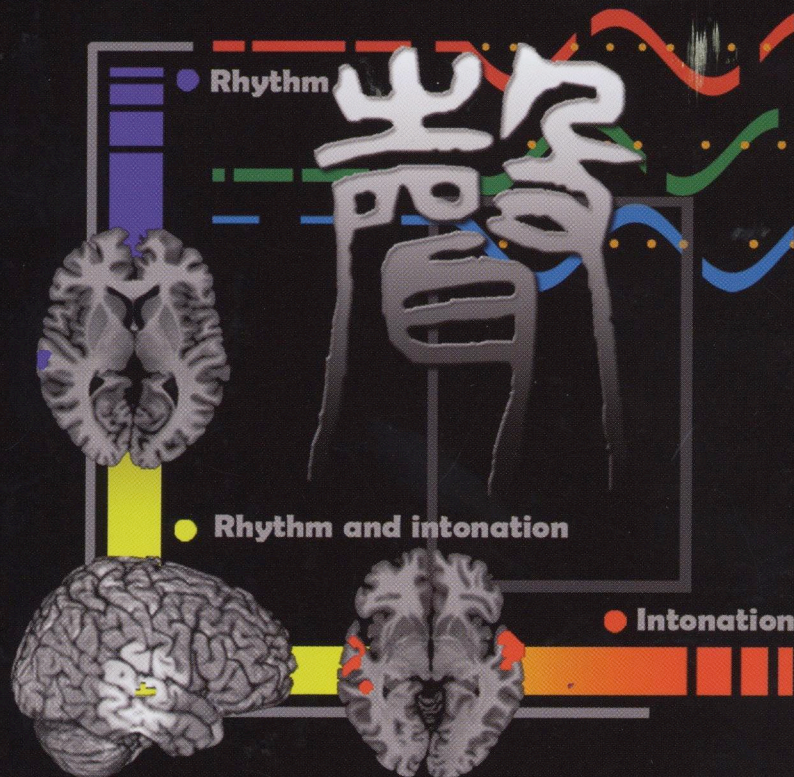


VOLUME 31, NUMBER 7, JULY 2010

Articles published online in Wiley InterScience,
27 October 2010 - 20 June 2010

HUMAN BRAIN MAPPING



 WILEY-BLACKWELL

ISSN 1065-9471

Common and Distinct Neural Substrates for the Perception of Speech Rhythm and Intonation

Linjun Zhang,^{1,2} Hua Shu,^{1*} Fengying Zhou,¹ Xiaoyi Wang,¹ and Ping Li^{3,4*}

¹State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

²College of Chinese Studies, Beijing Language and Culture University, Beijing, China

³Department of Psychology, Pennsylvania State University, University Park, Pennsylvania

⁴Center for Language Science, Pennsylvania State University, University Park, Pennsylvania

Abstract: The present study examines the neural substrates for the perception of speech rhythm and intonation. Subjects listened passively to synthesized speech stimuli that contained no semantic and phonological information, in three conditions: (1) continuous speech stimuli with fixed syllable duration and fundamental frequency in the standard condition, (2) stimuli with varying vocalic durations of syllables in the speech rhythm condition, and (3) stimuli with varying fundamental frequency in the intonation condition. Compared to the standard condition, speech rhythm activated the right middle superior temporal gyrus (mSTG), whereas intonation activated the bilateral superior temporal gyrus and sulcus (STG/STS) and the right posterior STS. Conjunction analysis further revealed that rhythm and intonation activated a common area in the right mSTG but compared to speech rhythm, intonation elicited additional activations in the right anterior STS. Findings from the current study reveal that the right mSTG plays an important role in prosodic processing. Implications of our findings are discussed with respect to neurocognitive theories of auditory processing. *Hum Brain Mapp* 31:1106–1116, 2010. © 2010 Wiley-Liss, Inc.

Key words: fMRI; speech rhythm; intonation; prosodic processing

INTRODUCTION

Contract grant sponsor: PCSIRT; Contract grant number: IRT0710 (to HS); Contract grant sponsor: Natural Science Foundation of China; Contract grant numbers: 60534080 (to HS), 30870758 (to HS); Contract grant sponsor: National Science Foundation; Contract grant number: BCS-0642586 (to PL); Contract grant sponsor: Fund for Foreign Scholars in University Research and Teaching Programs; Contract grant number: B07008 (to PL and HS).

*Correspondence to: Hua Shu, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China. E-mail: shuh@bnu.edu.cn; or Ping Li, Department of Psychology, Pennsylvania State University, University Park, PA 16802, USA. E-mail: pul8@psu.edu

Received for publication 6 May 2009; Revised 4 September 2009; Accepted 9 September 2009

DOI: 10.1002/hbm.20922

Published online 8 January 2010 in Wiley InterScience (www.interscience.wiley.com).

Rhythm and intonation are the two subsystems of prosody in natural speech. They are the so-called “suprasegmental” properties that typically span several phonemic segments (consonants and vowels) and are formed by changes in acoustic parameters such as pitch, amplitude, and duration. Rhythm and intonation carry important functions in speech comprehension and production, and they also play critical roles in language acquisition. While intonation contributes to both linguistic and emotional communication [Bänziger and Scherer, 2005; Cruttenden, 1986], rhythmic properties tend to shape speech segmentation and spoken-word recognition strategies [Cutler and Norris, 1988; Segui et al., 1990]. It also appears that rhythm and intonation are the two earliest dimensions of speech that infants use in the acquisition of native languages [Fernald and Kuhl, 1987; Ramus et al., 2000; see Kuhl [2004] for a review].

The neural mechanisms for the processing of prosody have been examined extensively in recent literature. It has been shown that intonation or the variation in pitch contours of an utterance is processed by a right lateralized fronto-temporal network, that is, the inferior and middle frontal gyrus and the superior temporal gyrus and sulcus [Friederici and Alter, 2004; Gandour et al., 2004; Hesling et al., 2005; Meyer et al., 2002, 2003, 2004; Plante et al., 2002]. High memory load tasks involving intonation also lead to asymmetric recruitment of the right frontal gyrus [Gandour et al., 2004; Plante et al., 2002]. More recently, research has demonstrated that patterns of activation and lateralization depend on the functional relevance of particular pitch patterns in the target language, irrespective of mechanisms underlying lower-level auditory processing [Gandour et al., 2000, 2002, 2003, 2004; Wong et al., 2004; Zatorre and Gandour, 2008]. For example, phonological processing of lexical tones by speakers of tonal languages always results in asymmetric activation in the left prefrontal, inferior parietal, and middle temporal regions, areas also involved in the processing of segmental consonants and vowels [Gandour et al., 2002; Hsieh et al., 2001]. Such findings indicate that the processing of intonation is supported by a large neural network in both the right and left hemispheres, depending on the task demand and the specific linguistic experience of the language user.

Most of the published studies so far, however, have focused on intonation or tone processing. There is a scarcity of research on the neural mechanisms of the processing of rhythm. Unlike intonation, which can be clearly defined by variations in pitch, speech rhythm has not been clearly specified, and researchers tend to use this term to refer to different things such as word stress [Böcker et al., 1999; Magne et al., 2007], prosodic structure of sentences [Pannekamp et al., 2005; Steinhauer et al., 1999], or regular meter verse rhythm [isochronous versus nonisochronous, Geiser et al., 2008; Riecker et al., 2002]. In other words, speech rhythm has been defined as a conglomerate of various suprasegmental cues such as duration, pitch, intensity, and pause.

Recent findings from speech typology and language acquisition have provided new evidence on the acoustic/phonetic characteristics of speech rhythm [Bunta and Ingram, 2007; Grabe et al., 1999; Grabe and Low, 2002; Nazzi and Ramus, 2003; Ramus et al., 1999]. In particular, speech or syllabic rhythm is defined in these studies as the way language is organized in time, and its acoustic correlates can be clearly identified by variations in vocalic and consonantal durations. Ramus and coworkers, among others, have conducted perceptual experiments with synthesized speech in which only rhythmic features were present and found that adults, infants, and even primates (e.g., cotton-top tamarins) could use speech rhythm information for language discrimination [Cho, 2004; Ramus, 2002; Ramus et al., 2000; Tincoff et al., 2005] and for accent perception [Low et al., 2000; Zhang, 2006]. These studies indicate that humans and primates are sensitive to the temporal variability of continuous speech.

An issue that has concerned the neuroimaging of language literature is whether overlapping or distinct regions of the brain are involved in the processing of speech prosody. Researchers have examined this issue with different aspects of prosody such as lexical tone versus sentence modality (declarative or interrogative intonation) and contrastive stress versus sentence modality [Gandour et al., 2003, 2004, 2007, Tong et al., 2005]. They showed that the processing of speech prosody could recruit either overlapping or distinct regions of the two hemispheres, depending on the acoustic properties and functions of the processed materials. For example, discrimination judgment of lexical tones and that of sentence modality by native speakers of Chinese share right-lateralized activation in the mSTS and mMFG and left-lateralized activation in the IPL, aSTG and pSTG, while at the same time, sentence modality elicits greater activity relative to tones in the left aMFG and right pMFG [Gandour et al., 2004]. However, no study has directly contrasted the processing of speech rhythm and that of intonation, and so the common versus distinct neural substrates issue here remains unclear.

In a recent fMRI study, Zhao et al. [2008] investigated the neural mechanisms underlying the interactions among prosodic (rhythm and intonation), phonological, and lexical-semantic information during language discrimination. They showed that patterns of cortical activity depend on the competitive advantage of the type of information available: stronger activation is seen in areas responsive to semantic processing rather than prosodic processing when both semantic and prosodic cues are present. When activation in the two prosodic conditions are contrasted, no areas of activation is found for the rhythm-only condition, but more areas of activation in the right STG/STS (BA42/22) and the left STG (BA41/22) are found for the rhythm + intonation condition. However, Zhao et al.'s study did not incorporate a baseline and the experimental design was nonparametric; therefore, no clear conclusion can be drawn from that study regarding the common or distinct neural correlates of speech rhythm and intonation.

Two additional studies that have used stimuli of pure tones and musical notes are also relevant to our current investigation. Griffiths et al. [1999] found the engagement of a right lateralized neural network including the fronto-temporal gyrus and cerebellum, for the processing of both music melody and rhythm. In a study with children, however, Overy et al. [2004] found stronger activation for melody processing compared to rhythm processing in the right superior temporal gyrus, indicating the existence of possibly different neural substrates for the perception of music melody versus rhythmic patterns. Given that there are similarities between music and speech rhythms and between music melody and speech intonation [Patel et al., 2006], we might ask, in light of the above conflicting findings, whether speech rhythm and intonation have overlapping or distinct neural substrates during processing.

The study of speech rhythm (temporal variations) and intonation (pitch variations) is of particular interest to the understanding of hemispheric functions in language

processing. The left hemisphere and the right hemisphere appear to play different roles in the processing of segmental/phonemic versus suprasegmental/prosodic information [Friederici and Alter, 2004; Gandour et al., 2003, 2004; Hesling et al., 2005; Meyer et al., 2002, 2004, 2005; Zaehle et al., 2004; Zhao et al., 2008], but whether and how rhythm and intonation engage the two hemispheres differentially is not yet clear. According to the “asymmetric sampling in time” (AST) hypothesis [Poeppel, 2003], different temporal integration windows are preferentially handled by the auditory system in the two hemispheres: the left hemisphere is more sensitive to rapidly changing (~20–40 ms) cues, while the right hemisphere is better adept at processing of slowly changing (~150–250 ms) cues. The AST hypothesis would thus suggest that both speech rhythm and intonation are related to right hemisphere functions, because both span a number of segments during a longer period. In contrast to the AST hypothesis, some theories suggest that temporal and spectral variations are encoded by cerebral networks in different hemispheres [Jamison et al., 2006; Meyer et al., 2005; Zaehle et al., 2004; Zatorre and Belin, 2001], according to which speech rhythm and intonation may be differentially related to left and right hemisphere functions. Given these contrasting hypotheses, it is important to examine speech rhythm and intonation together to identify the hemispheric lateralization of prosodic processing at the acoustic level.

To sum up, with respect to the neural substrates for the perception of prosody, the majority of the studies so far have focused on intonation and found that the perception of intonation involves widely distributed neural networks, modulated by task demands and functional properties. Speech rhythm has received much less attention and few studies have examined its neural correlates in the context of temporal organization of syllabic groups. Furthermore, no study has examined the perception of speech rhythm and that of intonation in a single experiment, thus leaving the question of common-versus-distinct neural correlates of speech rhythm and intonation unanswered. The present study aims to fill these gaps.

Our study makes the following predictions. First, the perception of speech rhythm and intonation should activate some overlapping brain areas in the auditory cortex, given the prosodic nature of both types of information. Second, because of differences in acoustic properties, speech rhythm and intonation might activate brain areas that are uniquely responsive to each type. Third, we hypothesize that the perception of speech rhythm and intonation are both supported preferentially by the right hemisphere, because they are suprasegmental properties spanning a longer period of time, based on the reasoning of the AST hypothesis as discussed earlier.

METHODS

Participants

Fifteen neurologically healthy volunteers (12 females, 3 males; aged 18–25, mean age = 20.9) with normal hearing

and minimal musical experience participated in the study. All participants were native speakers of Chinese, and all were college students at Beijing Normal University. They were all right handed according to a handedness questionnaire adapted from a modified Chinese version of the Edinburgh Handedness Inventory [Oldfield, 1971]. Participants gave written consent before they took part in the experiment. The experiment was approved by the ethics review board at Beijing Normal University’s Imaging Center for Brain Research.

Stimuli

The stimuli consisted of a total of 90 items in 3 conditions, 30 items in each condition: monotonous plus isochronous condition, which was used as a baseline and also called “standard” condition in the following, rhythm condition, and intonation condition.

As mentioned in the “Introduction,” intonation is the pitch variation of an utterance, the acoustic parameter of which is the fundamental frequency (F0). Speech rhythm can be defined as the way language is organized in time, and Ramus et al. [1999] and Low et al. [2000] have provided acoustic measures that can successfully capture the rhythmic characteristics of different languages. Here, we followed Ramus et al. and Low et al. in defining speech rhythm as the temporal variability of syllables, in particular, variation in vocalic durations, corresponding to ΔV and $nPVI$ of their rhythmic metrics.¹

To manipulate the test materials so that the acoustic cues at the two prosodic levels can be separated and to prevent influences of high-level lexical and phonological information, synthesized speech materials were used in our study, modeled after the stimuli used in previous studies of prosody [Cho, 2004; Dogil et al., 2002; Ramus, 2002; Riecker et al., 2002; Zhao et al., 2008]. The stimuli were continuous streams of speech and each one lasted 1,200 ms, consisting of six CV monosyllables (e.g., /sasasasasa/). In the standard condition, each item had a fixed F0 (175 Hz) and a fixed syllable duration (the duration of consonants and vowels in each syllable is 50 and 150 ms, respectively). It provided a monotonous and isochronous baseline. In the rhythm condition, the items were identical to the standard condition except that the durations of syllables varied randomly. In this condition, the syllable durations varied within the range of 140 to 320 ms; only the duration of vowels was varied and the duration of consonants was kept constant at 50 ms. Thus, the only difference between the rhythm condition and the standard condition is in the temporal structure of the speech stimuli. Finally, in the intonation condition, the items were identical to the standard condition except that different F0 contours were designed and each was applied to one of

¹ ΔV is defined as the standard deviation of vocalic intervals over a sentence, and $nPVI$ as the difference in duration between successive vocalic intervals.

the streams of syllables, replacing the original flat intonation contours. In this condition, all intonation contours included a regular decline toward their end and the average F0 of each “sentence” was 175 Hz (the pitch range was from 135 to 205 Hz), which is the same as that in the other two conditions. Thus, the only difference between the intonation condition and the standard condition is in the pitch patterns of the speech stimuli.

All the stimuli were synthesized by concatenation of diphones using the software MBROLA [Dutoit et al., 1996]. The duration and F0 information was fed into MBROLA using a French male database from which five consonants (/s/,/f/,/k/,/p/,/t/) and six vowels (/a/,/i/,/o/,/e/,/u/,/y/) were chosen for the synthesis.² The materials were matched in terms of average intensity (70 dB). The stimuli in the three conditions are illustrated in Figure 1.

Procedure

We used a passive listening task in which participants listened to synthesized speech materials that contained the relevant rhythmic and intonational information. This passive listening task was adopted because of concerns with potential top-down influences and the resulting activities in the frontal and temporal regions of the brain [Geiser et al., 2008; Plante et al., 2002; Poeppel et al., 1996; Reiterer et al., 2005; Tervaniemi and Hugdahl, 2003; Tong et al., 2005]. According to this reasoning, studies intended to explore low level acoustic/phonetic processing of speech rhythm and intonation should avoid participants’ active processing of stimuli that are phonologically relevant to the participant’s native language.

Each subject participated in two functional imaging sessions. In each session, the test materials were pseudorandomly presented in a fast event-related design. There was a total of 60 test trials in each session: 30 from the standard stimuli items, 15 from the rhythm items, and 15 from the intonation items. Each stimulus (six CV syllable) lasted 1,200 ms and a pure tone of 500 Hz lasting 200 ms was

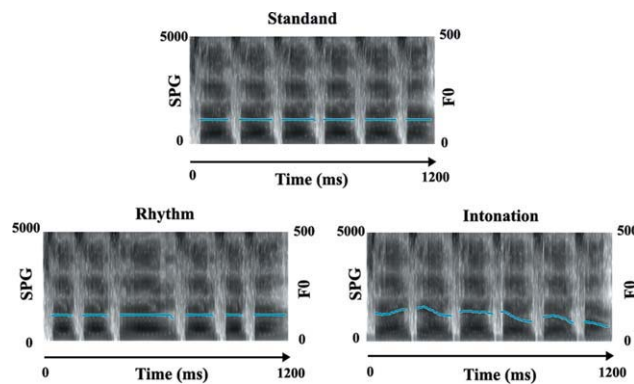


Figure 1.

Schematic illustration of an example stimulus from each of the three conditions. Broad-band spectrogram (SPG: 0–5 kHz) and voice fundamental frequency contours (F0: 0–500 Hz) are displayed. Stimuli in the standard condition (top) consisted of six CV monosyllables with fixed duration and fundamental frequency (F0 at about 175 Hz). The duration of consonants and vowels is 50 and 150 ms, respectively, in each syllable. It provided a monotonous and isochronous baseline. In the rhythm condition (bottom left), the fundamental frequency remained fixed at 175 Hz, but syllable durations varied by changing intervals of the vocalic portions. In the intonation condition (bottom right), the temporal patterns of the syllable sequences were the same as in the standard condition, but the F0 varied to form a changing pitch contour (the average F0 is about 175 Hz, the same as in the other two conditions). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

added before each trial. The interval between the tone and the trial varied from 200 to 600 ms (mean was 400 ms). A variable TOA design was used, in which the stimuli were presented at different trial onset asynchronies (TOA = 4–12 s; mean was 8 s). Subjects were instructed to lie still in the scanner and listen to the sounds, and press the left or the right button at the end of each stimulus item presented (counterbalanced for left versus right between the sessions). Subjects were reminded to maintain attention to the stimuli even though it was a passive listening task, and the required button pressing at the end of each stimulus was a way to ensure that they did pay attention to the stimuli (although without actively discriminating or identifying stimulus properties). This passive listening task with voluntary button presses has been used by several researchers to study the cognitive processes inherent in automatic perceptual analyses including early auditory processing and executive functions mediating attention and arousal [Gandour et al., 2004, 2007; Tong et al., 2005].

Imaging Acquisition

Scanning was performed on a Siemens MAGNETOM Trio (a Tim System) 3-T scanner equipped with a standard

²There are several reasons for why our synthesized stimuli were based on the MBROLA database (with French male voice) rather than natural speech from Mandarin Chinese. First, in this study, we attempt to derive stimuli that are orthogonal with respect to properties of rhythm and intonation (i.e., duration and pitch), and use of native phonological patterns from Mandarin will not allow us to do so. Second, stimuli based on natural speech from Mandarin could introduce confounds; for example, tonal information could affect prosodic processing results. Finally, the MBROLA database has been a widely used database in speech processing and it offers several distinct advantages, most clearly in its flexibility in manipulating suprasegmental features (e.g., pitch, duration) independently of segmental features. Given that the focus of our study is prosody, use of the French database as the phonemic/segmental basis should not adversely affect the results in our experiments.

quadrature head coil at the BNU Imaging Center for Brain Research. Participants heard the stimuli on both ears with a comfortable and equal volume. Each subject's head was aligned to the center of the magnetic field. They were instructed to relax, keep their eyes closed, and refrain from moving their head.

Functional data were acquired using a gradient-echo EPI pulse sequence (axial slices = 32) with the following parameters: thickness = 4 mm, FOV = 200 mm × 200 mm, matrix = 64 × 64, TR = 2,000 ms, TE = 30 ms, FA = 90°. High-resolution anatomic images (32 slices) of the entire brain were obtained using a 3D MPRAGE sequence after the functional images were acquired with the following parameters: TR = 2,530 ms, TE = 3.39 ms, FA = 7°, matrix = 512 × 512, FOV = 200 mm × 200 mm, thickness = 1.33 mm. A total of 254 volumes were acquired for each session and the session lasted 8 min. The entire experiment with two sessions took about 20 min, including about 4 min of anatomic scanning.

Data Analysis

Image analysis was conducted using the AFNI software package [Cox, 1996]. The first seven scans were excluded from data processing to minimize the transit effects of hemodynamic responses. Functional images were motion-corrected by aligning all volumes to the eighth volume using a six-parameter rigid-body transformation [Cox and Jesmanowicz, 1999]. Statistical maps were spatially smoothed with a 6-mm FWHM Gaussian kernel.

The preprocessed images were submitted to individual voxelwise deconvolution and multiple linear regression analyses to estimate the individual statistical *t*-maps for each condition. Using deconvolution and multiple regression analyses of AFNI's 3dDeconvolve, the estimated impulse response function (IRF) of each condition (standard, intonation, and rhythm) was obtained. Using multiple regression techniques (with motion-corrected regressors included), a regressor that specifies where each event occurs was first defined. This regressor was then time-shifted by 1 TR to create a new regressor that specified events 1 TR later. This was repeated up to 7 TRs (14 s, 2–14 s poststimulus). The deconvolution procedure then calculated a scaling value for each regressor using a least-squares multiple regression algorithm and these scalar values defined the IRF for each voxel. The estimated hemodynamic shape for each voxel was converted to a percentage area-under-the-curve score by expressing the area under the hemodynamic curve as a percentage of the area under the baseline, which has been used in many recent studies [e.g., Murphy and Garavan, 2005; Zhao et al., 2008]. In the current study, we extracted and averaged the values from 4 to 10 s poststimulus for each condition and each subject for subsequent across-subjects analysis of variance (ANOVA) and ROI analyses. Individual anatomic images, mean percentage signal change maps, and *t*-maps were

coregistered to the standard Talairach and Tournoux [1988] space. All images were resampled to 3 × 3 × 3 mm³ voxels. In a random effects analysis, Talairach-warped maps of percentage signal change (mean of 4–10 s poststimulus) for each condition and each participant were entered into a two-way, mixed-factor ANOVA with stimulus type as fixed factor and participant as random factor.

To address the multiple comparison correction and adjust for false positive errors on an area of activation basis, the Monte Carlo simulation was used as the basis of our statistical corrections. The group maps were thresholded at a voxel-level of $t > 2.976$ ($P < 0.01$) with an automated cluster detection in the contrasts of rhythm versus standard, intonation versus standard, and intonation versus rhythm, and a clusterwise threshold was set at corrected $P < 0.05$ with cluster size 23 (volume = 621 mm³), as determined by AFNI's AlphaSim (<http://afni.nimh.nih.gov>) for all intracranial voxels in the imaged volume.

The standard, rhythm, and intonation conditions were directly contrasted to identify the brain areas differentially responsive to increased variations in the intonation and rhythm domains. ROI analyses were then conducted to test lateralization in those regions identified. ROIs were defined functionally, on the basis of activation clusters from the group analysis.

We selected the peak activation coordinates from the cluster of the contrast analysis as the center of each ROI. These ROIs are uniform in size, with a sphere of 6 mm in radius. Percent signal change of the ROI was calculated by taking the means of all clusters in the sphere at the range of 4–10 s poststimulus in the previous IRF curves of each condition. Once selected, ROIs of a specific type were employed as masks to extract the mean percent signal change (averaged over the ROI) in the blood oxygen level-dependent (BOLD) response.

To determine which area was sensitive to the items in both the rhythm and intonation conditions as compared with the standard condition, a conjunction analysis was also performed based on the previous ANOVA results. After the group maps were obtained for the contrasts of rhythm versus standard, intonation versus standard, which were thresholded at a voxel-level of $t > 2.976$ ($P < 0.01$) with an automated cluster detection, the clusterwise threshold was set at corrected $P < 0.05$ (volume = 621 mm³) for all intracranial voxels in the imaged volume.

RESULTS

Whole-Brain Analysis

Table I presents a summary of all activation clusters, significant at corrected $P < 0.05$ in the planned comparisons: intonation versus standard, rhythm versus standard, and intonation versus rhythm. For all the planned comparisons, the areas activated were in the STG and STS areas of the auditory association cortex.

TABLE I. Areas of significant activation in planned comparisons, thresholded at a voxel-wise $P < 0.01$ ($t > 2.976$)

Region	BA	Peak voxel coordinates			Voxels	t
		x	y	z		
Intonation > standard						
Right STG/STS	38/21/22	52.5	-1.5	-6.5	162	5.66
Left STG/STS	38/21/22	-58.5	-10.5	-6.5	151	13.63
Right STS	21	46.5	-34.5	-0.5	26	4.36
Rhythm > standard						
Right STG	22	58.5	-16.5	2.5	25	4.04
Intonation > rhythm						
Right STS	21	55.5	-4.5	-9.5	24	3.98

Cluster level activated volume $\geq 621 \text{ mm}^3$ ($P < 0.05$, corrected). Note: Coordinates are in Talairach and Tournoux (1988) space, where the voxel with the maximum intensity for the cluster lies.

Below we focus on the planned comparisons between pairs of different stimulus conditions.

Intonation versus standard

When the intonation condition was contrasted with the standard condition, no additional areas of activation were found for the standard condition, but significantly more areas of activation were found for the intonation condition (Table I). The significant activations for the intonation condition were in the STG/STS areas bilaterally. The peak activation was in both hemispheres extending along the STG. In the right hemisphere, we found an additional significant cluster in the posterior STS.

Rhythm versus standard

When the rhythm condition was contrasted with the standard condition, no additional areas of activation were found for the standard condition, but significantly more areas of activation were found for the rhythm condition. The significant activations for the rhythm condition were in the right STG area (Table I).

Intonation versus rhythm

When the intonation condition was contrasted with the rhythm condition, no additional areas of activation were found for the rhythm condition, but significantly more areas of activation were found for the intonation condition. The significant activations for the intonation condition were in the right anterior STS area (Table I).

ROI Analysis

To further localize the effects of intonation and rhythm processing and the lateralization patterns, we conducted an ROI analysis. Given the significant clusters extending along the STG/STS areas, we selected the peak activation coordinates from seven clusters of the contrast analysis above as the center of each ROI. For the clusters that were activated only in one hemisphere, the center of left ROI and right ROI was flipped along the x axis to ensure that they were symmetrical in location. The centers of the ROIs were listed in Table II.

Five ROIs were selected for the intonation versus the standard condition (Table II). The activations were bilateral in the temporal pole (TP), aSTG, and mSTS, but tended to be right lateralized in the mSTG ($P = 0.077$; Fig. 2a) and pSTS ($P < 0.01$; Fig. 2b). For the rhythm versus the standard condition, the activation in the mSTG was significantly greater in the right hemisphere ($P < 0.05$; Fig. 2c). For the intonation versus the rhythm condition, the

TABLE II. Results of ROI analysis for the intonation > standard, rhythm > standard, and intonation > rhythm contrasts

	Left hemisphere				Right hemisphere				P
	x	y	z	Mean β value	x	y	z	Mean β value	
Intonation > standard									
TP	-43.5	16.5	-9.5	0.05	49.5	22.5	-12.5	0.07	0.273
aSTG	-52.5	1.5	-0.5	0.07	52.5	-1.5	-6.5	0.08	0.759
mSTS	-58.5	-10.5	-6.5	0.09	55.5	-16.5	-6.5	0.09	0.852
mSTG	-58.5	-22.5	2.5	0.05	58.5	-22.5	2.5	0.09	0.077
pSTS	-46.5	-34.5	-0.5	0.01	46.5	-34.5	-0.5	0.04	0.005
Rhythm > standard									
mSTG	-58.5	-16.5	2.5	0.02	58.5	-16.5	2.5	0.05	0.017
Intonation > rhythm									
aSTS	-55.5	-4.5	-9.5	0.02	55.5	-4.5	-9.5	0.08	0.003

Note: Coordinates are in Talairach and Tournoux (1988) space, where the voxel with the maximum intensity for the cluster lies.

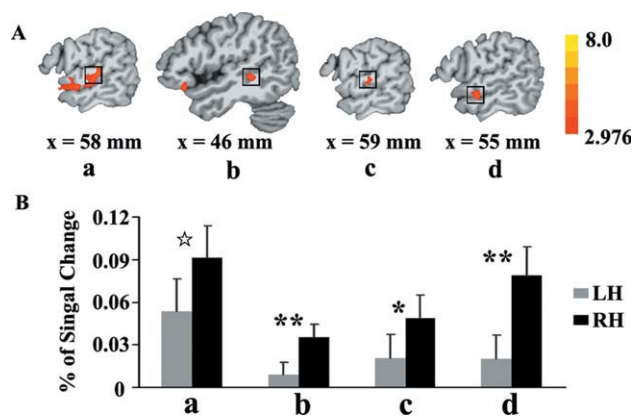


Figure 2.

(A) Clusters that were right lateralized in the ROI analysis for the three contrasts between the intonation condition and the standard condition (a, b), between the rhythm condition and the standard condition (c), and between the intonation condition and the rhythm condition (d), thresholded at a voxelwise $P < 0.01$ ($t > 2.976$). Cluster level activated volume ≥ 621 mm³. (B) Results of ROI analysis. Mean percent BOLD signal change extracted from ROIs based on functionally defined activation clusters in the left and right hemispheres. (a) mSTG; (b) pSTS; (c) mSTG; (d) aSTS. ☆ $P < 0.08$; * $P < 0.01$; ** $P < 0.01$.

activation in the aSTS was also right lateralized ($P < 0.01$; Fig. 2d).

Conjunction Analysis

A conjunction analysis was done to find out the overlapping brain areas activated in both rhythm and intonation conditions [Friston et al., 1999]. For this purpose, we identified a significant cluster which showed common activation in both the rhythm versus standard and the intonation versus standard contrast. The conjunction analysis revealed that a region in the right mSTG ($x = 60$, $y = -21$, $z = 3$) was activated (corrected $P < 0.05$) in both the rhythm and intonation conditions (Fig. 3).

DISCUSSION

Speech perception involves a series of processes in decoding segmental and suprasegmental information in the unfolding acoustic stream. A critical issue in the cognitive neuroscience of language is the understanding of the neural mechanisms in the processing of distinct features of speech sounds. In the present study, the neural correlates of two prosodic cues, that is, speech rhythm and intonation, have been investigated. Our results showed that speech rhythm elicited neural responses in the right mSTG, whereas intonation activated the bilateral STG/STS and the right pSTS regions. Conjunction analysis also revealed that rhythm and intonation activated a common

area in the right mSTG, but intonation elicited additional activations in the right aSTS. These results have significant implications for our understanding of the common versus distinct neural substrates for speech perception.

Role of STG/STS in the Perception of Speech Rhythm and Intonation

Active tasks such as recognition, comparison, and discrimination may recruit brain areas that are responsible for higher-order executive functions (e.g., attention and memory), most clearly in the frontal regions of the brain [Stephan et al., 2003; Tervaniemi and Hugdahl, 2003]. Our study adopted a passive perception task in which participants were not required to make overt responses according to characteristics of the stimuli, and their perception of speech rhythm and intonation resulted in activation in the STG/STS areas with no activation in fronto-parietal regions. Our results are consistent with the conclusion that the superior temporal region is implicated in passive auditory perception [Jamison et al., 2006; Plante et al., 2002; Zatorre et al., 1994; Zatorre and Belin, 2001].

A particularly important area in the processing of both speech rhythm and intonation is the right mSTG, as shown in the conjunction analysis. Previous imaging studies have found that the mSTG or nearby regions play an important role in the processing of human voices and animal vocalizations [Altmann et al., 2007; Belin et al., 2000; Lewis et al., 2005]. These areas, especially in the left hemisphere, were also reported to be sensitive to words more than to tones and to tones more than to white noise [Binder et al., 2000; Lewis et al., 2004; Wessinger et al., 2001]. As a result, the mSTG has been proposed to represent “intermediate” stages of auditory processing that are primarily involved in the encoding of structural acoustic attributes as opposed to high-level phonological and semantic information [Binder et al., 1997; Tranel et al., 2003]. The mSTG has also been found to be involved in the perception of temporal

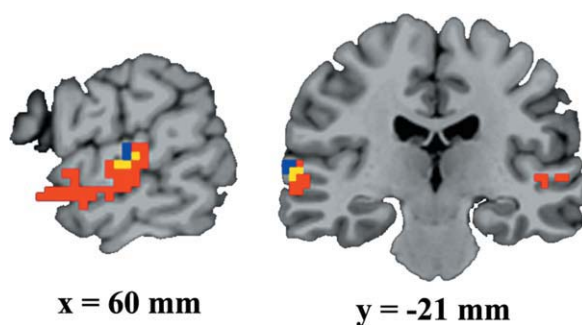


Figure 3.

Conjunction analysis revealed a common area in the middle of the right superior temporal gyrus (mSTG) for the perception of both speech rhythm and intonation (red: intonation > standard; blue: rhythm > standard; yellow: overlap).

and spectral aspects of non-speech stimuli [Boemio et al., 2005; Bueti et al., 2008; Molholm et al., 2005; Warren et al., 2003]. In the present study, speech rhythm and intonation involve temporal (duration) and spectral (fundamental frequency) variations, and the neural activities in the mSTG area are consistent with previous findings.

While differing from monotonous speech by temporal or spectral variations, speech rhythm and intonation differ between themselves too in their acoustic characteristics. Speech rhythm involves variations mainly in duration (temporal variation), while intonation involves variations in dynamic pitch (spectral variation). Lesion studies and functional imaging studies have already highlighted the differences due to the perceptual features of duration versus pitch. In the present study, compared to rhythm, intonation elicits additional stronger right hemispheric activation in the aSTS. This pattern accords well with the argument that the right anterior temporal cortex is particularly responsive to intonation and voice pitch information [Humphries et al., 2005; Lattner et al., 2005]. More generally, our result agrees with the proposal that the right STS specializes in the processing of spectral information [Jamison et al., 2006; Obleser et al., 2008; Overath et al., 2008; Warren et al., 2005].

Scott and Wise [2004] have outlined a neuroanatomical framework for auditory processing in terms of two distinct streams, the what pathway and the how pathway of auditory processing. Specifically, the what pathway encompasses the STG/STS regions, running lateral and anterior to the primary auditory cortex, with the right side involved in the processing of dynamic pitch variation seen in intonation and melody while the left side involved in the processing of sound-to-meaning mappings. Our findings from the current study appear to square evenly with Scott and Wise's framework, especially the right what pathway, in that the mSTG and aSTS regions are preferentially activated for prosodic and rhythmic processing in our results. More specifically, the projection from the primary auditory cortex to mSTG is involved in the processing of both duration and pitch variation, accounting for common neural activities of intonation and rhythm, while further projections from the right mSTG to STS are more responsive to dynamic pitch variation.

Hemispheric Lateralization in the Perception of Speech Rhythm and Intonation

A long-standing question in speech perception has been that of hemispheric lateralization for the processing of prosodic information. The above discussion already indicates hemispheric differences with respect to prosodic and lexical processing. While the general functional differences between the two hemispheres have been established [Gandour et al., 2000, 2002, 2004; Tong et al., 2005; Wong et al., 2004], the precise mechanism underlying the functional asymmetry at the acoustic level remains a matter of

debate. As discussed earlier, different studies have made different predictions regarding hemispheric functions for acoustic and prosodic processing. The AST model [Poepel, 2003] assumes that the processing of both speech rhythm and intonation is right lateralized because both types of cues span longer time period. By contrast, a number of researchers [Jamison et al., 2006; Zatorre and Belin, 2001] have argued that speech rhythm and intonation are encoded by different neural networks in the two hemispheres because duration is preferentially handled by the left while pitch by the right hemisphere.

In our present study, perception of intonation elicited bilateral activation in the STG/STS areas and our ROI analysis showed a significant right lateralization in the mSTG and pSTS. These results are consistent with the findings of previous imaging studies of the lateralization of prosodic processing [Gandour et al., 2004; Hesling et al., 2005; Meyer et al., 2002, 2004]. For example, Meyer et al. [2002, 2004] observed stronger right hemispheric activations for low-pass-filtered speech than for natural speech, and Hesling et al. [2005] observed right lateralized activation in STG when speech with a wider range of pitch variations was compared to speech with low degrees of prosodic information. Such findings indicate that the right auditory areas are dominant in the perception of intonation, a pattern similar to those found in studies of pitch perception using nonspeech and musical stimuli [e.g., Jamison et al., 2006; Penagos et al., 2004; Zatorre et al., 1994; Zatorre and Belin, 2001].

Compared with the well-established contribution of the right hemisphere to the processing of intonation, hemispheric lateralization for the processing of sound duration and timing, important characteristics of speech rhythm, has not been established. Some studies found a left hemispheric advantage [Belin et al., 1998; Brancucci et al., 2008; Giraud et al., 2005; Ilvonen et al., 2001; Molholm et al., 2005], others found a right hemispheric bias [Belin et al., 2002; Griffiths et al., 1999; Pedersen et al., 2000; Rao et al., 2001], and still others reported no asymmetries [Inouchi et al., 2002; Jancke et al., 1999; Takegata et al., 2004]. In our study, the perception of speech rhythm led to significant activation in the right mSTG, and the asymmetry was confirmed by the ROI analysis.

Many factors could have contributed to the conflicting findings reported in the literature, such as task demands, task difficulty, and stimulus properties. One contributing factor that we believe to be crucial is the nature of the stimuli used in previous studies, especially the duration of stimuli. Most of the studies supporting the left hemispheric advantage used stimuli with a duration range of tens of milliseconds, representative of the main differences between voiced and voiceless consonants and less characteristic of true rhythmic differences. In contrast, most of the studies supporting the right hemispheric bias had much longer stimulus duration, with a range of hundreds of milliseconds, more representative of the rhythmic differences in speech. In the current study, we used stimuli

with a duration range of 140–320 ms, consistent with studies that used longer durations [e.g., Boemio et al., 2005], and therefore right hemispheric advantage for speech rhythm has been observed.

Given our findings that both speech rhythm and intonation led to right lateralized activation in the auditory association cortex, the current study does not support the view that the left hemisphere is preferentially sensitive to temporal information (duration) while the right hemisphere is adept at processing of spectral information (pitch) [Jamison et al., 2006; Zatorre and Belin, 2001]. In contrast, our results are consistent with the AST model [Poeppel, 2003], according to which both rhythm and intonation are prosodic features of speech, with their acoustic characteristics spanning a longer time range and therefore both relating to right hemisphere functions. More recently, Glasser and Rilling [2008] used DTI tractography to detect asymmetry of human brain's language pathways with respect to both structure and function. They found that the left and right STG/MTG areas were both connected to the frontal lobe via the arcuate fasciculus to form dual pathways. According to this framework, the left pathway is involved in phonological and semantic processing, while the right pathway is involved in prosodic processing. Our results are also consistent with this dual-path framework in that the perception of both prosodic dimensions, rhythm and intonation, are related to right hemisphere functions in our study.

One issue that arises with the current study is whether the native language features of the participants (Chinese, a tonal language) would have affected the findings observed in this study. It can be reasoned that the life-long experience of the participants with a tonal language might affect neural responses to pitch patterns that are inherent in intonation, and therefore our findings might be confined to the specific speakers used in the study.³ While this possibility cannot be ruled out, so far there has not been neuroimaging evidence that demonstrates that experience with tones directly influences neural responses to low-level acoustic processing of pitch patterns, patterns as examined in this study that are not characteristic of the native language of the speakers. Previous research by Gandour and coworkers has shown that tonal language experience provides the speaker with enhanced sensitivity to linguistically relevant variations in pitch, but such effects occur only when stimuli reflect the contours representative of the target language tones [Gandour et al., 2004; Krishnan et al., 2009; Xu et al., 2006]. Future studies are needed to test speakers of nontonal languages in the processing of speech rhythm and intonation as in this study.

In conclusion, in this study, we have identified common neural substrates in the STG/STS area (more specifically the mSTG area) for the perception of both speech rhythm and intonation while isolating a particular role played by

the right anterior STS in the processing of intonation. Our results indicate that on the one hand, the perception of speech rhythm and intonation involves common neural mechanisms due to the acoustic properties associated with both prosodic dimensions, and on the other, certain brain regions may be selectively more responsive to specific acoustic features such as spectral variation in intonation. Our study also points to a more broad issue by suggesting complex mapping relations rather than one-to-one correspondences between neural activities in given brain regions and cognitive functions in linguistic processing. The exact correspondences between brain regions and cognitive processing are weighted by the specific acoustic properties, the speaker's linguistic experience, and the processing dynamics involved in the task at hand, as is being demonstrated in the fast growing literature on the neurobiology of language.

ACKNOWLEDGMENTS

We would like to thank Guoqing Xu and Wan Sun for their assistance in running the fMRI experiment.

REFERENCES

- Altmann CF, Doehrmann O, Kaiser J (2007): Selectivity for animal vocalizations in the human auditory cortex. *Cereb Cortex* 17:260–268.
- Bänziger T, Scherer KR (2005): The role of intonation in emotional expressions. *Speech Commun* 46:252–267.
- Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y (1998): Lateralization of speech and auditory temporal processing. *J Cogn Neurosci* 10:536–540.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000): Voice-selective areas in human auditory cortex. *Nature* 403:309–312.
- Belin P, McAdams S, Thivard L, Smith B, Savel S, Zilbovicius M, Samson S, Samson Y (2002): The neuroanatomical substrate of sound duration discrimination. *Neuropsychologia* 40:1956–1964.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997): Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17:353–362.
- Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, Possing J (2000): Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512–528.
- Böcker KBE, Bastiaansen MCM, Vroomen J, Brunia CHM, De Gelder B (1999): An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology* 36:706–720.
- Boemio A, Fromm S, Braun A, Poeppel D (2005): Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
- Brancucci A, D'Anselmo A, Martello F, Tommasi L (2008): Left hemisphere specialization for duration discrimination of musical and speech sound. *Neuropsychologia* 46:2013–2019.
- Bueti D, van Dongen EV, Walsh V (2008): The role of superior temporal cortex in auditory timing. *PLoS ONE* 3:e2481.
- Bunta F, Ingram D (2007): The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *J Speech Lang Hear Res* 50:999–1014.

³We are grateful to an anonymous reviewer for pointing out this issue.

- Cho MH (2004): Rhythm typology of Korean speech. *Cogn Process* 5:249–253.
- Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Cox RW, Jesmanowicz A (1999): Real-time 3D image registration for functional MRI. *Magn Reson Med* 42:1014–1018.
- Cruttenden A (1986): Intonation. New York: Cambridge University Press.
- Cutler EA, Norris D (1988): The role of strong syllables in segmentation for lexical access. *J Exp Psychol Hum Percept Perform* 14:113–121.
- Dogil G, Ackermann H, Grodd W, Haider H, Kamp H, Mayer J, Riecker A, Wildgruber D (2002): The speaking brain: A tutorial introduction to fMRI experiments in the production of speech, prosody, and syntax. *J Neurolinguistics* 15:59–90.
- Dutoit T, Pagel V, Pierret N, Bataille F, van der Vrecken O (1996): The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In: *The Proceedings of the Fourth International Conference on Language Processing (ICSLP'96)*. pp 1393–1396.
- Fernald A, Kuhl P (1987): Acoustic determinants of infant preference for motherese speech. *Infant Behav Dev* 10:279–293.
- Friederici AD, Alter K (2004): Lateralization of auditory language functions: A dynamic dual pathway model. *Brain Lang* 89:267–276.
- Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ (1999): Multisubject fMRI studies and conjunction analyses. *Neuroimage* 10:385–396.
- Gandour J, Wong D, Hsieh L, Weinzapfel B, Van Lancker D, Hutchins GD (2000): A crosslinguistic PET study of tone perception. *J Cogn Neurosci* 12:207–222.
- Gandour J, Wong D, Lowe M, Dziedzic M, Sathamnuwong N, Xu Y, Li X (2002): A crosslinguistic fMRI study of spectral and temporal cues underlying phonological processing. *J Cogn Neurosci* 14:1076–1087.
- Gandour J, Wong D, Dziedzic M, Lowe M, Tong Y, Li X (2003): 3 crosslinguistic fMRI study of perception of intonation and emotion in Chinese. *Hum Brain Mapp* 18:149–157.
- Gandour J, Tong Y, Wong D, Talavage T, Dziedzic M, Xu Y, Li X, Lowe M (2004): Hemispheric roles in the perception of speech prosody. *Neuroimage* 23:344–357.
- Gandour J, Tong Y, Talavage T, Wong D, Dziedzic M, Xu Y, Li X, Lowe M (2007): The neural basis of first and second language processing in sentence-level linguistic prosody. *Hum Brain Mapp* 28:94–108.
- Geiser E, Zaehle T, Jancke L, Meyer M (2008): The neural correlate of speech rhythm as evidenced by metrical speech processing. *J Cogn Neurosci* 20:541–552.
- Giraud K, Demonet JF, Habib M, Marquis P, Chauvel P, Liegeois-Chauvel C (2005): Auditory evoked potential patterns to voiced and voiceless speech sounds in adult developmental dyslexics with persistent deficits. *Cereb Cortex* 15:1524–1534.
- Glasser MF, Rilling JK (2008): DTI tractography of the human brain's language pathways. *Cereb Cortex* 18:2471–2482.
- Grabe E, Low EL (2002): Durational variability in speech and the rhythm class hypothesis. In: Gussenhoven C, Warner N, editors. *Laboratory Phonology VII*. Berlin: Mouton de Gruyter. pp 515–546.
- Grabe E, Post B, Watson I (1999): The acquisition of rhythmic patterns in English and French. In: *Proceedings of the International Congress of Phonetic Sciences*. pp 1201–1204.
- Griffiths TD, Johnsrude I, Dean JL, Green GGR (1999): A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport* 10:3825–3830.
- Hesling I, Clement S, Bordessoules M, Allard M (2005): Cerebral mechanisms of prosodic integration: Evidence from connected speech. *Neuroimage* 24:937–947.
- Hsieh L, Gandour J, Wong D, Hutchins G (2001): Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang* 76:227–252.
- Humphries C, Love T, Swinney D, Hickok G (2005): Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum Brain Mapp* 26:128–138.
- Ilvonen TM, Kujala T, Tervaniemi M, Salonen O, Naatanen R, Pekkonen E (2001): The processing of sound duration after left hemisphere stroke: Event-related potential and behavioral evidence. *Psychophysiology* 38:622–628.
- Inouchi M, Kubota M, Ferrari P, Roberts TP (2002): Neuromagnetic auditory cortex responses to duration and pitch changes in tones: Crosslinguistic comparisons of human subjects in directions of acoustic changes. *Neurosci Lett* 31:138–142.
- Jamison HL, Watkins KE, Bishop DV, Matthews P (2006): Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb Cortex* 16:1266–1275.
- Jancke L, Buchanan T, Lutz K, Specht K, Mirzazade S, Shah NJ (1999): The time course of the BOLD response in the human auditory cortex to acoustic stimuli of different duration. *Brain Res Cogn Brain Res* 8:117–124.
- Krishnan A, Swaminathan J, Gandour J (2009): Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J Cogn Neurosci* 21:1092–1105.
- Kuhl PK (2004): Early language acquisition: cracking the speech code. *Nat Rev Neurosci* 5:831–843.
- Lattner S, Meyer ME, Friederici AD (2005): Voice perception: Sex, pitch, and the right hemisphere. *Hum Brain Mapp* 24:11–20.
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA (2004): Human brain regions involved in recognizing environmental sounds. *Cereb Cortex* 14:1008–1021.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005): Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 25:5148–5158.
- Low EL, Grabe E, Nolan F (2000): Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Lang Speech* 43:377–401.
- Magne C, Astésano C, Aramaki M, Ystad S, Kronlund-Martinet R, Besson M (2007): Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cereb Cortex* 17:2659–2668.
- Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY (2002): fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17:73–88.
- Meyer M, Alter K, Friederici AD (2003): Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. *J Neurolinguistics* 16:277–300.
- Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY (2004): Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang* 89:277–289.
- Meyer M, Zaehle T, Gountouna V-E, Barron A, Jancke L, Turk A (2005): Spectral-temporal processing during speech perception involves left posterior auditory cortex. *Neuroreport* 16:1985–1989.

- Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ (2005): The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. *Cereb Cortex* 15:545–551.
- Murphy K, Garavan H (2005): Deriving the optimal number of events for an event-related fMRI study based on the spatial extent of activation. *Neuroimage* 27:771–777.
- Nazzi T, Ramus F (2003): Perception and acquisition of linguistic rhythm by infants. *Speech Commun* 41:233–243.
- Obleser J, Eisner F, Kotz SA (2008): Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci* 28:8116–8123.
- Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia* 9:97–113.
- Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008): Encoding of spectral correlation over time in auditory cortex. *J Neurosci* 28:13268–13273.
- Overy K, Norton A, Cronin K, Gaab N, Alsop D, Schlaug G (2004): Imaging melody and rhythm processing in young children. *Neuroreport* 15:1723–1726.
- Pannekamp A, Toepel U, Alter K, Hahne A, Friederici AD (2005): Prosody-driven sentence processing: An event-related brain potential study. *J Cogn Neurosci* 17:407–421.
- Patel AD, Iversen, JR, Rosenberg, JC (2006): Comparing the rhythm and melody of speech and music: The case of British English and French. *J Acoust Soc Am* 119:3034–3047.
- Pedersen CB, Mirz F, Ovesen T, Ishizu K, Johannsen P, Madsen S, Gjedde A (2000): Cortical centres underlying auditory temporal processing in humans: A PET study. *Audiology* 39:30–37.
- Penagos H, Melcher JR, Oxenham AJ (2004): A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci* 24:6810–6815.
- Plante E, Creusere M, Sabin C (2002): Dissociating sentential prosody from sentence processing: Activation interacts with task demands. *Neuroimage* 17:401–410.
- Poeppl D (2003): The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun* 41:45–255.
- Poeppl D, Yellin E, Phillips C, Roberts TP, Rowley HA, Wexler K, Marantz A (1996): Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Brain Res Cogn Brain Res* 4:231–242.
- Ramus F (2002): Language discrimination by newborns: Teasing apart phonotactic, rhythmic and intonational cues. *Annu Rev Lang Acquis* 2:85–115.
- Ramus F, Nespor M, Mehler J (1999): Correlates of linguistic rhythm in the speech signal. *Cognition* 73:265–292.
- Ramus F, Hauser MD, Miller C, Morris D, Mehler J (2000): Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288:349–351.
- Rao SM, Mayer AR, Harrington DL (2001): The evolution of brain activation during temporal processing. *Nat Neurosci* 4:317–323.
- Reiterer SM, Erb M, Droll CD, Anders S, Ethofer T, Grodd W, Wildgruber D (2005): Impact of task difficulty on lateralization of pitch and duration discrimination. *Neuroreport* 16:239–242.
- Riecker A, Wildgruber D, Dogil G, Grodd W, Ackermann H (2002): Hemispheric lateralization effects of rhythm implementation during syllable repetitions: An fMRI study. *Neuroimage* 16:169–176.
- Scott SK, Wise RJ (2004): The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92:13–45.
- Segui J, Djupoux E, Mehler J (1990): The role of the syllable in speech segmentation, phoneme identification and lexical access. In: Altmann G, Sillcock R, editors. *Cognitive Models of Speech Processing*. Cambridge: MIT Press. pp 263–280.
- Steinhauer K, Alter K, Friederici AD (1999): Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat Neurosci* 2:191–196.
- Stephan KE, Marshall JC, Friston KJ, Rowe JB, Ritzl A, Zilles K, Fink GR (2003): Lateralized cognitive processes and lateralized task control in the human brain. *Science* 301:384–386.
- Takegata R, Nakagawa S, Tonoike M, Naatanen R (2004): Hemispheric processing of duration changes in speech and non-speech sounds. *Neuroreport* 15:1683–1686.
- Talairach J, Tournoux P (1988): *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. New York: Thieme.
- Tervaniemi M, Hugdahl K (2003): Lateralization of auditory-cortex functions. *Brain Res Rev* 43:231–246.
- Tincoff R, Hauser N, Tsao F, Spaepen G, Ramus F, Mehler J (2005): The role of speech rhythm in language discrimination: Further tests with a non-human primate. *Dev Sci* 8:26–35.
- Tong Y, Gandour J, Talavage T, Wong D, Dziedzic M, Xu Y, Li X, Lowe M (2005): Neural circuitry underlying sentence-level linguistic prosody. *Neuroimage* 28:417–428.
- Tranel D, Damasio H, Eichhorn GR, Grabowski TJ, Ponto LLB, Hichwa, RD (2003): Neural correlates of naming animals from their characteristic sounds. *Neuropsychologia* 41:847–854.
- Warren JD, Uppenkamp S, Patterson RD, Griffiths TD (2003): Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci USA* 100:10038–10042.
- Warren JD, Jennings AR, Griffiths TD (2005): Analysis of the spectral envelop of sounds by the human brain. *Neuroimage* 24:1052–1057.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauchschecker JP (2001): Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1–7.
- Wong PC, Parsons LM, Martinez M, Diehl RL (2004): The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *J Neurosci* 24:9153–9160.
- Xu Y, Krishnan A, Gandour J (2006): Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport* 17:1601–1604.
- Zaehle T, Wüstenberg T, Meyer M, Jäncke L (2004): Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fMRI study. *Eur J Neurosci* 20:2447–2456.
- Zatorre RJ, Belin P (2001): Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Zatorre RJ, Gandour JT (2008): Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* 363:1087–1104.
- Zatorre RJ, Evans AC, Meyer E (1994): Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.
- Zhang L (2006): *Rhythm in Chinese and its acquisition*. Unpublished Doctoral Dissertation, Beijing Normal University, Beijing.
- Zhao J, Shu H, Zhang L, Wang X, Gong Q, Li P (2008): Cortical competition during language discrimination. *Neuroimage* 43:624–633.