

Neural Systems for Auditory Perception of Lexical Tones

Veronica P.Y. Kwok¹, Guo Dan^{2,3}, Kofi Yakpo¹, Stephen Matthews¹, & Li Hai Tan^{2,3}

¹State Key Laboratory of Brain and Cognitive Sciences and Department of Linguistics
University of Hong Kong, Pokfulam Road, Hong Kong

²School of Biomedical Engineering, Shenzhen University Health Science Center, Shenzhen
518060, China

³Guangdong Key Laboratory of Biomedical Information Detection and Ultrasound Imaging,
Shenzhen 518060, China

Please send correspondence to:

Guo Dan or Li Hai Tan

School of Biomedical Engineering

Shenzhen University Health Science Center

Shenzhen 518060, China.

Email: danguo@szu.edu.cn or tanlh@szu.edu.cn .

Abstract

Previous neuroimaging research on cognitive processing of speech tone has generated dramatically different patterns of findings. Even at the basic perception level, brain mapping studies of lexical tones have yielded inconsistent results. Apart from the data inconsistency problem, experimental materials in past studies of tone perception carried little or minimal lexical semantics, an important dimension that should not be dispensed with because speech tones serve to distinguish lexical meanings. The present study sought to examine the neural correlates of the perception of speech tone using lexically meaningful experimental stimuli. A simple lexical tone perception task was devised in which native Mandarin speakers were asked to judge whether or not the two syllables of an auditorily presented Chinese bisyllabic word had the same tone. We selected bisyllabic words as experimental stimuli because Chinese monosyllables often convey little or very vague meanings due to rampant homophony. We found that the left inferior frontal gyrus, the right middle temporal gyrus and bilateral superior temporal gyri are responsible for basic perception of linguistic pitches. Our interpretation of the data sees the left superior temporal gyrus as engaged in primary acoustic analysis of the auditory stimuli, while the right middle superior temporal gyrus and the left inferior frontal region are involved in both tonal and semantic processing of the language stimuli.

Key words: fMRI, language, pitch, temporal cortex, tone

In a tonal language, lexical tone is used to distinguish lexical or grammatical meanings. Since tonal languages account for around 50% of the world's languages (Maddieson, 2014), understanding the neural substrates underlying speech tone perception in native speakers has been a central question in electrophysiological and neuroimaging research on spoken language (Luo et al., 2006; Nan, et al., 2009; Ren et al., 2009, 2013; Tsang et al., 2011; Xi et al., 2010; Yang et al., 2015; Zhang et al., 2010, 2011). While existing research has yielded important findings on how lexical tone is neuroanatomically represented, the experimental tasks employed in brain mapping studies tap dramatically different cognitive processes, including perception and working memory (Gandour et al., 2002, 2003; Hsieh et al., 2001; Klein et al., 2001; Li et al., 2003, 2010; Nan and Friederici, 2013; Wang et al., 2003; Xu et al., 2006; Zhang et al., 2010, 2011), and the neural correlates of perception of speech tone in native listeners have not been precisely pinpointed.

Klein et al. (2001) used PET to identify the brain systems subserving auditory lexical tone perception in Mandarin native speakers. In their tone discrimination task, pairs of monosyllabic Chinese words were presented auditorily. Within each pair, the syllables were identical; half of them carried the same tone (e.g. /tou2/ and /tou2/) and the other half had different tones (e.g. /fei2/ and /fei4/). Subjects were required to make same-different lexical tone judgments of the syllable pairs. Peak activation was found in bilateral superior temporal gyri (STG), bilateral parietal areas and bilateral cerebellum in native Mandarin speakers.

Similarly, Gandour et al. (2002) used fMRI to examine how the brain processes linguistically relevant pitch patterns (spectral vs. temporal cues) in tonal language speakers. The tone discrimination task they used is the same as in Klein et al.'s study, where subjects were required to make same-different judgments on the syllable pairs they heard. The authors of this study used the same stimuli for the spectral (i.e. lexical tone) and temporal (i.e. vowel length) conditions, and the syllable pairs were monosyllabic Thai pseudowords which either (1) had the same tone with different vowel duration (e.g., p^hin^R and p^hiin^R), (2) had different tones with the same vowel duration (e.g., haaj^M and haaj^H), or (3) had same tone and same vowel duration. They found that native Thai speakers showed greater activation in left inferior prefrontal gyrus in discriminating Thai tone relative to nonspeech pitch. Wong et al. (2004) used PET to compare the neural correlates underlying lexical tone perception between Mandarin natives and native English speakers. Their task is a tone judgment on Chinese syllable pairs (e.g. /fei2/ and /wei2/). Relative to passive listening of Mandarin syllables, lexical tone judgment in native Mandarin speakers induced stronger activation in the left insular cortex, putamen, thalamus, fusiform gyrus, and medial frontal gyrus. Activation in the right hemisphere was also observed in middle frontal gyrus and postcentral gyrus. In an fMRI study, Xu et al. (2006) found that when native Mandarin speakers performed tone discrimination tasks on two different Chinese syllables (e.g., /bai2/ and /yao2/) or stimuli resynthesized by superimposing Thai tones onto Chinese syllables, stronger activity in the left

planum temporale (PT) was seen in response to native compared with non-native tones.

The tasks used by Klein et al., Gandour et al., Wong et al., and Xu et al. are highly similar and are all perception-based, but their findings are inconsistent. This inconsistency may be related to the different languages used in the studies. Indeed, in Gandour et al.'s study (2002), native Chinese exhibited stronger activation in the left anterior superior temporal cortex in identifying Chinese speech tones relative to the nonspeech baseline.

Gandour and associates conducted several functional imaging studies to elucidate the neural mechanisms dedicated to Mandarin lexical tone processing. In their experiments, participants were presented a list of three to five monosyllables consecutively (Gandour, et al., 2003; Hsieh, et al., 2001; Li, et al., 2003, 2010); they were asked to make tone judgments of the two syllables located first and last in the sequence on each trial (Gandour et al., 2003; Hsieh et al., 2001), to match the last item in the sequence with the probe (Li et al., 2003), or to match the probe with the target syllable within the sequence in random positions (Li et al., 2010). Tonal processing yielded greater brain activations in bilateral frontal-parietal networks, including the inferior prefrontal cortex, posterior inferior frontal gyrus (IFG), middle frontal gyrus (MFG) and the inferior parietal lobule. Nan and Friederici (2013) compared pitch processing of tonal language and music in native Mandarin speakers. Experimental stimuli involved sequences of Chinese four-syllable phrases and four-note musical phrases. Half of the Chinese phrases were semantically meaningful, and the other half were similar except that

the pitch contours of the last syllables of the phrases were manipulated, making these phrases semantically and syntactically incongruent. Subjects were asked to judge whether the four-syllable phrases presented sounded congruous. Nan and Friederici found that processing of Mandarin pitch congruities involved greater cortical activations in bilateral STG and left IFG. In the tasks of Gandour et al. and Nan and Friederici, working memory is required to hold linguistic items in mind during the task, and therefore, lexical tone processing may occur at a late short-term memory stage, instead of at the perceptual phase.

In summary, previous research on speech tone processing has generated markedly different patterns of findings, partly because the tasks employed in those studies measured different levels of cognitive processing. As a matter of fact, even in neuroimaging experiments on basic perception of lexical tones in native Mandarin speakers, the findings are highly inconsistent too. Thus, it is worthwhile to further address the neuroanatomical representation issue of speech tone.

Apart from the data inconsistency problem, we have also noted that experimental stimuli in past studies of tone perception carried little or minimal lexical semantics, an important facet that should not be neglected because speech tones serve, after all, to distinguish lexical meanings. For example, when the syllables /fei²/ and /fei⁴/ are used in a tone decision task, the two syllables have little or very vague meanings because of the rampant homophony in Mandarin Chinese. In this case, the neural basis for tone processing may be related to general

tone analysis (e.g., pure tone) but is hardly associated with natural, meaning-related lexical tone. Perception of lexical tones includes both acoustic processing and semantic processing of the pitch signal. Yet, it seems that most of the past studies have emphasized on the processing of acoustic/ phonological information carried by lexical tones, and little focus is put on the neural substrates associated with the semantic processing aspect. Nan and Friederici (2013) used Chinese word phrases as experimental materials, in which word meaning of each lexical item was very precise. They observed that the left inferior language region (BA45) was activated in Mandarin speakers while performing the Mandarin tone task. Previous studies have also found that the inferior frontal gyrus (BA 44, 45, 47) was responsible for semantic processing of Chinese characters (Chee et al., 2000; Chou et al., 2009; Ding et al., 2003; Price et al., 2012; Tan et al., 2001). Since lexical tones are closely linked to the processing of lexical semantics, the left inferior frontal and surrounding language regions might also play a role in processing linguistic information of lexical tones.

In the present fMRI study, we used an auditory tone judgment task in which Mandarin natives decided whether or not the two syllables heard in a meaningful Chinese bisyllabic word on each trial had the same lexical tone. This design was simple in that subjects performed the task with minimal effort. Also, in this task, the lexical semantics of the bisyllabic word should be automatically activated (Neely, 1977; Tan and Perfetti, 1998). We used two-syllable words, instead of monosyllables, as stimuli, also because most words in

Mandarin are disyllabic (DeFrancis, 1984; Duanmu, 2000).

Materials and Methods

Subjects

Eleven Beijing college students with normal hearing participated in this functional magnetic resonance imaging (fMRI) study (6 males and 5 females; average age= 21.2 years, SD= 0.41 years). All subjects were healthy, native speakers of Mandarin. They were strongly right-handed as judged by the handedness inventory devised by Snyder and Harris (1993). We adopted nine unimanual tasks (which could only be done by one hand at a time) including writing, drawing a picture, throwing a ball, holding chopsticks, hammering a nail, brushing teeth, cutting with scissors, opening a door and striking a match. A 5-point Likert-type scale was used, with “1” representing exclusive left-hand use, and “5” representing exclusive right-hand use. All subjects scored not less than 43 in total. Written consent was obtained from all participants. The study was approved by the Institutional Review Board of Beijing MRI Center for Brain Research.

Stimuli and procedure

A list of 48 meaningful Chinese bisyllabic words was produced by a 25-year-old female native Mandarin speaker and digitally edited using Adobe Audition software 7.0 (Adobe Systems, Inc.) to achieve a constant duration (1000 ms). In 24 stimulus words, the tone of the

two syllables was identical (e.g. /hu1/ /xi1/, breath). For the rest, the tones of the two syllables were different (e.g. /fang1/ /fa3/, method). For the same-tone trials, Tone 3 words were excluded to avoid any tone confusion induced by the tone sandhi effect (i.e., when two or more consecutive syllables were third tones in a row, only the last syllable would be pronounced as third tone while the initial syllable(s) would be shifted to tone 2 obligatorily). All four tones were included in different-tone trials. All selected words were frequently used verbs, nouns or adjectives, with a frequency of occurrence not fewer than 15 per million according to the Modern Chinese Frequency Dictionary (Wang, 1986). All syllables in the disyllabic words were also commonly used, with frequency of occurrences of more than 20 per million.

Participants judged whether the bisyllabic words carried the same lexical tone. In each trial, a bisyllabic word was presented through MRI-compatible earphones for 1000 ms, followed by a 1500 ms blank interval. Participants responded by pressing the corresponding key with their left or right index finger. If no response was made during the response interval, the trial was considered as incorrect. The baseline condition was silence: subjects were instructed to relax and no overt response was required. Task instructions were presented for 2 s before each block. Prior to the scan, participants were given sufficient practice; stimulus words used in the practice section did not reoccur during the scan.

The experiment was conducted in a single run. It began with a 6-s fixation crosshair,

followed by a 2-s task instruction then a block of 12 lexical tone discrimination trials. The experiment consisted of four blocks of tone discrimination and the baseline. Different Chinese bisyllabic words were presented in each block to avoid any practice effect, and the presentation order was randomized across subjects. Sound stimuli were presented binaurally via a pair of MRI-compatible earphones. Sound levels were adjusted to comfortable levels for all subjects before the experiment began. All participants reported that they could hear the practice stimuli in the scanner before the experiment started.

Image Acquisition

fMRI data was acquired on a 3T Siemens scanner at Beijing 306 hospital using a T2*-weighted, gradient-echo echo planar imaging (EPI) sequence (TE = 30 ms, TR = 2000 ms, flip angle = 90°, field of view = 21 cm, slice thickness = 4 mm, and image matrix = 64 x 64). Thirty axial slices were collected.

Image Analysis

The fMRI data analysis was performed with MATLAB software (Version 7.10; Mathworks, Natick, MA) and SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>, Wellcome Department of Cognitive Neurology University College London, London). After data conversion, participants' data were preprocessed in batch model one by one. The first three

volumes of each participant's scan were discarded, and the remaining functional images were slice-timing corrected (in ascending order, with reference slice 15) and realigned. Functional images were then co-registered with the anatomical image and segmented. The data were then spatially normalized to the Montreal Neurological Institute (MNI) stereotaxic template, resampled into 3 x 3 x 3 mm cubic voxels and spatially smoothed with a full-width half maximum (FWHM) isotropic Gaussian kernel of 8 mm. Participants' data were high-pass filtered at 128 s to remove low-frequency components. First level analysis was conducted on an individual basis. Data from each subject was entered into a general linear model using an event-related analysis procedure. Group analysis was done by obtaining contrast images using a second-level random-effects model. Activation patterns were evaluated by the lexical tone discrimination of bisyllabic words > silent baseline contrast with a one-sample t-test. Brain regions were estimated from Talairach and Tournoux, after adjustments for differences between MNI and Talairach coordinates (Talairach and Tournoux, 1988).

Results

Behavioral performance

Participants were highly accurate in the experimental condition. Mean accuracy was 88.9% (SD = 0.08) for the auditory lexical tone judgment of bisyllabic words (91 % for the same-tone pairs and 88% for different-tone pairs). After excluding incorrect trials, the average

reaction time was 1278 ms (SD = 322).

Imaging results

Images of group maps (11 subjects) for bisyllabic words tone judgment compared to silent baseline are shown in [Figure 1](#). With the significance threshold set at $P < 0.05$ familywise error (FWE) correction at the cluster level ($P < 0.001$ uncorrected at voxel level), three clusters reached significance for multiple comparisons. Significant brain activations are listed in [Table 1](#). These clusters were located in right middle/ superior temporal gyrus (BA21; $x = 49, y = -13, z = -5$; BA22; $x = 57, y = -25, z = 5$; BA22; $x = 57, y = -7, z = 0$; $p = 0.007$, cluster size = 159 voxels) and left inferior frontal gyrus (BA44/45; $x = -27, y = 19, z = 7$; BA44/45; $x = -32, y = 16, z = 12$; $p = 0.025$, cluster size = 116 voxels). Activation in left superior temporal gyrus was marginally significant (BA22; $x = -51, y = -16, z = 5$; BA42; $x = -59, y = -25, z = 10$; $p = 0.057$, cluster size = 90 voxels). Insert Table 1 about here. Insert Figure 1 about here.

Discussion

This study aimed to examine the brain basis of auditory perception of lexical tones by using natural, meaning-related experimental stimuli with a simple, explicit tone discrimination task. Despite the simple baseline task used in this study, only four sites of activation are observed (i.e. left inferior frontal gyrus, right middle/superior temporal gyrus and left superior temporal gyrus). Both the left and the right hemispheric regions are engaged

in processing tones of Chinese bisyllabic words. Recently, researchers have proposed a more comprehensive hypothesis regarding the hemispheric lateralization of lexical tones. They suggested that perception of lexical tones involves the processing of two types of pitch information carried by the tonal signal, namely, the acoustic information and semantic information. Therefore, both hemispheres participated and interacted in processing lexical tones (Gandour et al., 2004; Gandour, 2006; Yu et al., 2014). Our results are consistent with this assumption that both hemispheres work together to achieve lexical tone perception. Also, this activation pattern is similar to that observed by Nan and Friederici (2013), in which meaningful word phrases were used as stimuli and lexical semantics of the syllables was processed along with the tone discrimination task.

The left inferior frontal region (BA44/45) identified in this study is very close to the anterior insula. Both the inferior frontal gyrus and insular cortex have been shown to be implicated in pitch processing (Flagmeier et al., 2014; Jacquemot et al., 2003; Riecker et al., 2000) and lexical-semantic processing (Chan et al., 2004; Mummery et al., 1999; Price, 2012; Rodríguez-Fornells, 2009; Rossell et al., 2001; Rumsey et al., 1997; Tan et al., 2001). Moreover, Nan and Friederici (2013) found significant activation in the left inferior frontal region in Mandarin speakers who performed tone discrimination of Chinese word phrases relative to music phrases. In the present study, meaning is involved in the tone judgment task. We assume that the inferior frontal gyrus was responsible for the semantic processing of both

the lexical tone and the bisyllabic word stimuli. Superior temporal gyrus activation occurred bilaterally, though activation in the right temporal regions was much stronger than in the left (cluster size of 159 voxels vs. 90 voxels). In the left hemisphere, marginally significant activation in the temporal cortex was focused on the superior temporal gyrus (BA 22) and extended medially to its neighboring region, the posterior transverse temporal area (BA42). Previous studies have indicated that processing of simple acoustic stimuli, such as frequency-modulated tones and sound with discontinuous acoustic patterns, activates BA 42 (Binder et al., 2000; Mirz et al., 1999). BA 42 has also been shown to be critical for auditory lexical tone processing in a recent meta-analysis (Kwok et al., 2015). The left STG has long been seen to be implicated in the basic processing of both speech and non-speech sounds in auditory research (Binder et al., 2000; Hickok and Poeppel, 2007; Price et al., 2005; Wang et al., 2003; Wong et al., 2004; Xu et al., 2006; Zatorre and Belin, 2001; Yang et al., 2015; Zhang et al., 2010, 2011). Our findings suggest that the left superior temporal region is involved in the initial processing of auditory stimuli that may not be speech-related, as reflected in its relatively weak activation.

The activation cluster seen in the right temporal lobe is composed of middle and superior temporal gyrus (BA21 and BA22). The right STG has been repeatedly shown to be critical to perceptual pitch perception, vocal pitch error detection, and voice control in previous literature (Flagmeier et al., 2014; Johnsrude et al., 2000; Robin et al., 1990; Zatorre et al.,

2002; Zatorre and Belin, 2001; Zhang et al., 2010, 2011). Moreover, several lexical tone studies have demonstrated right-lateralized cortical activity in processing tones. In a phonological recognition task, direct contrasts of Mandarin tones relative to consonants and rhymes showed stronger brain activation in right fronto-parietal network. Liu et al., (2006) found that the production of Mandarin tones elicited stronger activation in the right hemisphere than the production of vowels. Moreover, the observed activation in right MTG in this study was consistent with findings in a previous ERP study on the pre-attentive processing of Mandarin tones (Luo et al., 2006). Luo et al. found that Mandarin native-speakers showed right-lateralized activity in MTG in perceiving tones, while left hemispheric dominance was found in response to the processing of consonants. We believe that the right middle superior temporal gyrus in our study was the key region for the processing of lexical tone information. In general, our results lend support to the hypothesis that spectral processing of tones (i.e. with longer duration approximately at 150-250 ms) is lateralized to the right hemisphere (Poeppel, 2003; Zatorre et al., 1994; Zatorre and Belin, 2001). As mentioned earlier, stronger brain activation was observed in right temporal regions than that in the left. One possibility was that the right temporal areas were responsible for numerous roles in auditory lexical tone perception. The middle temporal gyrus has been implicated in lexical semantic processing according to the dual-stream model of speech processing (Hickok and Poeppel 2000, 2004, 2007) and in idiom comprehension studies

(Mashal et al., 2008; Zemleni et al., 2007; Zhao et al., 2013). Thus, apart from acoustic processing of lexical tones, we believe that the right middle superior temporal gyrus was also involved in lexical-semantic processing.

Thus, our study has contributed to the identification of the neural systems for the basic auditory perception of lexical tones. The left inferior frontal/insula region is seen as responsible for semantic processing of tonal stimuli. The left superior temporal gyrus is involved in the primary acoustic analysis of the auditory stimuli, whereas the right middle superior temporal gyrus participates in the more complex auditory analysis of lexical tone information, as well as lexical-semantic processing. Methodologically, our study introduces a new, meaning-related tone discrimination paradigm in examining the perception of lexical tones with least effort. Further investigation is needed to investigate the connectivity of cortical regions critical for speech tone perception in order to unveil the correlations between the regions and the influence that one cortical region exerts over another.

Acknowledgments

This study was supported by the National Strategic Basic Research Program (“973” Program) of the Ministry of Science and Technology of China (Grant No.: 2012CB720701).

References

Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, Possing J (2000): Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex* 10:512-528.

Chan AH, Liu HL, Yip V, Fox PT, Gao JH, Tan LH (2004): Neural systems for word meaning modulated by semantic ambiguity. *Neuroimage*, 22(3):1128-1133.

Chee MW, Weekes B, Lee KM, Soon CS, Schreiber A, Hoon JJ, Chee M (2000): Overlap and dissociation of semantic processing of Chinese characters, English words, and pictures: evidence from fMRI. *Neuroimage* 12(4):392-403.

Chou TL, Chen CW, Wu MY, Booth JR (2009): The role of inferior frontal gyrus and inferior parietal lobule in semantic processing of Chinese characters. *Experimental brain research* 198(4): 465-475.

DeFrancis J (1984): *The Chinese language: Fact and fantasy*. University of Hawaii Press.

Ding G, Perry C, Peng D, Ma L, Li D, Xu S, Luo Q, Xu D, Yang J (2003): Neural mechanisms underlying semantic and orthographic processing in Chinese-English bilinguals. *NeuroReport* 14(12): 1557-1562.

Duanmu S (2000): *The Phonology of Standard Chinese*. Oxford University Press.

Flagmeier SG, Ray KL, Parkinson AL, Li K, Vargas R, Price LR, Laird AR, Larson CR, Robin DA (2014): The neural changes in connectivity of the voice network during voice pitch perturbation. *Brain and Language* 132:7-13.

Gandour J, Wong D, Hutchins G (1998): Pitch processing in the human brain is influenced by language experience. *NeuroReport* 9:2115-2119.

Gandour J, Wong D, Hsieh L, Weinzapfel B, Van Lancker D, Hutchins GD (2000): A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience* 12:207-222.

Gandour J, Wong D, Talavage T, Dziedzic M, Sathamnuwong N, Tong Y (2002): A cross linguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience* 14:1076-1087.

Gandour J, Xu Y, Wong D, Dziedzic M, Lowe M, Li X, Tong Y (2003): Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping* 20:185-200.

Hickok G, Poeppel D (2000): Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* 4:131–138.

Hickok G, Poeppel D (2004): Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67–99.

Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nature Reviews Neuroscience* 8:393-402.

Hsieh L, Gandour J, Wong D, Hutchins GD (2001): Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language* 76:227-252.

Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003): Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience* 23:9541-9546.

Johnsrude IS, Penhune VB, Zatorre RJ (2000): Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain: A Journal of Neurology* 123(1):155-163.

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.

Kwok VPY, Wang T, Chen S, Yakpo K, Zhu L, Fox PT, & Tan LH (2015): Neural signatures of lexical tone reading. *Human brain mapping*, 36(1): 304-312.

Li X, Gandour J, Talavage T, Wong D, Dziedzic M, Lowe M, Tong Y (2003): Selective attention to lexical tones recruits left dorsal frontoparietal network. *Neuroreport* 14:2263-2266.

Li X, Gandour JT, Talavage T, Wong D, Hoffa A, Lowe M, Dziedzic M (2010): Hemispheric asymmetries in phonological processing of tones versus segmental units. *Neuroreport* 21:690-694.

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.

Luo H, Ni J, Li Z, Li X, Zhang D, Zeng G, Chen L (2006): Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences*, 103(51):19558-19563.

Maddieson I (2013): Tone. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/13>, Accessed on 2015-01-28.)

Mashal N, Faust M, Hendler T, & Jung-Beeman M (2008): Hemispheric differences in processing the literal interpretation of idioms: Converging evidence from behavioral and fMRI studies. *Cortex*, 44(7): 848-860.

Mirz F, Ovesen T, Ishizu K, Johannsen P, Madsen S, Gjedde A, Pedersen CB (1999): Stimulus-dependent central processing of auditory stimuli: a PET study. *Scandinavian Audiology* 28:161–169.

Mummary CJ, Shallice T, Price CJ (1999): Dual-process model in semantic priming: a functional imaging perspective. *Neuroimage* 9(5):516-525.

Nan Y, Friederici AD, Shu H, Luo YJ (2009): Dissociable pitch processing mechanisms in lexical and melodic contexts revealed by ERPs. *Brain Research* 1263:104-113.

Nan Y, Friederici AD (2013): Differential roles of right temporal cortex and Broca's area in pitch processing: evidence from music and Mandarin. *Human Brain Mapping* 34(9):2045-2054.

Neely JH (1977): Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General* 106:226-254.

Poeppel D (2003): The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication* 41(1):245-255.

Price C, Thierry G, Griffiths T (2005): Speech-specific auditory processing: where is it? *Trend in Cognitive Sciences* 9:271-276.

Price CJ (2012): A review and synthesis of the first 20 years of PET and fMRI studies of

heard speech, spoken language and reading. *NeuroImage*, 62(2): 816–47.

Ren GQ, Yang Y, Li X (2009): Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience* 162:87-95.

Ren GQ, Tang YY, Li XQ, Sui X (2013): Pre-attentive Processing of Mandarin tone and intonation: Evidence from Event-Related Potential.

Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W (2000): Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport* 11:1997-2000.

Robin DA, Tranel D, Damasio H (1990): Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language* 39(4): 539-555.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 364(1536), 3711–35. doi:10.1098/rstb.2009.0130

Rossell SL, Bullmore ET, Williams SC, David AS (2001): Brain activation during automatic and controlled processing of semantic relations: a priming experiment using lexical-decision. *Neuropsychologia* 39(11):1167-1176.

Rumsey JM, Horwitz B, Donohue BC, Nace K, Maisog JM, Andreason P (1997): Phonological and orthographic components of word recognition. A PET-rCBF study. *Brain* 120(5):739-759.

Snyder PJ, Harris LJ (1993): Handedness, sex and familiar sinistrality effects on spatial tasks. *Cortex* 29:115-134.

Talairach J, Tournoux P (1988): *Co-planar Stereotactic Atlas of the Human Brain*. Thieme Medical, New York.

Tan LH, Liu HL, Perfetti CA, Spinks JA, Fox PT, Gao JH (2001): The neural system underlying Chinese logograph reading. *Neuroimage* 13(5):836-846.

Tan LH, Perfetti CA (1998): Phonological codes as early sources of constraint in Chinese

word identification: A review of current discoveries and theoretical accounts. In *Cognitive processing of the Chinese and the Japanese languages* (pp. 11-46). Springer Netherlands.

Tsang YK, Jia S, Huang J, Chen HC (2011): ERP correlates of pre-attentive processing of Cantonese lexical tones: the effects of pitch contour and pitch height. *Neuroscience Letters* 487:268-272.

Wang, H (1986): *Modern Chinese Frequency Dictionary*. Beijing: Beijing Language Institute Press.

Wang Y, Sereno JA, Jongman A, Hirsch J (2003): fMRI evidence for cortical modification during learning of Mandarin lexical tone, *Journal of Cognitive Neuroscience* 15:1019-1027.

Wong PCM, Parsons LM, Martinez M, Diehl RL (2004): The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience* 24:9153-9160.

Wong P, Perrachione TK, Parrish TB (2007): Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping* 28(10):995-1006.

Xi J, Zhang L, Shu H, Zhang Y, Li P (2010): Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience* 170:223-231.

Xu Y, Gandour J, Talavage T, Wong D, Dziedzic M, Tong Y, Li X, Lowe M (2006): Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping* 27:173-183.

Yang J, Gates KM, Molenaar P, Li P (2015): Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*, 33, 29-49.

Yip M (2002): *Tone*. Cambridge University Press.

Yu K, Wang R, Li L, Li P (2014): Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity. *Frontiers in human neuroscience*, 8.

Zatorre RJ, Evans AC, Meyer E (1994): Neural mechanisms underlying melodic perception

and memory for pitch. *The Journal of Neuroscience* 14(4):1908-1919.

Zatorre RJ, Belin P, Penhune VB (2002): Structure and function of auditory cortex: music and speech. *Trends in Cognitive Science* 6:37-46.

Zatorre RJ, Belin P (2001): Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* 11:946-953.

Zatorre RJ, Gandour J (2008): Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:1087-1104.

Zempleni MZ, Haverkort M, Renken R, & Stowe LA (2007): Evidence for bilateral involvement in idiom comprehension: An fMRI study. *Neuroimage*, 34(3):1280-1291.

Zhang L, Shu, H, Zhou, F, Wang, X, Li, P (2010): Common and distinct neural substrates for the perception of speech rhythm and intonation. *Human Brain Mapping* 31:1106–1116.

Zhang L, Xi J, Xu G, Shu H, Wang X, Li P (2011): Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS ONE* 6, e20963.

Zhao Q, Zhou Z, Xu H, Chen S, Xu F, Fan W, & Han L (2013): Dynamic neural network of insight: a functional magnetic resonance imaging study on solving Chinese ‘chengyu’ riddles. *PloS one*, 8(3): e59351.

Figure Legend

Figure 1. Brain regions with significant activity of auditory perception of lexical tones. Figure 1a: Axial sections; Figure 1b: Lateral view. The significant threshold is $P < 0.001$ uncorrected at voxel level, $P < 0.05$ FWE-corrected at cluster level. The functional maps (in color) are overlaid on the corresponding T1 images (in gray scale). Planes are axial sections, labeled with the height (mm) relative to the bicommissural line. L=the left hemisphere; R=the right hemisphere.