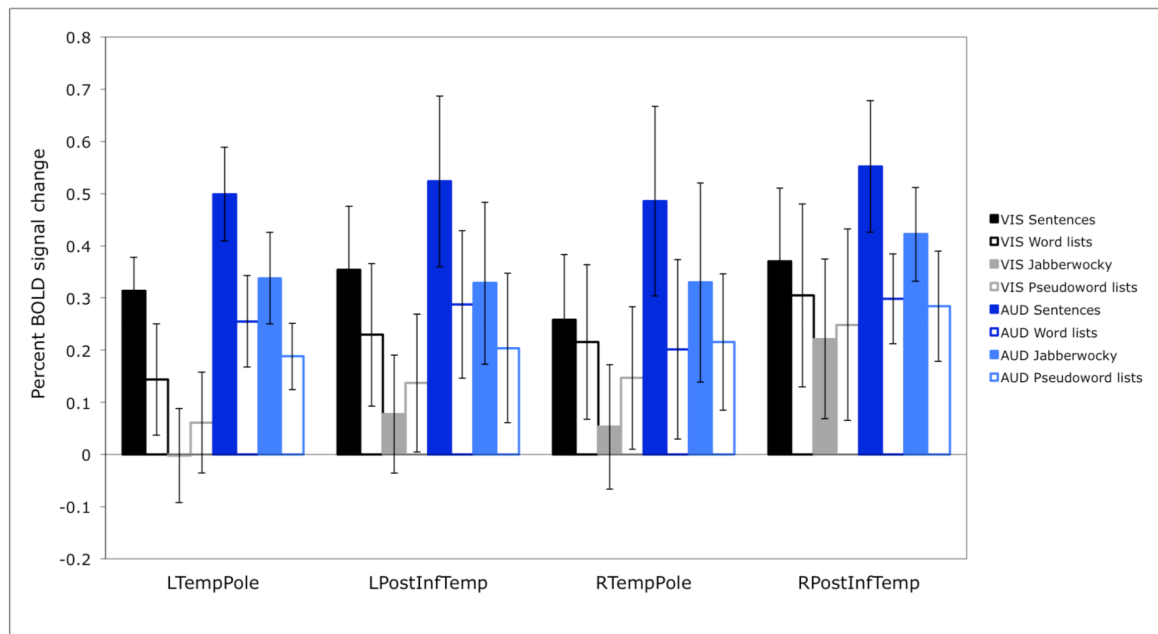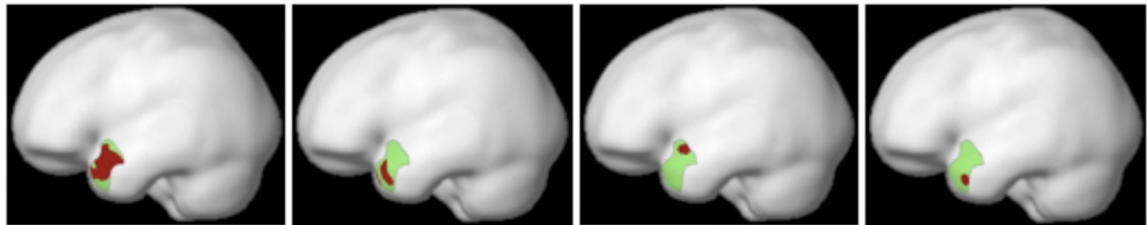

Figure A2.
Responses of the four regions shown in Figure 2 to the individual experimental conditions.
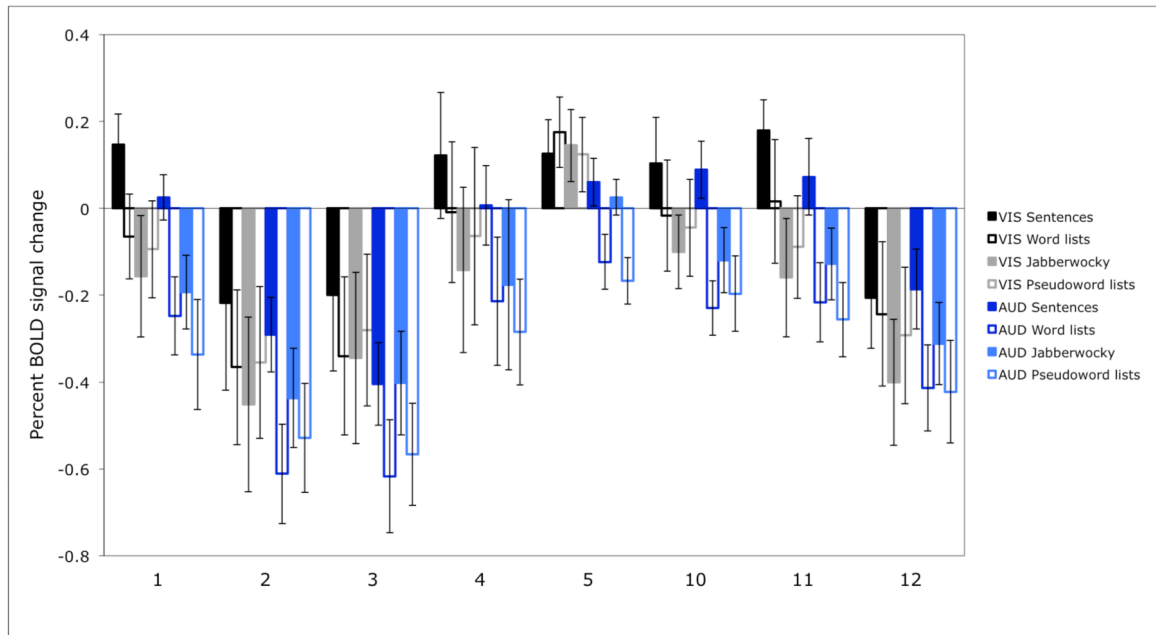
Figure A3.
Responses of the regions not shown in Figure 2 (i.e., regions #1-5, 10-12 in Fig. A1) to the individual experimental conditions. Region numbers correspond to the numbers of the regions in Figure A1.

Appendix B

Figure B1.
Sample individual fROIs for the Left Temporal Pole region. The parcel is shown in green, and the individual activations are shown in red.

Appendix C

Pitch tracks for the four conditions of two sample items
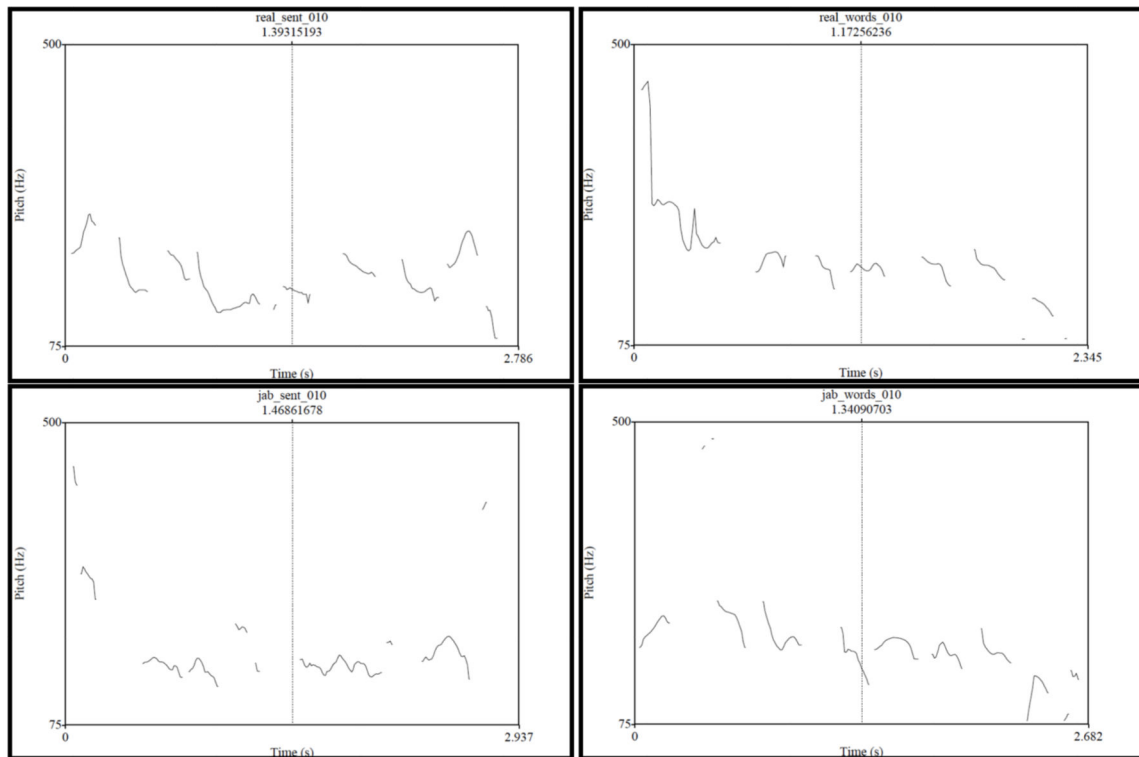
Figure C1a.
Pitch tracks for Item #010 (real_sent = sentence; real_words = word list; jab_sent = Jabberwocky sentence; and jab_words = pseudoword list).
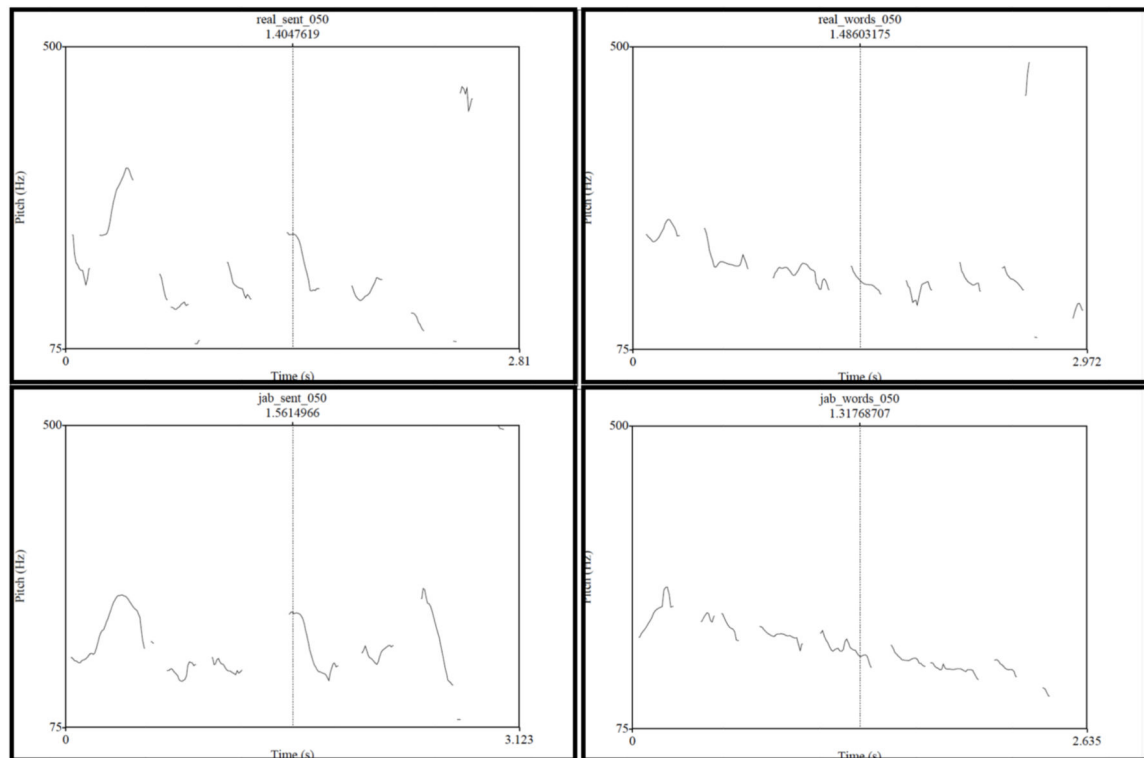
Figure C1b.
Pitch tracks for Item #050 (real_sent = sentence; real_words = word list; jab_sent = Jabberwocky sentence; and jab_words = pseudoword list).

Acoustic analyses of the experimental materials

Methods

The analysis was performed in three steps:

1. The audio files were aligned with the corresponding text files to identify the word/ pseudoword boundaries in each audio file using the Prosodylab Aligner tool (Gorman et al., 2011).

The alignment was performed following the instructions provided at http://prosodylab.org/ tools/aligner/. The aligner uses the CMU pronouncing dictionary (available at http:// www.speech.cs.cmu.edu/cgi-bin/cmudict). A few words from the Sentences (n=4) and Word-lists (n=6) conditions were not in the dictionary. And, as expected, a large number of pseudowords in the Jabberwocky sentences and Pseudoword-lists conditions were not in the dictionary (544 and 547, respectively). Consequently, a total of 686 unique words were added to the dictionary. Their pronunciation was generated by the tools available at http:// www.speech.cs.cmu.edu/tools/lextool.html.

To align the audio files to the text files, the aligner takes as input a *.wav file and a *.lab file (i.e., the text file with the orthographic words that are matched to the items in the

dictionary), and it outputs a *.TextGrid file, which has the forced time divisions of the *.wav file for each word (and phoneme) in the *.lab file.

The aligner was unsuccessful in forcing the alignment for a small number of items in each condition (1 in the Sentences condition, 1 in the Word-lists condition, 18 in the Jabberwocky sentences condition, and 17 in the Pseudoword-lists condition). A plausible reason for this is that the acoustic pattern has to be fairly close to the phonetic form of the word in the dictionary, and some of the pseudowords may have been pronounced quite differently from the pronunciation generated by the dictionary. All the analyses were thus performed on 127 sentences, 127 word lists, 110 Jabberwocky sentences, and 111 pseudoword lists.

2. A set of acoustic features was extracted from each word/pseudoword in each aligned audio file using Praat (Boersma & Weenink, 2005).

The extraction was performed with a Praat script written by Michael Wagner. This script takes as input a *.wav file and a *.TextGrid file and outputs a set of pre-specified acoustic measures.

We extracted the following features for each word/pseudoword:

Duration measure:

- duration (s) [word/pseudoword duration]

Pitch measures:

- mean pitch (Hz) [mean fundamental frequency (F0) of the word/pseudoword]

- maximum pitch (Hz) [highest F0 observed in the word/pseudoword]

- minimum pitch (Hz) [lowest F0 observed in the word/pseudoword]

- initial pitch (Hz) [mean F0 of the first 5% of the word/pseudoword]

- center pitch (Hz) [mean F0 of the 5% of the word/pseudoword centered on the midpoint of the word/pseudoword]

- final pitch (Hz) [mean F0 of the last 5% of the word/pseudoword]

Other measures:

- mean intensity (dB) [mean loudness level in the word/pseudoword]

- maximum intensity (dB) [highest dB level observed in the word/pseudoword]

- minimum intensity (dB) [lowest dB level observed in the word/pseudoword]

- power (Pascal^2) [power across the entire word/pseudoword]

3. Several statistical tests were performed on the extracted values.

Analyses reported here were conducted with the lme4 package (Bates, Maechler, & Dai, 2008) for the statistical language R (R Development Core Team, 2008).

We asked the following two questions:

i. Averaging across the eight words/pseudowords in a sentence/word list, do conditions differ with respect to any of the acoustic features above? In particular, is there a difference between structured conditions (Sentences + Jabberwocky sentences) and unstructured conditions (Word lists + Pseudoword lists)? For exploratory purposes, we also examined the changes in the acoustic features over the course of the sentence / word list.

ii. Do conditions differ with respect to how much they vary over the course of the sentence / word list in any of the acoustic features above?

Results

The figures that summarize the results of the analyses appear below.

First, for each acoustic feature we show the mean value (averaged across words/ pseudowords within a stimulus, and then averaged across trials within each condition) in a bar graph on the left, and the word-by-word values in a line graph on the right in Figures C2a-i. To quantify the differences, we conducted a series of linear mixed-effects regressions with the two experimental factors (structure: sentences vs. word lists; and lexical content: real words vs. pseudowords) as fixed effects, and items as random effects (we included both an intercept and a slope for each item following Barr et al., 2013). Of most interest to the current investigation are potential differences between structured and unstructured conditions. (We find a few differences between real words and pseudowords, but we don't discuss these here given that all of the key fMRI analyses focus on the structured vs. unstructured contrast collapsing across real words and pseudowords as in Fig. 2, or treating them separately as in Figs. 4a-b.)

It is worth noting that the creation of the materials was successful in that the structured and unstructured conditions are quite well matched in a number of ways. In particular, they do not differ in terms of mean pitch, initial, center or final pitch, and maximum or minimum intensity (all ts<1, except for center pitch where t= |1.73|). Furthermore, the changes in various features over the course of the stimulus (see line graphs in the figures below) also look similar across structured and unstructured conditions, at least in their general trends. For example, the pitch and intensity values decrease over the course of the sentence / word list and the duration values increase at the end of the sentence / word list.

However, we do find a few differences between structured and unstructured conditions, and these differences plausibly contribute to the effects we observe in fMRI. Because we are examining 11 acoustic features, a correction for multiple comparisons is needed. We chose the most stringent, Bonferroni, correction: with 11 tests, the p value has to be smaller than $0.05/11 = 0.0045$. With respect to pitch, we find that words/pseudowords in the structured conditions are produced with higher maximum pitch (t= |5.71|) and lower minimum pitch (t=|6.32|). Furthermore, words/pseudowords in the structured conditions are produced with less power than in the unstructured conditions (t= |3.88|). With respect to duration, we find that on average, words/pseudowords in the structured conditions are longer, but the effect doesn't reach significance at the corrected p level (p<.02), and in any case, as can be seen in

Fig. C2a, structured-ness appears to interact with real word / pseudoword dimension, so this difference is unlikely to contribute to the observed differences in fMRI.

In all the graphs below, the error bars represent standard errors of the mean over items. For the line graphs, the error bars were computed as follows: the standard deviation for all the structured (or unstructured, for the unstructured conditions) items was divided by the square root of the number of structured (or unstructured) items. Note also that in all the line graphs below, we averaged condition means for the real words and pseudoword conditions to get each of structured and unstructured values. The reason for doing that (instead of averaging across items) is that the number of items differs slightly between conditions – see Methods part 1 above.
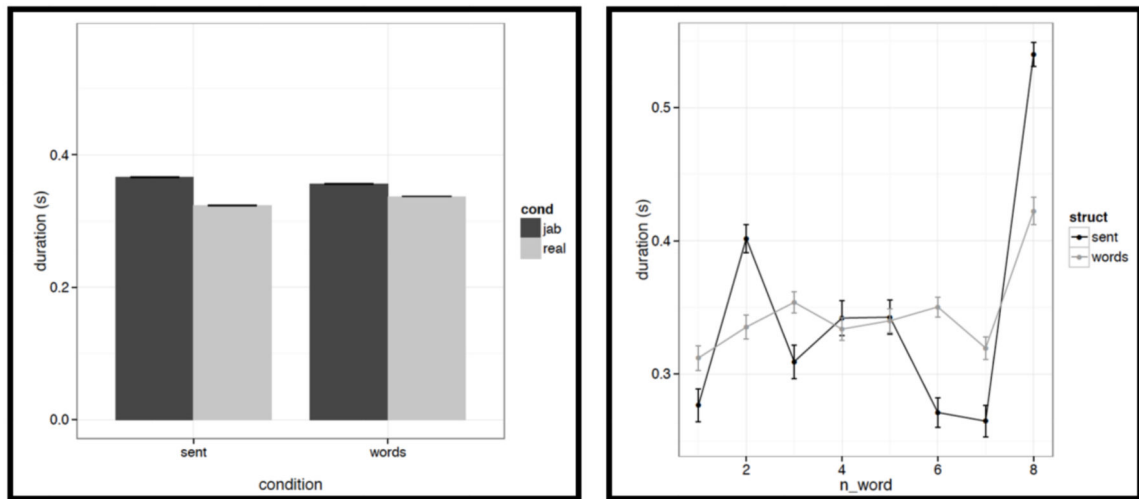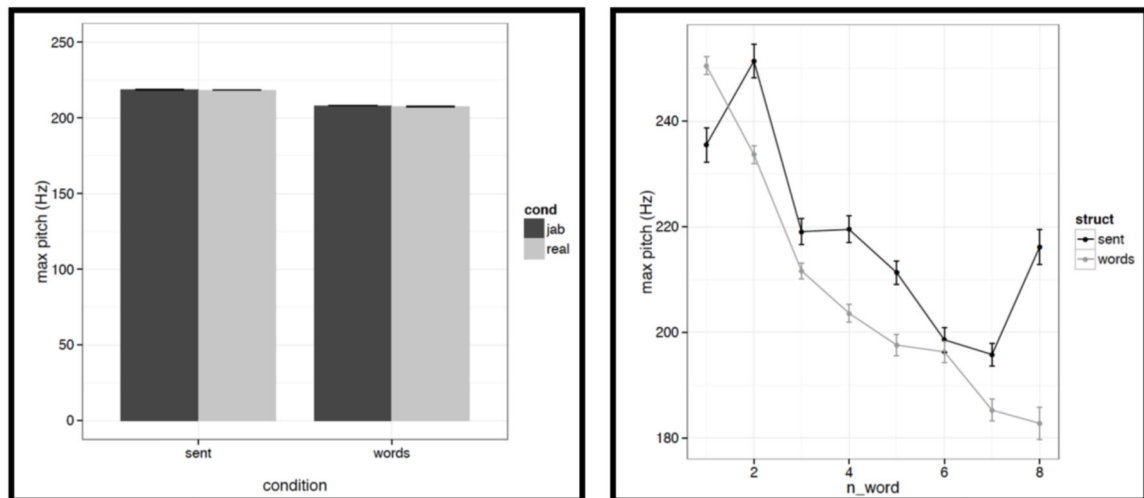

Figure C2a.
Left: Average <u>duration</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>duration</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
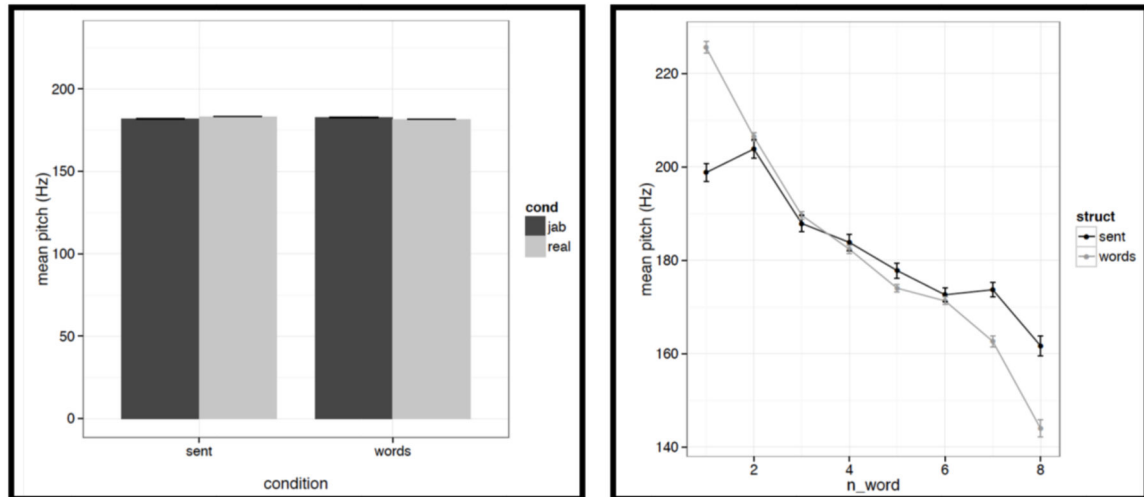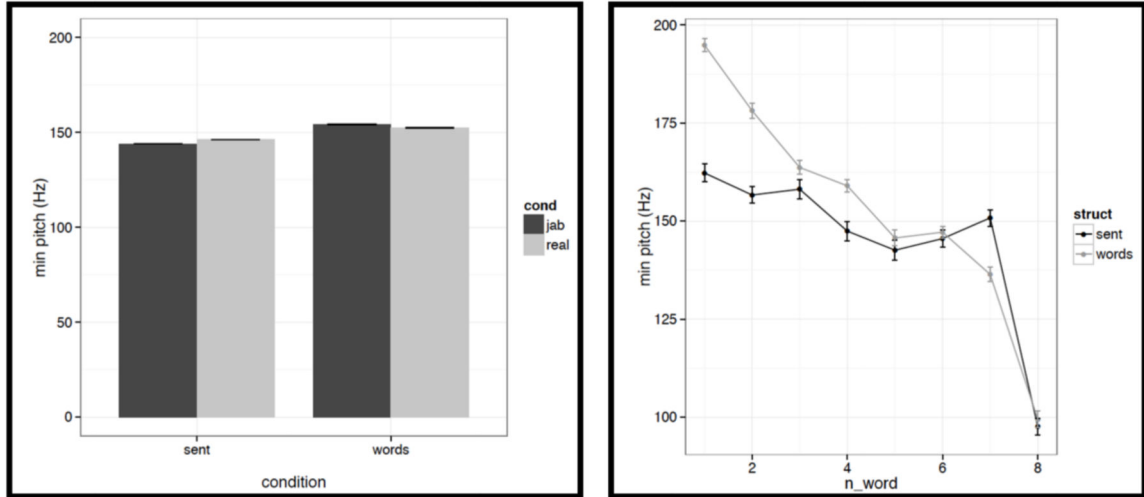
Figure C2b.
Left: Average <u>mean pitch</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>mean pitch</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
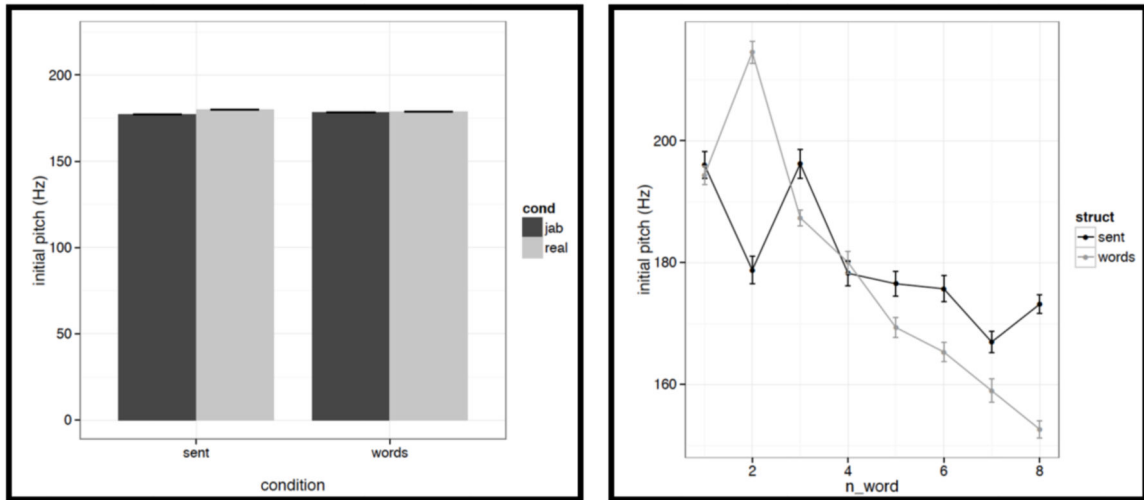

Figure C2c.
Left: Average <u>maximum pitch</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>maximum pitch</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
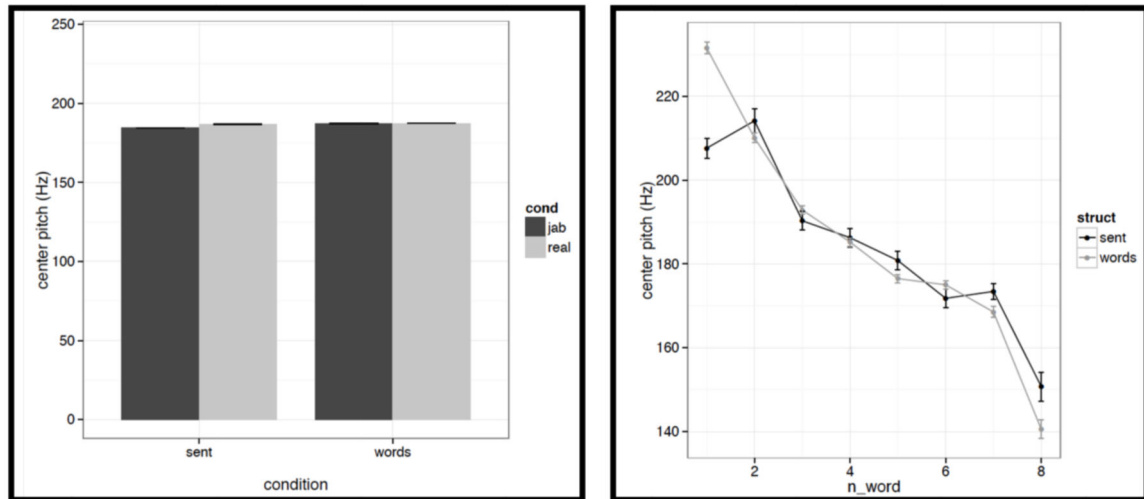
Figure C2d.
Left: Average <u>minimum pitch</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>minimum pitch</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
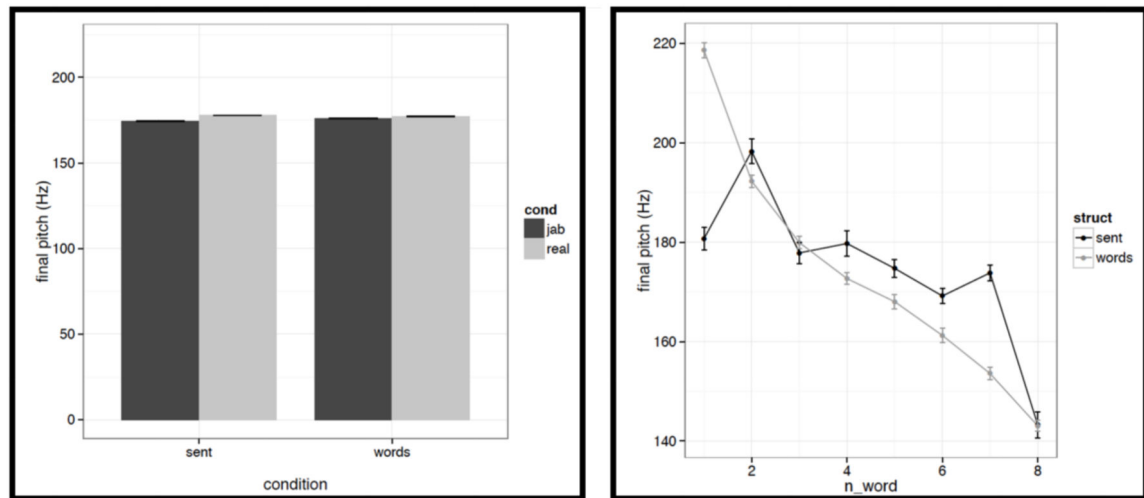

Figure C2e.
Left: Average <u>initial pitch</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>initial pitch</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
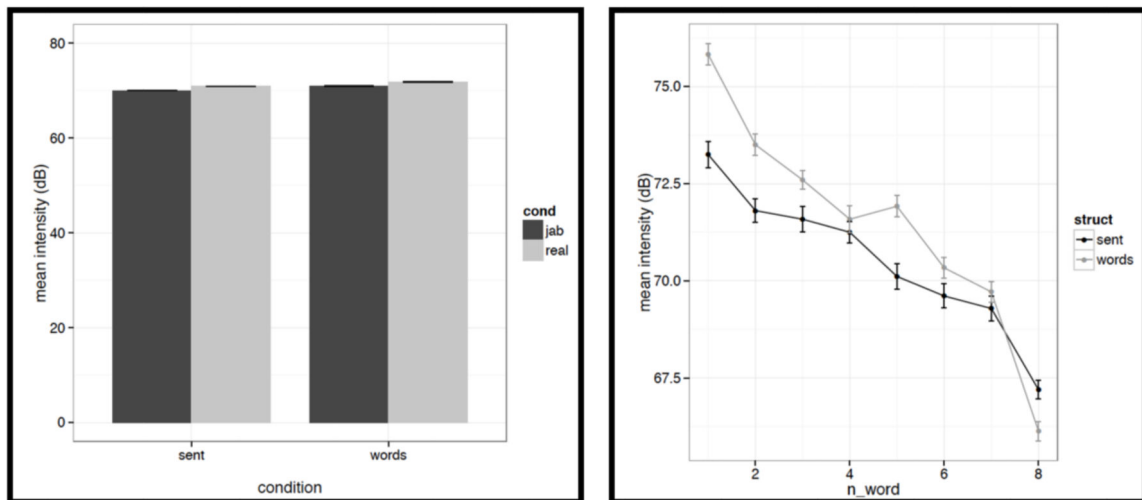
Figure C2f.
Left: Average <u>center pitch</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>center pitch</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
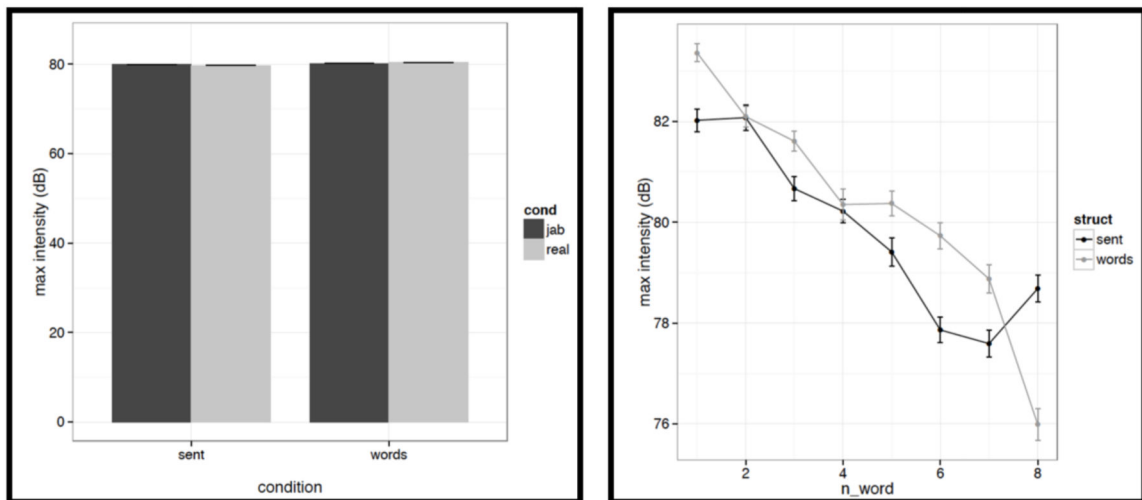

Figure C2g.
Left: Average <u>final pitch</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>final pitch</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
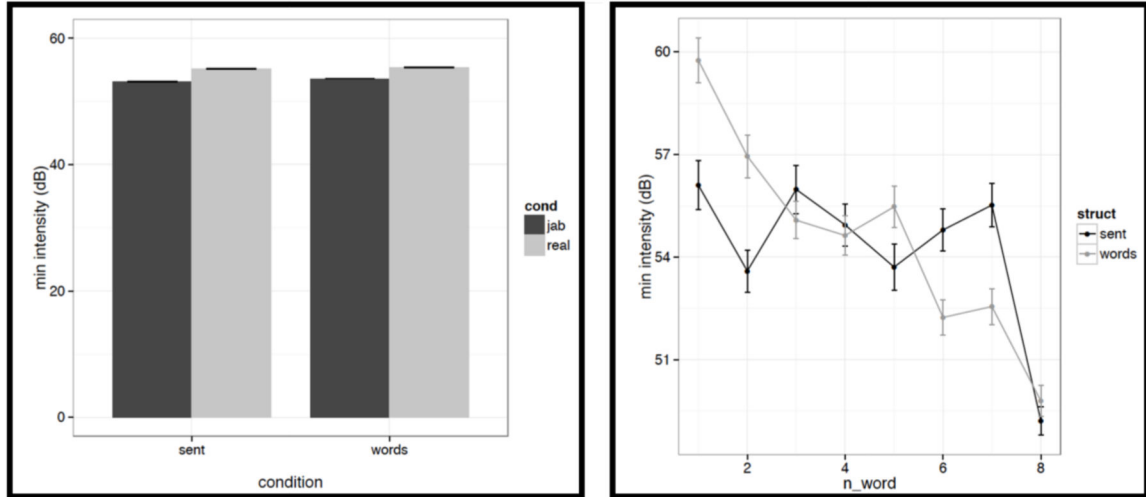
Figure C2h.
Left: Average <u>mean intensity</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>mean intensity</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.

Figure C2i.
Left: Average <u>maximum intensity</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>maximum intensity</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.
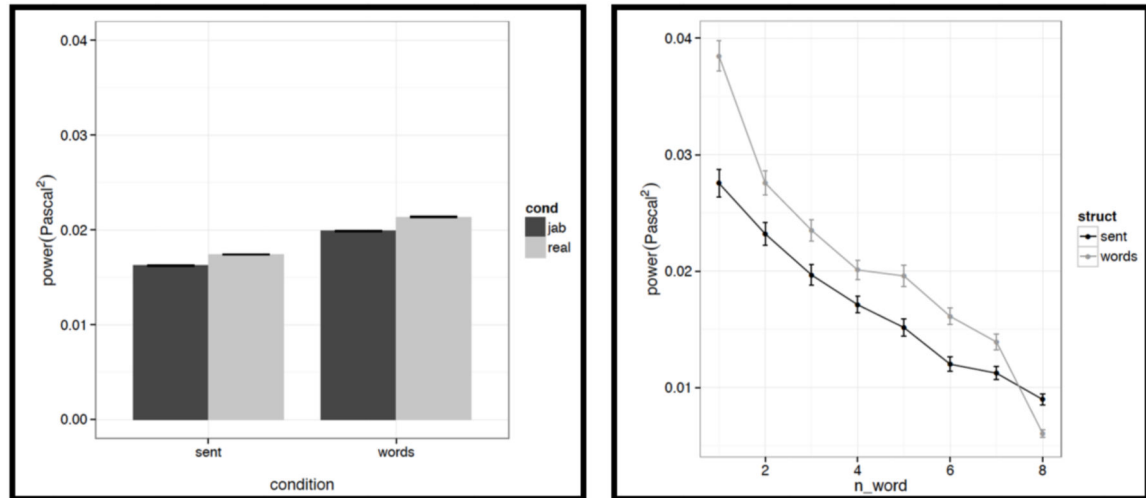
Figure C2j.
Left: Average <u>minimum intensity</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>minimum intensity</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.

Figure C2k.
Left: Average <u>power</u> values for words/pseudowords – averaged across the eight positions within a stimulus – across the four conditions (sentences with real words, Jabberwocky sentences, lists of real words, lists of pseudowords). Right: Average <u>power</u> values for words/pseudowords at each of the eight positions for the structured (black) vs. unstructured (grey) conditions.

And second, we also find some differences between structured and unstructured conditions in terms of the amount of variability across the eight words/pseudowords. As in the first

analysis, because we are examining 11 acoustic features, we are applying a Bonferroni correction, and the *p* value has to be smaller than 0.05/11 = 0.0045. With respect to duration, we find that words/pseudowords in the structured conditions are more variable over the course of the stimulus (t= |10.33|). With respect to pitch, the structured and unstructured conditions do not differ significantly with respect to variability in mean pitch, minimum pitch, initial pitch, or final pitch (all *t*s<1.5). However, the words/pseudowords in the structured conditions are more variable in terms of maximum pitch (t= |4.28|) and center pitch (t= |5.41|). None of the three intensity measures reveal significant differences between structured and unstructured conditions, although there is a trend for more variability in minimum intensity in the structured conditions. Finally, we find that words/pseudowords in the structured conditions are less variable over the course of the stimulus in terms of power (t= |4.59|).

References

Adank P. The neural bases of difficult speech comprehension and speech production: Two activation likelihood estimation (ALE) meta-analyses. Brain and Language. 2012; 122(1):42–54. [PubMed: 22633697]

Alba-Ferrara L, Hausmann M, Mitchell R, Weis S. The neural correlates of emotional prosody comprehension: Disentangling simple from complex emotion. PLoS ONE. 2011; 6(12):e28701. [PubMed: 22174872]

Albritton D, McKoon G, Ratcliff R. Reliability of prosodic cues for resolving syntactic ambiguity. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996; 22:714–735.

Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HBM, Zilles K. Broca's region revisited: Cytoarchitecture and inter-subject variability. Journal Comparative Neurology. 1999; 412:319–341.

Bader, M. Prosodic influences on reading syntactically ambiguous sentences. In: Fodor, J.; Ferreira, F., editors. Reanalysis in sentence processing. Kluwer; Dordrecht: 1998. p. 1-46.

Baker CI, Liu J, Wald LL, Kwong KK, Benner T, Kanwisher N. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. PNAS. 2007; 104(21): 9087–9092. [PubMed: 17502592]

Barr DJ, Levy R, Scheepers C, Tily HJ. Random-effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 2013; 68:255–278.

Bates, D.; Maechler, M.; Dai, B. [Accessed June 1, 2012] Lme4: Linear mixed-effects models using S4 classes. 2008. Available at http://lme4.r-forge.r-project.org/

Baum S, Daniloff R, Daniloff J, Lewis J. Sentence comprehension by Broca's aphasics: effects of some suprasegmental variables. Brain and Language. 1982; 17:261–271. [PubMed: 7159835]

Baum S, Pell M. The neural bases of prosody: insights from lesion studies and neuroimaging. Aphasiology. 1999; 13:581–608.

Behrens SJ. The role of the right hemisphere in the production of linguistic stress. Brain and Language. 1988; 33:104–127. [PubMed: 3342315]

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Thomas P. Human brain language areas identified by functional Magnetic Resonance Imaging. The Journal of Neuroscience. 1997; 17(1): 353–362. [PubMed: 8987760]

Blumstein S, Cooper WE. Hemispheric Processing of Intonation Contours. Cortex. 1974; 10(2):146–158. [PubMed: 4844467]

Blumstein S, Goodglass H. The perception of stress as a semantic cue in aphasia. Journal of Speech and Hearing Research. 1972; 15:800–806. [PubMed: 4672047]

Boersma, P.; Weenink, D. Praat: doing phonetics by computer (Version 5.3.59) [Computer program]. 2005. Retrieved from http://www.praat.org/

Bowers D, Coslett HB, Bauer RM, Speedie LJ, Heilman KH. Comprehension of emotional prosody following unilateral hemispheric lesions: processing defect versus distraction defect. Neuropsychologia. 1987; 25:317–328. [PubMed: 3601038]

Bradvik B, Dravins C, Holtas S, Rosen I, Ryding E, Ingvar D. Disturbances of speech prosody following right hemisphere infarcts. Acta Neurologica Scandinavica. 1991; 84:114–126. [PubMed: 1719738]

Braze D, Mencl WE, Tabor W, Pugh KR, Constable RT, Fulbright RK, Magnuson JS, Van Dyke JA, Shankweiler DP. Unification of sentence processing via ear and eye: An fMRI study. Cortex. 2011; 47(4):416–431. [PubMed: 20117764]

Breen M, Fedorenko E, Wagner M, Gibson E. Acoustic correlates of information structure. Language and Cognitive Processes. 2010; 25(7):1044–1098.

Brodmann, K. Vergleichende lokalisationslehre der grosshirnrinde in ihren prinzipien dargestellt auf grund des zellenbaues. 1909. Barth Leipzig

Bryan K. Language prosody and the right hemisphere. Aphasiology. 1989; 3:285–299.

Buchanan T, Lutz K, Mirzazade S, Specht K, Shah N, Zilles K, Jancke L. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Cognitive Brain Research. 2000; 9:227–238. [PubMed: 10808134]

Caltagirone C, Ekman P, Friesen W, Gainotti G, Mammucari A, Pizzamiglio L, Zoccolotti P. Posed emotional expression in unilateral brain damaged patients. Cortex. 1989; 25:653–663. [PubMed: 2612182]

Cohen L, Dehaene S, Naccache L, Lehericy S, Dehaene-Lambertz G, Henaff MA, Michel F. The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain. 2000; 123:291–307. [PubMed: 10648437]

Couper-Kuhlen, E. An Introduction to English Prosody. Edward Arnold and Niemeyer; London and Tuebingen: 1986.

Coutanche MN, Thompson-Schill SL. The advantage of brief fMRI acquisition runs for multi-voxel pattern detection across runs. Neuroimage. 2012; 61:1113–1119. [PubMed: 22498658]

Darby D. Sensory aprosodia: a clinical clue to lesion of the inferior division of the right middle cerebral artery? Neurology. 1993; 43:567–572. [PubMed: 8451003]

Davis, C.; Molitoris, J.; Newhart, M.; Bahrainwala, Z.; Khurshid, S.; Heidler-Gary, J.; Ross, E.; Hillis, A. Areas of Right Hemisphere Ischemia Associated with Impaired Comprehension of Affective Prosody in Acute Stroke; Clinical Aphasiology Conference; Isle of Palms, SC. 2010;

Doherty CP, West WC, Dilley LC, Shattuck-Hufnagel S, Caplan D. Question/statement judgments: an fMRI study of intonation processing. Human Brain Mapping. 2004; 23:85–98. [PubMed: 15340931]

Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N. Domain Specificity in Visual Cortex. Cerebral Cortex. 2006; 16(10):1453–61. [PubMed: 16339084]

Duncan J. Brain mechanisms of attention. Quarterly Journal of Experimental Psychology. 2006; 59(1): 2–27.

Duncan J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends in Cognitive Sciences. 2010; 14:172–179. [PubMed: 20171926]

Duncan J, Owen A. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends in Neuroscience. 2000; 23:475–483.

Dykstra K, Gandour J, Stark R. Disruption of prosody after frontal lobe seizures in the non-dominant hemisphere. Aphasiology. 1995; 9:453–476.

Emmorey KD. The neurological substrates for prosodic aspects of speech. Brain & Language. 1987; 30:305–329. [PubMed: 3567552]

Ethofer T, Anders S, Erb M, Droll C, Royen L, Saur R, Reiterer S, Grodd W, Wildgruber D. Impact of voice on emotional judgment of faces: an event-related fMRI study. Human Brain Mapping. 2006a; 27:707–714. [PubMed: 16411179]

Ethofer T, Anders S, Erb M, Herbert C, Wiethoff S, Kissler J, Grodd W, Wildgruber D. Cerebral pathways in processing of emotional prosody: a dynamic causal modelling study. NeuroImage. 2006b; 30:580–587. [PubMed: 16275138]

Fedorenko, E. The role of domain-general cognitive control mechanisms in language comprehension. under review

Fedorenko E, Kanwisher K. Neuroimaging of language: why hasn't a clearer picture emerged? Language & Linguistics Compass. 2009; 3(4):839–865.

Fedorenko E, Hsieh P-J, Nieto-Castañon A, Whitfield-Gabrieli S, Kanwisher N. A new method for fMRI investigations of language: Defining ROIs functionally in individual subjects. Journal of Neurophysiology. 2010; 104:1177–94. [PubMed: 20410363]

Fedorenko E, Behr M, Kanwisher N. Functional specificity for high-level linguistic processing in the human brain. PNAS. 2011

Fedorenko E, Nieto-Castanon A, Kanwisher N. Syntactic processing in the human brain: What we know, what we don't know, and a suggestion for how to proceed. Brain and Language. 2012a; 120:187–207. [PubMed: 21334056]

Fedorenko E, Duncan J, Kanwisher N. Language-selective and domain-general regions lie side by side within Broca's area. Current Biology. 2012b; 22:2059–2062. [PubMed: 23063434]

Fedorenko E, Duncan J, Kanwisher N. Broad domain-generality in focal regions of frontal and parietal cortex. Proceedings of the National Academy of Sciences USA. 2013

Fedorenko E, McDermott J, Norman-Haignere S, Kanwisher N. Sensitivity to musical structure in the human brain. Journal of Neurophysiology. 2012c; 108:3289–3300. [PubMed: 23019005]

Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, et al. Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. Cognition. 1995; 57(2): 109–28. [PubMed: 8556839]

Fodor JD. Learning to parse? Journal of Psycholinguistic Research. 1998; 27:285–319.

Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Reviews Neuroscience. 2007; 8:701–711.

Friederici AD, Rüschemeyer S-A, Hahne A, Fiebach CJ. The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. Cerebral Cortex. 2003; 13(2):170–177. [PubMed: 12507948]

Friederici AD, Alter K. Lateralization of auditory language functions: A dynamic dual pathway model. Brain and Language. 2004; 89:267–276. [PubMed: 15068909]

Friston KJ, Zarahn E, Josephs O, Henson RN, Dale AM. Stochastic designs in event-related fMRI. Neuroimage. 1999; 10:607–619. [PubMed: 10547338]

Frost MA, Goebel R. Measuring structural-functional correspondence: Spatial variability of specialized brain regions after macro-anatomical alignment. Neuroimage. 2011; 59(2):1369–81. [PubMed: 21875671]

Gandour J, Wong D, Dzemidzic M, Lowe M, Tong Y, Li X. A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. Human Brain Mapping. 2003; 18:149–157. [PubMed: 12599272]

George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P, Post RM. Understanding emotional prosody activates right hemisphere regions. Arch. Neurol. 1996; 53:665–670. [PubMed: 8929174]

Geschwind N, Levitsky W. Human brain: Left-right asymmetries in temporal speech region. Science. 1968; 161:186–187. [PubMed: 5657070]

Gleitman, LR.; Wanner, E. Language acquisition: The state of the art. In: Wanner, E.; Gleitman, LR., editors. Language acquisition: The state of the art. University Press; Cambridge: 1982.

Goodglass H, Calderon M. Parallel processing of verbal and musical stimuli in right and left hemispheres. Neuropsychologia. 1977; 15:397–407. [PubMed: 854158]

Gorman, K.; Howell, J.; Wagner, M. Prosodylab-Aligner: A tool for forced alignment of laboratory speech; Proceedings of Acoustics Week in Canada; Quebec City. 2011;

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat. Neurosci. 2005; 8:145–146. [PubMed: 15665880]

Heilman K, Bowers D, Speedie L, Coslet H. Comprehension of affective and non-affective prosody. Neurology. 1984; 34:917–921. [PubMed: 6539867]

Heilman K, Leon S, Rosenbek J. Affective Aprosodia from a Medial Frontal Stroke. Brain and Language. 2004; 89:411–416. [PubMed: 15120533]

Heilman K, Scholes R, Watson R. Auditory affective agnosia: disturbed comprehension of speech. J. Neurol. Neurosurg. Psychiatry. 1975; 38:69–72. [PubMed: 1117301]

Herrero JV, Hillix WA. Hemispheric performance in detecting prosody. A competitive dichotic listening task. Perceptual and Motor Skills. 1990; 71:479–486. [PubMed: 2251084]

Hickok G, Okada K, Serences J. Area Spt in the human planum temporale supports sensory-motor interaction for speech processing. Journal of Neurophysiology. 2009; 101:2725–2732. [PubMed: 19225172]

Hirose Y. Recycling prosodic boundaries. Journal of Psycholinguistic Research. 2003; 32:167–195. [PubMed: 12690830]

Huettel, SA.; Song, AW.; McCarthy, G. Functional Magnetic Resonance Imaging. Sinauer Associates; 2008.

Hughlings-Jackson, J. Selected writings of John Hughlings-Jackson. Hodder and Stoughton; London: 1931.

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]

Imaizumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, Itoh K, Kato T, Nakamura A, Hatano K, Kojima S, Nakamura K. Vocal identification of speaker and emotion activates different brain regions. Neuroreport. 1997; 8:2809–2812. [PubMed: 9295122]

Juch H, Zimine I, Seghier ML, Lazeyras F, Fasel JH. Anatomical variability of the lateral front lobe surface: implication for intersubject variability in language neuroimaging. Neuroimage. 2005; 24:504–514. [PubMed: 15627592]

Julian J, Fedorenko E, Webster J, Kanwisher N. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. Neuroimage. 2012; 60:2357–2364. [PubMed: 22398396]

Jusczyk PW, Cutler A, Redanz NJ. Preference for the predominant stress patterns of English words. Child Development. 1993; 64:675–687. [PubMed: 8339688]

Jusczyk, PW. The discovery of spoken language. MIT Press; Cambridge, MA: 1997.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: A module in human extrastriate cortex specialized for face perception. J Neurosci. 1997; 17(11):4302–11. [PubMed: 9151747]

Kanwisher N, Tong F, Nakayama K. The Effect of Face Inversion on the Human Fusiform Face Area. Cognition. 1998; 68:B1–B11. [PubMed: 9775518]

Kawahata N, Nagata K, Shishido F. Alexia with agraphia due to the left posterior inferior temporal lobe lesion – neuropsychological analysis and its pathogenic mechanisms. Brain and Language. 1988; 33:296–310. [PubMed: 3258777]

Kotz SA, Schwartze M. Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends in Cognitive Science. 2010; 14(9):392–399.

Kotz S, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD. On the lateralization of emotional prosody: an event-related functionalMR investigation. Brain & Language. 2003; 86:366–376. [PubMed: 12972367]

Kotz SA, Meyer M, Paulmann S. Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. Prog Brain Res. 2006; 156:285–294. [PubMed: 17015086]

Kowner R. Laterality in facial expressions and its effect on attributions of emotion and personality: a reconsideration. Neuropsychologia. 1995; 33:539–559. [PubMed: 7637852]

Kraljic T, Brennan SE. Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? Cognitive Psychology. 2005; 50:194–231. [PubMed: 15680144]

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience – the dangers of double dipping. Nature Neuroscience. 2009; 12(5):535–40.

Ley RG, Bryden MP. A dissociation of right and left hemispheric effects for recognizing emotional tone and verbal content. Brain & Cognition. 1982; 1:3–9. [PubMed: 6927552]

Marslen-Wilson WD, Tyler LK, Warren P, Grenier P, Lee CS. Prosodic effects in minimal attachment. Quarterly Journal of Experimental Psychology. 1992; 45A:73–87.

Mattys SL, Jusczyk PW, Luce PA, Morgan JL. Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology. 1999; 38:465–494. [PubMed: 10334878]

Mayer, J.; Wildgruber, D.; Riecker, A.; Dogil, G.; Ackermann, H.; Grodd, W. Prosody production and perception: Converging evidence from fMRI studies. 2002.

McCandliss BD, Cohen L, Dehaene S. The visual word form area: Expertise for reading in the fusiform gyrus. Trends in Cognitive Sciences. 2003; 7(7):293–299. [PubMed: 12860187]

Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY. Functional MRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Human Brain Mapping. 2002; 17:73–88. [PubMed: 12353242]

Miller MB, Van Horn JD, Wolford GL, Handy TC, Valsangkar-Smyth M, Inati S, Grafton S, Gazzaniga MS. Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. Journal of Cognitive Neuroscience. 2002; 14(8):1200–1214. [PubMed: 12495526]

Mitchell RL, Elliott R, Barry M, Cruttenden A, Woodruff PW. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia. 2003; 41:1410–1421. [PubMed: 12757912]

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage. 2001; 13(4):684–701. [PubMed: 11305897]

Nakamura K, Honda M, Okada T, Hanakawa T, Toma K, Fukuyama H, Konish J, Shibasaki H. Participation of the left posterior inferior temopral cortex in writing and mental recall of kanji orthography: A functional MRI study. Brain. 2000; 123:954–967. [PubMed: 10775540]

Newman AJ, Suppalla T, Hauser PC, Newport E, Bavelier D. Prosodic and narrative processing in American Sign Language: An fMRI study. Neuroimage. 2010; 52:669–676. [PubMed: 20347996]

Nieto-Castanon, A.; Fedorenko, E. Functional localizers increase sensitivity and selectivity. submitted

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences. 2006; 10(9):424–430. [PubMed: 16899397]

O'Craven K, Kanwisher N. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. Journal of Cognitive Neuroscience. 2000; 12:1013–1023. [PubMed: 11177421]

Olson IR, Plotzker A, Ezzyat Y. The enigmatic temporal pole: A review of findings on social and emotional processing. Brain. 2007; 130(Pt 7):1718–31. [PubMed: 17392317]

Ono, M.; Kubik, S.; Abernathey, CD. Atlas of the cerebral sulci. Thieme Verlag; Stuttgart: 1990.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. Neuron. 2002; 36:767–776. [PubMed: 12441063]

Patterson K, Nestor P, Rogers T. Where do you know what you know? The representation of semantic knowledge in the human brain. Nature Reviews Neuroscience. 2007; 12:976–987.

Peretz I, Coltheart M. Modularity of music processing. Nature Neuroscience. 2003; 6(7):688–691.

Pell M, Baum S. The ability to perceive and comprehend intonation in linguistic and affective contexts by brain damaged adults. Brain & Language. 1997; 57:80–99. [PubMed: 9126408]

Peretz I, Zatorre RJ. Brain organization for music processing. Annual Review of Psychology. 2005; 56:89–114.

Petersen SE, Fox PT, Snyder AZ, Raichle ME. Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. Science. 1990; 249:1041–1044. [PubMed: 2396097]

Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, Williams SCR, Bullmore ET, Brammer M, Gray JA. Neural responses to facial and vocal expressions of fear and disgust. Proc R Soc Lon. 1998; 265:1809–1817.

Poldrack R. Can cognitive processes be inferred from neuroimaging data. Trends in Cognitive Sciences. 2006; 10(2):59–63. [PubMed: 16406760]

Polk TA, Farah M. The neural development and organization of letter recognition; Evidence from functional neuroimaging, computational modeling, and behavioral studies. PNAS. 1998; 95:847–852. [PubMed: 9448250]

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2008.

Rapcsak SZ, Beeson PM. The role of left posterior inferior temporal cortex in spelling. Neurology. 2004; 62(12):2221–2229. [PubMed: 15210886]

Ross ED, Mesulam M. Dominant language functions of the right hemisphere? Prosody or emotional gesturing. Archives of Neurology. 1979; 35:144–148. [PubMed: 435134]

Ross ED, Thompson RD, Yenkosky J. Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. Brain and Language. 1997; 56:27–54. [PubMed: 8994697]

Ross ED. The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. Archives of Neurology. 1981; 38:561–569. [PubMed: 7271534]

Sackheim HA, Gur RC, Saucy MC. Emotions are expressed more intensely on the left side of the face. Science. 1978; 202:434–436. [PubMed: 705335]

Saxe R, Kanwisher N. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". Neuroimage. 2003; 19(4):1835–42. [PubMed: 12948738]

Saxe R, Brett M, Kanwisher N. Divide and conquer: A defense of functional localizers. Neuroimage. 2006; 30(4):1088–96. discussion 1097-9. [PubMed: 16635578]

Schafer AJ, Speer SR, Warren P, White SD. Intonational disambiguation in sentence production and comprehension. Journal of Psycholinguistic Research. 2000; 29:169–182. [PubMed: 10709182]

Schlanger B, Schlanger P, Gerstmann L. The perception of emotionally toned sentences by right hemisphere damaged and aphasic subjects. Brain Lang. 1976; 3:396–403. [PubMed: 949593]

Schmitt JJ, Hartje W, Willmes K. Hemispheric asymmetry in the recognition of emotional attitude conveyed by facial expression, prosody and propositional speech. Cortex. 1997; 33:65–81. [PubMed: 9088722]

Schubiger, M. English Intonation: Its form and function. Max Niemeyer Verlag; Tuebingen: 1958.

Schwarzlose R, Baker C, Kanwisher N. Separate face and body selectivity on the fusiform gyrus. J of Neuroscience. 2005; 25:11055–59.

Selkirk, E. Phonology and syntax: The relation between sound and structure. MIT Press; Cambridge, MA: 1984.

Seron X, Van der Kaa MA, Vanderlinden M, Remits A, Feyereisen P. Decoding paralinguistic signals: effect of semantic and prosodic cues on aphasics' comprehension. Journal of Communication Disorders. 1982; 15:223–231. [PubMed: 7096619]

Shapiro B, Danly M. The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. Brain and Language. 1985; 25:19–36. [PubMed: 4027566]

Snedeker J, Trueswell J. Using prosody to avoid ambiguity: Effects of speaker awareness and referential contest. Journal of Memory and Language. 2003; 48:103–130.

Soma Y, Sugishita M, Kitamura K, Maruyama S, Imanaga H. Lexical agraphia in the Japanese language: pure agraphia for kanji due to left posteroinferior temporal lesions. Brain. 1989; 112:1549–61. [PubMed: 2597996]

Speedie LJ, Coslett HB, Heilman KM. Repetition of affective prosody in mixed transcortical aphasia. Archives of Neurology. 1984; 41:268–270. [PubMed: 6696645]

Starkstein SE, Federoff JP, Price TR, Leiguarda RC, Robinson RG. Neuropsychological and neuroradiologic correlates of emotional prosody comprehension. Neurology. 1994; 44:515–522. [PubMed: 8145924]

Strauss E, Moscovitch M. Perception of facial expression. Brain and Language. 1981; 13:308–332. [PubMed: 7260576]

Swets B, Desmet T, Hambrick D, Ferreira F. The role of working memory in syntactic ambiguity resolution: A working memory approach. Journal of Experimental Psychology: General. 2007; 136:64–81. [PubMed: 17324085]

Tahmasebi AM, David MH, Wild CJ, Rodd JM, Hakyemez H, Abolmaesumi P, Johnsrude IS. Is the link between anatomical structure and function equally strong at all cognitive levels of processing? Cerebral Cortex. 2011 doi:10.1093/cercor/bhr205.

Tomaiuolo F, MacDonald JD, Caramanos Z, Posner G, Chiavaras M, Evans AC, Petrides M. Morphology, morphometry and probability mapping of the pars opercularis of the inferior frontal gyrus: an in vivo MRI analysis. European Journal of Neuroscience. 1999; 11:3033–3046. [PubMed: 10510168]

Tucker D, Watson R, Heilman K. Discrimination and evocation of affectively intoned speech in patients with right parietal disease. Neurology. 1977; 27:947–950. [PubMed: 561908]

Van Lancker, D.; Breitenstein, C. Emotional dysprosody and similar dysfunctions. In: Bougousslavsky, J.; Cummings, JL., editors. Disorders of behavior and mood in focal brain lesions. Cambridge University Press; Cambridge, UK: 2000. p. 326-368.

Van Lancker D, Sidtis JJ. 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]

Van Lancker Sidtis D, Pachana N, Cummings JL, Sidtis JJ. Dysprosodic speech following basal ganglia insult: Toward a conceptual framework for the study of the cerebral representation of prosody. Brain and Language. 2006; 97:135–153. [PubMed: 16271755]

Van Lancker D. Cerebral lateralization of pitch cues in the linguistic signal. Papers in Linguistics: International Journal of Human Communication. 1980; 13:201–277.

Vul E, Kanwisher N. Begging the question: The non-independence error in fMRI data analysis. Foundations and Philosophy for Neuroimaging. 2010

Watson D, Gibson E. Making Sense of the Sense Unit Condition. Linguistic Inquiry. 2004; 35:508–517.

Weintraub S, Mesulam M-M, Kramer L. Disturbances in prosody: A right-hemisphere contribution to language. Archives of Neurology. 1981; 38:742–744. [PubMed: 7316838]

Wertz RT, Henschel CR, Auther LL, Ashford JR, Kirshner HS. Affective prosodic disturbance subsequent to right hemisphere stroke: A clinical application. Journal of Neurolinguistics. 1998; 11:89–102.

Wiethoff S, Wildgruber D, Kreifelts B, Becker H, Herbert C, Grodd W, Ethofer T. Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. Neuroimage. 2008; 39:885–893. [PubMed: 17964813]

Wildgruber D, Ackermann H, Kreifelts B, Ethofer T. Cerebral processing of linguistic and emotional prosody: fMRI studies. Progress in Brain Research. 2006; 156:249–268. [PubMed: 17015084]

Wildgruber D, Ethofer T, Kreifelts B, Grandjean D. Cerebral processing of emotional prosody: A network model based on fMRI studies. Speech Prosody. 2008

Witterman J, van IJzendoorn MH, van de Velde D, van Heuven VJJP, Schiller N. The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. Neuropsychologia. 2011; 49:3722–3738. [PubMed: 21964199]

Wohlschläger AM, Specht K, Lie C, Mohlberg H, Wohlschläger A, Bente K, Pietrzyk U, Stöcker T, Zilles K, Amunts K, Fink GR. Linking retinotopic fMRI mapping and anatomical probability maps of human occipital areas V1 and V2. Neuroimage. 2005; 26(1):73–82. [PubMed: 15862207]

Wong PCM. Hemispheric specialization of linguistic pitch patterns. Brain Research Bulletin. 2002; 59:83–95. [PubMed: 12379438]

Zhao J, Shu H, Zhang L, Wang X, Gong Q, Li P. Cortical competition during language discrimination. NeuroImage. 2008; 43:624–633. [PubMed: 18694836]

Zilles K, Schleicher A, Langemann C, Amunts K, Morosan P, Palomero-Gallagher N, Schormann T, Mohlberg H, Bürgel U, Steinmetz H, Schlaug G, Roland PE. Quantitative analysis of sulci in the human cerebral cortex: Development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. Human Brain Mapping. 1997; 5(4):218–221. [PubMed: 20408218]

Zurif EB, Mendelsohn M. Hemispheric specialization for the perception of speech sounds: The influence of intonation and structure. Perception and Psychophysics. 1972; 72:329–332.
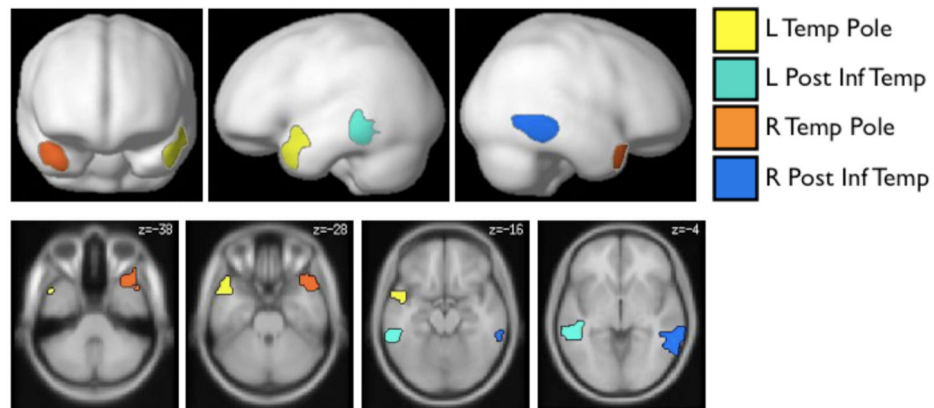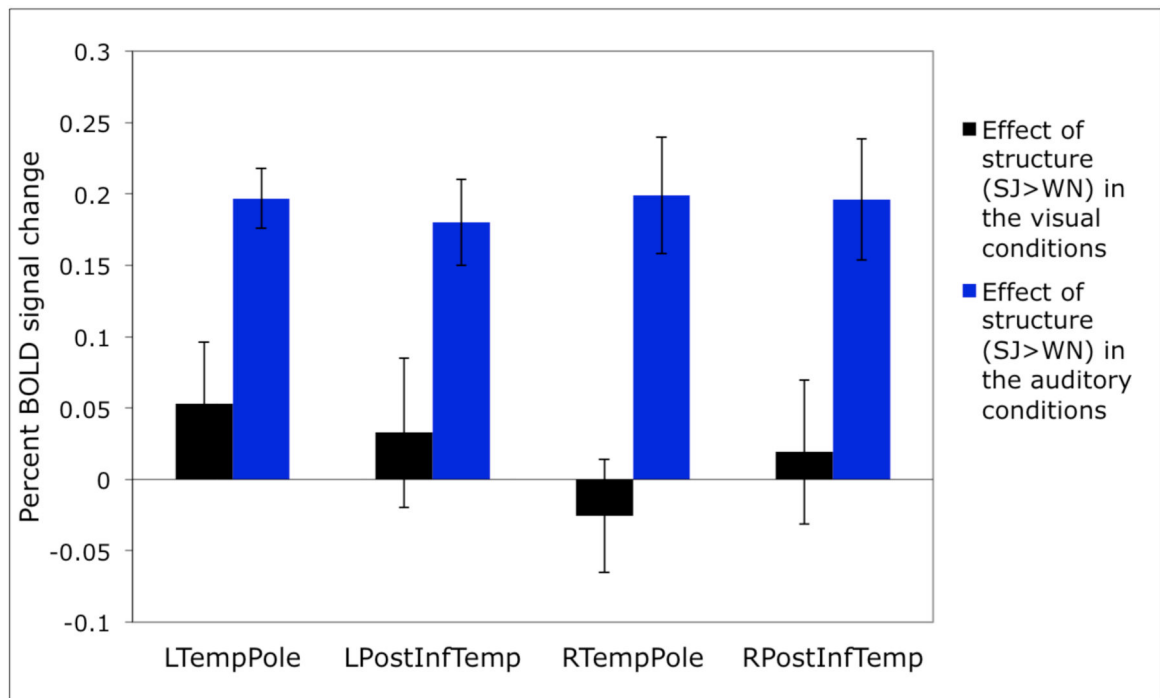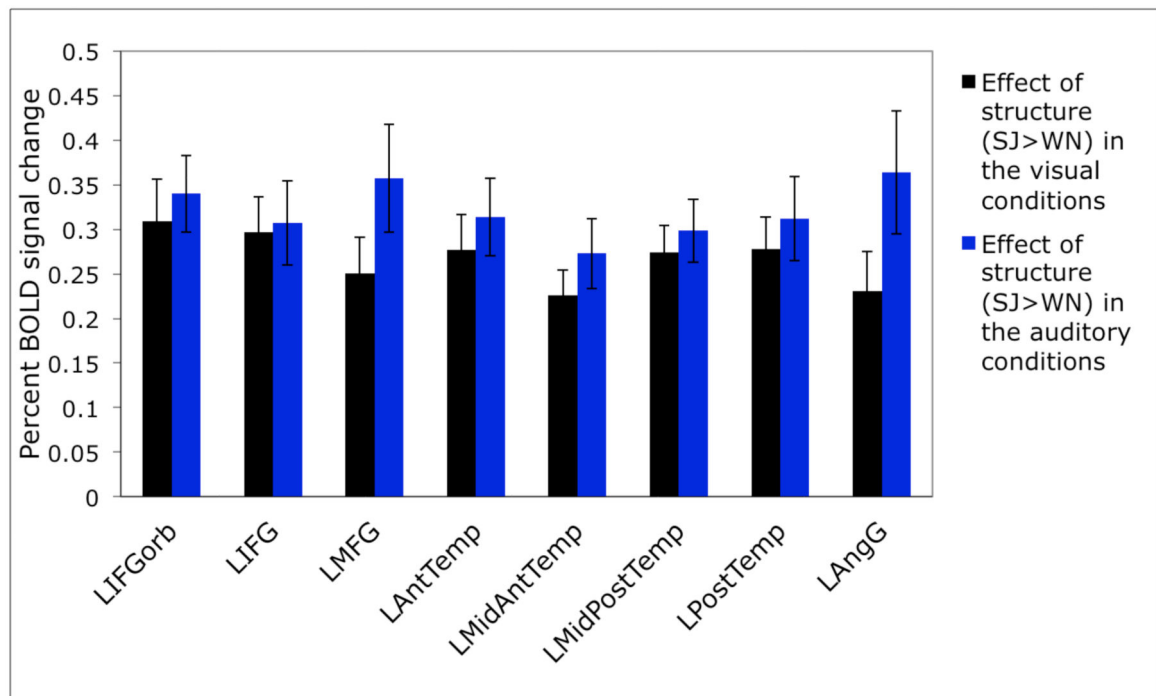
Figure 1.
Top: Prosody-sensitive parcels projected onto the cortical surface. The parcels show the locations where most subjects showed activation for the relevant contrasts (i.e., an effect of structure for the auditory, but not for the visual, conditions; see Methods for details). Bottom: Parcels projected onto axial slices (color assignments are the same in both views). [These parcels are available from the first author upon request and will soon be made available at: http://web.mit.edu/evelina9/www/funcloc.html.]

Figure 2.
Sensitivity to structure in the visual vs. auditory conditions in individually defined prosody-sensitive regions. (The responses are estimated using n-fold cross-validation, as discussed in Methods, so that data to define the fROIs and estimate the responses are independent.)

Figure 3.
Sensitivity to structure in the visual vs. auditory conditions in brain regions sensitive to high-level linguistic processing (defined in individual subjects using the sentences > pseudoword lists contrast; Fedorenko et al., 2010). (The responses are estimated using n-fold cross-validation, so that data to define the fROIs and estimate the responses are independent.)
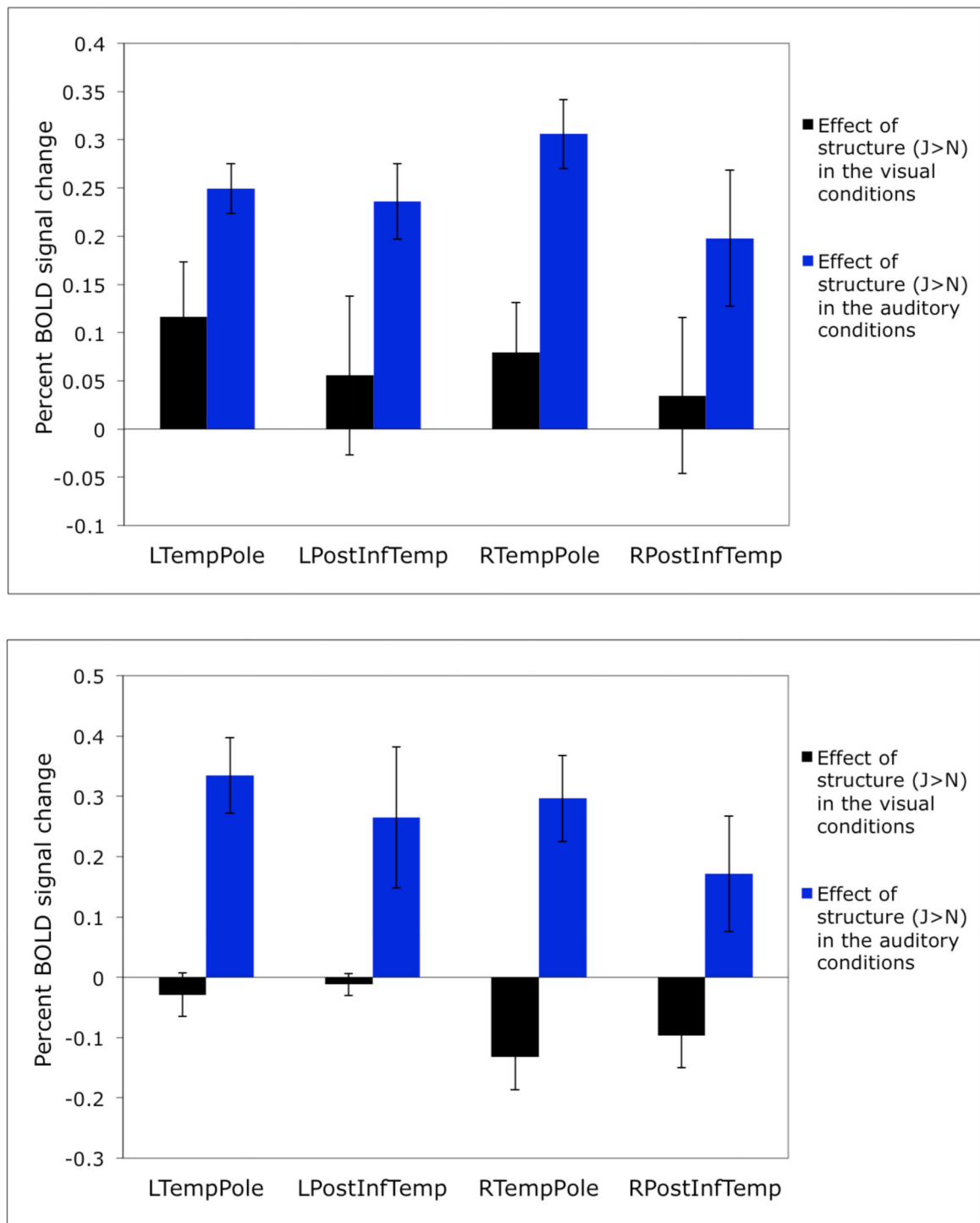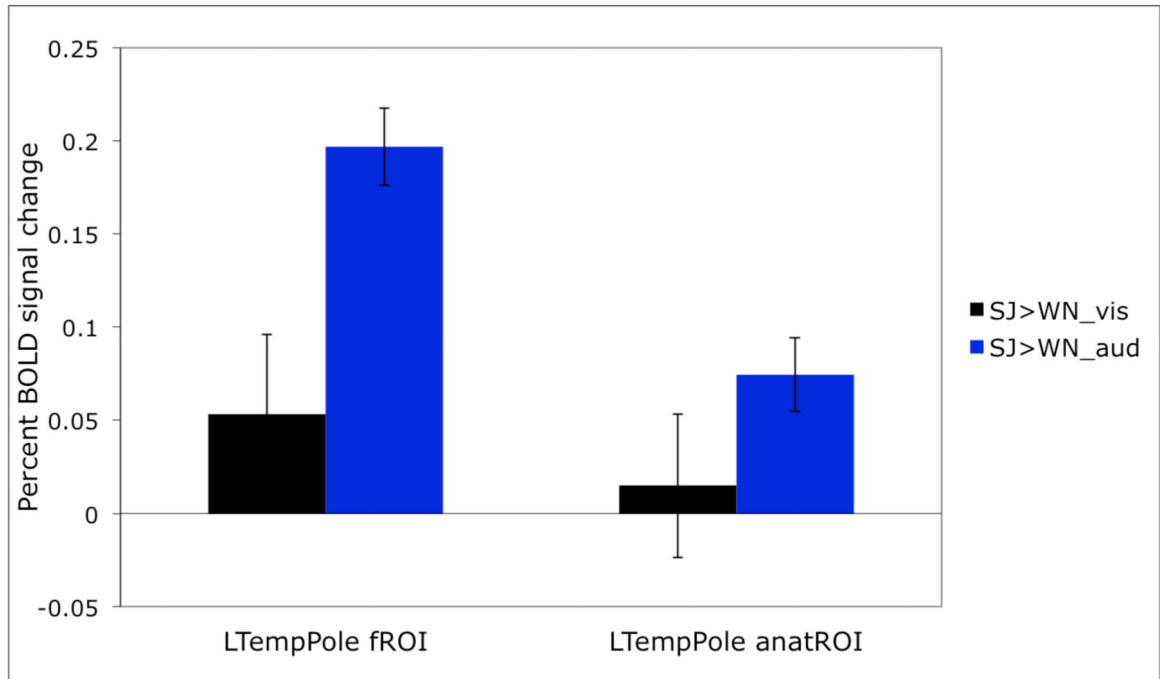
Figure 4.

a: Sensitivity to structure in the visual vs. auditory conditions in individually defined prosody-sensitive regions, for the subset of conditions consisting of real words: sentences and word lists). The fROIs are defined by a conjunction of two contrasts: i) a greater response to sentences than word lists in the auditory conditions, and ii) no difference

between sentences and word lists in the visual conditions. (Here, as in all the other analyses, the responses are estimated using n-fold cross-validation, as discussed in Methods, so that data to define the fROIs and estimate the responses are independent.) The Sentences > Word lists contrast is significant for the auditory conditions in all four ROIs (ps<.0005, except for RPostInfTemp whose p<.05). The Sentences > Word lists contrast is not significant for the visual conditions (except for LTempPole where it reaches significance at p<.05).

b: Sensitivity to structure in the visual vs. auditory conditions in individually defined prosody-sensitive regions, for the subset of conditions consisting of pseudowords: Jabberwocky sentences and pseudoword lists). The fROIs are defined by a conjunction of two contrasts: i) a greater response to Jabberwocky sentences than pseudoword lists in the auditory conditions, and ii) no difference between Jabberwocky sentences and pseudoword lists in the visual conditions. The Jabberwocky > Pseudoword-lists contrast is significant for the auditory conditions in three of the four ROIs (ps<.05) and marginal in RPostInfTemp (p=.062). The Jabberwocky > Pseudoword-lists contrast is not significant for the visual conditions in any of the ROIs.

Figure 5.
A comparison of the effects in the left temporal pole for individually defined functional ROIs vs. for the subject-independent anatomical ROI.

Table 1

Basic information on the prosody-sensitive brain regions. The units for the parcel sizes and the average individual fROI sizes are 2 mm isotropic voxels.

	Left Hemisphere		Right Hemisphere	
	Temp Pole	Post Inf Temp	Temp Pole	Post Inf Temp
Present in n/12 subjs	10	9	9	9
Parcel size	723	889	765	885
Average size of individual fROI	41	46	36	47

Table 2

Effects of sensitivity to structure in the visual and auditory conditions. We report uncorrected *p* values, but the effects of structure in the auditory conditions would remain significant after a Bonferroni correction for the number of regions (even including all 12 regions discovered by the GSS method).

	Left hemisphere		Right hemisphere	
	TempPole	**PostInfTemp**	**TempPole**	**PostInfTemp**
df	9	8	8	8
1. Effect of structure in the visual conditions	t<1.24; n.s.	t<1; n.s.	t<1; n.s.	t<1; n.s.
2. Effect of structure in the auditory conditions	t=9.42; p<.0001	t=5.98; p<.0005	t=4.85; p<.001	t=4.63; p<.001