

# Differential Roles of Right Temporal Cortex and Broca's Area in Pitch Processing: Evidence from Music and Mandarin

Yun Nan<sup>1\*</sup> and Angela D. Friederici<sup>2</sup>

<sup>1</sup>National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

<sup>2</sup>Department of Neuropsychology, MPI for Human Cognitive and Brain Sciences, Leipzig, Germany

---

**Abstract:** Superior temporal and inferior frontal cortices are involved in the processing of pitch information in the domain of language and music. Here, we used fMRI to test the particular roles of these brain regions in the neural implementation of pitch in music and in tone language (Mandarin) with a group of Mandarin speaking musicians whose pertaining experiences in pitch are similar across domains. Our findings demonstrate that the neural network for pitch processing includes the pars triangularis of Broca's area and the right superior temporal gyrus (STG) across domains. Within this network, pitch sensitive activation in Broca's area is tightly linked to the behavioral performance of pitch congruity judgment, thereby reflecting controlled processes. Activation in the right STG is independent of performance and more sensitive to pitch congruity in music than in tone language, suggesting a domain-specific modulation of the perceptual processes. These observations provide a first glimpse at the cortical pitch processing network shared across domains. *Hum Brain Mapp* 34:2045–2054, 2013. © 2012 Wiley Periodicals, Inc.

**Key words:** fMRI; musicians; tone

---

## INTRODUCTION

Pitch, as the percept of the fundamental frequency, is a vital element of music and auditory language. It forms essential variations in melodies and it establishes the basic intonational patterns in speech. Melodic pitch in language has two functional parts, one at the level of syntax indicating intonational phrase boundaries and at the same time syntactic boundaries [Meyer et al., 2004; Steinhauer et al.,

1999], and one at the lexical level determining word meaning in tonal languages [Nan et al., 2010]. The latter form of pitch bears an important resemblance with musical melody: pitch variations occur in small units.

The abilities to process musical pitch and linguistic tone are closely linked. Pitch-related expertise seems transferable across domains [Bidelman et al., 2011; Magne et al., 2006; Moreno et al., 2009; Schon et al., 2004; Wong et al., 2007]. Music training facilitates pitch perception as reflected by behavioral and event-related brain potential measures in both music and language [Schon et al., 2004]. Follow-up training studies with 8-year-old children observed similar effects [Magne et al., 2006] in linguistic pitch discrimination and a general enhancement in reading [Moreno et al., 2009]. Compared to nonmusicians, musicians have more accurate brainstem pitch tracking of linguistic tones [Wong et al., 2007]. Experience with a tone language, similar to that with music, could greatly improve the accuracy of subcortical linguistic and musical pitch tracking [Bidelman et al., 2011]. Along the same line, there is a behavioral study showing that pitch in speech

---

Contract grant sponsor: National Natural Science Foundation of China; Contract grant number: 30700225.

\*Correspondence to: Yun Nan, National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, 19 Xin-Wai St., Hai-Dian District, Beijing 100875, People's Republic of China. E-mail: dr.yunnan@gmail.com

Received for publication 27 April 2011; Revised 22 November 2011; Accepted 2 January 2012

DOI: 10.1002/hbm.22046

Published online 19 March 2012 in Wiley Online Library (wileyonlinelibrary.com).

and music might be stored in the same short-term memory [Semal et al., 1996]. Conversely, pitch deficits affect not only music, but also sometimes language, tonal [Nan et al., 2010], and nontonal languages [Liu et al., 2010] alike. Although congenital amusia is widely accepted as a neuro-genetic disorder that selectively affects the processing of musical pitch [Peretz, 1993], recent studies have found subgroups of amusic individuals with parallel pitch deficits in language, showing impairments in the discrimination and identification of Mandarin lexical tones among tone language speakers [Nan et al., 2010] or speech intonation problems among nontone language speakers [Liu et al., 2010]. A recent study suggests that the pitch deficits of amusia may even compromise the ability to process and learn tone languages for nontone language speakers [Tillmann et al., 2011].

These findings could be best explained by a common mechanism encoding pitch processing across domains. However, strong evidence for a common neural basis is still lacking [but see Nan et al., 2009]. Conventionally, melodic pitch processing seems to be right lateralized [see Limb, 2006 for a review; Zatorre et al., 1992], mostly centered in fronto-temporal regions [Hyde et al., 2006, 2007, 2008, 2010; Loui and Schlaug, 2009]. In the language domain, however, accumulating evidence suggests a much wider network participating in processing of sentential prosody and lexical tone, including both left- and right-hemispheric regions [for reviews, see Friederici and Alter, 2004; Wong et al., 2009]. Linguistically relevant lexical tones are reported to normally recruit left superior temporal and frontal language areas [Gandour et al., 1998, 2000; Liang and van Heuven, 2004]. However, the right fronto-temporal network which is implicated in musical pitch and sentence-level prosody also seems to play a certain role in forming lexical tone perception [Klein et al., 2001; Luo et al., 2006; Liu et al., 2006].

What adds complexity to this question is the malleability of the auditory system [Kraus and Banai, 2007]. The neural representation of pitch is largely shaped by related experiences in each domain. Compared to nonmusicians, musicians demonstrated more faithful frequency following responses toward linguistic pitch contours [Wong et al., 2007]. In addition, tone language experience also demonstrates similar effects upon representations of pitch in the brainstem [Bidelman et al., 2011]. To understand the brain mechanism underlying pitch processing in music and tone language, it is thus important to consider participants whose pertaining experiences in pitch are similar across domains.

Chinese musicians are a special group of people, with expertise in both music and tone language (Mandarin). They are the best biological models for observing the possible neural interaction between linguistic and musical pitch. With Chinese musicians as participants, the current fMRI study adopted a pitch congruity judgment task from Nan et al. (2009), in which the participants were required to first listen to musical or linguistic phrases ending in

either congruous or incongruous pitches then judge the congruity of the phrases.

We predict that areas in the right auditory regions should constitute the general pitch network. Given that the left inferior frontal gyrus (IFG) is not only engaged in linguistic incongruity processing [for a review, see Friederici, 2002; Friederici and Kotz, 2003], but also sensitive to incongruities in the musical phrases [Maess et al., 2001; Koelsch et al., 2002; Sammler et al., 2011], we hypothesize that the left IFG is another common candidate region for pitch processing in language and music.

## MATERIALS AND METHODS

### Participants

Eighteen healthy young adult female musicians (mean age, 20.8 years; SD, 1.7 years) from China Conservatory of Music (recruited on campus through flyers), with normal hearing abilities were paid for participation in the current study. All participants were right-handed, spoke Mandarin as their first language, and met standard MRI safety criteria. None of them reported to be an absolute pitch possessor. All these musicians started to play musical instruments at an age around 6 years old ( $6.0 \pm 2.1$ ). Eleven out of 18 participants reported to be able to play at least two instruments. The present main instruments were piano (14), accordion (1), violin (1), lute (1), and Guzheng (1). On average, these musicians practiced 2.6 hours ( $2.6 \pm 0.9$ ) on a daily basis. Informed consents were obtained from all participants before the study. The experimental protocol was approved by the Ethics Committee of Beijing Normal University.

### Materials and Procedure

The stimulus materials and experimental paradigm were adopted from Nan et al. (2009). To avoid long measurement time under the scanner, a set of two thirds of the original 240 stimuli [Nan et al., 2009] were chosen for the current study, which included 80 quadrisyllabic Chinese phrases and 80 four-note musical phrases. Each phrase lasted 2000 ms, with four equal-length units. All the Chinese phrases were produced by a female native Mandarin speaker. Half of the Chinese phrases were semantically meaningful. The other half was acoustically similar to the first half except ending in incongruous tonal contours as a result of cross-splicing. This resulted in 40 incongruous language phrases that are either semantically incorrect (incongruous meaning, 12 of 40) or combined semantically/syntactically incorrect (incongruous meaning and incorrect syntactic word category, 28 of 40). Similarly, the 80 musical phrases consisted of 40 congruous and 40 incongruous phrases, with minimal acoustic differences in the ending notes between the two types. As compared to the congruous musical phrases, the incongruous counterparts



Figure 1.

Example of stimuli used in the experiment. In Mandarin Chinese, lexical tones determine word meaning. The first panel illustrates a pair of congruous and incongruous language phrases. When the phrase ends with a third tone [i.e., bi(3)], it is congruous and means “catch one thing and lose another.” Whereas when

the phrase ends with a second tone [i.e., bi(2)], it is incongruous and means “catch one thing and lose one’s nose.” Numbers in parentheses are lexical tone labels. The second panel shows musical notations for a pair of congruous and incongruous musical phrases.

contained a different fourth note which formed incongruous intervals with the previous tones, and thus causing syntactical violations in musical syntax. In sum, the congruous and incongruous phrase pairs in music and language both contained pitch manipulations in the ending units which resulted in syntactic errors in music and semantic or semantic and syntactic errors in language. Figure 1 lists one pair of Chinese phrases and one pair of musical phrases [please see Nan et al., 2009 for more examples].

Recordings of the language phrases were made in a soundproof booth with a Sony 60EC digital recorder and a NT1 microphone with a Samson MDR8 mixer. Musical phrases were created using a YAMAHA DGX-620 keyboard with a grand piano timbre. All the language and musical stimuli were digitized at a 16bit/44.1 kHz sampling rate and were root mean square amplitude normalized using Praat 4.4.26 [Boersma, 2001] to 70 dB.

All participants completed two experimental runs with mixed musical and Chinese phrases presented binaurally in a pseudo-random order with a task of congruity judgment. Each run contained 80 phrases with short rest periods (each 2 s in duration) in between. Participants controlled the resting duration between runs as needed. Before scanning, a detailed instruction and practice session was given to ensure that participants understood the task well. The experimental design was event-related. A trial sequence normally consisted of a 2 s of initial fixation, a 2 s of stimulus presentation followed by a 1 s of pause, a 1.5 s of response collection time (with a visual prompt using a question mark), a 0.5 s of visual feedback (with a visual prompt of “correct” or “wrong”), at last ended with a 1 s of final fixation. For some of the trials within the run, the initial fixation could be randomly shifted to 0 s with a paired final fixation lasted 3 s. As a result, each stimulus onset followed the previous one 4 s, 6 s, or 8 s later. The participants were asked to press buttons (“congruous” or “incongruous”) to indicate the congruity of the stimulus. The response buttons and hands were counterbalanced across participants.

## MR Imaging

Twenty-eight axial slices (field of view =  $1200 \times 1200$  mm, with a  $64 \times 64$  matrix, 4 mm thickness, 1 mm interslice distance) were collected from a Siemens 3T scanner (Siemens TRIO, Erlangen, Germany). The images were parallel to the AC-PC plane and covering the whole brain, acquired using a single shot gradient recalled EPI sequence (repetition time = 2000 ms, echo time = 30 ms, flip angle =  $90^\circ$ ). One functional run consisted of 320 scans, with each scan sampling over 28 slices. One-hundred and sixty time points were collected for each condition. High-resolution ( $1.0 \times 1.0 \times 1.0$  mm)  $T_1$ -weighted images (128 sagittal slices, field of view =  $256 \times 256$  mm, data matrix =  $256 \times 256$ ) were collected to coregister the functional scans before two functional runs.

Participants were placed comfortably in a supine position in the scanner. They received binaurally presented auditory stimuli via specially constructed, MR compatible headphones (Commander XG, Resonance Technology, Northridge). To attenuate the scanner noise without reducing the quality of the sound, we combined soundproof headphones and special ear plugs. All the sound stimuli were adjusted so that participants could perceive the stimuli with no problem.

## Data Analysis

### fMRI data analysis

The fMRI data analysis was performed with SPM5 (Wellcome Department of Cognitive Neuroscience, London, UK) using the general linear model in an event-related design manner. To correct the potential slice timing errors caused by the interleaved acquisition sequence, each slice of the functional data were first synchronized in time to the middle slice within the same volume. Images were then spatially realigned, normalized, smoothed (with a 6 mm Gaussian kernel full-width half maximum) and filtered (high pass filter of 128 s).

Two sets of contrast images were created for each participant, representative of two main effects: (1) congruous versus incongruous phrases (CONGRUITY) and (2) language versus music phrases (DOMAIN). These contrast images were then entered into a second-level random effect analysis for each contrast. The generated statistical parametric maps (SPMs) depicting main effects (CONGRUITY and DOMAIN) were thresholded at  $P < 0.05$  adjusted for multiple comparisons using the false discovery rate (FDR), with a minimal cluster-size of 10 contiguous voxels.

### Conjunction analysis

We conducted a conjunction analysis with SPM5, based on four separate sets of contrast images each representing an experimental condition in comparison to rest. The resulted SPMs were thresholded at  $P < 0.05$  adjusted for multiple comparisons using the FDR, with a minimal cluster-size of 10 contiguous voxels.

### Region of interest analysis

We selected a set of ROIs in regions which were functionally defined from the two main effects based on the group data, as well as those from the conjunction analysis. All significant voxels within a 7 mm radius of any local focus among the activated voxels defined an ROI. For each of these ROIs, we then extracted the normalized mean  $\beta$  weights from each of the four conditions in each individual.

### Group statistics

Both the behavioral data and the ROI-based  $\beta$  values for each individual across different conditions were subject to two-way repeated measure ANOVAs with the factors DOMAIN (language, music) and CONGRUITY (congruous, incongruous). Respective  $P$  values were adjusted with Bonferroni correction.

## RESULTS

### Behavioral Results

All 18 musicians performed well above chance, with mean correct response rates more than 90%: congruous language phrases,  $94.4\% \pm 3.4\%$  (mean correct response rate  $\pm$  SD); incongruous language phrases  $91.6\% \pm 5.0\%$ ; congruous music phrases  $93.3\% \pm 4.2\%$ ; incongruous music phrases  $96.4\% \pm 2.3\%$ . Both hit rates and false alarms were taken into account to generate  $d'$ -scores: hits were the correct answers to congruous phrases; false alarms were the wrong answers to the incongruous ones. The correct response rates and the  $d'$ -scores were further analyzed with CONGRUITY (congruous vs. incongruous)  $\times$  DOMAIN

(music vs. language) ANOVAs. No main effect or interaction was found.

### Imaging Results

All 18 participants showed extensive bilateral temporal and frontal activations in response to music and language phrases and were all included in the final imaging analysis. Different analyses were conducted. Whole-brain analysis obtained a main effect of DOMAIN, but no main effect of pitch variations (CONGRUITY effect). Conjunction analysis identified common areas of activation to both pitch variations across music and language. Supplementary ROI analysis performed on the peak sites of the DOMAIN effect and conjunction results revealed brain areas showing interaction between DOMAIN and CONGRUITY factors, as well as CONGRUITY effect. The results from these analyses are discussed in more detail later.

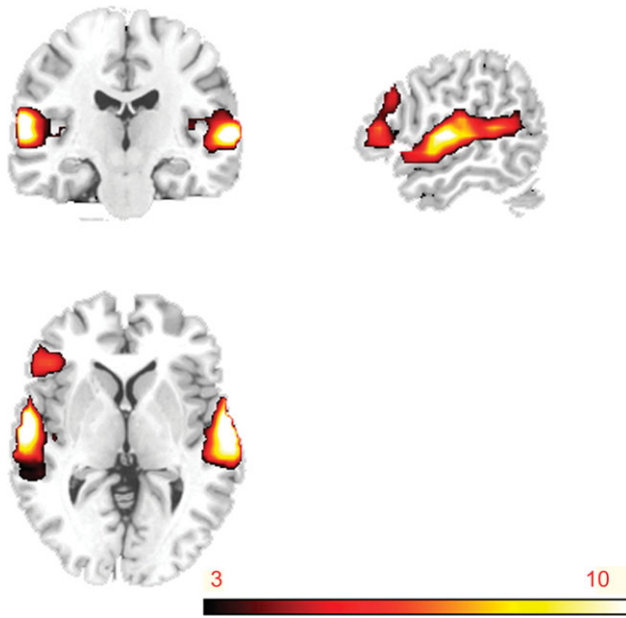
### Whole Brain Analysis

The whole brain analysis revealed a main effect of DOMAIN in several regions. The bilateral temporal cortices, although extensively activated for both music and language phrases, exhibited both a stronger left- than right-hemispheric peak activation and a larger right- than left-hemispheric cluster size when comparing language with music (Fig. 2, Table I). This stronger network for language as compared to music included the anterior portion of the left STG extending inferiorly to the left middle temporal gyrus, and the anterior portion of the right STG. It also included the triangular part of Broca's area (BA 45). These results are in line with the notion that relative to music, auditory language processing recruits bilateral temporal cortices and Broca's area. No significant activity was found for the inverse contrast in which we compared music to language.

### Results of the Conjunction Analysis

Bilateral superior temporal gyri, including bilateral Heschl's Gyri, showed overlapping responses to both pitch variations across music and language (blue and green regions in Fig. 3). This suggests that bilateral STG was a prerequisite for processing auditory language and music. The shared bilateral STG activation in language and music perception is likely due to those acoustic features that are similar across domains such as pitch and rhythm. Overlapping activity was not observed in other brain regions outside of primary auditory cortex. The analysis, moreover, revealed that parts of these shared brain areas also showed a DOMAIN effect, i.e., activating more for language than music (green regions in Fig. 3). This is probably due to the additional processing load (such as phonological analysis and the more various pitch trajectories) in language compared to music phrases. In general,





**Figure 2.**

Brain network shows increased activity for language phrases compared to music phrases ( $X = 144, Y = 109, Z = 72$ ). Brain areas demonstrating stronger activity for language phrases than music phrases include bilateral temporal cortices and the left inferior frontal region. The color bar indicates  $t$  value. All the activated voxels reached voxelwise threshold of  $P < 0.05$  (FDR), with minimal cluster size exceeding 10 voxels.

language activations were higher than music activations across the STG bilaterally and in the left IFG (red regions in Fig. 3).

This may be due to different underlying processing strategies. Although the task was to judge the pitch congruity in both music and language phrases, it is possible that participants might have employed different strategies completing the task in the respective domains. Incongruous pitches led to syntactic errors in music and either semantic or combined semantic/syntactic errors in language. In Chinese, syntactic processes always rely on semantic comprehension [Zhang et al., 2010] and cannot always be strictly separated from semantic processing. It thus is reasonable to assume that semantic processes were involved in judging the congruity of language phrases but not in music. This difference may have contributed to the observed overlapping as well as different brain networks between music and language.

### Results of ROI Analysis

All the ROIs were functionally defined based on either their responses to DOMAIN factor or their significantly overlapping activation to both music and language phrases. Follow-up exploratory analyses with these ROIs

were conducted to examine their responses to CONGRUITY factor. As such, only those with significant CONGRUITY effect (main effect or interaction involving CONGRUITY) were reported. The other ROIs with no effect of CONGRUITY but only DOMAIN effect (such as those based on DOMAIN effect) replicated their original functional definition. These ROIs were thus not repeatedly reported.

For the right STG (MNI coordinates: 63,  $-6 -5$ ; BA41), a significant interaction between DOMAIN (language vs. music) and CONGRUITY (congruous vs. incongruous) was detected,  $F_{(1,17)} = 6.715, P = 0.019$  (Fig. 4). Post hoc comparisons suggested a significant CONGRUITY effect for music phrases ( $P = 0.003$ ), with statistically increased activation for congruous music phrases relative to incongruous ones. Pitch variations in language phrases did not produce the similar pattern ( $P = 0.233$ ). No significant interaction or main effect of CONGRUITY was found for ROIs within the left STG.

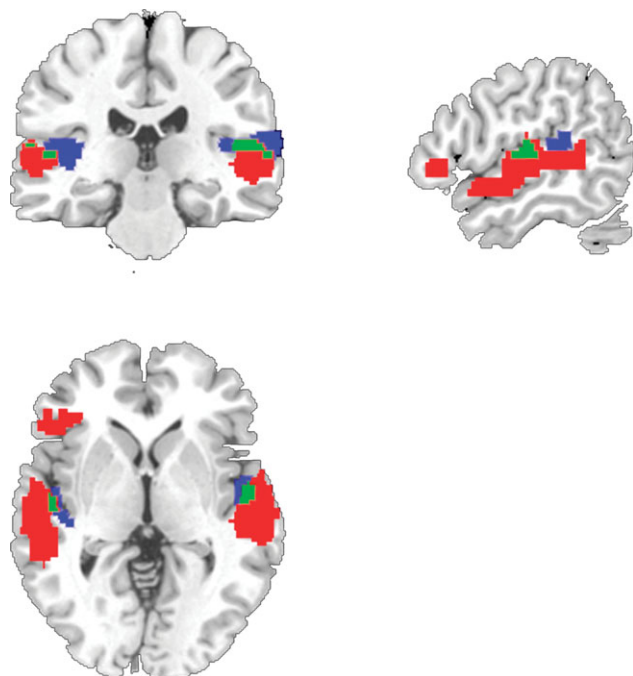
Within the triangular part of Broca’s area (MNI coordinates:  $-54, 27, 0$ ; BA 45), a significant CONGRUITY effect was observed,  $F_{(1,17)} = 5.046, P = 0.038$ . According to probabilistic cytoarchitectonic maps [Eickhoff et al., 2005] these coordinates assign the activation to BA 45 with a probability of 60% and to BA 44 with 10%. In this brain region phrases with congruous endings were associated with significantly higher activation than phrases with incongruous ending pitches (Fig. 5). Moreover, mean performance measured by  $d'$  (averaged across language and music conditions) was significantly correlated with the mean activation in the BA 45,  $r(18) = 0.605, P = 0.008$ . Namely, higher behavioral performance was associated with higher signal change in the BA 45 for pitch processing (Fig. 5). No such correlation was found for the right STG,  $r(18) = 0.378, P = 0.122$ .

Both the right STG and left BA 45 showed DOMAIN effects in the ROI analysis, with greater responses to language phrases relative to music phrases. These aspects of

**TABLE I. Regions of significant activations where language yielded stronger responses than music phrases**

Voxel	X	Y	Z	L/R	Location	BA	$Z_{\max}$
570	63	$-6$	$-5$	R	Superior temporal G	41	5.61
	57	$-30$	5	R	Superior temporal G	21	5.31
	63	$-24$	0	R	Superior temporal G	22	5.35
766	$-54$	$-12$	$-5$	L	Superior temporal G	42	5.19
	$-60$	$-18$	5	L	Superior temporal G	22	5.40
	$-54$	27	0	L	Inferior frontal G	45	5.03

The coordinates shown in the table are MNI coordinates for peak activations within each cluster. All the activated voxels reached voxelwise threshold of  $P < 0.05$  (FDR), with minimal cluster size exceeding 10 voxels. The anatomical locations and Brodmann areas were estimated based on the MNI coordinates using SPM Anatomy Toolbox [Eickhoff et al., 2005].



**Figure 3.**

The relationship between the conjunction map and DOMAIN effect map ( $X = 37, Y = 105, Z = 72$ ). The conjunction map (blue) shows overlapped responses to both pitch variations across music and language. These common brain areas are situated on bilateral superior temporal gyri, including bilateral Heschl's Gyri,  $P < 0.05$  (FDR). The conjunction map (blue) and DOMAIN effect map (red) overlap within bilateral STG, as shown in green. These areas were not only commonly activated by music and language phrases for both types of pitch variations, but also showing more activation for language than music phrases.

the data confirmed the original functional definition of these ROIs and were already reported earlier.

In summary, our results showed that pitch processing in music and Mandarin was associated with the right STG and the pars triangularis (BA 45). Congruous ending pitches in both music and tone language phrases elicited greater activation in left BA 45 relative to the incongruous ending pitches. The strength of the activity in left BA 45, but not in the right STG, is positively related to behavioral performance in pitch congruity judgment across domains. The right STG although showing activation for both language and music phrases, demonstrates a sensitivity to pitch variations only in music phrases, but not in language.

## DISCUSSION

The current study directly compared pitch processing in music and tone language (Mandarin) in Mandarin speak-

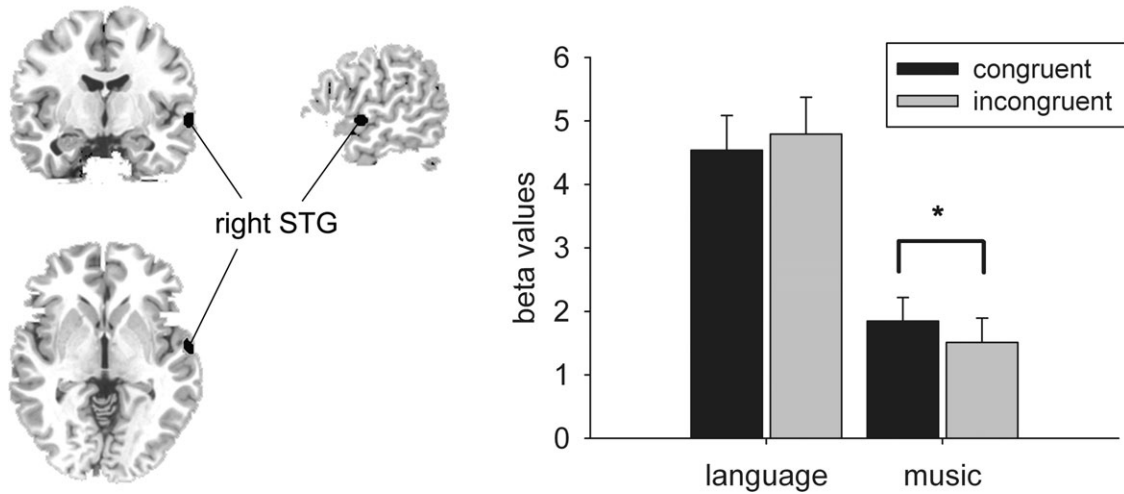
ing musicians with fMRI. This is the first time where the neural substrates underlying pitch processing are examined with a group of subjects whose expertise in pitch is similar across music and tone language. Our results demonstrate a common pitch processing network across music and tone language, including the triangular part of Broca's area and the right STG with the latter region differentially modulated by the different domains.

### Broca's Area

Congruous ending pitches in both music and tone language phrases elicited greater activation in the pars triangularis (BA 45) as part of Broca's area relative to the incongruous ending pitches. The strength of the activation in Broca's area is positively related to behavioral performance in pitch congruity judgment, suggesting that the integrity of Broca's area is mandatory for pitch processing performance. Broca's area in the left hemisphere has long been known to be involved in language processing with the more anterior portion mainly holding responsible for semantic aspects and the more posterior portion for syntactic aspects [for reviews see Bookheimer, 2002; Friederici, 2002] and for the processing of musical syntax [Koelsch et al., 2002; Maess et al., 2001].

It should be noted that in the current study, congruous pitches were associated with higher activation in left BA 45 as compared to incongruous ones in both music and language, whereas a prior study showed more activation for incongruous harmonic continuation in music [Koelsch et al., 2002]. This discrepancy might be attributed either to the fact that participants in the two studies came with different musical expertise, with high expertise in the current study as musicians, but with less expertise as nonmusicians in the Koelsch et al. (2002) study, or to the fact that different experimental tasks were used. The first difference may only be of minor importance as qualitative similar electrophysiological patterns were observed for the processing of musical violations in musicians and nonmusicians [Koelsch et al., 1999, 2000]. Moreover, differences in brain activation patterns between the musicians and nonmusicians were reported mainly for structures involved in control, programming and planning of actions [Schulze et al., 2011]. Thus, the task-related aspect may be more important. In the current study, the subjects were asked to respond explicitly to the congruity and incongruity of language and music phrases. Congruous and incongruous conditions were both task-relevant. In the fMRI study by Koelsch et al. (2002) the participants' task was to detect deviant instruments and dissonant tone-clusters and to indicate deviant instruments by button press. Thus, only one of the incongruous conditions, the cluster, was task-relevant, whereas the other incongruous conditions (the modulation) and the congruous condition (in-key chords) were both task-irrelevant. Within a single study, task-relevant and task-irrelevant conditions are associated with

## Right STG



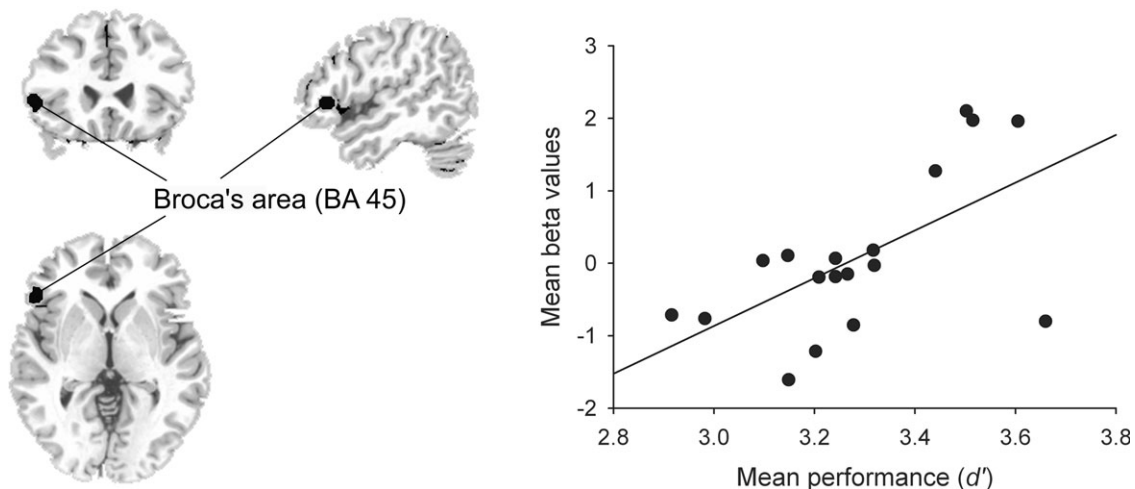
**Figure 4.**

Music CONGRUITY effect in the right STG (MNI coordinates: 63, -6, -5). Average  $\beta$  values for each condition in the ROI located in the right STG (shown on the left). Error bars indicate SDs (\* $P < 0.05$ ).

differential effects of attention: task-relevant items are more attended to than task-irrelevant ones, and yield more brain activation [Downar et al., 2001]. These task-induced attentional effects could be of importance. Although the final results by Koelsch et al. (2002) showed generally

greater BOLD responses for the incongruous conditions as compared to the congruous condition, but for the incongruous conditions, the one that was relevant to the task (the cluster) yielded bigger response than the one that was task-irrelevant (the modulation), suggesting that the

## Broca's area (BA 45)



**Figure 5.**

Neural activity in the triangular part of Broca's area (BA 45, MNI coordinates: -54, 27, 0) correlated with behavioral performance. Scatter plot between the mean behavioral performance (measured by  $d'$ ) and the mean activity in Broca's area (shown on the left) across four experimental conditions. There was a positive and reliable correlation between the mean performance and the mean  $\beta$  values in Broca's area across participants ( $r = 0.605$ ,  $P < 0.01$ ).

imaging results could be greatly influenced by the degree of task-relevance.

The engagement of Broca's area in the common network for music and language [Koelsch et al., 2002] might reflect a role for this region in task-relevant processes involved in both music and tone language. Broca's area has long been associated with task-related lexical processes [Thompson-Schill et al., 1997] and lexical pitch processing in tonal language speakers [Gandour et al., 1998, 2004; Wong et al., 2004]. Damage to Broca's area caused by left-sided brain lesion resulted in greatly compromised identification of lexical tone and thereby the word's semantic meaning [Liang and van Heuven, 2004]. Similarly, brain anomalies in bilateral inferior frontal areas have been associated with congenital amusia – the neurogenetic condition of music pitch deficit [Hyde et al., 2007, 2010; Mandell et al., 2007]. Moreover, congenital amusia has been shown to affect sentence-level prosodic information processing [Liu et al., 2010]. The involvement of Broca's area in a common pitch network is also consistent with the observed frontally distributed late positive component for pitch violations across music and language domains in our previous ERP study [Nan et al., 2009]. It is thus consistent with our data to suggest that part of Broca's area contributes to the general cognitive processes such as verbal and tonal working memory [Rogalsky et al., 2011; Schulze et al., 2011] as well as cognitive control [Norvick et al., 2010] that are essential for the judgment of pitch information in musical and language phrases.

### Right Superior Temporal Cortex

The right STG which was also identified as part of the common pitch processing network, on the other hand, might be responsible for the early perceptual acoustic analysis of pitch. For music, congruous ending pitches elicited higher activation in the right STG relative to the incongruous ones. This corroborates previous findings about the right STG in perceptual analysis of melodies and retention of pitch [Zatorre et al., 1994]. The malfunction of the right STG might indicate amusia, via a neural network that includes the right inferior frontal area [Hyde et al., 2007, 2010]. Lexically relevant pitch variations in tone language did not produce similar pattern as was observed in musical phrases: for pitch variations resulting in semantic or combined semantic/syntactic incongruity no activation difference was found in the right STG. This is in line with previous fMRI results indicating a left-localization of lexical pitch only when language processing beyond the auditory analysis is required [Gandour et al., 2004]. It is also consistent with a previous ERP work [Luo et al., 2006] which allows a more fine grained temporal resolution of the neural activation underlying the different aspects of pitch processing. Using ERP it was found that early auditory processing of lexical tones as reflected by the auditory mismatch negativity is lateralized to the right STG [Luo et al.,

2006]. Thus, in the current study the involvement of the right STG in lexical tone processing independent of semantic congruity (and sometimes combined semantic/syntactic congruity) might be related to the analysis of prelexical acoustic features [Luo et al., 2006]. It is only until a later stage of processing, most possibly under the involvement of Broca's area, that pitch information in language phrases is finally mapped onto its lexical representations. The specific difference of the right STG's response to congruity between tone language and music indicates that after initial acoustic analysis of an auditory stimulus' form, the function of the stimulus determines brain activation. Thus the right STG responded to musical congruence, but not to lexical congruence. Furthermore, the lack of correlation between behavioral performance and the right STG activation, in contrast to the tight link between pitch performance and activation in Broca's area, provides additional evidence that the right STG supports perceptual processes which are not directly related to judgment performance.

### Temporo-Frontal Network

Taken together, our results support a model whereby pitch information in music and tone language undergoes multiple stages of processing [Hickok and Poeppel, 2007; Obleser et al., 2007], each of which relates to a distinct neural structure. Most importantly, within each of these different processing stages, we found shared neural resources engaged in pitch processing for music and tone language. Specifically, the right STG might reflect perceptual stage essential for acoustic analysis of pitch information, whereas Broca's area most likely represents a performance-related processing stage.

This common neural network could help us understand the positive transfer effects of pitch processing across domains [Magne et al., 2006; Moreno et al., 2009; Schon et al., 2004; Wong et al., 2007]. Life long experience with pitch in one domain might strengthen the related neural structures and provide benefits to the other domain, via this common neural network.

Similarly, the disruption of this common circuitry might cause deficits in pitch processing which are not limited to one domain [Nan et al., 2010]. The current results thus suggest that our previously observed common pitch deficits in music and tone language might actually have its common neural underpinnings. We assume that the core deficit common to music and tone language pitch processing observed in Mandarin speaking amusics may be caused by the malfunction of the right STG, which is crucial for pitch analysis. Broca's area, however, might not be causally involved, since lexical tone production was reserved in these amusics [Nan et al., 2010]. This assumption is also supported by the present data showing that activation in Broca's area is performance related, rather than perception related. Further evidence comes from recent findings of associated speech intonation deficits in British amusics [Liu et al., 2010],



whose native language does not involve lexical tone as implicated in the left inferior frontal area.

An important caveat in music and language research is observed gender effects. Previous results suggested that females had less left lateralized hemispheric involvement in pitch-related tasks compared to males [Gaab et al., 2003]. Because of the precaution of gender confound, we chose only female musicians as our participants. Given this, however, it is not clear whether the observed interhemispheric network for pitch processing in the current study could be generalized across genders. A recent study with children observing a similar network involving the right STG and left IFG for language rhyming task showed no gender effect [Bitan et al., 2010]. Future work will have to test how gender might modulate the pitch processing network. Moreover, at this point, there is not enough evidence to support the generality of the current findings to other groups of people, such as nonmusicians or other tone language speakers. Recent language-music comparative studies have found that, compared to musicians, nonmusicians showed a greater degree of neural overlap as well as behavioral similarity in tonal and verbal working memory tasks [Schulze et al., 2011; Williamson et al., 2010]. Future research may focus on how musical and different tone language experiences affect the current observed common pitch network.

## CONCLUSIONS

Cross-domain comparisons provide unique insights into the specific functional role of particular brain areas across processing domains. Here, we used fMRI to test whether there is a common neural implementation of pitch processing in music and tone language (Mandarin). Our findings obtained from Mandarin speaking musicians, i.e., from individuals with a similarly high expertise in pitch processing in both tone language and music, demonstrate shared neural circuits for pitch processing across domains, including part of the right STG and the pars triangularis of Broca's area. Within this network, Broca's area's pitch sensitivity is tightly linked to the behavioral performance of pitch congruity judgment thereby reflecting controlled aspects of pitch processing, whereas the right STG is independent of performance most likely reflecting purely perceptual processes. These observations provide a first glimpse at the cortical pitch processing network shared across domains.

## ACKNOWLEDGMENTS

The authors thank Litao Zhu and Hui Wu for technical assistance and Yanan Sun for help with data collection. They thank Zonglei Zhen for help in an early version of analysis. They also thank the two anonymous reviewers for their insightful comments.

## REFERENCES

- Bidelman GM, Gandour JT, Krishnan A (2011): Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J Cogn Neurosci* 23:425–434.
- Bitan T, Lifshitz A, Breznitz Z, Booth JR (2010): Bidirectional connectivity between hemispheres occurs at multiple levels in language processing but depends on sex. *J Neurosci* 30:11576–11585.
- Boersma P (2001): Praat, a system for doing phonetics by computer. *Glott Int* 5:341–345.
- Bookheimer S (2002): Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci* 25:151–188.
- Downar J, Crawley AP, Mikulis DJ, Davis KD (2001): The effect of task relevance on the cortical response to changes in visual and auditory stimuli: An event-related fMRI study. *NeuroImage* 14:1256–1267.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005): A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25:1325–1335.
- Friederici AD (2002): Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 6:78–84.
- Friederici AD, Alter K (2004): Lateralization of auditory language functions: A dynamic dual pathway model. *Brain Lang* 89:267–276.
- Friederici AD, Kotz SA (2003): The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage* 20:S8–S17.
- Gaab N, Keenan JP, Schlaug G (2003): Effects of gender on the neural substrates of pitch memory. *J Cogn Neurosci* 15:810–820.
- 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, Wong D, Hsieh L, Weinzapfel B, Van LD, Hutchins GD (2000): A crosslinguistic PET study of tone perception. *J Cogn Neurosci* 12:207–222.
- Gandour J, Wong D, Hutchins G (1998): Pitch processing in the human brain is influenced by language experience. *Neuroreport* 9:2115–2119.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hyde KL, Lerch JP, Zatorre RJ, Griffiths TD, Evans AC, Peretz I (2007): Cortical thickness in congenital amusia: When less is better than more. *J Neurosci* 27:13028–13032.
- Hyde KL, Peretz I, Zatorre RJ (2008): Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46:632–639.
- Hyde KL, Zatorre RJ, Griffiths TD, Lerch JP, Peretz I (2006): Morphometry of the amusic brain: A two-site study. *Brain* 129:2562–2570.
- Hyde KL, Zatorre RJ, Peretz I (2011): Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb Cortex* 21:292–299.
- Klein D, Zatorre RJ, Milner B, Zhao V (2001): A cross-linguistic PET study of tone perception in mandarin Chinese and English speakers. *NeuroImage* 13:646–653.
- Koelsch S, Gunter TC, von Cramon DY, Zysset S, Lohmann G, Friederici AD (2002): Bach speaks: A cortical 'language-network' serves the processing of music. *NeuroImage* 17:956–966.

- Koelsch S, Gunter T, Friederici AD, Schröger E (2000): Brain indices of music processing: "Nonmusicians" are musical. *J Cogn Neurosci* 12:520–541.
- Koelsch S, Schröger E, Tervaniemi M (1999): Superior pre-attentive auditory processing in musicians. *NeuroReport* 10:1309–1313.
- Kraus N, Banai K (2007): Auditory-processing malleability - Focus on language and music. *Curr Direct Psychol Sci* 16:105–110.
- Liang J, van Heuven VJ (2004): Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain Lang* 91:282–293.
- Limb CJ (2006): Structural and functional neural correlates of music perception. *Anat Rec A Discov Mol Cell Evol Biol* 288:435–446.
- Liu F, Patel AD, Fourcin A, Stewart L (2010): Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain* 133:1682–1693.
- Liu L, Peng D, Ding G, Jin Z, Zhang L, Li K, Chen C (2006): Dissociation in the neural basis underlying Chinese tone and vowel production. *Neuroimage* 29:515–523.
- Loui P, Schlaug G (2009): Investigating musical disorders with diffusion tensor imaging: A comparison of imaging parameters. *Ann NY Acad Sci* 1169:121–125.
- Luo H, Ni JT, Li ZH, Li XO, Zhang DR, Zeng FG, Chen L (2006): Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc Natl Acad Sci USA* 103:19558–19563.
- Maess B, Koelsch S, Gunter TC, Friederici AD (2001): Musical syntax is processed in Broca's area: An MEG study. *Nat Neurosci* 4:540–545.
- Magne C, Schon D, Besson M (2006): Musician children detect pitch violations in both music and language better than non-musician children: Behavioral and electrophysiological approaches. *J Cogn Neurosci* 18:199–211.
- Mandell J, Schulze K, Schlaug G (2007): Congenital amusia: An auditory-motor feedback disorder? *Restor Neurol Neurosci* 25:323–334.
- 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.
- Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M (2009): Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cereb Cortex* 19:712–723.
- Nan Y, Friederici AD, Shu H, Luo YJ (2009): Dissociable pitch processing mechanisms in lexical and melodic contexts revealed by ERPs. *Brain Res* 1263:104–113.
- Nan Y, Sun Y, Peretz I (2010): Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain* 133:2635–2642.
- Novick JM, Trueswell JC, Thomsson-Schill SL (2010): Broca's area and language processing: Evidence for the cognitive control connection. *Lang Linguist Compass* 4:906–924.
- Obleser J, Zimmermann J, Van MJ, Rauschecker JP (2007): Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb Cortex* 17:2251–2257.
- Peretz I (1993): Auditory atonalia for melodies. *Cogn Neuropsychol* 10:21–56.
- Rogalsky C, Rong F, Saberi K, Hickok G (2011): Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *J Neurosci* 31:3843–3852.
- Sammler D, Koelsch S, Friederici AD (2011): Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex* 47:659–673.
- Schon D, Magne C, Besson M (2004): The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41:341–349.
- Schulze K, Zysset S, Mueller K, Friederici AD, Koelsch S (2011): Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum Brain Mapp* 32:771–783.
- Semal C, Demany L, Ueda K, Halle PA (1996): Speech versus non-speech in pitch memory. *J Acoust Soc Am* 100:1132–1140.
- Steinhauer K, Alter K, Friederici AD (1999): Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat Neurosci* 2:191–196.
- Tillmann B, Burnham D, Nguyen S, Grimault N, Gosselin N, Peretz I (2011): Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Front Psychol* 2:120.
- Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997): Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc Natl Acad Sci USA* 94:14792–14797.
- Williamson VJ, Baddeley AD, Hitch GJ (2010): Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Mem Cognit* 38:163–175.
- 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.
- Wong PC, Perrachione TK, Gunasekera G, Chandrasekaran B (2009): Communication disorders in speakers of tone languages: Etiological bases and clinical considerations. *Semin Speech Lang* 30:162–173.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N (2007): Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10:420–422.
- Zatorre RJ, Evans AC, Meyer E (1994): Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992): Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849.
- Zhang Y, Yu J, Boland JE (2010): Semantics does not need a processing license from syntax in reading Chinese. *J Exp Psychol Learn Mem Cogn* 36:765–781.