CATEGORICAL PERCEPTION OF LEXICAL TONES IN CHINESE REVEALED BY MISMATCH NEGATIVITY

J. XI,a,b L. ZHANG,c H. SHU,a Y. ZHANGd* AND P. LIa*
*aState Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, PR China
bState Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, PR China
cCollege of Chinese Studies, Beijing Language and Culture University, Beijing, PR China
dDepartment of Speech-Language-Hearing Sciences and Center for Neurobehavioral Development, University of Minnesota, MN 55455, USA

Abstract—The present study investigated the neurophysiological correlates of categorical perception of Chinese lexical tones in Mandarin Chinese. Relative to standard stimuli, both within- and across-category deviants elicited mismatch negativity (MMN) in bilateral frontal-central recording sites. The MMN elicited in the right sites was marginally larger than in the left sites, which reflects the role of the right hemisphere in acoustic processing. At the same time, relative to within-category deviants, the across-category deviants elicited larger MMN in the left recording sites, reflecting the long-term phonemic traces of lexical tones. These results provide strong neurophysiological evidence in support of categorical perception of lexical tones in Chinese. More important, they demonstrate that acoustic and phonological information is processed in parallel within the MMN time window for the perception of lexical tones. Finally, homologous nonspeech stimuli elicited similar MMN patterns, indicating that lexical tone knowledge influences the perception of nonspeech signals. © 2010 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: mismatch negativity (MMN), categorical perception, lexical tones, Mandarin Chinese.

Categorical perception is one of the most extensively studied phenomena in speech perception. It refers to the ability that human listeners can perceive continuous acoustic signals as discrete linguistic representations. While most of the early categorical perception studies focused on segmental features (consonants and vowels) (Fry et al., 1962; Harnad, 1987; Liberman et al., 1957, 1961), there has been a recent surge of interests in whether categorical perception applies to the suprasegmental level, including vowel duration contrasts of quantity languages (e.g., Nenonen et al., 2003; Ylinen et al., 2005) and lexical tone contrasts of tonal languages (Francis et al., 2003; Hallé et al., 2004; Xu et al., 2006). For example, behavioral results of lexical tone perception show that continua ranging from one tone to another are not perceived categorically unless the tonal continua involve contour tones (Abramson, 1979; Francis et al., 2003; Hallé et al., 2004; Wang, 1976; Xu et al., 2006). In Mandarin Chinese there are four lexical tones, only one of which (high level tone, Tone1) does not involve a contour (Howie, 1976). Thus, native perception of Chinese tonal continua is expected to be categorical with better sensitivity to across-category distinction relative to within-category differences.

Advances in brain imaging and neurophysiological studies have allowed researchers to address fundamental questions about the neural mechanisms underlying behavioral responses for lexical tone perception (Chandrasekaran et al., 2007, 2009; Gandour et al., 2000, 2002, 2004; Gandour, 2006; Hsieh et al., 2001; Krishnan et al., 2009; Ren et al., 2009; Wang et al., 2003; Wong et al., 2004). The neural correlates of categorical perception of lexical tones, however, have not been carefully examined. The only exception was Chandrasekaran et al. (2009), in which across-category (Tone1/Tone2) and within-category (Tone2/Tone2L) tonal contrasts were investigated with an electrophysiological measure, and a larger across-category than within-category mismatch negativity (MMN) was found. However, Chandrasekaran’s study used stimuli from different tonal continua and the physical differences for the two contrasts in comparison were not equated, and therefore, their conclusions must remain tentative.

The neurophysiological study of categorical perception of lexical tones has particular theoretical significance for the understanding of the dynamic interplay between acoustic and phonological processing at the suprasegmental level. Recent neuroimaging studies have demonstrated that the two hemispheres play different roles in lexical tone perception; specifically, the right hemisphere is involved in auditory/acoustic processing, whereas the left hemisphere is responsive to linguistic/phonological functions (Gandour et al., 2000, 2002, 2004; Gandour, 2006; Hsieh et al., 2001; Wang et al., 2003; Wong et al., 2004). But the interplay between auditory and phonological processing has not been addressed in these imaging studies. There are at least two fundamental questions of interest: (a) whether the lower-level auditory processing precedes the higher-level phonological processing, and (b) whether the two processes interact with each other during processing. Zhao et al. (2008) provided some functional magnetic
Acoustic and phonological processing of both speech and nonspeech stimuli.

**EXPERIMENTAL PROCEDURES**

**Participants**

Sixteen neurologically healthy volunteers (eight females, eight males; aged 20–26, mean age = 22) with normal hearing and minimal musical experience (less than 1 year of total musical training and no musical training within the past 5 years) 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. Data from two female participants were excluded from further analyses due to insufficient number of acceptable trials as a result of strong electroencephalogram (EEG) artifacts including excessive blinking.

**Stimuli**

Two sets of stimuli were constructed for this experiment: speech and nonspeech. The speech stimuli were Chinese monosyllables /pa/ that differed in their lexical tones (the high rising tone, Tone2 and the falling tone, Tone4). The original stimuli were recorded at a sampling rate of 44.1 kHz from a female native Chinese speaker. The monosyllables were digitally edited using SoundForge (SoundForge9, Sony Corporation, Japan), each having a duration of 200 ms. In order to isolate the lexical tones and keep the rest of the acoustic features identical, pitch tier transfer was performed using the Praat software (http://www.fon.hum.uva.nl/praat). This procedure generated two stimuli, /pa2/ and /pa4/, which were identical with each other except for the pitch contour difference. The /pa2/ and /pa4/ stimuli were then taken as the endpoint stimuli to create a 10-interval lexical tone continuum. A morphing technique was performed in Matlab (Mathworks Corporation, USA) using STRAIGHT (Kawahara et al., 1999) in 10 equal intervals. All 11 stimuli were normalized in RMS intensity. The 11 stimuli in the /pa2/-/pa4/ lexical tone continuum were examined in a behavioral test (identification and discrimination tasks) on 16 native Chinese speaking adults who did not participate in the subsequent event-related potential (ERP) study. The identification task required participants to identify each stimulus. In the session, 40 trials for each stimulus were randomly presented in isolation. The AX discrimination task required participants to judge whether pairs of presented stimuli were the same or different. Stimulus presentation was randomized and both directions of presentation order were tested, each for 20 trials. Based on the behavioral test results (Fig. 1), an across-category stimulus pair (3 and 7) and a within-category stimulus pair (7 and 11) were chosen for the ERP oddball paradigm experiment. In particular, stimulus 7 in the continuum was used as the standard stimulus, and stimuli 3 (an across-category deviant) and 11 (a within-category deviant) were used as deviants.2

The nonspeech stimuli were harmonic tones with the same pitch, amplitude, and duration parameters as the speech stimuli.
The only difference between the speech and nonspeech stimuli was in the spectral components. The nonspeech stimuli were composed of six equal amplitude harmonics (1, 3, 6, 7, 8, 12) of the F0. Harmonics 2, 4, 5, 9, 10, and 11 were omitted to increase perceptual dissimilarity with the speech stimuli. All the stimuli were normalized to have equal average RMS intensity (Fig. 2).

**ERP procedure**

Two oddball blocks were presented to each participant with counterbalanced sequence among the participants. The within-category deviant and across-category deviant occurred pseudo-randomly among standards with a probability of 10% respectively, and any two adjacent deviants were separated by at least three standards for a total presentation of 1000 stimuli in each presentation block. The stimulus-onset-asynchrony (SOA) was 800 ms.

Participants were seated comfortably in an acoustically and electrically shielded chamber and instructed to ignore the presented sounds while watching a self-selected movie. The movie was presented in mute mode with subtitles on a screen of 15 cm × 10 cm positioned approximately 1 m in front of the seat. Participants were informed to complete a questionnaire about the movie after the experiment. Sound stimuli were presented binaurally via an insert Sony earphone. The right and left acoustic channels of the insert earphone were calibrated for equal and comfortable loudness (70 dB SPL) prior to the experiment. The experimental session lasted approximately 1.5 h including preparation, data acquisition, breaks, and cleanup.

**Electroencephalogram (EEG) recording**

Continuous EEG was recorded using a HydroCel Geodesic Sensor Net (HCGSN) consisting of 128 electrodes evenly distributed.
across the scalp and referenced against the vertex electrode (Tucker, 1993). The GSN also includes electrodes next to, and below, the eyes for recording horizontal and vertical eye-movements. The impedance of each electrode was kept below 5 kΩ.

**Data analysis**

Off-line signal processing was carried out using Netstation software (Version 4.2). The raw data were first digitally filtered with a 0.3–20 Hz bandpass filter and segmented for 700 ms starting 100 ms prior to the onset of stimuli. Data were then re-referenced to the average of all the electrodes, and baseline corrected. Recorded trials with eye blinks or other activities beyond the range of ±50–50 μV were rejected.

Based on previous studies and visual inspection of the grand-average waveforms, two recording sites were selected for statistical analyses: left frontal (sites F3, channels 19, 23 and 24) and right frontal (sites F4, channels 3, 4 and 124) (Díaz et al., 2008; Kaan et al., 2007; Kirmse et al., 2008; Luo et al., 2006; Oken and Chiappa, 1986). Only the standard before the deviant was used for averaging and subtraction. Difference waves for MMN were obtained by subtracting the averaged standard from the averaged deviant. The MMN peak latency for each subject was found within a 60 ms time window that was defined by the grand-average waveforms at Fz. The calculation of mean amplitude was conducted with a moving window technique: first the negative peak within the 60 ms time window was found for each subject, then the value of a time window which extended ±40 ms surrounding the MMN peak was averaged. Statistical analysis only included those participants with at least 80 accepted deviant trials in each condition.

**RESULTS**

The grand average waveforms elicited by the standards and the deviants are shown in Fig. 3. Negative peaks were observed in the deviant-minus-standard difference waves for the speech (lexical tone) and nonspeech (harmonic tone) conditions (see Fig. 4). Two three-way condition (lexical tone/harmonic tone) x deviant type (within-category/across-category) x hemisphere (left/right) repeated measures ANOVAs were conducted for mean amplitude and peak latency respectively. For all analyses, degrees of freedom were adjusted according to the method of Greenhouse–Geisser when appropriate.

**MMN peak latency**

The mean peak latencies at the electrodes (F3, F4) were plotted in the upper panel of Fig. 5. Results from the omnibus ANOVA yielded a significant main effect of condition ($F(1, 13) = 17.186$, $P = 0.001$, nonspeech > speech), showing that speech was processed faster than nonspeech. The main effect of deviant type or hemisphere failed to reach significance. None of the two- or three-way interactions reached significance.

**MMN mean amplitude**

For the MMN amplitudes (lower panel of Fig. 5), the main effect of condition was not significant ($F(1, 13) = 1.498$, $P = 0.262$).
The main effect of deviant type was significant \( (F(1, 13) = 7.382, P = 0.018, \text{across-category} > \text{within-category}) \) and the main effect of laterality was marginally significant \( (F(1, 13) = 3.693, P = 0.077, \text{right hemisphere} > \text{left hemisphere}) \). There was also a significant interaction between deviant type and hemisphere \( (F(1, 13) = 5.790, P = 0.032) \), indicating that the amplitude of the response to across-category deviant was greater than response to within-category deviant in the left hemisphere \( (F(1, 13) = 20.80, P = 0.001) \), but not in the right hemisphere \( (F(1, 13) = 0.16, P = 0.694) \). No other effects reached statistical significance.

**DISCUSSION**

The present study examined the neurophysiological, particularly MMN, indices of categorical perception of lexical tones in native speakers of Mandarin Chinese. It was found that across-category contrast elicited a larger MMN than within-category distinction for Chinese listeners. Because physical intervals for the two contrasts were equated, the study provided strong neurophysiological evidence in support of categorical perception of lexical tones. Previous studies using a similar design showed that the MMN response is enhanced by the listener’s detection of across-category differences in vowels (Winkler et al., 1999), place of articulation (Dehaene-Lambertz, 1997), voice onset time (VOT) (Sharma and Dorman, 2000) and vowel length of quantity languages (Nenonen et al., 2003). Our results are consistent with these studies, supporting the view that when listeners process a speech signal to recover the phonological

![Fig. 4. Mismatch negativities (MMNs) recorded in speech and nonspeech conditions. (A) Grand average traces of MMN evoked by within- and across-category changes from the F3, Fz, and F4 electrode locations; (B) Maps display the topographic distribution of the mean amplitudes in the MMN analysis window.](image-url)
representations of their native language, they automatically extract the categorical linguistic information and show reduced sensitivity towards within-category acoustic differences.

With regard to laterality of lexical tone perception, two competing hypotheses have been proposed in the literature. The acoustic hypothesis claims that speech perception is cue-dependent and that pitch is mediated by the right hemisphere regardless of linguistic functions (Klouada et al., 1988; Zatorre and Belin, 2001). In contrast, the functional hypothesis claims that speech perception is task-dependent and that pitch patterns carrying linguistic functions are lateralized to the left hemisphere while those carrying no linguistic functions lateralized to the right hemisphere (Van Lancker, 1980; Wong, 2002). Both hypotheses have found empirical support in the past (Gandour et al., 2000, 2002; Wong et al., 2004; Zatorre and Belin, 2001). In recent years, a more comprehensive model that integrates the acoustic hypothesis and the functional hypothesis has been put forward by Gandour and his colleagues (Gandour et al., 2004; Tong et al., 2005; Zatorre and Gandour, 2008), according to which the right hemisphere is sensitive to low-level acoustic processing and the left hemisphere is sensitive to high-level linguistic processing.

Data from the present study provided neurophysiological evidence in testing these models. We found that relative to standard stimuli, both within- and across-category deviants elicited larger MMNs in the right frontal-central recording sites, which is consistent with previous studies (Luo et al., 2006; Ren et al., 2009). The right lateralized response presumably reflects the role of the right hemisphere in acoustic processing. At the same time, relative to within-category deviants, the across-category deviants elicited larger MMN, which probably reflects the long-term phonemic traces of lexical tones because the ERP topography data showed that the enhancement was dominant in the left frontal-central recording sites (Näätänen et al., 1997). These results are consistent with the hypothesis of Gandour et al. (2004) with regard to the differential functional roles of the two hemispheres. Given the low spatial resolution of ERP, however, the results should be considered preliminary. Magnetoencephalography (MEG) study with high temporal resolution and acceptable spatial resolution is needed in order to clearly identify the temporal dynamics of cortical activation in different brain regions involved in the processing of Chinese lexical tones.

Previous models, however, have not addressed the time course issue with regard to whether auditory/acoustic processing precedes high-level linguistic/phonological processing or whether the two types of processing occur in parallel. Our data provided insights into this issue. Findings from our study indicate that acoustic processing and linguistic processing interact at an early stage, in that acoustic and phonological information is activated in parallel for lexical tone perception within the short MMN time window. This finding provides counter evidence to the two-stage model for lexical tone processing (Luo et al., 2006), according to which lower-level auditory processing precedes higher-level phonological processing.
Another issue that our data help to resolve is the question of how linguistic experience affects the perception of nonspeech sounds. One view is that there exist separate modules for speech and nonspeech processing, and that exposure to the sound structure of a particular language has no influence on nonspeech auditory processing (Burnham et al., 1996; Liberman and Mattingly, 1989; Miyawaki et al., 1975; Zhang et al., 2005). The contrasting view is that there are shared speech and nonspeech processing mechanisms, and that language learning exerts strong influence on the perception of nonspeech sounds (Bent et al., 2006; Francis et al., 2003). Such discrepant views could be due to the fact that speech contrasts such as lexical tone categories, vowels, and approximants in the cited studies are cued by markedly different spectral and temporal features—some acoustic features in speech may be directly analyzed on the basis of general auditory processing capacity, and others may reflect language-specific processing of the neural system. Our findings on lexical tone processing are consistent with the second view—the MMN amplitude data for nonspeech stimuli were similar to the results for speech stimuli. Our study accords well with studies of other languages, for example, by Sussman et al. (2004) who found similar patterns of MMN for speech and nonspeech counterparts for Finnish speakers. Despite the MMN amplitude similarity between speech stimuli and nonspeech control, our analysis of MMN latencies revealed differences in speech vs. nonspeech processing. The MMN latency data showed that speech was processed faster than nonspeech. This latency effect could be attributed to listeners’ linguistic processing or familiarity with the phonetic segments (i.e. /p/ and /a/) of the speech stimuli. Although speech and nonspeech stimuli in our study have the same pitch, amplitude, and duration parameters, the difference in spectral components clearly distinguishes the two. Therefore listeners’ phonetic processing of the /pa/ syllable might have led to faster change detections.

An alternative view is that categorical perception of pitch contours results from short-term categorical memory, which runs parallel to fine-grain sensory encoding (Xu et al., 2006). The MMN difference between the within- and across-category contrasts, therefore, might be due to the rising versus falling pitch patterns in terms of feature abstraction rather than high-level phonological representation. This view can also potentially account for our findings that the across-category deviants elicited larger MMN than the within-category deviants for both speech and nonspeech sounds. According to this view, for nonnative listeners, across-category deviants should also elicit larger MMN than within-category deviants, especially for the nonspeech stimuli.

Evidence from a separate ERP experiment that we conducted does not agree with this view. Using the same stimuli and procedure as in the current study, we tested Korean listeners who had no experience with Chinese lexical tones and found that both within- and across-category deviants elicited larger MMN in the right frontal-central recording sites, similar to the results of the Chinese participants. However, these listeners showed no difference in the MMNs elicited by the within- and across-category deviants, which is different from the results of the Chinese native listeners (see the supplementary materials for further details). These results therefore provide more convincing evidence that the MMN differences elicited by the within- and across-category deviants for the Chinese listeners in the present study could be attributed to high-level phonemic processing rather than low-level auditory processing. In recent fMRI studies, it has been found that categorical perception of phonemic sound patterns is mediated by the left middle superior temporal sulcus (Joanisse et al., 2007; Liebenthal et al., in press). Future fMRI studies on categorical perception of Chinese lexical tones could further help to clarify whether it is attributable to long-term phonological representations or to short-term categorical memory.

CONCLUSION

In conclusion, the present study used neurophysiological (MMN) indices to identify the neural correlates of categorical perception of Chinese lexical tones. Our findings have shown spatial (hemispheric) differences in auditory processing versus phonological processing but no temporal differences between the two types of processing—the two types of processing occur in parallel within the MMN time window and show early interaction between the two hemispheres. Our data also suggest common mechanisms underlying the processing of speech and nonspeech materials for native speakers, in which language-specific tonal experience with the perception of lexical tones influences the processing of nonspeech signals.

Acknowledgments—The research was supported by the Program for Changjiang Scholars and Innovative Research Team at Beijing Normal University, by grants from the Natural Science Foundation of China (7092051) to HS and in part from the National Science Foundation of USA (BCS#0642586) to PL. YZ was supported by a Single Semester Leave Award and two CLA Brain Imaging Research Awards from the University of Minnesota to work on the project. The authors thank four anonymous reviewers for suggestions to improve the manuscript.

REFERENCES


Burnham D, Francis E, Webster D (1996) The development of tone perception: cross-linguistic aspects and the effect of linguistic con-


APPENDIX

Supplementary data
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroscience.2010.06.077.