Temporal hierarchy of intonation and tone processing in Mandarin Chinese

Jia-qi Li*

State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research,

Beijing Normal University, Beijing, China 100875

Running Head: temporal hierarchy of intonation and tone processing

---

*To whom correspondence should be sent: jiaqi.li7@outlook.com

ABSTRACT

Intonation and lexical tone both rely on F0 in tonal languages. While the former one can indicate the speaker's intention (e.g., raise a question or state a fact), the latter helps determine the meaning of a word. These two interact with each other in sentence, but the access order of them by the brain is unknown. The current study aimed at investigating the access order of intonation and lexical tones in Mandarin Chinese. Participants were required to listen to Mandarin sentences containing both intonation and lexical tone in the same syllable and judge the intonation (in the intonation task) and tone (in the tone task). Their behavioral reactions and mean amplitudes in ERP were recorded and analyzed. Results showed that in the intonation task, the participants responded to intonation first and then the lexical tone. In the tone task, participants were sensitive to both intonation and tone in an early interaction beginning from 100ms. Participants were alert to intonation and would process it no matter in which task, indicating the important role of intonation in human communication.

KEY WORDS：Intonation, lexical tone, Mandarin

1. Introduction

Pitch is an important feature of language. At the syntax level, termed as intonation, it indicates intonational phrase boundaries as well as syntactic boundaries (Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004; Steinhauer, Alter, & Friederici, 1999). At the lexical level, termed as lexical tone, it determines word meaning in tone languages (e.g., Nan & Friederici, 2013). Given that both intonation and lexical tone are based on F0 cues, these two inevitably interact during speech processing (see Connell, Hogan, & Rozsypal, 1983; Yuan, 2006; Yuan, 2011 for research on Mandarin ; Ma, Ciocca, & Whitehill, 2011 for research on Cantonese). Previous research has examined tone recognition and found it quite stable under changes in pitch and intonation contour (Connell, Hogan, & Rozsypal, 1983; Liu, Tsao, & Kuhl, 2007), but a recent study indicated that toddlers' lexical-tone based word recognition might be influenced by intonation using a preferential looking paradigm (Singh & Chee, 2016). As for the perception of intonation, past research has found that lexical tone of the last syllable affects the identification of questions (Yuan, 2006; Yuan, 2011). For instance, Yuan (2011) showed that question intonation was easier to identify on a sentence ending with Tone4 and more difficult to identify when ending with Tone2. There were also studies that focused on the interaction between intonation and tone though did not instruct participants to identify intonation or tone. These studies revealed that stimuli with a rising question intonation and a low lexical tone (e.g., tone4 in Mandarin) would evoke P600 effect (Kung, Chwilla, & Schriefers, 2014) and MMN effect (Ren et al., 2009; Ren et al., 2013).

As intonation and tone processing unfolds over time, this temporal dynamic should be included in the examination of the intonation and tone identification as well as the interaction between them. However, the temporal dynamics of how intonation and tone would be processed in an utterance and the access order of them still remains unclear. As one of the tone languages, Mandarin provides a good chance to investigate the temporal dynamics of intonation and tone processing for two reasons. First, in Mandarin, question and statement intonations can be differentiated by the pitch difference on the final syllable (Ho, 1977), making it possible that both intonation information and lexical tone information can be carried by the same syllable. Second, intonation variations in Mandarin do not necessarily require any syntactical or lexical change, which increases listeners' dependence on pitch in the final syllable to recognize intonation, and thus provides a perfect language environment to examine the online processing of intonation and lexical tone that both rely on pitch cues in the same syllable. In this study, the temporal hierarchy of intonation and tone processing in Mandarin was investigated in order to explore the dynamic process of speech prosody. To predict the order of intonation and tone processed by speakers of tonal languages, two hypotheses are proposed and tested in this study: task dependent and intonation-dominant.

It has long been established that online speech processing by our brain is task dependent (Zatorre, Evans, Meyer, & Gjedde, 1992). Using the same speech signal (consonant-vowel-consonant speech syllable pairs), phonetic discrimination task resulted in increased activity in the left Broca's area, whereas pitch discrimination task led to activation of the right prefrontal cortex (Zatorre et al., 1992). If the temporal hierarchy of intonation and tone processing is task dependent in nature, it is expected that the effects of tone would show up the earliest in the tone identification task, while intonation effects are the top priority in the intonation task.

On the contrary, the intonation-dominant hypothesis posits that intonation might have higher temporal hierarchy than tone in online speech processing. There are four reasons supporting this proposition. First, prosodic cues are language independent and may have deeper evolutionary roots than the lexical tones, which seem have arisen out of the loss of certain prior voicing contrasts in non-tonal languages (Yip, 2002). Prosodic discrimination is believed to be a predominant ability for Java sparrows, who paid attention to differences in prosody rather than sentence contents, both of which they were able to discriminate in the Japanese language (Naoi, Watanabe, Maekawa, & Hibiya, 2012). Spierings and ten Cate suggested that zebra finches are sensitive to the same set of prosodic cues (pitch, duration and amplitude) as employed in human speech. Moreover, pitch cues were treated more saliently by zebra finches than the other prosodic features similar to the way human perceive speech prosody (Spierings & ten Cate, 2014). A cross-fostering experiment (Araki, Bandi, & Yazaki-Sugiyama, 2016) revealed that the temporal structure in birdsong prosody is innate for zebra finches. Second, the prosodic skills, emerged at a very early age, are central to the newborn's innate abilities to learn speech and may have scaffolded language acquisition (Fernald, 1989), supported by the findings that 5-month-old infants could respond differentially to approval and disapprovals (Fernald, 1993) and positive and negative vocal expressions (Fernald, 1995), and that infants developed preferences for the prosodic characteristics between six and nine months and may rely on prosodic cues (e.g., stress patterns) to segment words from fluent speech (Jusczyk, 1999) and thus to facilitate auditory word learning (Shukla, White, & Aslin, 2011). In contrast, although the sensitivity to lexical tones also emerged quite early during infancy, the stabilization of perception and production skills of lexical tones takes a lengthier learning curve, often spanning several years even beyond early childhood (Singh & Chee, 2016). Third, universally shared by different languages (Hirst & Di Cristo, 1998), intonation conveys not only linguistic information, but also paralinguistic information in human communication, such as speakers' emotion, attitude and intention (Cruttenden, 1997; Banse & Scherer, 1996; Hirst & Di Cristo, 1998; Ladd, 1996). People can even decode emotional prosody in unfamiliar Western and non-Western languages (Thompson & Balkwill, 2006). Such social information might be assessed with higher priority by the human brain than the lexical information cued by lexical tones because of its important role for social interaction and human survival (e.g., Brancucci, Lucci, Mazzatenta, & Tommasi, 2009). Fourth, previous research has found evidence of categorical perception for lexical tones (Xu, Gandour, & Francis, 2006; Halle, Chang, & Best, 2004), but not for intonation (Liu & Rodriguez, 2012), suggesting that subtle changes in intonation is important to human beings given that our brain priorities limited resources to most important tasks. In their study, Liu and Rodriguez (2012) examined categorical perception of intonation of English speech (statement versus question) among native English and Chinese listeners. Chinese listeners showed steeper intonation identification functions and different intonation boundaries than English listeners. However, the functions of discrimination were not typical for categorical perception with no or at most modest discrimination peaks for both groups (Liu & Rodriguez, 2012).

Considering the above characteristics of intonation compared to lexical tones, it is natural to hypothesize that our brain might be more alert and sensitive to intonations than tones during online speech processing. However, most of previous research focused on how intonation and tone interact with each other, with a lack of understanding in the fine-grained temporal details. It is widely believed that lexical tones affect the recognition of intonation identities. Questions are easier to identify when the final syllable is tone4 than tone2, while no significant effects of tone on the identification of statement intonation was found (Yuan, 2006; Yuan, 2011). The falling slope of tone4—which would be flattened by the question intonation—not only makes question more obvious perceptually and easier to be recognized, but also strengthen the question-statement contrast, supported by findings in previous ERP studies (see Ren et al., 2009; Ren et al., 2013 for MMN effects of statements and questions in tone4 rather than tone2; Liu et al., 2016 for P300; Kung et al., 2014 for P600). Taking into account the strong intervention of the falling tone on the question intonation processing, Yuan regarded intonation as tone-dependent in a local perspective, and concluded that the tone identification might precede—or at least be processed at the same time with—the intonation identification (Yuan, 2011). As for the influence of intonation on lexical tones, it was believed that in Mandarin, the intonation does not affect the shape of tones (Ho, 1977) and has little effect on tone perception for adult Mandarin speakers (Connell et al., 1983). Chinese infant-directed speech with exaggerated intonation contours does not distort the lexical tones, either (Liu, Tsao, & Kuhl, 2007). However, research in Cantonese has found that question intonation made a large proportion of low tones be misperceived as high rising tone (Ma, Ciocca, & Whitehill, 2006), and research in toddlers has found the effect of intonational variation on lexical-tone based word recognition in Mandarin (Singh & Chee, 2016), which both indicated the possible influence from intonation-induced changes in F0 contour on the recognition of tone identities.

The current study set out to examine the temporal dynamics of intonation and tone processing during online speech perception in Mandarin. We focused on the micro-structure of the entire dynamic process as it occurs in real time. Event-related potentials (ERPs) are optimal since it can track neuronal brain activities with fine-grained temporal resolution. Using sentences where the lexical tones and intonation for the final words were independently manipulated, we employed both tone and intonation identification tasks during the ERP measurements. Tone and intonation identification tasks were based on the same set of sentences, the acoustic characteristics between two tasks were thus maximally matched. Note that Liu and her colleagues also conducted tone and intonation tasks in a ERP study (Liu et al., 2016), but they mixed the ERP data from two tasks, thus were unable to differentiate the temporal status of intonation and tone identification for native Mandarin speakers. Besides, we need to adopt a stimulus set that could present both intonation information and tone information in the same segment. Mandarin Chinese can serve this purpose because both the lexical tone and intonation could be carried by the exact same final syllable in Mandarin. As a matter of fact, previous research showed that the intonation perception relies more on the F0 characteristics of the final syllables than the overall pitch level in Mandarin Chinese (Liang & van Heuven, 2007). Thus, we have Intonation (question versus statement) and Tone (tone2 versus tone4) as independent variables. It is hypothesized that (1) Intonation takes priority in Mandarin sentence processing, no matter in the identification task of itself or of lexical tones; (2) While the acoustic cues of lexical tones have influence on the identification of intonation, intonation also has effects on tone identification task.

2. Method

2.1 Participants

Fourteen native speakers of Mandarin Chinese (five males) participated in the experiment. All participants were college students from universities in Beijing recruited through internet advertising, and they were all right-handed according to the Edinburgh Handedness Inventory

(Oldfield, 1971). Age of the participants ranged from 19 to 26 (Mean ± SD: 22.75 ± 2.12). The audiometric thresholds of all the participants are lower than 20 dB HL for octaves ranging from 250 Hz to 8000 Hz. All participants had not received any formal music training and showed normal music ability when assessed using the Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod, & Hyde, 2003), which consisted of three pitch-based subtests, two time-based subtests, and one memory subtest. They also showed normal lexical tone discrimination and identification ability measured by the lexical tone perception tests (Nan, Sun, & Peretz, 2010).

2.2 Materials

Thirty-nine pairs of monosyllabic words were selected. Each pair contained two words that had the same vowel and consonant and only differed in tone (either tone2 or tone4). These monosyllabic words were embedded in the final position of a five-syllable sentence, i.e., zhe4 ge4 zi4 nian4 X (This word is pronounced X), which was semantically neutral. The sentences were produced by a female native Mandarin speaker with either a statement or a question intonation, making up a total of 156 sentences (39 word pairs x 2 tones x 2 intonations).

The speaker's recordings were then manipulated in PRAAT (Boersma & Weenink, 2004). We cut out the final critical word of every sentence and calculated the average duration of the four words in a word group (tone2 and tone4 words produced as question or statement). Then we manipulated the duration of words in the same group to be the same as the average value. The duration of critical words ranges from 405ms to 584ms. A same carrier sentence (i.e., zhe4 ge4 zi4 nian4) produced as statement intonation was used for each of the 156 critical final words. In order to examine the quality of these sentences, eleven Mandarin speakers were invited to identify the intonation (question/statement) and the final tone (tone2/tone4) of each sentence. The order of the two tasks (intonation task and tone task) for each subject was randomized. For any sentence with intonation or tone identification rates lower than 0.72, the related pairs (including its intonation pairs and tone pairs) were removed from the experiment materials. In total 28 sentences were removed and 128 sentences remained (32 word pairs x 2 tones x 2 intonations). The mean intonation identification rate of all the 128 sentences was 0.93 (minimum: 0.73; maximum: 1.00; std. deviation= 0.09) and the mean tone identification rate was 0.98 (minimum: 0.82; maximum: 1.00; std. deviation= 0.05). No significant difference in word frequency was found in T-Test of words in tone2 compared with words in tone4 ($t(57) = 1.617$, $p=0.111$). The acoustic analysis revealed that for each word, its question version and statement version were significantly different in F0 level ($t(63) = 25.046$, $p<0.001$) and intensity ($t(63)=6.783$, $p<0.001$). There were four stimuli conditions in total (q2: question in tone2, q4: question in tone4, s2: statement in tone2, s4: statement in tone4). Their mean F0 level, intensity and duration are shown in Table 1. F0 contours of sample stimuli ("fei") were displayed in Fig. 1.

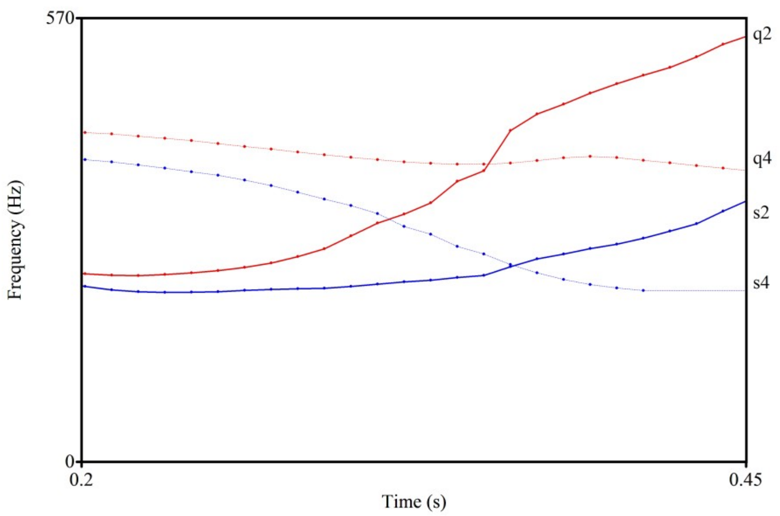


Fig. 1. Voice fundamental frequency contours (F0: 0–570 Hz) of sample stimuli ("fei"). F_0

contours are displayed for four conditions of this final critical word consisting of two tones (i.e., tone2, tone4) and two intonations (i.e., question, statement).

Additionally, F0 interval for each word was measured by calculating the difference in semitone of the 0% and the 100% time points (Point 100% - point 0%). For both F0 interval and F0 level, ANOVA of all final words showed a significant main effect of Intonation ($F(1,31)=592.736$, $p<0.001$, $\eta_p^2=0.950$ for F0 level; $F(1,31)=82.936$, $p<0.001$, $\eta_p^2=0.728$ for F0 interval) and a significant main effect of Tone ($F(1,31)=539.493$, $p<0.001$, $\eta_p^2=0.946$ for F0 level; $F(1,31)=431.720$, $p<0.001$, $\eta_p^2=0.933$ for F0 interval) (see Fig. 2). No significant interaction effect was revealed by the pairwise comparison of the interaction between Intonation and Tone. For both tones, statements have lower F0 level (ps<0.001) and F0 interval (ps<0.001) than questions. For both intonations, tone2 has lower F0 level (ps<0.001) and higher F0 interval (ps<0.001) than tone4 (see Fig.3).

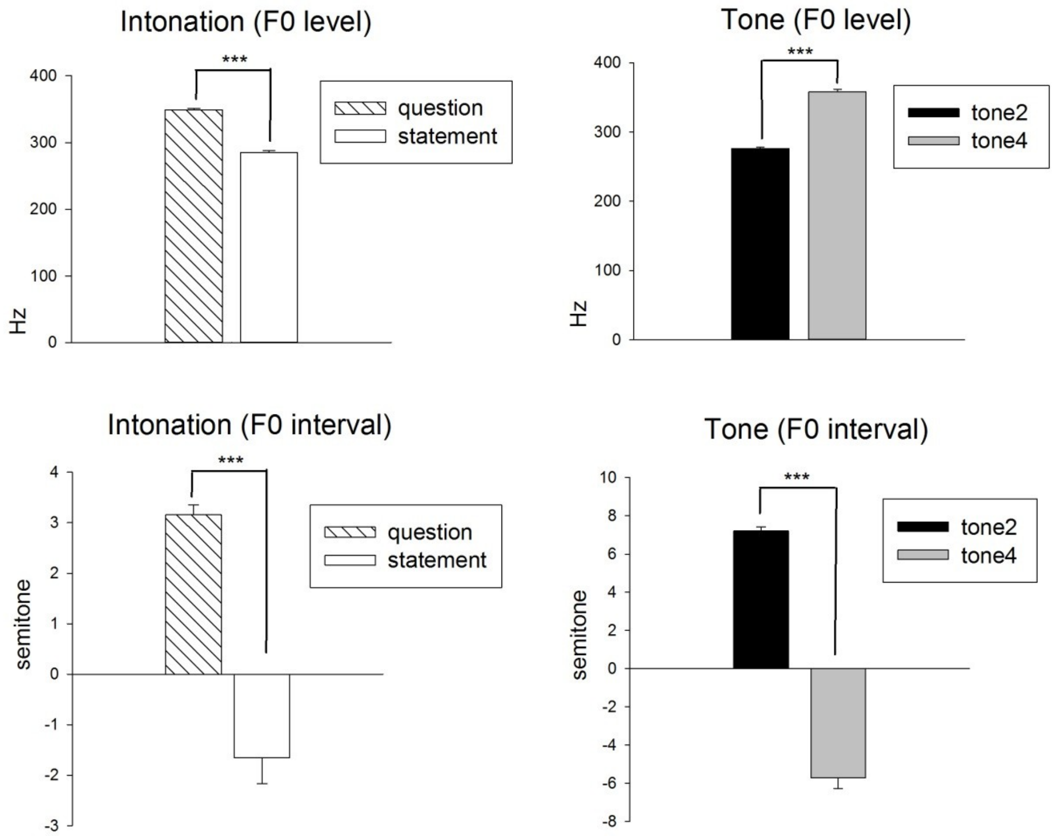
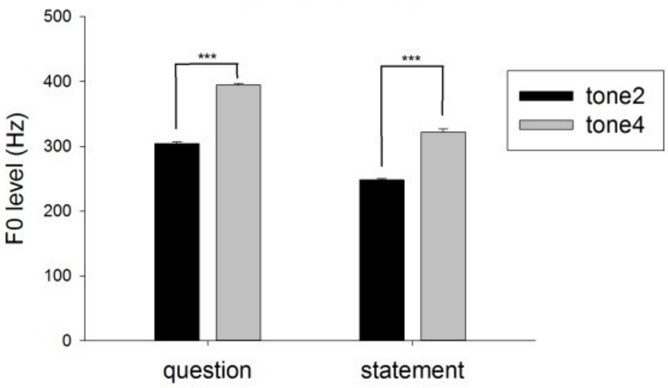


Fig. 2. Significant main effect of Intonation and Tone in F0 level (top line) and F0 interval (bottom line). ***, $p<0.001$. Error bar indicates standard error.



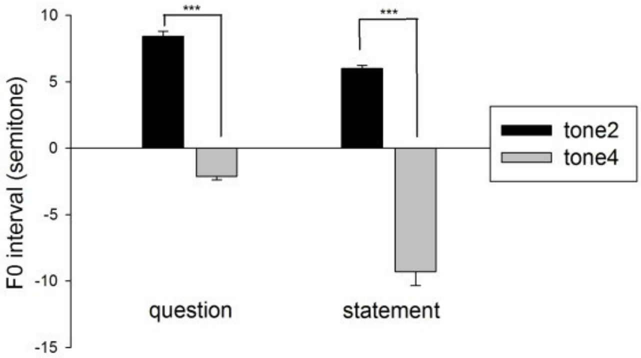


Fig. 3. No meaningful interaction between Intonation and Tone in F0 level (top line) and F0 interval (bottom line). *n.s.* means p>0.05; *, p<0.05; **, p< 0.01; ***, p<0.001. Error bar indicates standard error.

Table 1

Acoustic features of final words in the four stimuli conditions (2 tones x 2 intonations).

Stimulus	F0 level (Hz)		F0 interval (semitone)		Intensity (dB)		Duration (s)		N
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	
q2	304.33	12.54	8.43	2.12	73.71	1.97	0.49	0.04	32
q4	394.33	17.32	-2.11	1.49	76.89	2.40	0.49	0.04	32
s2	248.30	10.32	5.99	1.33	72.47	2.05	0.49	0.04	32
s4	322.00	28.76	-9.30	5.93	75.40	2.62	0.49	0.04	32

2.3 Procedure

Participants were asked to perform an intonation identification task and a tone identification task, which were presented in a counterbalanced order across participants. Participants were required to identify whether the previously presented sentence was a question or a statement, and whether the final tone of the sentence was tone2 or tone4, respectively. Stimuli used in the two tasks were the same 256 sentences and each was presented twice in both tasks (512 sentences in total in each task). Participants were asked to press the corresponding buttons indicating a question or a statement (a tone2 or tone4 in tone identification task) within a three-second time limit. The behavioral identification tasks were presented 1200ms after the critical word onset.

In each task, instructions were given on screen, and then one practice session containing 10 trials were presented with feedback ("correct"/ "incorrect"/ "unrecorded")to familiarize the participants with the task. "Unrecorded" indicated that the participant did not press any button within the time limit. In a task, there were four experiment sessions, each containing 128 trials presented in a random order. Between each session, the participants could take a break (usually 2-5 minutes), and continue the next session when they were ready. In each trial, the participants were required to gaze on the fixation which lasted 250ms, and then they would hear the sentence (890ms for the carrier sentence and 405-584ms for the final word) from the headphone at a comfortable listening level. The instruction of response ("please press the button") would appear on the screen 1200ms after the beginning of the final word. Then the participants could press the button within the time limit of 3000ms. The feedback ("recorded"/ "unrecorded") without any indication of correctness would be presented on the screen immediately after the participants made their choices. Then the next trial would automatically start.

2.4 EEG recording

In current study, participants were tested individually in a quiet room and each wore a Quick-Cap with 64 scalp channels placed according to the international 10–20 system. The EEGs were recorded using a SynAmps EEG amplifier from a 64-channel (Ag/AgCl) NeuroScan system with a 1000-Hz sampling rate and a bandpass filter of 0.05–200 Hz. The reference electrode was placed on the tip of the nose, and the ground electrode was placed between FPz and Fz. Horizontal electrooculograms (HEOGs) were recorded from a pair of electrodes placed next to the outer canthi of both eyes, and vertical electrooculograms (VEOGs) were recorded above and below the left eye. The impedance of each channel was kept below 5 kΩ.

2.5 Data analysis

For the off-line signal processing of EEG data in both tasks, eye blinks were corrected using computational algorithms in the NeuroScan software (Semlitsch, Anderer, Schuster, & Presslich, 1986), and then the data were filtered (low-pass filter of 30 Hz, zero-phase, and 24 dB/octave) and segmented into epochs of 1400ms (a 200ms baseline before the critical word and 1200ms after the onset of the critical word). The trials in which participants did not answer correctly were excluded from the analysis. After baseline correction, epochs with voltage changes exceeding ±150 μV at any channel were excluded. The remaining epochs were separately averaged for each stimuli condition. The average accepted number of trials across stimuli conditions were all above 40 in the intonation task and above 60 in the tone task. The mean amplitudes at nine electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4) were selected for statistical analysis. Area (frontal: F3, Fz, F4; central: C3, Cz, C4; parietal: P3, Pz, P4) and Hemisphere (the left: F3, C3, P3; the midline ：Fz, Cz, Pz; the right: F4, C4, P4) were included as within-subjects factors in the repeated measures analyses of variance (ANOVAs) described later in ERP results.

3. Results

3.1 Behavioral Results

For both tasks, the behavioral performance (i.e., identification rate) was calculated for all the four stimuli conditions. Repeated measures ANOVA with Task (intonation identification task versus tone identification task), Intonation (question versus statement) and Tone (Tone2 versus Tone4) as the within-subject factors revealed a significant main effect of **Task** ($F(1,13)=23.042$, $p<0.001$, $\eta_p^2=0.639$). The tone identification task had higher identification accuracy than the intonation identification task. The main effect of **Intonation** also reached significance ($F(1,13)=10.390$, $p=0.007$, $\eta_p^2=0.444$). Question had a much lower identification rate than statement in both tasks (Fig. 4). A significant interaction of **Task×Tone** ($F(1,13)=7.675$, $p=0.016$, $\eta_p^2=0.371$) was found (Fig. 5). Identification rate of sentences with final Tone2 was lower than that of Tone4 in intonation identification task ($p=0.020$), but no significant difference between Tone2 and Tone4 was observed in tone identification task ($p=0.140$). Similar ANOVA analysis of d-prime in two tasks also revealed the same findings.

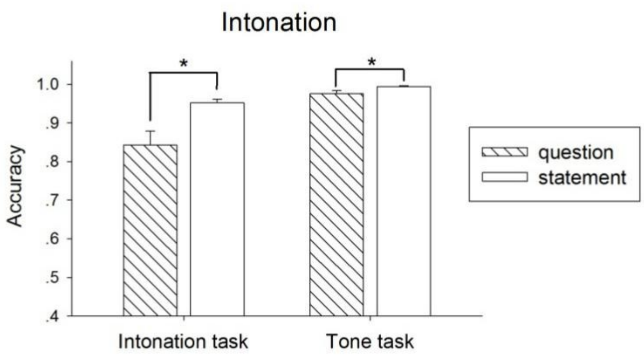


Fig. 4. The main effect of Intonation. *n.s.* means $p>0.05$; *, $p<0.05$; **, $p< 0.01$; ***, $p<0.001$. Error bar indicates standard error.

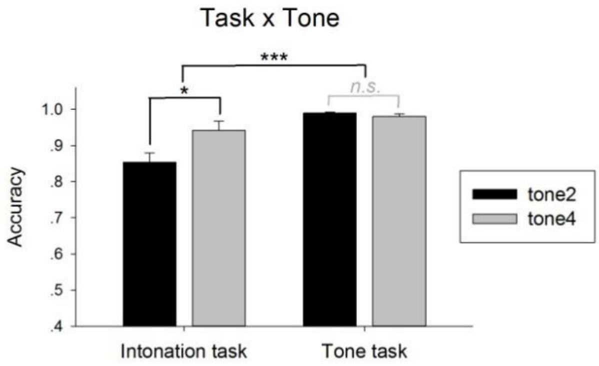


Fig.5. Two-way interaction Task × Tone. *n.s.* means p>0.05; *, p<0.05; **, p< 0.01; ***, p<0.001. Error bar indicates standard error.

To explore the relationship between the acoustic features of the final word and the identification rate in two tasks, a linear regression analysis was conducted, including F0 level and F0 interval as the independent variables. Results showed that for intonation identification, the regression model with F0 interval included and F0 level excluded was significant (p<0.001). On the contrary, for tone identification, the regression model that was statistically significant included F0 level and excluded F0 interval (p<0.001). It suggested that although both F0 level and F0 interval could differentiate intonations and tones as reflected by the acoustic analysis, when listening to those stimuli combining the intonation and tone, listeners' identification of intonation might primarily be influenced by F0 interval, while the identification of tones might connect more with the F0 level than F0 interval.

3.2 ERP results

Repeated measures analyses of variance (ANOVAs) with Intonation (question and statement), Tone (tone2 and tone4), Area (frontal, central, and parietal) and Hemisphere (left, midline, and right) as within-subjects factors were performed for the mean amplitudes in participants' ERP responses in continuous 50ms time windows. Results are presented in Table 2. The consecutive time windows that have the same effects according to the ANOVA results were combined. Only the effects that could be found in more than two consecutive 50ms time windows were analyzed. Findings below are reported according to the combined time windows.

In the intonation task, the overall ANOVA for the mean amplitude of the nine electrodes in the time window of 150–250ms revealed a significant **Intonation×Area** effect ($F(2,26)=33.089$, $p<0.001, \eta_p^2=0.718$). In parietal area, question has a larger negative component ($p=0.008$). No other significant interaction or main effects were found. In 250-400ms, the significant effect of **Intonation×Area** was also found ($F(2,26)=15.677$, $p=0.001$, $\eta_p^2=0.547$). In parietal area, question has a larger negative component ($p=0.008$). In this time window, a significant interaction of **Tone×Area** was observed ($F(2,26)=4.385$, $p=0.049$, $\eta_p^2=0.252$). In parietal area, tone2 has a larger negative component ($p=0.033$). ANOVA in the time window of 450-550ms revealed a significant interaction of **Intonation×Tone** ($F(1,13)=7.184$, $p=0.019$, $\eta_p^2=0.356$). For stimuli ending in tone4, statements have a smaller positive component than questions ($p=0.005$). For stimuli ending in tone2, no significant difference in mean amplitudes was found between questions and statements ($p=0.649$). In the next time window 550-750ms, main effect of **Intonation** ($F(1,13)=8.916$, $p=0.011$, $\eta_p^2=0.407$) and **Tone** ($F(1,13)=21.517$, $p<0.001$, $\eta_p^2=0.623$) were found significant. Tone2 has a smaller positive component compared to tone4, and statements have a smaller positive component than questions. All the significant effects are shown in Fig. 6. Since most of the effects were found in the parietal area, mean amplitudes in the parietal area (an average of P3, Pz, P4 electrodes) and analysis of them are presented in this figure (and the Fig. 7).

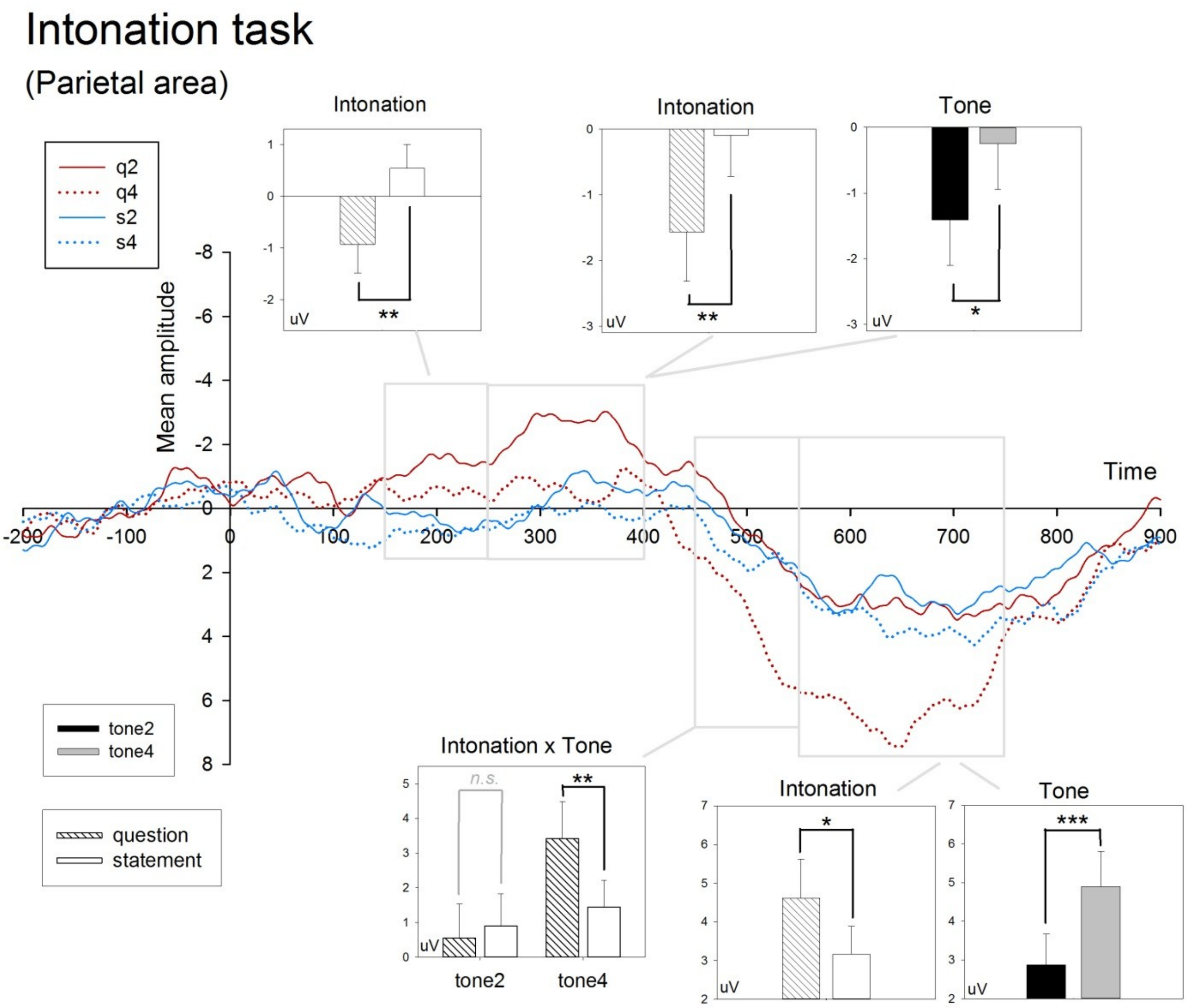


Fig. 6. The significant effects in the intonation task. *n.s.* means $p>0.05$; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$. Error bar indicates standard error.

In tone identification task, the overall ANOVA for the mean amplitude of the nine electrodes in the time window of 100–250ms revealed a significant three-way interaction of **Intonation×Tone×Area** ($F(2,26)=8.410$, $p=0.008$, $\eta_p^2=0.393$). For questions, in parietal area, tone4 has a larger negative component than tone2 ($p=0.017$). For statements, no significant differences between tones were found ($p>0.05$). No other interaction or main effects were found significant in ANOVA. The result is presented in Fig. 7.

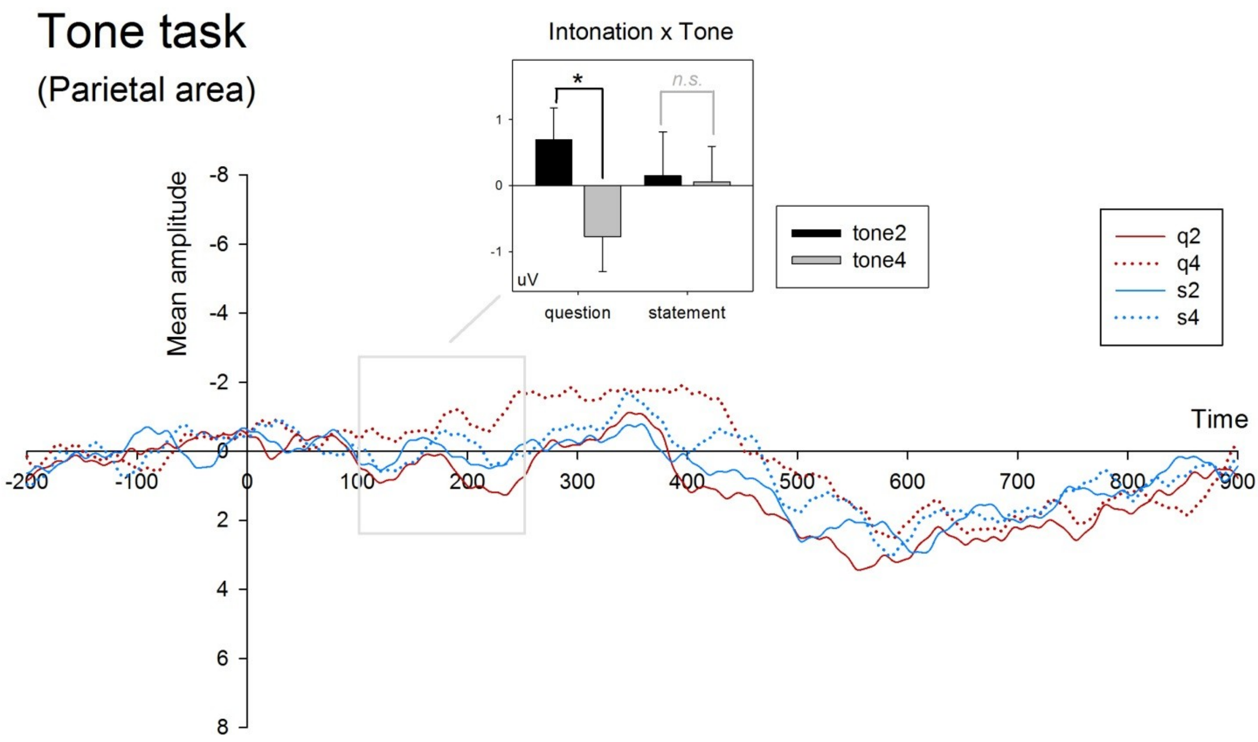


Fig. 7. The significant effects in the tone task. *n.s.* means *p*>0.05; *, *p*<0.05; **, *p*< 0.01; ***, *p*<0.001. Error bar indicates standard error.

Table 2

Significant ANOVA effects of the continuous 50ms time windows from 0 to 800ms.

Intonation identification task		Tone identification task	
Time window	**Significant effects**	**Time window**	**Significant effects**
0-50	×	0-50	×
50-100	×	50-100	×
100-150	×	100-150	Intonation x Tone x Area *
150-200	Intonation x Area ***	150-200	Intonation x Tone x Area **
200-250	Intonation x Area ***	200-250	Tone *
250-300	Intonation x Area ***	250-300	×
300-350	Intonation x Area **	300-350	×
	Tone *		
350-400	Intonation x Area *		×
400-450	×	400-450	Intonation x Tone*Area **
450-500	Intonation x Tone *	450-500	×
500-550	Intonation x Tone *	500-550	Tone x Area **
550-600	Tone x Area *	550-600	×
	Intonation x Area x Hemi **		

600-650	Intonation x Area x Hemi *** Intonation x Tone *	600-650	Tone x Area *
650-700	Tone x Area *	650-700	×
700-750	Tone x Area **	700-750	×
750-800	×	750-800	Intonation x Area x Hemi *

Significant effects that include the variables Intonation or Tone were presented. When there is an interaction contains a variable, the main effect of it would not be shown. *, p<0.05; **, p< 0.01; ***, p<0.001.

4. Discussion

The present study examined the temporal hierarchy of intonation and tone processing in Mandarin Chinese. We recorded and analyzed the participants' electrophysiological responses to Mandarin sentences that carried intonation and tone information in the same syllable in both an intonation identification task and a tone identification task. In the intonation task, according to the ERP results, the participants responded to intonation first (150–250ms). Tone effects appeared later after 250ms. In the tone task, participants were sensitive to both intonation and tone in an early interaction beginning from 100ms. For questions, stimuli with tone4 would evoke a larger negative component than tone2, while for statements, no difference in mean amplitudes was found between stimuli with different lexical tones. This interaction indicated our brain's different response to tone2 and tone4 was based on intonation, since for syllables with different intonation (question/statement), the response to lexical tones was different. If the task-dependent hypothesis holds true, the effect of tone alone should be detected at the first place in the tone identification task. However, intonation plays a role in lexical tone identification and participates in the processing of lexical tones in a very early time window. This result suggested that participants are very sensitive to intonation, and is more in line with the second hypothesis—intonation has higher temporal hierarchy than tone, having early effects on the auditory process of both intonation identification and tone identification.

The early emergence of intonation ERP effects in both intonation identification and tone identification task is in relation to the deep evolutionary root and cross-language nature of intonation (Hirst & Di Cristo, 1998; Naoi, Watanabe, Maekawa, & Hibiya, 2012), as well as the importance of its paralinguistic information to human social communication (Chun, 1988). People are sensitive to paralanguage such as pitch and intensity. They will match their interlocutor's pitch range (Chun, 1988) and intensity (Natale, 1975), and the extent to which people will match others' intensity was found positively related to their social desirability (Natale, 1975). These findings highlight intonation's fundamental social and interactional functions such as promoting cooperation between speakers (Chun, 1988), which helps explain why intonation effects emerge so early in participants' speech processing in current study.

Similar to the interaction in 100-250ms in tone identification task, an Intonation x Tone effect was also found in the intonation identification task, which revealed a larger positive component of questions compared to statements for stimuli ending in tone4 in the time window of 450-550ms. These ERP results could be caused by the conflict between a rising intonation and a falling tone, supported by plenty of previous research findings (e.g., Kung et al., 2014; Yuan, 2006; Yuan, 2011; Ren et al., 2009; Ren et al., 2013; Liu et al., 2016). The main effect of Intonation and Tone in the next time window 550-750ms, which indicated the larger positive components of question compared to statement and tone4 compared to tone2, might also be caused by the conflict between question intonation and tone4. The positive component found in this time window was regarded as a P600 component, which was known as the reflection of a reprocessing (Kolk, Chwilla, Van Herten, & Oor, 2003; Kolk & Chwilla, 2007) and also detected in previous research on the interaction of lexical tone and intonation (Kung et al., 2014). P600 reflects a process when participants reanalyzed and checked the sentence to resolve some conflicts or unexpected events that they detected (Kolk et al., 2003). In the present study, because of the strong conflict between tone4 and question, before making an identification choice, participants might have the need to reprocess the sentence they that have either question intonation or tone4 at the final syllable in order to resolve the potential conflict and avoid judgment errors. These two interaction effects in intonation and tone tasks were not found in behavioral results of identification rates, which could be due to the ceiling effect of the identification tasks. In tone task, the average tone identification rates of q2, q4, s2, s4 were 0.99, 0.97, 0.99 and 0.99, with Std. Deviation as 0.02, 0.05, 0.01, 0.01, respectively. In intonation task, although the identification rate of q2 was lower than other conditions (average accuracy= 0.78, Std. Deviation=0.17), identification rates of q4, s2, s4 were 0.90, 0.93 and 0.98 (Std. Deviation = 0.19, 0.06, 0.03, respectively). The tasks might be too simple for those native Mandarin speakers so that some effects could only be seen in ERP amplitudes. These findings of interaction between intonation and tone in ERP responses can be caused by both the bottom-up and top-down processing. As for the bottom-up processing related to the acoustic features of stimuli, F0 level has been found significant in the regression model for tone identification rate, and F0 interval was significant in the regression model for intonation task, but the two-way interaction was supported by the acoustic analysis of neither F0 level nor F0 interval. As shown in Fig.3, F0 level of tone4 was higher than tone2 in both questions and statements. So did the F0 interval. Thus, the interaction cannot be explained by the pure button-up processing of stimuli. Taking the top-down modulation into account, the significantly larger early negative component of q4 stimuli in tone identification task and the larger positive component of q4 stimuli in 450-550ms in intonation identification task could be explained by the huge conflict between question intonation and tone4. This conflict greatly changes the F0 contour of the tone4 by flattening the falling slope of it (see Fig. 1; also see Liu et al., 2016), making it harder to identify the lexical tone but making the question intonation easier to be detected.

Besides the interaction effect, significant main effects were also found in intonation task. In time window 150ms-250ms and 250ms-400ms, question has a larger negative component than statement in parietal area, which could be explained by the behavioral result that questions are harder to identify than statements. As Yuan suggested (Yuan, 2006), identification of statements is easier than questions perhaps because statement might be a default intonation while question intonation is a marked intonation type which requires some specific features to identify. Previous ERP research has found the link between the enhanced early negative component and the selective attention at an early stage of processing. The negative component in early time window such as 80-110ms (known as N1; Hillyard, Hink, Schwent, & Picton, 1973) and 100-250ms (Coch, Sanders, & Neville, 2005) was larger for the attended stimuli compared to unattended stimuli. In this study, participants might focus on questions, which are difficult to identify and needs specific features, instead of statements, and thus lead to questions' enhanced negative components. Tone2 with a larger negative component in parietal area than tone4 in 250-400ms can also be caused by the larger difficulty in intonation identification when the sentence ends with tone2 than tone4.

It is interesting to notice that most of our ERP results are found in the parietal area, which is consistent with the previous findings of intonation and tone effects in centro-parietal area (Kung et al., 2014; Liu et al., 2016). In Liu and her colleagues' study, the amplitude difference between questions and statements ending with tone4 was more significant at the posterior area than the central area (Liu et al., 2016). In previous fMRI (functional magnetic resonance imaging) studies on speech processing, a dorsal frontoparietal network of lexical tone processing was found (Li et al., 2003), and parietal lobe activation was observed when comparing active conditions (participants paid attention to intonation or lexical tones and made discrimination judgments) to passive-listening-to-speech condition (Gandour et al., 2003). The parietal area might play an important role in intonation and tone processing. A fMRI study (Iacoboni et al., 2004) has found increased activity in the medial parietal (precuneus) cortex when participants were watching movie clips of two people interacting with each other. Parietal cortex's response to social relationships might be related to its participation in intonation and tone identification in the current study.

To our knowledge, the present study provides the first ERP evidence that enables us to compare the access order of intonation and tone by our brain when detected in the same speech syllable. The findings of intonation's early effects on both intonation identification and tone identification process emphasize the role of both intonation and the social meanings carried by it to human beings. More research and various approaches are urged to examine the access order of intonation and tone in individuals with different language background and hearing experience. For instance, whether intonation precedes lexical tone in speech processing in other tonal languages such as Cantonese is worth investigating, and might further examine the influence of tonal experience on speech processing order. Besides, future research could look at participants who wear cochlear implants or tone-deaf individuals to explore whether pitch impairment would make people less sensitive to intonation cues and thus influence the access order of intonation and tone.

References

Araki, M., Bandi, M. M., & Yazaki-Sugiyama, Y. (2016). Mind the gap: Neural coding of species identity in birdsong prosody. *Science, 354,* 1282-1287.

Banse, R. & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology, 70,* 614.

Boersma, P., & Weenink, D. (2004). Praat [Computer Software], Version 4.3. *Amsterdam, Netherlands: Institute of Phonetic Sciences, University of Amsterdam*.

Brancucci, A., Lucci, G., Mazzatenta, A., & Tommasi, L. (2009). Asymmetries of the human social brain in the visual, auditory and chemical modalities. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364,* 895-914.

Chun, D. M. (1988). The neglected role of intonation in communicative competence and proficiency. *The modern language journal, 72*(3), 295-303.

Coch, D., Sanders, L. D., & Neville, H. J. (2005). An event-related potential study of selective auditory attention in children and adults. *Journal Of Cognitive Neuroscience, 17*(4), 605-622.

Connell, B. A., Hogan, J. T., & Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of phonetics*.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child development*, 1497-1510.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child development, 64,* 657-674.

Fernald, A. (1995). Human maternal vocalizations to infants as biologically relevant signals: an evolutionary perspective. In J.H.Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391). Oxford University Press, USA.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L. et al. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and language, 84,* 318-336.

Halle, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of phonetics, 32,* 395-421.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*(4108), 177-180.

Hirst, D. & Di Cristo, A. (1998). *Intonation systems : a survey of twenty languages*. Cambridge University Press.

Ho, A. T. (1977). Intonation Variation in a Mandarin Sentence for Three Expressions: Interrogative, Exclamatory and Declarative. *Phonetica*, *34*, 446-457.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage*, *21*(3), 1167-1173.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in cognitive sciences*, *3*, 323-328.

Kolk, H. & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and language*, *100*, 257-261.

Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and language*, *85*, 1-36.

Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: an ERP study on Cantonese Chinese. *Neuropsychologia*, *53*, 293.

Ladd, D. R. (1996). *Intonational Phonology*. (vols. 79) Cambridge University Press.

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

Liang, J. & van Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. *Tones and tunes*, *2*, 27-61.

Liu, C. & Rodriguez, A. (2012). Categorical perception of intonation contrasts: Effects of listeners' language background. *The Journal of the Acoustical Society of America*, *131*, EL427-EL433.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental psychology*, *43*, 912-917.

Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia*, *91*, 307-317.

Ma, J. K. Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America*, *120*, 3978-3987.

Ma, J. K. Y., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *Journal of the Acoustical Society of America*, *129*, 1012.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain Lang*, *89*(2), 277-289. doi: 10.1016/S0093-934X(03)00350-X

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain A Journal of Neurology*, *133*, 2635.

Nan, Y., & Friederici, A. D. (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. doi: 10.1002/hbm.22046

Naoi, N., Watanabe, S., Maekawa, K., & Hibiya, J. (2012). Prosody discrimination by songbirds (Padda oryzivora). *PLoS One*, *7*, e47446.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97-113.

Peng, S. C., Lu, N., & Chatterjee, M. (2009). Effects of Cooperating and Conflicting Cues on Speech

Intonation Recognition by Cochlear Implant Users and Normal Hearing Listeners. *Audiology and Neurotology*, *14*, 327-337.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58.

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

Ren, G. Q., Tang, Y. Y., Li, X. Q., & Sui, X. (2013). Pre-Attentive Processing of Mandarin Tone and Intonation: Evidence from Event-Related Potentials.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, *23*(6), 695-703.

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038-6043.

Spierings, M. J. & ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*, 20140480.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat Neurosci*, *2*(2), 191.

Thompson, W. F. & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica*, *2006*, 407-424.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*, 1063-1074.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *Journal of the Acoustical Society of America*, *130*, 4063-4069.

Yuan, J. (2006). Mechanisms of Question Intonation in Mandarin. In *Chinese spoken language processing* (pp. 19-30). Berlin, Heidelberg: Springer.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846-849.