Tone and intonation processing: from ambiguous acoustic signal to linguistic representation

Liu, M.

Chapter 2

Online processing of tone and intonation in Standard Chinese: Evidence from ERPs[2]

[2] A version of this chapter is published as: Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia, 91*, 307-317.

Abstract

Event-related potentials (ERPs) were used to investigate the online processing of tone and intonation in Standard Chinese at the attentive stage. We examined the behavioral and electrophysiological responses of native Standard Chinese listeners to Standard Chinese sentences, which contrast in final tones (rising Tone2 or falling Tone4) and intonations (Question or Statement). A clear P300 effect was observed for the question-statement contrast in sentences ending with Tone4, but no ERP effect was found for the question-statement contrast in sentences ending with Tone2. Our results provide ERP evidence for the interaction of tone and intonation in Standard Chinese, confirming the findings from behavioral metalinguistic data that native Standard Chinese listeners can distinguish between question intonation and statement intonation when the intonation is associated with a final Tone4, but fail to do so when the intonation is associated with a final Tone2. Our study extends the understanding of online processing of tone and intonation 1) from the pre-attentive stage to the attentive stage and 2) within a larger domain (i.e., multi-word utterances) than a single word utterance.

Keywords: Tone2, Tone4, intonation, Standard Chinese, attentive processing, P300

2.1 Introduction

In spoken language processing, different aspects of linguistic information are involved, such as lexical, semantic, syntactic and prosodic information (Friederici, 2002; Isel, Alter, & Friederici, 2005). Among these aspects, prosodic information, especially pitch information, has been shown to be indispensable for spoken language processing in tonal languages such as Standard Chinese (e.g., Li, Chen, & Yang, 2011). Tone and intonation have been considered the two most significant prosodic features of Standard Chinese speech (Tseng & Su, 2014). At the lexical level, F0 is employed to differentiate the four lexical tones (Tone1 - high-level, Tone2 - mid-rising, Tone3 - low-dipping and Tone4 - high-falling), which contrast lexical meanings (Cutler & Chen, 1997; Yip, 2002). At the sentential level, F0 is also used to convey post-lexical information, for example, intonation types (e.g., question intonation, statement intonation) (Ladd, 2008). Although other acoustic correlates (such as duration, intensity and phonation) have also been shown to contribute to cue tonal and intonational contrasts (Garellek, Keating, Esposito, & Kreiman, 2013; Hu, 1987; Shi, 1980; Xu, 2009; Yu & Lam, 2014), F0 has been identified as the primary acoustic correlate of both tone and intonation in Standard Chinese (Ho, 1977; Shen, 1985; Wu, 1982; Xu & Wang, 2001; Xu, 2004). It may therefore not be surprising that tone and intonation interact with each other both in production and perception.

The interaction of tone and intonation in Standard Chinese has aroused great interests among researchers, and several models or theories have been put forward based mainly on acoustic data since the first studies on the topic by Chao (Chao, 1929, 1933). Of these acoustic studies, a general belief is that question intonation has higher F0 than statement intonation (Cao, 2004; Gårding, 1987; Shen, 1989; Wu, 1996). However, there is controversy about the temporal scope of such higher F0 in question intonation. Two alternative views have been established. One holds that there is an overall F0 rising of sentences in questions compared to statements (Ho, 1977; Shen, 1989; Yuan, 2011). The other claims that the F0 difference between questions and statements is more pronounced towards the end of the sentences (Kratochvil, 1998; Liu & Xu, 2005; Xu, 2005; Peng et al., 2005).

Different from the above acoustic studies, Liang and Van Heuven (2007) conducted intonation perception experiments with a seven-syllable sentence containing merely high-level tone syllables. They manipulated both the overall pitch level of the sentence and the pitch level of the final tone. Results showed that manipulating the final rise has a much stronger effect on the perception of intonation type than manipulation of the overall pitch level, indicating that the F0 of the final tone is more important than that of the whole sentence for intonation perception.

Not unique to Standard Chinese, the final rise has been shown to be a language-universal perceptual cue for question intonation (Gussenhoven & Chen, 2000). In a made-up language, Gussenhoven and Chen (2000) tested the perceptual cues for question intonation across three different language groups. All listeners tended to take the higher peak, the later peak and the higher end rise as cues for question intonation perception. In Cantonese, another representative language other than Standard Chinese within the Sinitic language family, Ma, Ciocca, and Whitehill (2011) also found that the perception of questions and statements relies primarily on the F0 characteristics of the final syllables.

Apart from studies on the temporal domain of perceptual cues of intonation, there has also been research, though regretfully little, on the effect of intonation on tone perception and vice versa. Connell, Hogan, and Rozsypal (1983) ran a tone perception experiment in Standard Chinese and found that intonation-induced F0 has little effect on tone perception. Tone identity is maintained in question intonation. With regard to the effect of tone on intonation perception, Yuan (2011) found that in Standard Chinese, questions ending with Tone4 (falling tone) were easier to identify than questions ending with Tone2 (rising tone). Three mechanisms were proposed for question intonation: an overall higher phrase curve, higher strengths of sentence-final tones and a tone-dependent mechanism. The tone-dependent mechanism conflicts with the strength mechanism on the final Tone2, possibly accounting for the difficulty of question identification in sentences ending with Tone2. In sentences ending with Tone4, the tone-dependent mechanism flattens the falling slope of the

final falling, making question intonation perceptually more salient for falling tone (Yuan, 2006).

Unlike in Standard Chinese, the intonation-induced F0 affects tone perception in Cantonese. Low tones (21, 23, 22) (tone values in 5-point scale notation, each tone is described by the initial and the end point of the pitch level) were misperceived as the mid-rising tone (25) at the final positions of questions (Fok-Chan, 1974; Kung, Chwilla, & Schriefers, 2014; Ma et al., 2011). This is probably because with a rising tail superimposed on all tone contours by question intonation (Ma, Ciocca, & Whitehill, 2006), the F0 contour of the low tones in questions resembles that of a mid-rising tone in questions. As for the effect of tone on intonation perception, native listeners were least accurate of all the six tones in Cantonese in distinguishing statements and questions for sentences ending with Tone 25 (Ma et al., 2011), suggesting that listeners confused the rising contour of Tone 25 with the final rise of question intonation.

Taken together, potential conflicts exist between tone and intonation in Standard Chinese and Cantonese, causing processing difficulties at the behavioral level. However, the underlying neural mechanisms leading to the eventual behavioral decisions are not yet clear. To shed light on this issue, research is needed to investigate the online processing of tone and intonation.

In recent years, a number of neurophysiological studies in regard to pitch processing have emerged, mainly with lesion, dichotic listening and functional neuroimaging techniques (Gandour et al., 1992; Klein, Zatorre, Milner, & Zhao, 2001; Van Lancker & Fromkin, 1973; Wang, Sereno, Jongman, & Hirsch, 2003). However, due to the low temporal resolution of these techniques, event-related potentials (ERPs), a high temporal resolution measure was introduced to pitch processing, offering more precise temporal information of online processing.

The majority of ERP studies relevant to pitch processing focus on the neural mechanisms of tone processing at the pre-attentive stage, where participants are directed to watch a silent movie or read a book and to ignore the auditory input (Fritz, Elhilali, David, & Shamma, 2007). In these studies, the ERP component of interest is the Mismatch-Negativity (MMN), an indicator of acoustic change detection (Näätänen, 2001; Pulvermüller & Shtyrov, 2006). Only two studies

examined the online processing of both tone and intonation in Standard Chinese, to our knowledge. Ren, Yang, and Li (2009) constructed an oddball sequence. A word with lexical Tone4 (i.e., /gai4/[3]), uttered with statement intonation was presented as the standard stimulus, and /gai4/ with question intonation was presented as the deviant stimulus to native Standard Chinese listeners. Their results showed a clear MMN effect when subtracting the waveform of the standard from that of the deviant. In another study, Ren, Yang, Li, and Sui (2013) adopted a three-stimuli oddball paradigm. The standard stimulus was /lai2/ with statement intonation. The deviant stimuli included an intonation deviant (/lai2/ with question intonation) and a lexical tone deviant (/lai4/ with statement intonation). Results showed an MMN for the tone deviant but not for the intonation deviant. As the MMN is linked to higher order perceptual processes underlying stimulus discrimination (Pulvermüller & Shtyrov, 2006), the above two studies suggest that at the pre-attentive stage, native listeners can tease apart question intonation from statement intonation when the intonation is combined with Tone4, but they are not able to tease apart the two types of intonation when the intonation is combined with Tone2, just as what Yuan (2011) has reported with behavioral perceptual judgment data. This correspondence of the online MMN results with the offline behavioral results validates the initial ERP evidence of the interaction of tone and intonation in Standard Chinese.

In addition to Standard Chinese, ERP evidence of online interplay of tone and intonation is also revealed in Cantonese (Kung et al., 2014). In this study, Cantonese participants were asked to perform a lexical-identification task, i.e., choosing the right word they heard from six Cantonese words on the screen in the form of Chinese characters, and the six words were tonal sextuplets of the critical word. ERP analyses revealed a P600 effect for low tone in questions relative to low tone in statements. The P600 effect was explained as an indicator of reanalysis, in the presence of a strong conflict of two competing

[3] The number following the letters in Standard Chinese Pinyin represents Standard Chinese tone. "1" is Tone1 (high-level tone); "2" is Tone2 (mid-rising tone); "3" is Tone3 (low-dipping tone), and "4" is Tone4 (high-falling tone).

representations activated in questions ending with low tones. The two representations are a lexical representation with a low tone on the one hand and a lexical representation with a high rising tone on the other. Special attention should be paid to the fact that Kung et al. (2014) found a P600 effect in the semantically neutral sentence context. In their subsequent study, when introducing a highly constraining semantic context to the target words, the P600 disappeared, suggesting that semantic context plays a role in resolving the online conflict between intonation and tone.

Several remaining issues may be noticed given the above ERP studies on the processing of tone and intonation. First, the MMN studies of Standard Chinese restricted their attention to the interaction of tone and intonation in a one-syllable-sentence domain. Given that intonation is a feature at the sentential level that typically spans over several lexical items, it would be not only interesting but also necessary to investigate how tone and intonation information are processed when the length of an utterance is extended from one syllable to several syllables. Specifically, the question that arises here is whether native Standard Chinese listeners can disentangle intonation information from tone information over a broader sentence domain. Second, the extant ERP studies (Ren et al., 2009, 2013) investigated tone and intonation processing at the pre-attentive stage where the attention of the participants was directed to elsewhere. In this kind of design, there is no way to measure the behavioral effects of the deviant acoustic stimulus in the stream, making it impossible to distinguish between automatic neural responses arising from acoustic variability and responses related to "attention capture" (Fritz et al., 2007). Therefore, the present study aimed to extend the online processing of tone and intonation from the pre-attentive stage to the attentive processing stage. Moreover, we are interested in whether the processing of tone and intonation at the attentive stage differs from that at the pre-attentive stage. Third, the ERP study on Cantonese (Kung et al., 2014) extended the online processing of tone and intonation to a broader sentence domain. However, this study employed a lexical identification task rather than a pitch identification task *per se*. In this way, a potential concern is that the interaction of tone and intonation is not directly examined. Also, as stated earlier, intonation distorts

tone identity in Cantonese but not in Standard Chinese. There arises the question of whether the mechanisms underlying tone and intonation processing in Standard Chinese are different from that in Cantonese. Fourth, semantic context affects the processing of tone and intonation in Cantonese. It has also been proven that in Standard Chinese a constraining semantic context facilitates the processing of tone (Ye & Connine, 1999) and intonation (Liu, Chen, & Schiller, 2016a). In this study, we therefore took semantic context as a control variable and set it to be neutral so that it can serve as a baseline comparison for further research. In short, the present study was designed to investigate the online processing of tone and intonation in Standard Chinese over a broader sentence domain at the attentive stage under neutral semantic context.

The ERP component that is of our particular interest in the present study is the P300 (the P3b in particular). The P300 is a positive-going deflection peaking at around 300 ms in a time window of about 250 to 500 ms, or even to 900 ms (Patel & Azzam, 2005). It is thought to be elicited in the process of decision making (Hillyard, Hink, Schwent, & Picton, 1973; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Rohrbaugh, Donchin, & Eriksen, 1974; Smith, Donchin, Cohen, & Starr, 1970; Verleger, Jaśkowski, & Wascher, 2005), reflective of processes involved in stimulus evaluation or categorization (Azizian, Freitas, Watson, & Squires, 2006; Frenck-Mestre et al., 2005; Johnson & Donchin, 1980; Kutas, McCarthy, & Donchin, 1977).

In the present study, we examined participants' behavioral and electrophysiological responses to Standard Chinese sentences contrasting in final tones (Tone2 or Tone4) and intonations (Question or Statement). We employed two pitch identification tasks. Participants were asked to categorize the final tone or intonation of the stimuli. A two-alternative forced choice (2AFC) task was adopted in this study. 2AFC is considered a highly simplified decision making condition, in which a choice must be made between two responses based on limited information about which is correct (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). It best captures the essence of decision-making, and therefore, fits for the purpose of eliciting P300. To decrease fast guesses in the task, a 0.5/0.5 probability of each category was implemented, as in Pfefferbaum, Ford, Johnson, Wenegrat, and Kopell (1983). The same 2AFC

task was performed on all the stimuli, not just on one specific category to avoid selective tuning, which has been proved to be unnecessary and insufficient for P300 enhancement (Hillyard et al., 1973; Rohrbaugh et al., 1974).

We hypothesized that under neutral semantic context, at the attentive stage, native Standard Chinese listeners should be able to disentangle question intonation from statement intonation when the intonation concurs with a final Tone4. Behaviorally, this should be reflected in high identification accuracy. Electrophysiologically, we expect a P300 effect for questions ending with Tone4 relative to statements ending with Tone4. In the case of Tone2, due to the difficulty in teasing apart intonation information from tone information for participants, the behavioral performance is expected to show a lower accuracy. No clear P300 is expected between questions ending with Tone2 and statements ending with Tone2.

2.2 Method

2.2.1 Participants

Twenty right-handed native speakers of Standard Chinese from Northern China were paid to participate in the experiment. They were undergraduate or graduate students at Renmin University. Five of the participants were excluded from the analysis because of excessive artifacts in their EEG data. Age of the remaining 15 participants (7 male, 8 female) ranged from 20 to 28 ($M \pm SD$: 23.8 ± 2.8). None of them had received any formal musical training or had reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

2.2.2 Materials

Forty monosyllabic minimal word pairs varying by tone (Tone2 or Tone4) with otherwise identical segments were selected. Each minimal Tone2_Tone4 word pair contains words of comparable word frequency, homophone density, and syntactic word category. To avoid any word frequency effect, only frequent words with more than 4,500 occurrences in a corpus of 193 million words were used (Da, 2004). Following Chen, Vaid, and Wu (2009), homophone density

was defined as the number of homophone mates of a word, i.e., words that contain exactly the same phonetic segments and lexical tones. Tone2 words have similar homophone densities as their Tone4 equivalents. The forty word pairs comprise mainly pairs of nouns (32 pairs), but pairs of verbs (6 pairs) and adjectives (2 pairs) were also included to guarantee sufficient number of stimuli.

All the critical words were embedded in the final position of a five-syllable carrier sentence, i.e., *ta1 gang1gang1 shuo1 X* (English: She just said X), produced with either a statement or a question intonation. Only high-level tones (Tone1) were contained in the carrier sentence. This is to avoid downstep effect and to minimize the contribution of tone to the observed F0 movement (Shih, 2000). The carrier sentence was semantically meaningful but offered neutral semantic context to the target stimuli. By using the semantically neutral carrier, intonation information was successfully elicited. On the other hand, potential confound of semantic context with sentence prosody was excluded (Kung et al., 2014).

In total, 160 target sentences (40 Syllables × 2 Tones × 2 Intonations) were designed (see Table 1 for an example). Moreover, another 240 filler sentences were constructed. They resembled the target sentences in carrier sentence structure but differed from them in critical syllables in terms of either segmental composition or lexical tone (e.g., Tone1/Tone3). Pooling target sentences and filler sentences resulted in 400 sentences, which were uttered for the perception experiment.

Table 1. *An example of the experimental design.*

Condition		Example				
Tone	Intonation					
Tone2	Statement	Characters	她	刚刚	说	**X (财)。**
		Pinyin	ta1	gang1gang1	shuo1	**cai2**
		IPA	[tʰA1]	[kɑŋ1 kɑŋ1]	[ʂuo1]	**[tsʰai2]**
		English	She	just	said	**money.**
Tone2	Question	Characters	她	刚刚	说	**X (财)?**
		Pinyin	ta1	gang1gang1	shuo1	**cai2**
		IPA	[tʰA1]	[kɑŋ1 kɑŋ1]	[ʂuo1]	**[tsʰai2]**
		English	She	just	said	**money?**
Tone4	Statement	Characters	她	刚刚	说	**X (菜)。**
		Pinyin	ta1	gang1gang1	shuo1	**cai4**
		IPA	[tʰA1]	[kɑŋ1 kɑŋ1]	[ʂuo1]	**[tsʰai4]**
		English	She	just	said	**vegetable.**
Tone4	Question	Characters	她	刚刚	说	**X (菜)?**
		Pinyin	ta1	gang1gang1	shuo1	**cai4**
		IPA	[tʰA1]	[kɑŋ1 kɑŋ1]	[ʂuo1]	**[tsʰai4]**
		English	She	just	said	**vegetable?**

Note. The critical syllables are in bold.

2.2.3 Recording and stimuli preparation

One female native speaker of Standard Chinese, who was born and raised in Beijing, recorded the sentences. The recordings took place in a soundproof recording booth at the Phonetics Lab of Leiden University. Sentences were randomly presented to the speaker and recorded at 16-bit resolution and a sampling rate of 44.1 kHz. To eliminate paralinguistic information, the speaker was instructed to avoid any exaggerated emotional prosody during the recording.

This female speaker's recordings were chosen for the experiment for the clarity and consistency of the articulation. More importantly, the acoustic results of her recordings (see Figures 1 and 2) showed comparable F0 realization of tone and intonation to a prior study (Yuan, 2006) and were

therefore taken as the prototypical patterns for the perception study. In the subsequent perception experiment, the amplitude of all the sentences was normalized in Praat (Boersma & Weenink, 2015).

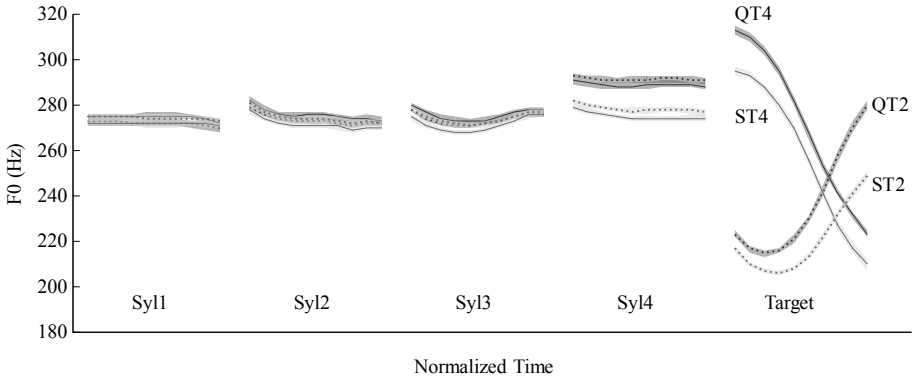

Figure 1. *F0 contours of the four experimental conditions. Each experimental condition is a combination of the levels of the factors Tone (Tone2, Tone4) and Intonation (Question, Statement), for example, QT2 refers to questions ending with Tone2. Syl1 to Syl4 are the carrier syllables, whereas Syl5 is the critical syllable. Dark solid lines indicate the mean F0 contours of QT4 (dark grey areas for ±1 SD of mean), and light solid lines indicate the mean F0 contours of ST4 (light grey areas for ±1 SD of mean). The corresponding dark dotted lines and light dotted lines indicate the mean F0 contours of QT2 and ST2, respectively.*

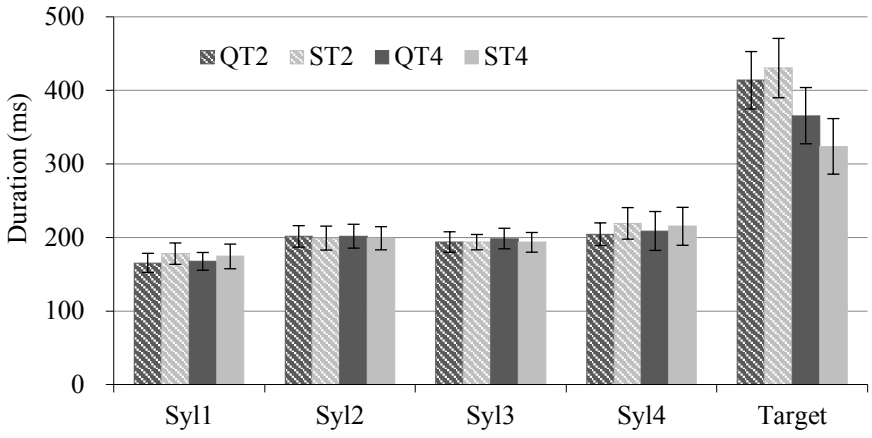

Figure 2. *Duration means ± 1 SD for each syllable of the four experimental conditions.*

The acoustic results (see Table 2) of the speech files showed that over our target stimuli pairs (which are five syllables long), the F0 and duration of the first three syllables revealed no differences between the two types of intonation for sentences ending with both Tone2 and Tone4 (all *p*s > .05). However, intonation affected the fourth syllable with a significantly raised F0 for the level tone over a question in comparison with that over a statement in both the Tone2 ($t(39) = 7.22$, $p < .001$, Cohen's $d = 1.64$) and the Tone4 ($t(39) = 7.17$, $p < .001$, Cohen's $d = 1.64$) conditions. When it comes to the critical syllable (i.e., the fifth syllable of the utterance), question with a final Tone2 had a significantly wider F0 range than its statement counterpart ($t(39) = 9.87$, $p < .001$, Cohen's $d = 1.90$). The minimum F0 showed a slight increase ($t(39) = 6.37$, $p < .001$, Cohen's $d = 1.27$), while a sharp rise of the maximum F0 was observed in the question condition relative to the statement condition ($t(39) = 11.47$, $p < .001$, Cohen's $d = 2.30$), suggesting a trend of sharper final rising. For questions with a final Tone4, there was an overall higher F0 contour than statements with a final Tone4. Though the F0 range was comparable between the two conditions ($t(39) = 1.24$, $p = .22$, Cohen's $d = 0.30$), the initial F0 ($t(39) = 8.65$, $p < .001$, Cohen's $d = 1.40$) and the final F0 ($t(39) = 4.48$, $p < .001$, Cohen's $d = 0.80$) of the high-falling tone were significantly higher in the question intonation (both *p*s < .001). Consistent with previous findings (Cao, 2004; Wu, 1996), our data showed that the pitch contour of Tone2 and Tone4 is maintained in both question and statement intonations, but pitch height differs between the two intonations across final tone identities. In addition to F0, duration also played a role in making a distinction between question intonation and statement intonation. Consistent with Yuan (2006), final Tone4 syllables in question intonation had significantly longer duration than those in statement intonation ($t(39) = 6.73$, $p < .001$, Cohen's $d = 1.11$). However, final Tone2 syllables tended to be slightly shorter in question intonation than in statement intonation ($t(39) = -3.10$, $p < .01$, Cohen's $d = -0.41$), whereas similar duration was found between the two intonation types for final Tone2 syllables in Yuan (2006). Considering the temporal scope of the interaction of tone and intonation, our data lend no support to the global or to the strictly local theory of question intonation, but are sympathetic towards the final-rising

theory given that F0 did not start to increase significantly until the pre-final syllable.

Table 2. *Acoustic properties of the experimental materials (SDs in parentheses).*

Parameter	Syllable	Tone2, Statement	Tone2, Question	Tone4, Statement	Tone4, Question
Duration (ms)	Syl1	178 (15)	165 (13)	174 (17)	167 (12)
	Syl2	199 (16)	201 (15)	199 (16)	202 (16)
	Syl3	194 (10)	194 (14)	193 (13)	199 (14)
	Syl4	219 (21)	204 (15)	215 (26)	209 (26)
	Syl5	430 (40)	414 (39)	324 (38)	366 (38)
Mean F0 (Hz)	Syl1	274 (11)	272 (10)	272 (9)	275 (8)
	Syl2	274 (9)	274 (8)	272 (7)	276 (8)
	Syl3	274 (8)	274 (7)	272 (8)	276 (8)
	Syl4	278 (7)	292 (9)	275 (8)	289 (9)
Max F0 (Hz)	Syl5	250 (12)	280 (14)	296 (10)	313 (13)
Min F0 (Hz)	Syl5	205 (6)	214 (8)	210 (16)	223 (11)
F0 range (Hz)	Syl5	45 (10)	66 (12)	86 (20)	90 (11)

2.2.4 Task

Participants were asked to perform two pitch identification tasks. The tone identification task consisted of half of the trials (1 sentence in a trial, 400 sentences in total), whereas the intonation identification task consisted of the other half. The specific task was randomly allocated from trial to trial. Task types were indicated by tone and intonation marks in Standard Chinese Pinyin system. When " ˊ ˋ " marks (" ˊ " stands for Tone2; " ˋ " stands for Tone4) appeared on the screen, participants were asked to identify the final tone of the sentence they heard. When "。？" marks appeared ("。" stands for statement intonation; "？" stands for question intonation), they were asked to identify whether the previously presented sentence bears a statement intonation or a question intonation. To complete the tasks, participants were requested to press the corresponding button within a two-second time limit. The tone and intonation marks used here are acquired at a very early age by native Standard

Chinese speakers. No participants had reported difficulty in understanding the tasks.

2.2.5 Procedure

Participants were tested individually in a soundproof booth. Stimuli were randomly presented using E-Prime 2.0 software through loudspeakers at a comfortable listening level of a 75 dB sound pressure level at source. Instructions were given to participants visually on screen and orally by the experimenter in Standard Chinese before the experiment.

The whole experiment included one practice block and four experiment blocks. The practice block contained 12 trials. Each experiment block encompassed 100 trials. Between each block, there was a 3-minute break. An experiment trial started with a 100 ms warning beep, followed by a 300 ms pause. After that an auditory sentence was presented while a red fixation cross appeared on the screen. Participants were instructed to gaze on the fixation cross and not to blink or move during the presentation of the sentence. In the meantime, participants were instructed to pay special attention to the final tone and the intonation of the sentence. One second after the offset of the stimuli, they were asked to perform either a tone identification task or an intonation identification task as accurately as possible within a two-second time limit. By doing so, the ERP effects of interest can be prevented from being confounded by motor-related processes (Kung et al., 2014; Salisbury, Rutherford, Shenton, & McCarley, 2001). The Inter Stimulus Interval (ISI) was 500 ms.

2.2.6 EEG data recording

EEG was recorded from 64 Ag/AgCI electrodes mounted in an elastic cap (Neuroscan system) with a sampling rate of 500 Hz. The right mastoid served as the reference electrode and AFz as the ground. Electrooculograms (EOGs) were monitored vertically and horizontally. Vertical EOGs were recorded by electrodes placed above and below the left eye, while horizontal EOGs were recorded by electrodes placed at the outer canthi of the left and right eye. The impedance of all electrodes was kept below 5KΩ.

2.2.7 Behavioral data analysis

Given that the behavioral responses were performed one second after the presentation of the stimuli, these delayed reaction time measurements were not further analyzed. Only Identification Rate (IR) was analyzed. IR was defined as the percentage of correct identification of tone in the tone identification task, and as the percentage of correct identification of intonation in the intonation identification task.

Statistical analyses were carried out with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.1.2 (R Core Team, 2015). Binomial logistic regression models were constructed for the dependent variable Response (Correct or Incorrect) with Task, Tone, Intonation and their interactions as fixed factors, and Subjects and Items as random factors. The fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. To capture the binary nature of the dependent variable, a logistic link function was applied. The estimate (β), z and p-values are reported.

2.2.8 EEG data analysis

The EEG data were analyzed with Brain Version Analyzer (Version 2.0). A 0.05-20 Hz band-pass filter was applied offline to the original EEG data. ERP epochs were defined in a 1,200 ms interval from −400 ms to 800 ms time-locked to the onset of the critical word. The baseline was calculated from −400 ms to −200 ms. In our acoustic data, F0 differences among the experimental conditions have been observed from the pre-final syllable. We therefore defined the time interval before the pre-final syllable as the baseline. Epochs with excessive eye movements and blinks were discarded. The criteria for artifact rejection were a maximal sudden voltage change of 25 μV in 100 ms, a maximal amplitude difference of 100 μV in a time window of 200 ms and a low amplitude activity within a range of 0.5 μV in a time window of 100 ms.

Prior to averaging, trials with incorrect behavioral responses were excluded. However, ERP data were collapsed across task types during averaging because on the one hand, the tone identification task evoked very few response errors,

and on the other hand, we did not observe differences in the ERP waveforms between the tone identification task and the intonation identification task under each experimental condition. To gain more statistical power, we aggregated all the correctly identified artifact-free trials from the tone identification task and the intonation identification task in the ERP analyses. As a result, a total of 30% of the data points were rejected. We found a clear peak in only one of the experimental conditions. Thus, in the following, we will exclusively report mean amplitudes.

A set of 27 electrodes was used for analyses, including 3 midline electrodes (Fz, Cz and Pz) and 24 lateral electrodes (F3/4, F1/2, FC3/4, FC1/2, C3/4, C1/2, CP3/4, CP1/2, P3/4, P1/2, PO5/6, PO3/4). The lateral electrodes were divided into six areas comprising four electrodes each (see Figure 3). These six areas were Left Frontal (F3, F1, FC3, FC1), Right Frontal (F2, F4, FC2, FC4), Left Central (C3, C1, CP3, CP1), Right Central (C2, C4, CP2, CP4), Left Posterior (P3, P1, PO5, PO3) and Right Posterior (P2, P4, PO4, PO6). For each area, the mean amplitude of the four electrodes was calculated and used in the following analyses. Due to the different numbers of the midline electrodes and the lateral electrodes, we decided to run repeated measures ANOVAs on the midline electrodes and the lateral electrodes separately.

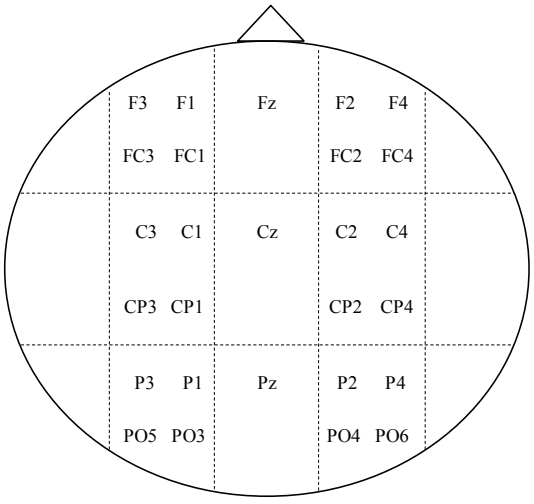

Figure 3. *Electrode areas used in the analyses. For the lateral electrodes, the amplitude of the four electrodes within each area was averaged.*

To establish the exact onset and range of the ERP effects, we ran repeated measures ANOVAs for 16 successive 50 ms time windows starting from the onset of the critical word up to 800 ms (following Schirmer, Tang, Penney, Gunter, & Chen, 2005). For the midline electrodes, within-subject variables included Tone (Tone2, Tone4), Intonation (Question, Statement) and Region (Frontal, Central, Posterior). For the lateral electrodes, within-subject variables included Tone (Tone2, Tone4), Intonation (Question, Statement), Region (Frontal, Central, Posterior) and Hemisphere (Left, Right). Statistical significance was computed using the Greenhouse-Geisser correction when the assumption of sphericity was violated. Corrected *p*-values are reported.

2.3 Results

2.3.1 Behavioral results

Figure 4 presents the identification rate of the four experimental conditions under different tasks (see also Table A1 in Appendix A for details). Tone stands for the tone identification task, and Intonation stands for the intonation identification task.

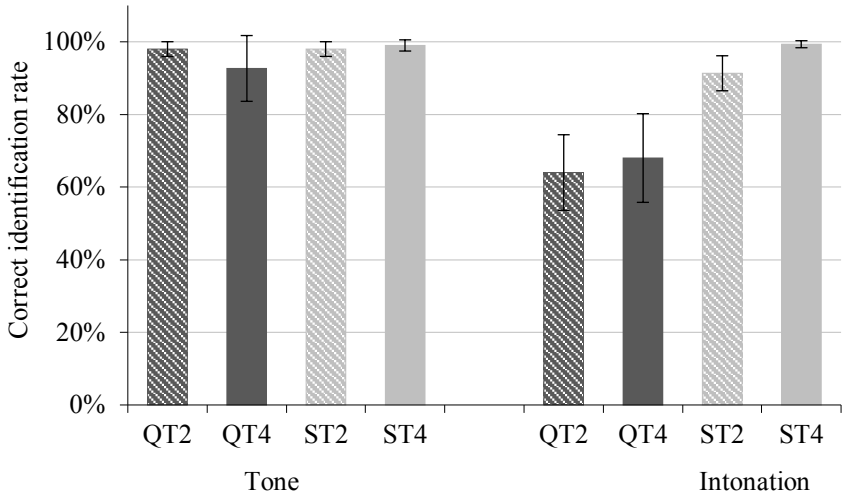

Figure 4. *Identification rate for each experimental condition under different tasks. Tone indicates the tone identification task; Intonation indicates the intonation identification task.*

Results showed a significant main effect of Task (β = 0.69, z = 2.61, p < .01) and Intonation (β = 1.99, z = 5.45, p < .01) on the odds of correct responses over incorrect responses. Two-way interactions, i.e., Task × Tone, Task × Intonation, and Tone × Intonation, also reached significance (all ps < .01). There was no three-way interaction of Task, Tone and Intonation (p > .1). Separate models for subset data of Tone and Intonation revealed that the effects of Task were manifested in that the tone identification task had much higher identification rate than the intonation identification task in questions ending with Tone2 and Tone4, and also in statements ending with Tone2 (all ps < .05). Due to the near-ceiling level of identification performances in both tasks, no task difference was observed for statements ending with Tone4 (p > .05). Apparently, the tone identification task was much easier than the intonation identification task for the participants.

Separate models were also constructed for subset data of different tasks. For the tone identification task, a significant interaction of Tone × Intonation (β = 2.17, z = 2.50, p < .05) was found. Identification rate of Tone4 was lower than that of Tone2 in questions (β = −1.54, z = −2.95, p < .01), but no difference was found between the two in statements (β = 0.71, z = 1.00, p > .05). No intonation effect was found in either Tone2 or Tone4 sentence pairs (both ps > .05). Overall, tone identification almost reached ceiling level across conditions. This suggests that the identity of tone was not hindered by intonation information.

With respect to the intonation identification task, a significant main effect of Intonation (β = 2.06, z = 5.14, p < .01) and an interaction of Tone × Intonation were discovered (β = 2.61, z = 2.95, p < .01). Question intonation had a much lower identification rate than statement intonation regardless of the final tone identities (both ps < .01). Despite the fact that no difference was found for IR of question intonation between questions ending with Tone2 and questions ending with Tone4 (β = 0.20, z = 0.51, p > .05), a higher identification rate of statement intonation was indeed discovered for statements ending with Tone4 relative to statements ending with Tone2 (β = 2.68, z = 3.61, p < .01). This suggests that participants had more difficulties perceiving question intonation than statement intonation.

2.3.2 ERP results

Figure 5 shows the grand average ERP waveforms time-locked to the onset of the critical syllables for 9 electrodes. Figure 6 presents the topographic maps obtained in all the 64 electrodes. Since the focus of interest of this paper is tone and intonation effects, only tone-related and intonation-related effects and the corresponding interactions are discussed below.

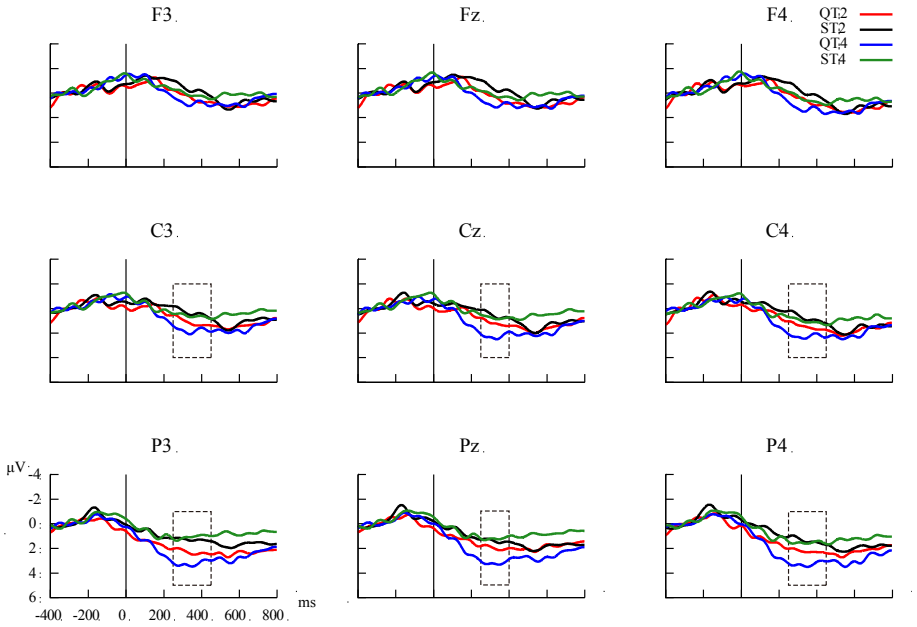

Figure 5. *Grand average waveforms time-locked to the onset of the critical syllables with a baseline from −400 ms to −200 ms for nine representative electrodes. Negativity is plotted upwards. The boxes with dash lines mark the P300 time-window for the questions ending with Tone4 condition.*

A summary of the time-course analyses for the midline electrodes and the lateral electrodes are presented in Table A2 and Table A3 (see Appendix A), respectively. Regions of Interest (ROIs) were identified as the time period when effects were consistently significant in two or more consecutive 50 ms time windows. Visual inspection of the waveforms also served as a complementary tool for the identification of ROIs. Consequently, we chose a ROI of 250-400

ms for the midline electrodes. For the lateral electrodes, a larger time window of 250-450 ms was identified as the ROI.

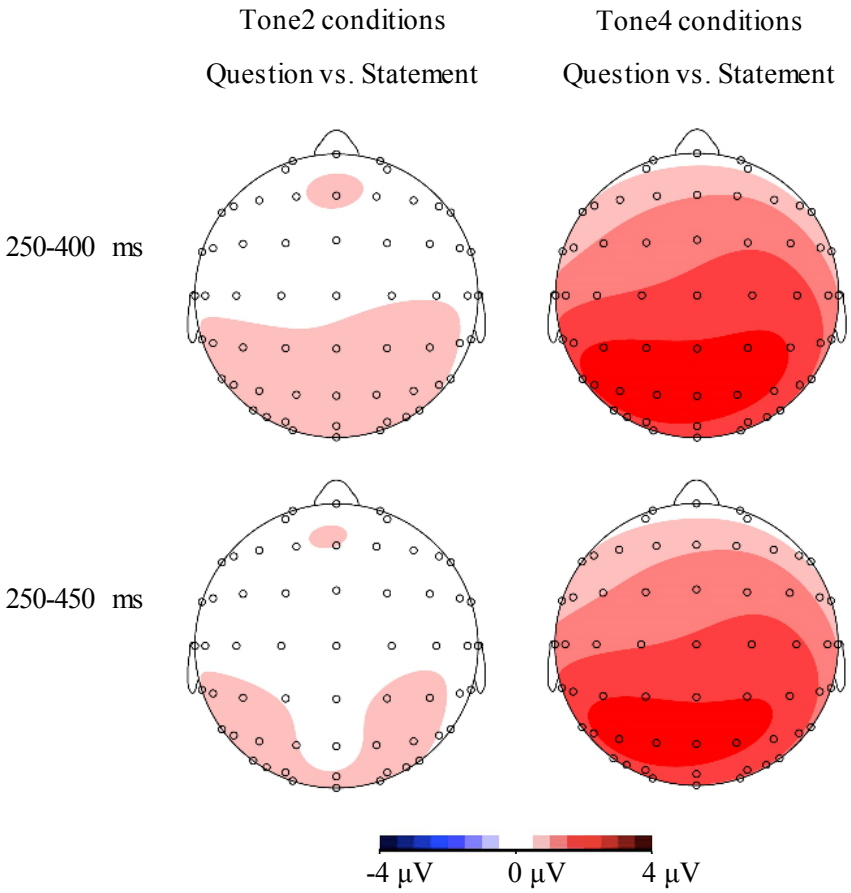

Figure 6. *Topographic maps obtained from all 64 electrodes. The maps were calculated by subtracting the waveforms in statements from those in questions for the Tone2 conditions (the left column) and the Tone4 conditions (the right column), respectively. The upper row shows the topographic maps in a time window of 250-400 ms, where the midline electrodes show P300 effect. The bottom row shows the topographic maps in a larger time window of 250-450 ms, where the lateral electrodes show P300 effect.*

The overall ANOVA for the mean amplitude of the midline electrodes in the time window of 250-400 ms revealed a main effect of Intonation ($F(1, 14)$

= 8.89, $p < .05$, $\eta_p^2 = .39$), and a three-way interaction of Tone × Intonation × Region ($F(1.56, 21.79) = 4.55$, $p < .05$, $\eta_p^2 = .25$). Follow-up ANOVAs were then performed for each level of Tone. Comparisons between QT2 and ST2 revealed neither a main effect of Intonation nor an interaction of Intonation × Region (both ps > .05). However, the analysis comparing QT4 and ST4 yielded a significant main effect of Intonation ($F(1, 14) = 6.81$, $p < .05$, $\eta_p^2 = .33$) and a significant interaction of Intonation × Region ($F(1.85, 25.83) = 9.05$, $p < .01$, $\eta_p^2 = .39$). Separate ANOVAs for each level of Region revealed a significant main effect of Intonation at the central ($F(1, 14) = 6.55$, $p < .05$, $\eta_p^2 = .32$) and posterior sites ($F(1, 14) = 13.49$, $p < .01$, $\eta_p^2 = .49$), with a larger positivity for QT4 than for ST4. No effect of Intonation was found at the frontal sites ($p > .05$).

As for the lateral electrodes, the overall ANOVA for the mean amplitude in the time window of 250-450 ms revealed a main effect of Intonation ($F(1, 14) = 10.18$, $p < .01$, $\eta_p^2 = .42$), a three-way interaction of Tone × Intonation × Region ($F(1.44, 20.08) = 4.87$, $p < .05$, $\eta_p^2 = .26$), and also a three-way interaction of Intonation × Region × Hemisphere ($F(1.56, 21.88) = 3.84$, $p < .05$, $\eta_p^2 = .22$). Follow-up ANOVAs for each level of Tone yielded no effects between QT2 and ST2 (all ps > .05), but a significant main effect of Intonation ($F(1, 14) = 4.78$, $p < .05$, $\eta_p^2 = .25$), a significant two-way interaction of Intonation × Region ($F(1.95, 27.27) = 12.08$, $p < .01$, $\eta_p^2 = .46$) and a significant three-way interaction of Intonation × Region × Hemisphere ($F(1.60, 22.35) = 8.09$, $p < .05$, $\eta_p^2 = .37$) between QT4 and ST4. Subsequent separate ANOVAs for each level of Region between QT4 and ST4 showed a significant main effect of Intonation at the central ($F(1, 14) = 4.66$, $p < .05$, $\eta_p^2 = .25$) and posterior sites ($F(1, 14) = 12.31$, $p < .01$, $\eta_p^2 = .47$), with QT4 eliciting more positivity than ST4 at these regions. Despite the statistical insignificance ($p > .05$), it is worth emphasizing that the amplitude difference between QT4 and ST4 was more prominent at the posterior sites than at the central sites. At the frontal sites, however, no effect of Intonation was found ($p > .05$).

In sum, the results of ANOVAs did not reveal any effect for QT2 versus ST2. In contrast, an ERP effect was observed for QT4 versus ST4. The ERP effect took place in different time windows for the midline electrodes and the

lateral electrodes, with a central-posterior distribution from 250-400 ms for the midline electrodes and a central-posterior distribution from 250-450 ms for the lateral electrodes. Through visual inspection of the waveforms, we identified a positive-going waveform peaking at about 300 ms after the onset of the critical word in the QT4 condition versus the ST4 condition in both the midline electrodes and the lateral electrodes. Taking together the polarity and the topographical distribution of the effect, we conclude that a P300 effect was found for QT4 versus ST4, whereas no effect was present for QT2 versus ST2.

2.4 General discussion

The present study investigated the online processing of tone and intonation in Standard Chinese at the attentive stage. We examined the behavioral and electrophysiological responses of native Standard Chinese listeners to Standard Chinese sentences, which contrast in final tones (Tone2 or Tone4) and intonations (Question or Statement). The context of these sentences was manipulated to be semantically neutral. Our behavioral results showed that while the identification of tone was not hindered by intonation, the identification of intonation was greatly impeded by tone. In the Tone4 conditions, question intonation was rather difficult to be correctly identified, whereas identification of statement intonation almost showed no difficulty at all. In the Tone2 conditions, question intonation was still difficult to identify, while identification of statement intonation also tended to be problematic. Regarding ERP results, we found a clear P300 effect for questions ending with Tone4 relative to statements ending with Tone4. No ERP difference was found between questions ending with Tone2 and statements ending with Tone2.

According to previous studies, P300 reflects neurophysiological mechanisms of decision-making and categorical processing (Azizian et al., 2006; Courchesne E., Hillyard, & Courchesne R., 1977; Kotchoubey & Lang, 2001; Kutas et al., 1977). When the categorization becomes more difficult, P300 amplitudes become smaller (Polich, 2007; Verleger, Gasser, & Möcks, 1985). The P300 requires attention. Previous studies also suggested that P300 amplitude is larger when participants devoted more effort to a task (Johnson, 1984, 1986; Isreal, Chesney, Wickens, & Donchin, 1980). Intuitively, one may expect that an

increase in task difficulty leads to investment of more effort and should thus elicit large P3 amplitude (Kok, 2001). However, the P300 amplitude decreases when tasks become perceptually or cognitively more demanding (Luck, 2005). Therefore, our ERP results above suggest that at the attentive processing stage, the question-statement contrast in Tone4 conditions is easier to categorize, whereas categorization of the question-statement contrast in the Tone2 conditions is much more demanding for native Standard Chinese listeners. These results are highly consistent with the MMN studies examining the online processing of tone and intonation in Standard Chinese at the pre-attentive stage. In those two studies (Ren et al., 2009, 2013), listeners are able to perceive the difference between question and statement intonation when the final tone is Tone4 (reflected in an MMN effect), but they cannot make a distinction between question and statement intonation when the final tone is Tone2 (reflected in no MMN). The MMN studies used one-syllable sentences, while our study extended the length of the utterances from one syllable to five syllables. Results in our study seem to confirm that the online processing patterns of tone and intonation in Standard Chinese are maintained from the pre-attentive stage to the attentive stage over a longer utterance.

Though consensus has been reached that question intonation can be distinguished online from statement intonation when the sentences end with Tone4, one may question why the P300 effect is elicited in questions ending with Tone4 compared to statements ending with Tone4, rather than vice versa. In the above MMN studies (Ren et al., 2009, 2013) and many other P300 studies, a target and a non-target are preset in the design. Very often participants just respond to the target stimuli. The corresponding ERP effect is therefore detected in the target category relative to the non-target category. In our study, no target category is set in advance, and neither is it the case that participants responded only to one category. With equal probability of the two intonation types for Tone4, we also did not expect any bias due to selective tuning (Hillyard et al., 1973; Rohrbaugh et al., 1974). All in all, the two conditions hold equal possibility of eliciting P300 effect in principle. So what makes question ending with Tone4 outperform its statement counterpart? Evidence from the resource framework has shown that P300 amplitude is

sensitive to the allocation of resources. P300 amplitude is larger when participants devoted more effort to a task (Johnson, 1984, 1986; Isreal et al., 1980). The advantage of questions endings with Tone4 over statements ending with Tone4 in the present study, therefore, seems to suggest that participants devote more processing effort to the former compared to the latter. As mentioned earlier, P300 is typically elicited in the target condition relative to the non-target condition. Azizian et al. (2006) proposed that equally probable stimuli that were easily evaluated as non-target required less mental work for discrimination and produced no P300-like components. From this line of reasoning, with a final Tone4, question intonation seems to be evaluated as possessing target-like properties, whereas statement intonation is evaluated as possessing non-target-like properties. Interestingly, this assumption happens to coincide with the view that statement intonation is a default intonation type and question intonation is a marked intonation type (Peters & Pfitzinger, 2008; Yuan, 2011). It appears that as a default intonation type, statement intonation occupies less mental attention or effort to be identified, leading to attenuated or even diminished P300 amplitude. In contrast, question intonation requires extra mental attention in order to be identified, resulting in increased P300 amplitude.

Interestingly, a comparison of our ERP results with the behavioral results revealed a discrepancy. In sentences ending with Tone4, question and statement contrast can be perceived electrophysiologically (reflected in P300). Behaviorally, listeners still had difficulty identifying question intonation (reflected in an identification rate of 68%, which was only marginally significantly ($p < .05$) above chance level, i.e., 50%) from statement intonation. This discrepancy could possibly be ascribed to the different processing that the neural responses and the behavioral responses reflect. In our study, the recorded EEG was time-locked to the onset of the critical syllable, whereas the behavioral responses were collected one second after the presentation of the stimuli. With such setup, one would not expect an isomorphic mapping between the neural responses and the behavioral responses, as P300 reflects cognitive processing restricted to a limited set of scalp electrodes within a limited temporal window, whereas the behavioral responses reflect whole brain processing.

Despite the discrepancy between the ERP results and the behavioral results for Tone4 conditions, the behavioral results in the present study are not uninterpretable. Our results showed that tone identity was seldom affected by intonation, while intonation identification was greatly affected by tone. Specifically, question intonation had a much lower identification rate than statement intonation regardless of final tone identities. For statement intonation identification, statements ending with Tone4 showed a significantly higher identification rate (99.3%) than statements ending with Tone2 (91.3%). These results are in line with previous studies (Xu & Mok, 2012a; Yuan, 2011). However, our results concerning question intonation identification were different from what was reported in Yuan (2011). Yuan (2011) found that questions ending with Tone4 were easier to identify than questions ending with Tone2. In our study, no statistically significant difference was found between questions ending with Tone4 (68%) and questions ending with Tone2 (64%). A closer comparison between these studies led us to infer that neutral semantic context might pose great difficulty to question intonation identification in questions ending with Tone4. In another study (Liu et al., 2016a), we examined context effects on question intonation identification in the questions ending with Tone2 condition and the questions ending with Tone4 condition in two behavioral experiments. One experiment embedded the target syllables in a neutral semantic sentence context; the other embedded the target syllables in a highly constraining semantic sentence context. What we found is that in the neutral semantic context, questions ending with Tone4 were not easier than questions ending with Tone2 for question intonation identification. However, in the highly constraining semantic context, question intonation was much better identified in questions ending with Tone4 than in those ending with Tone2, as was in Xu and Mok (2012a) and Yuan (2011). Even more interesting is that in low-pass filtered speech context, Xu and Mok (2014) found that questions ending with Tone2 had a higher accuracy rate than questions ending with Tone4. From the hierarchy of question intonation identification in Tone2 and Tone4 conditions in the above studies, it seems that context plays a role in question intonation identification. The stronger the linguistic context is (highly constraining semantic context > neutral semantic context > low-pass filtered

context), the better the identification of question intonation in questions ending with Tone4.

The opposing pattern was observed for questions ending with Tone2, with better identification of question intonation for weaker linguistic context. We infer that with less semantic information, frequency code (Ohala, 1983), high or rising pitch to mark questions, and low or falling to mark statements are more likely to be applied to intonation identification, resulting in relatively better identification of questions ending with Tone2. However, under no circumstance could listeners disentangle question intonation from Tone2 easily.

Semantic contexts affect question intonation perception. Speech contexts, however, impact question intonation production. Acoustic analyses in Yuan (2006) revealed that question intonation was realized as higher F0 at the end and steeper F0 slope of the final Tone2 than statement intonation in sentences ending with Tone2, and as higher F0 at the end of the final Tone4 than statement intonation in sentences ending with Tone4. Our acoustic results are in agreement with Yuan's results for Tone2, but not for Tone4. We discovered not only a higher F0 at the end in questions than in statements as in Yuan (2006), but also a distinctively higher F0 at the initial contour of the final Tone4 for questions than statements in our data. These different acoustic realizations between Yuan (2006) and our study might result from different coarticulation patterns by the preceding tone contexts. In Yuan's speech materials, the target tone was preceded by a low tone. Tonal coarticulation causes F0 lowering at the initial part of the falling contour of Tone4. Question intonation thus has to be realized as rising in F0 at the end. In comparison, the target tone was preceded by high-level tones in our speech materials. Tonal coarticulation led to a rising in the initial F0 and made the high feature of Tone4 more prominent. Meanwhile, F0 at the end of final Tone4 maintained its rising trend in question intonation.

Finally, it is interesting to compare the present study with the one conducted in Cantonese (Kung et al., 2014). Both studies investigated the online processing of tone and intonation under neutral semantic context. Using similar designs, the present study discovered a P300 effect for questions ending with Tone4 relative to statements ending with Tone4, while Kung et al. (2014)

observed a P600 effect for low tone in questions relative to low tone in statements. The P300 effect in Standard Chinese reflected the ease with which question and statement intonation can be distinguished in sentences with a final Tone4. However, the P600 effect in Cantonese revealed the strong conflicts and processing difficulties when intonation-induced F0 changes lead to the activation of two competing lexical representations. The two ERP components revealed different realizations of interaction of tone and intonation in Standard Chinese and Cantonese. In Standard Chinese, tone identity is maintained with the presence of intonation. Intonation identification is, however, greatly susceptible to the final tone identity. Question intonation is easier to be distinguished from statement intonation if the sentences bear a final Tone4, whereas the difference between intonation types is harder to perceive if the sentences bear a final Tone2. In Cantonese, tone identity is heavily distorted by intonation. The F0 contour of the low tones obtain a rising tail in questions, making it resemble the F0 contour of the mid-rising tone and therefore, causing processing difficulties in lexical identification.

2.5 Conclusion

To conclude, the present study provides online evidence that listeners can distinguish between question intonation and statement intonation when the intonation is associated with a final Tone4, but fail to do so when the intonation is associated with a final Tone2 at the attentive stage of processing. This study extended the sentence scope from one syllable to several syllables, expanded ERP evidence from the pre-attentive stage of processing to the attentive stage of processing, and revealed different realizations of interaction of tone and intonation in Standard Chinese and Cantonese.