HHS Public Access
Author manuscript
Brain Res. Author manuscript; available in PMC 2015 March 24.

Published in final edited form as:
Brain Res. 2007 January 12; 1128(1): 148–156. doi:10.1016/j.brainres.2006.10.064.

Mismatch negativity to pitch contours is influenced by language experience

Bharath Chandrasekaran, **Ananthanarayan Krishnan**, and **Jackson T. Gandour**[*]
Department of Speech Language and Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA

Abstract

A cross-language study utilizing the mismatch negativity (MMN) evoked response was conducted to explore the influence of language experience on the preattentive cortical processing of linguistically relevant pitch contours. Chinese and English subjects were presented with Mandarin Chinese tones while the mismatch negativity (MMN) response was elicited using a passive oddball paradigm. Two oddball conditions were constructed with a common deviant, a low falling rising contour tone (T3). One condition consisted of two tones that are acoustically similar to one another (T2/T3: T2, high rising contour=standard). The other condition consisted of two tones that are acoustically dissimilar to one another (T1/T3: T1, high level=standard). These tonal pairs enabled us to assess whether different degrees of similarity between pitch movements exert a differential influence on preattentive pitch processing. Results showed that the mean MMN amplitude of the Chinese group was larger than that of the English group for the T1/T3 condition. No group differences were found for the T2/T3 condition. The mean MMN amplitude was larger for the T1/T3 relative to the T2/T3 condition for the Chinese group only. By virtue of these language group differences, we infer that early cortical processing of pitch contours may be shaped by the relative saliency of acoustic dimensions underlying the pitch patterns of a particular language.

Keywords

Mismatch negativity; Experience-dependent plasticity; Speech perception; Pitch; Lexical tone; Mandarin Chinese

1. Introduction

As measured by the electrical mismatch negativity (MMN; or its magnetic equivalent, MMNm) response, a frontocentrally distributed cortical event-related potential (ERP), it is well established that language experience influences the automatic involuntary processing of speech sounds (see Kraus and Cheour, 2000; Näätänen, 2001, reviews). Early work focused on the presence or absence of specific consonants or vowels, i.e., *segmental* information, in a language's phonemic inventory (Näätänen et al., 1997; Sharma and Dorman, 2000;

[*]*Corresponding author*. Fax: +1 765 494 0771. gandour@purdue.edu (J.T. Gandour).

Winkler et al., 1999). More recently, experience-dependent effects have been demonstrated for other segmental aspects of phonology including phonotactics (Dehaene-Lambertz et al., 2000), phoneme boundaries (Sharma and Dorman, 1999; Ylinen et al., 2006), and size of phoneme inventories (Hacquard et al., in press).

Although it is well known that modulation of *suprasegmental* features of an utterance (e.g., pitch, duration, loudness) may signal a variety of linguistic and paralinguistic functions (Lehiste, 1996), there is a paucity of MMN data on the early automatic processing of suprasegmental information. The MMN response was reported to be sensitive to changes in loudness associated with non-contrastive, word-level stress in Hungarian (Honbolygo et al., 2004) and changes in duration associated with contrastive vowel length (e.g., /tuli/ 'fire' vs. / tu:li/ 'wind') in Finnish (Nenonen et al., 2003). In response to *nonspeech* sounds, changes in duration, but not in frequency, elicited a larger MMN in Finnish than in German listeners (Tervaniemi et al., 2006). This is presumably because pitch variations are not phonologically contrastive at the word level in either Finnish or German. Regarding the use of pitch in the music domain, musically trained subjects exhibited a larger MMN than musically naïve subjects (Koelsch et al., 1999; Tervaniemi, 2001). As far as we know, experience-dependent effects on the MMN have yet to be investigated for the processing of pitch in the language domain.

Languages that exploit phonologically contrastive variations in pitch at the word or syllable level are called *tone languages* (Gandour, 1994; Yip, 2003). Such languages provide a unique window for exploring experience-dependent effects on the automatic processing of phonologically contrastive pitch. Mandarin Chinese is a tone language. In addition to consonants and vowels, Chinese has four tones: *ma*[1] 'mother' [T1], *ma*[2] 'hemp' [T2], *ma*[3] 'horse' [T3], *ma*[4] 'scold' [T4]. Tones 1 to 4 can be described phonetically as high level, high rising, low falling rising, and high falling, respectively (Howie, 1976). Voice fundamental frequency (F_0) contours provide the dominant cue for tone recognition (Fig. 3) (Xu, 1997); other acoustic cues include amplitude (Whalen and Xu, 1992) and duration (Fu et al., 1998). Perceptual data on Mandarin tones indicate that variation in F_0 yields high levels of recognition for isolated tones (Howie, 1976).

Auditory electrophysiology provides an optimal window for viewing the *time course* of tonal processing. In a study of the role of tone and segmental information in Cantonese word processing (Schirmer et al., 2005), the time course and amplitude of the N400 effect were comparable for tone and segmental violations. Upon presentation of words from native (Thai) and nonnative (Mandarin) tone languages to Thai listeners (Sittiprapaporn et al., 2003), the MMN elicited by a native Thai word was greater than that elicited by a nonnative Mandarin word. The above-mentioned studies, however, were directed to spoken *word recognition* instead of tonal processing per se. As far as we know, there are no ERP studies of tonal processing at the cortical level in the extant literature.

The perceptual dimensions of tone and the effect of linguistic experience on a listener's perception of tone have been explored using multidimensional scaling (MDS) analysis of paired comparison judgments of tonal dissimilarity (e.g., Gandour, 1983; Gandour and Harshman, 1978). Based on data from tone languages of the Far East (Thai, Cantonese, Mandarin, Taiwanese) and West Africa (Yoruba), in addition to a non-tone language (English), three dimensions related to pitch – height, contour, and direction – were found to underlie an aggregate group stimulus space. Across the board, native speakers of a tone language placed more emphasis on pitch contour than on height whereas just the opposite was true for speakers of a non-tone language (English). The direction dimension separated primarily rising vs. non-rising F_0 movements. However, the relative importance of these dimensions across languages was shown to vary depending on the presence of specific types of pitch patterns in a tonal inventory. Unlike the four tones of Mandarin, four of the six Cantonese tones exhibit relatively flat trajectories. Cantonese speakers consequently attached relatively more importance to the height dimension than Mandarin speakers. Tonal dissimilarity judgments may also be influenced by phonological rules that govern a relationship between two or more tones (e.g., Mandarin T2 and T3). In a perceptual tone space derived from MDS, the distance between T2 and T3 was reduced in Mandarin Chinese speakers relative to English speakers (Huang, 2004; Hume and Johnson, 2001, 2003). These MDS data lead us to ask (i) whether the MMN can provide a noninvasive window to view these different degrees of tonal dissimilarity at the level of the cerebral cortex, and moreover, (ii) whether this electrophysiological event-related response is sensitive enough to reveal differences in perceptual saliency of dimensions across languages.

Accordingly, the aim of this cross-language (Chinese, English) ERP experiment was to determine whether language-experience modulates the automatic preattentive cortical processing of linguistically relevant pitch contours. The mismatch negativity, reflective of automatic processing, was measured in response to Mandarin tones presented to native speakers of Mandarin Chinese and a non-tone language control group of English speakers using two oddball conditions. One condition contained T1 and T3, two tones that are acoustically dissimilar to one another with a low rate of perceptual confusion; the other, T2 and T3, two tones that are acoustically similar to one another with a high rate of perceptual confusion (see Broselow et al., 1987; Gandour, 1994, 1978, reviews). By using these two pairs of tones, we were able to assess whether different degrees of similarity between pitch movements exert a differential influence on preattentive pitch processing, as reflected in the MMN. By including two language groups, one native (Chinese), the other nonnative (English), we were able to determine the extent to which modulation of the MMN could be attributed to a listener's long-term familiarity with the lexical tones themselves. If early preattentive cortical processing is driven primarily by acoustics, then we would expect uniform MMN responses for both oddball conditions irrespective of a listener's language affiliation. If, on the other hand, the relative perceptual weighting of acoustic dimensions underlying pitch patterns is influenced by a listener's language experience (Gandour, 1983; Huang, 2004; Hume and Johnson, 2001), then we would expect to observe cross-language effects on the modulation of the MMN.

2. Results

2.1. Mismatch negativity mean amplitude

The Chinese and English grand-average MMN standard and deviant waveforms for the two conditions (T1/T3, T2/T3) at the three electrode sites (F3, Fz, F4) are shown in Fig. 1.

Irrespective of group, relative to the 100% probability sequences, tones elicited robust MMN responses when presented in low probability (p=0.15) conditions. A three-way [group × condition × electrode location] repeated measures ANOVA conducted on the mean MMN amplitude yielded main effects of group ($F_{1,20}$=8.140, p=0.012) and condition ($F_{1,20}$=9.30, p<0.01), and an interaction effect between group and condition ($F_{1,20}$=5.49, p=0.04). No main effect of electrode location was found ($F_{2,80}$=1.30, p=0.282). Comparing groups (Chinese vs. English), Chinese subjects showed a larger MMN mean amplitude relative to English subjects for the T1/T3 condition ($F_{1,20}$=13.06, p<0.01) (Fig. 2). In contrast, the between-group comparison on the T2/T3 condition failed to reach significance ($F_{1,20}$=0.19, p=0.64). Comparing conditions ($C_{T1/T3}$ vs. $C_{T2/T3}$), the MMN mean amplitude response for the Chinese group was significantly less in the T2/T3 condition than the T1/T3 condition ($F_{1,20}$=14.07, p<0.01). For the English group, there was no significant difference between the two conditions ($F_{1,20}$=0.32, p=0.58).

2.2. Mismatch negativity peak latency

Irrespective of group, the MMN peaked later for the T2/T3 condition (C=178 ms; E=164 ms) relative to the T1/T3 condition (C=201 ms; E=196) (Fig. 2). A repeated measures three-way ANOVA [group × condition × electrode location] conducted on the peak latency measure yielded a significant main effect of condition ($F_{1,20}$=16.40, p<0.01), indicating that the peak of the MMN negativity occurred later in time for the T2/T3 condition relative to the T1/T3 condition. No other main effects or interaction effects reached significance.

3. Discussion

The major finding of the present cross-language, electrophysiological study of pitch perception is that language experience influences the early cortical automatic processing of linguistically relevant suprasegmental pitch contours. The native group (Chinese) exhibits larger MMN responses to the high dissimilarity condition (T1/T3) than the non-native (English). In the case of the low dissimilarity condition (T2/T3), MMN responses are similar for both groups. Interestingly, MMN responses are larger in the T1/T3 condition than the T2/T3 condition for native Chinese listeners only.

3.1. MMN responses are sensitive to cross-language differences in saliency of perceptual dimensions of tone

It has been proposed that the MMN reflects language-specific, long-term permanent or semi-permanent memory traces that subserve recognition patterns or templates in speech perception (see Näätänen, 2001, for a review). In the current experiment, all three stimuli (T1, T2, T3) represent excellent phonetic exemplars of Mandarin tones produced in isolation. From the standpoint of prototypicality, both conditions (T1/T3, T2/T3) would be expected to elicit comparable MMN enhancements for native speakers relative to non-native speakers. Instead, as predicted for native speakers, the MMN amplitude is larger with easier discrimination (T1/T3) than with harder discrimination (T2/T3). Consequently, the observed interaction between language group and condition cannot simply be attributed to the activation of long-term memory traces of *individual tones* for native Chinese speakers. Rather it is likely that due to a relatively greater similarity between T2 and T3 than between

T1 and T3, native Chinese speakers perceive a relatively greater difference between T1 and T3 than between T2 and T3. On the other hand, based on relative pitch heights (e.g., F_0 onset/offset, average F_0), native English speakers perceive the difference between T1 and T3 to be comparable to that of T2 and T3. We therefore infer that this experience-dependent MMN effect reflects *relations between tones* in the Mandarin tone space.

In MDS of listeners' perception of linear F_0 ramps (level, falling, rising), cross-language comparisons show that the relative importance of the pitch height and direction dimensions varies depending on a listeners' familiarity with specific types of pitch patterns that occur in the tone space of their native language (Gandour, 1983; Gandour and Harshman, 1978). Pitch height distinguishes tones, either static or dynamic, that differ according to average F_0, while direction of change distinguishes primarily rising from non-rising tones. The perceptual saliency of the direction dimension is greater for native speakers of tone languages, including Mandarin Chinese, than for speakers of non-tone languages (English), while English listeners give greater weight to the height dimension than do tone language speakers. Such cross-language differences in perceptual weighting suggest that long-term familiarity with a given language directs listeners' attention to linguistically relevant pitch characteristics of the auditory signal. In this study, T2 and T3 are very similar in F_0 contour and direction (Table 1 and Fig. 3). For Chinese listeners, variation in MMN responses between conditions may reflect differences between the standard and the deviant in terms of *weighting of pitch dimensions*. English listeners, on the other hand, are unfamiliar with the phonology and phonetics of Mandarin tones. We would expect less sensitivity to pitch contour dimensions because they may be placing greater emphasis on height-based features (e.g., F_0 onset/offset, average F_0). Such height features can also separate T3 from both T1 and T2 (cf. Gandour, 1983; Gandour and Harshman, 1978). Thus, the MMN may be an index of pitch features (i.e., contour, direction) that are differentially weighted depending on a listener's experience with lexical tones and their acoustic dimensions within a particular tone space.

Our finding that the mismatch negativity to pitch contours is modulated by long-term experience with Mandarin Chinese tones is in agreement with previous cross-language studies showing that MMN is sensitive to the listeners' native language (Dehaene-Lambertz et al., 2000; Näätänen et al., 1997; Nenonen et al., 2003; Sharma and Dorman, 2000; Tervaniemi et al., 2006; Ylinen et al., 2006). This experience-dependent effect has been shown to encompass all aspects of a language's phonological system, not only the presence or absence of a sound in its phonemic inventory. For example, mismatch negativity responses are sensitive to the overall size of a phonemic inventory (Hacquard et al., in press). MMNs are also enhanced by a change across a phoneme category boundary relative to the same physical change within a category boundary (e.g., vowels: Winkler et al., 1999).

Our MMN data are also compatible with a recent cross-language (Chinese, English) investigation of *preattentive* pitch processing as reflected in the human frequency following response (FFR) in the rostral brainstem (Krishnan et al., 2005). Using stimuli identical to those in the current study, FFR pitch strength and pitch tracking accuracy are reported to be greater for Chinese than for English listeners across all four Mandarin tones. These data indicate that there may be automatic enhancing or priming of *linguistically relevant features* of the auditory stream even as early as the brainstem.

3.2. Effects of psycholinguistic variables

We acknowledge that our experimental design does not explicitly eliminate the possibility of a lexical–semantic confound. All stimuli are real words in Mandarin Chinese. No direct comparisons are made to sequences containing pseudo-words as deviant stimuli (Pulvermüller et al., 2001; Shtyrov and Pulvermüller, 2002). It is unlikely, however, that lexical semantics, word frequency, or homophony can explain why the MMN amplitude is smaller in the T2/T3 than the T1/T3 condition for the Chinese group only.

All three words or morphemes (*yi*^{1} ‘clothing’; *yi*^{2} ‘aunt’; *yi*^{3} ‘chair’) occur frequently in standard Mandarin Chinese spoken and written text (Yuan, 1990). Yet there is a difference in frequency of occurrence of the standard and the deviant in the T1/T3 (778 vs. 157 per million) as compared to the T2/T3 condition (139 vs. 157 per million). On the basis of lexical decision tasks, high-frequency words are reported to elicit lower ERP amplitudes than low-frequency words in the 100–200 ms range (Assadollahi and Pulvermüller, 2001; Hauk and Pulvermüller, 2004). In this study, using the identity MMN, i.e., subtracting the deviant from itself rather than the standard, eliminates the possibility of a potential confound arising from the traditional subtraction procedure. It has also been reported that word frequency differences do not induce ERP latency shifts (Hauk and Pulvermüller, 2004). In this study, the peak latency of the T2/T3 condition is shifted later when compared to the T1/T3 condition irrespective of language group. An explanation based on word frequency is untenable since these stimuli are familiar to Chinese subjects only, and not American English. Word frequency is actually a relatively insignificant factor driving MMN enhancement compared to the lexical status of the stimuli (Pulvermüller et al., 2001, 2004).

Although the MMN enhancement attributable to the word effect occurs as early as 100–150 ms (Pulvermüller et al., 2004), our data show that regardless of language group, the magnitudes of the MMN responses elicited by the T1/T3 and T2/T3 conditions reach their peaks at around 175 and 200 ms, respectively. Instead of a lexical effect, it is more likely that the language-neutral, delayed latency of the MMN for the T2/T3 condition is due to acoustic similarity between F_0 contours of T2 and T3 (Fig. 3 and Table 1).

In a study of spoken word recognition in Cantonese (Schirmer et al., 2005), the time course and amplitude of the N400 effect were comparable for tone, consonant, and vowel violations, indicating that for tone language speakers, both segmental and suprasegmental information are equally important in processing words. In this study, stimuli are distinguished minimally by the suprasegmental information only; segmental information is fixed. It is commonly accepted that words are recognized as they are heard. This recognition process depends crucially on other words in the language that share the same acoustic–phonetic properties with the critical word. The *recognition point* (Marslen-Wilson, 1987) of a spoken word is the earliest point in time at which it becomes distinct from its lexical competitors (Pulvermüller et al., 2006). In a passive oddball paradigm, there are, in principle, no lexical competitors for Chinese listeners. In the case of English listeners, the stimuli are not familiar words at all. The fact that the peak latency did not vary by language group in either condition leads us to conclude that this measure does not reflect the word recognition point, but rather the point at which the pitch contours were sufficiently discriminated from one another.

These three words are also similar in terms of number of homophones (*yi*[1], 12; *yi*[2], 17; *yi*[3], 13). The difference between the standard and the deviant in the T1/T3 and T2/T3 conditions is one and four homophones, respectively. Because of the comparatively high degree of homophony in Chinese dialects (Chen, 1992; Chen et al., 2000), words are dependent on sentence context in order to be correctly interpreted. The oddball sequence, however, is not a sentence. Therefore, it is not biased in favor of one semantic interpretation over another. Any lexical effects due to homophony are likely to be subtracted out in the difference waveforms for the Chinese group.

3.3. Speech-specificity of pitch processing in tone languages

The experimental design does not include a homologous nonspeech condition. Thus, we are unable to determine whether the experience-dependent modulation of the MMN amplitude is speech-specific or not. In the case of duration, it has recently been shown in a passive oddball paradigm that the MMN amplitude elicited in response to duration differences in nonspeech sounds is larger in Finnish than in German speakers (Tervaniemi et al., 2006). In Finnish, but not in German, the length of a consonant or vowel phoneme can signal changes in word meaning (e.g., /tuli/ 'fire', /tuuli/ 'wind', /tulli/ 'customs'). Interestingly, no group differences were observed in either neural or behavioral discrimination of frequency changes in nonspeech sounds. This is presumably due to the fact that neither Finnish nor German exploits changes in F_0 contrastively at the syllable level. Thus, it appears that preattentive cortical processing can be *selectively tuned* to those sound features of the auditory signal that are of phonological relevance in a particular language even in nonspeech sounds (Tervaniemi et al., 2006, p. 2541). In the case of F_0, the next step is to determining whether selective tuning extends to frequency as well as duration in nonspeech sounds.

3.4. Conclusion

In conclusion, the current study demonstrates that the mismatch negativity is an excellent tool for investigating the automatic processing of suprasegmental information in speech. The use of multiple language groups is important for showing language-related differences in the relative importance of perceptual dimensions that may influence the magnitude of the mismatch response to pitch contours. The convergence of ERP and earlier behavioral evidence further supports the notion that the relative saliency of perceptual dimensions underlying lexical tones may influence the automatic processing of pitch contours at early stages of speech perception.

4. Experimental procedures

4.1. Subjects

Eleven adult, native speakers of Mandarin Chinese (6 male; 5 female), and eleven adult, native speakers of American English (6 male; 5 female), participated in the ERP experiment. Both groups were closely matched with respect to age (Chinese: mean=23.2, SD=4.1;

English: mean=25.2, SD=3.2) and years of formal education (Chinese: mean=17.2, SD=2.3; English: mean=18.2, SD=2.2). Native speakers of Mandarin Chinese had no English instruction before the age of 11. No English participant had any previous exposure to Chinese or for that matter any other tone language. Both groups had normal hearing sensitivity (i.e., pure-tone air conduction thresholds of 25 dB HL or better in both ears) at frequencies of 0.5, 1.2, and 4 kHz. None of the participants had more than three years of formal musical training, and none had any musical training within the past five years. All participants were paid for their participation. They gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

4.2. Stimuli

Stimuli consisted of a set of four Mandarin Chinese words that are distinguished minimally by tonal contour (*pinyin* Roman transliteration): yi^1 'clothing' [T1]; yi^2 'aunt' [T2]; yi^3 'chair' [T3]; yi^4 'easy' [T4]. Using a cascade/parallel formant synthesizer (Klatt, 1980; Klatt and Klatt, 1990), synthetic versions were created so that all four syllables were identical except for their tonal contours. Synthesis parameters for voice fundamental frequency (F_0) and duration were obtained from words produced in citation form by an adult male speaker (Xu, 1997). F_0 contours of the stimuli are displayed in Fig. 3. Vowel formant frequencies were steady-state, and were held constant across the four syllables (in Hz): F1=300; F2=2500; F3=3500; and F4=4530 (Howie, 1976). Tonal durations were normalized to 250 ms. Voice amplitude was kept constant at 60 dB.

Only three of the four Mandarin Chinese tones (T1, T2, T3) were chosen for presentation in a passive oddball paradigm. This limitation restricted EEG recording time to 90 min, thus minimizing the risk of subject fatigue.

Our rationale for the selection of T1, T2, and T3 was based primarily on acoustic phonetic, perceptual, as well as phonological considerations (Fig. 3). These three tones differed in F_0 height as measured by onset, offset, and average F_0, and in F_0 contour and/or direction as measured by slope from tonal onset to offset, from onset to turning point, and from turning point to offset (Table 1). The locations of the turning points (in ms) for T2 and T3 were 106 and 144, respectively. T1 may be described as a high level tone; T2 and T3 as high rising and low falling rising contour tones, respectively. With regard to pitch contour and direction, T2 is acoustically more similar to T3 than to T1. Tonal discrimination and identification data confirm that T2 and T3, the two tones that are acoustically most similar, are more often confused with one another as compared to other tonal pairs (see Gandour, 1978, review). Both timing of the turning point and ΔF_0 (difference in F_0 between onset and turning point) have been reported to be perceptually relevant for the identification of T2 and T3 (Shen and Lin, 1991; Shen et al., 1993). From a phonological perspective, the distinction between T2 and T3 is said to be 'neutralized' by a *tone sandhi* rule (Chao, 1968), i.e., T3 changes to T2 when immediately followed by another T3 in running speech. Its perceptual validity has been confirmed in studies using tonal identification tests (Peng, 2000; Wang and Li, 1967). Using MDS analysis, it has been demonstrated that the perceptual distance is smaller between T2 and T3 in the Mandarin tone space for native speakers as compared to nonnative (Huang, 2004). This cross-language disparity in perceived similarity between T2 and T3 points to the interplay of speech perception and phonology (Hume and Johnson, 2001).

The experiment consisted of two oddball conditions, both of which used T3 as the deviant stimuli. In one condition, T1 was the standard; the pair of tones (T1/T3) represented a *dissimilar contrast*. In the other condition, T2 was the standard; the pair of tones (T2/T3) represented a *similar contrast*. By pairing T1 with T3, we were able to contrast a level tone with a bidirectional low falling rising (T3). By pairing T2 with T3, we were able to contrast a high rising (T2) with a low rising (T3) contour tone. In the T2/T3 condition, the paired tones are also related to one another by a phonological rule that applies in connected speech (Chao, 1968).

4.3. Experimental protocol

Subjects were seated in a recliner in an acoustically and electrically shielded booth facing a video monitor. They were instructed to ignore the auditory stimuli presented via earphones; to refrain from extraneous body movements; and to focus their attention exclusively on a self-selected, closed-caption silent movie. The subtitles were presented in the participant's native language to ensure that their attention was focused on the movie. Furthermore, they were also told that they would be required to provide a short synopsis of the movie at the end of the session.

Three different stimulus sequences with an interstimulus onset-to-onset interval of 667 ms were presented in random order. Two were oddball conditions made up of a frequent stimulus or standard (p=0.85) and an infrequent stimulus or deviant (p=0.15). T3 was used as the deviant in both the oddball conditions; the standards were T1 and T2, respectively. The remaining sequence were comprised of the deviant (T3) presented alone (p=1.00). Within the oddball sequences, the order of presentation of stimuli was pseudo-random, i.e., at least one standard stimulus preceded each deviant.

One-hundred artifact-free deviants were collected for each of the three oddball sequences from each participant. In the deviant-alone (p=1.00) sequences, 500 artifact-free trials of the T3 stimuli were collected from each participant. The whole experiment lasted approximately 2 h including subject preparation.

All stimuli were controlled by a signal generation and data acquisition system (Intelligent Hearing Systems–Smart EP). The stimulus files were routed through a digital to analog module, and presented binaurally at 75 dB SPL to each ear through magnetically shielded insert earphones (Biologic TIP-300).

4.4. Evoked potential recording

For each subject, silver chloride electrodes were mounted on frontal midline (Fz), left frontal (F3), and right frontal (F4) sites according to the 10–20 location system. These three electrodes were chosen because the MMN response is known to be most robust at the frontal electrode sites (Näätänen et al., 1997). The electrode impedance was kept below 5 kΩ. The right and left mastoids were linked and served as the reference electrode. The ground electrode was placed on the forehead. Electrodes monitoring vertical eye movements were used in removing eye blink artifacts, as defined by epochs with voltage changes exceeding 75 μV. The electrical signal was bandpass filtered at 1–30 Hz and recorded at a 1000 Hz sampling rate. A digital bandpass filter (1–30 Hz) was applied offline.

4.5. Preprocessing of ERP data

The baseline for the grand averaged waveforms was defined as the average amplitude between − 100 ms and 0ms (onset of stimuli). To calculate the MMN, the deviant waveform obtained from the oddball paradigm was subtracted from the deviant presented in the 100% probability condition. This subtraction process, also called an ‘identity MMN’, effectively controls for any acoustical differences between stimuli (Jacobsen et al., 2004; Kraus et al., 1995; Sharma and Dorman, 2000; Ylinen et al., 2006).

The responses from F3, Fz, and F4 were separately measured for each subject and condition. The MMN window was defined between 100 ms and 250 ms in the difference waveform. The MMN amplitude was measured as the mean voltage from a 100 ms window centered on the grand average MMN peak latency obtained from each difference waveform, within the predefined MMN window.

4.6. Statistical analysis

MMN mean amplitudes and peak latencies were analyzed using a three-way mixed model ANOVA (subject as random effect) for the effects of language group (Chinese, English), tonal conditions (T1/T3, T2/T3) and electrode locations (F3, Fz, F4).

Acknowledgments

Research supported in part by the Purdue Research Foundation (J. G.). Thanks to Xiaojian Li (South China Normal University) for assistance with the statistics of word frequency usage in Chinese; Alexander Francis (Purdue University) and two anonymous reviewers for insightful and helpful comments on an earlier version of the manuscript. B. C. is currently a predoctoral student in the Purdue University Life Sciences Integrative Neuroscience Program.

REFERENCES

Assadollahi R, Pulvermüller F. Neuromagnetic evidence for early access to cognitive representations. NeuroReport. 2001; 12:207–213. [PubMed: 11209922]

Broselow, E., et al. The perception of second language prosody. In: Ioup, G.; Weinberger, S., editors. Interlanguage Phonology: The Acquisition of a Second Language Sound System. New York: Newbury House; 1987. p. 350-364.

Chao, Y. A Grammar of Spoken Chinese. Berkeley, CA: University of California Press; 1968.

Chen, H-C. Reading comprehension in Chinese: implications from character reading times. In: Chen, H-C.; Tzeng, O., editors. Language Processing in Chinese. North-Holland, Amsterdam: 1992. p. 175-205.

Chen H-C, et al. Effects of antecedent order and semantic context on Chinese pronoun resolution. Mem. Cogn. 2000; 28:427–438.

Dehaene-Lambertz G, et al. Electrophysiological correlates of phonological processing: a cross-linguistic study. J. Cogn. Neurosci. 2000; 12:635–647. [PubMed: 10936916]

Fu QJ, et al. Importance of tonal envelope cues in Chinese speech recognition. J. Acoust. Soc. Am. 1998; 104:505–510. [PubMed: 9670541]

Gandour, JT. The perception of tone. In: Fromkin, VA., editor. Tone: A Linguistic Survey. New York: Academic Press; 1978. p. 41-76.

Gandour J. Tone perception in Far Eastern languages. J. Phon. 1983; 11:149–175.

Gandour, J. Phonetics of tone. In: Asher, R.; Simpson, J., editors. The Encyclopedia of Language and Linguistics. Vol. 6. New York: Pergamon Press; 1994. p. 3116-3123.

Gandour J, Harshman R. Crosslanguage differences in tone perception: a multidimensional scaling investigation. Lang. Speech. 1978; 21:1–33. [PubMed: 692240]

Hacquard V, et al. The effects of inventory on vowel perception in French and Spanish: an MEG study. Brain and Language. in press.

Hauk O, Pulvermüller F. Effects of word length and frequency on the human event-related potential. Clin. Neurophysiol. 2004; 115:1090–1103. [PubMed: 15066535]

Honbolygo F, et al. Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. Neurosci. Lett. 2004; 363:84–88. [PubMed: 15158003]

Howie, J. Acoustical Studies of Mandarin Vowels and Tones. Cambridge, UK: Cambridge Univ. Press; 1976.

Huang, T. Linguistics. Columbus, OH: Ohio State University; 2004. Language-specificity in auditory perception of Chinese tones.

Hume, E.; Johnson, K. A model of the interplay of speech perception and phonology. In: Hume, E.; Johnson, K., editors. The Role of Speech Perception in Phonology. New York: Academic Press; 2001. p. 3-25.

Hume, E.; Johnson, K. The impact of partial phonological contrast on speech perception. In: Sole, MJ., et al., editors. 15th International Congress of Phonetic Sciences. ICPhS Organizing Committee; Barcelona, Spain. 2003. p. 2385-2388.

Jacobsen T, et al. Pre-attentive auditory processing of lexicality. Brain Lang. 2004; 88:54–67. [PubMed: 14698731]

Klatt DH. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 1980; 67:971–995.

Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am. 1990; 87:820–857. [PubMed: 2137837]

Koelsch S, et al. Superior pre-attentive auditory processing in musicians. NeuroReport. 1999; 10:1309–1313. [PubMed: 10363945]

Kraus N, Cheour M. Speech sound representation in the brain. Audiol. Neuro-otol. 2000; 5:140–150.

Kraus N, et al. Central auditory system plasticity associated with speech discrimination training. J. Cogn. Neurosci. 1995; 7:25–32. [PubMed: 23961751]

Krishnan A, et al. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cogn. Brain Res. 2005; 25:161–168. [PubMed: 15935624]

Lehiste, I. Suprasegmental features of speech. In: Lass, NJ., editor. Principles of Experimental Phonetics. St. Louis: Mosby; 1996. p. 226-244.

Marslen-Wilson WD. Functional parallelism in spoken word recognition. Cognition. 1987; 25:71–102. [PubMed: 3581730]

Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology. 2001; 38:1–21. [PubMed: 11321610]

Näätänen R, et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature. 1997; 385:432–434. [PubMed: 9009189]

Nenonen S, et al. Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. Brain Res. Cogn. Brain Res. 2003; 16:492–495. [PubMed: 12706229]

Peng, S. Lexical versus ‘phonological’ representations of Mandarin Sandhi tones. In: Broe, MB.; Pierrehumbert, JB., editors. Papers in Laboratory Phonology: V. Acquisition and the Lexicon. New York: Cambridge Univ. Press; 2000. p. 152-167.

Pulvermüller F, et al. Memory traces for words as revealed by the mismatch negativity. NeuroImage. 2001; 14:607–616. [PubMed: 11506534]

Pulvermüller F, et al. Word-specific cortical activity as revealed by the mismatch negativity. Psychophysiology. 2004; 41:106–112. [PubMed: 14693005]

Pulvermüller F, et al. Tracking speech comprehension in space and time. NeuroImage. 2006; 31:1297–1305. [PubMed: 16556504]

Schirmer A, et al. Brain responses to segmentally and tonally induced semantic violations in Cantonese. J. Cogn. Neurosci. 2005; 17:1–12. [PubMed: 15701235]

Sharma A, Dorman MF. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. J. Acoust. Soc. Am. 1999; 106:1078–1083. [PubMed: 10462812]

Sharma A, Dorman MF. Neurophysiologic correlates of cross-language phonetic perception. J. Acoust. Soc. Am. 2000; 107:2697–2703. [PubMed: 10830391]

Shen XS, Lin M. A perceptual study of Mandarin tones 2 and 3. Lang. Speech. 1991; 34:145–156.

Shen XS, et al. F_0 turning point as an F_0 cue to tonal contrast: a case study of Mandarin tones 2 and 3. J. Acoust. Soc. Am. 1993; 93:2241–2243.

Shtyrov Y, Pulvermüller F. Neurophysiological evidence of memory traces for words in the human brain. NeuroReport. 2002; 13:521–525. [PubMed: 11930174]

Sittiprapaporn W, et al. Preattentive processing of lexical tone perception by the human brain as indexed by the mismatch negativity paradigm. Ann. N. Y. Acad. Sci. 2003; 999:199–203. [PubMed: 14681141]

Tervaniemi M. Musical sound processing in the human brain. Evidence from electric and magnetic recordings. Ann. N. Y. Acad. Sci. 2001; 930:259–272. [PubMed: 11458833]

Tervaniemi M, et al. Selective tuning of cortical sound-feature processing by language experience. Eur. J. Neurosci. 2006; 23:2538–2541. [PubMed: 16706861]

Wang WS, Li KP. Tone 3 in Pekinese. J. Speech Hear. Res. 1967; 10:629–636. [PubMed: 6081948]

Whalen DH, Xu Y. Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica. 1992; 49:25–47. [PubMed: 1603839]

Winkler I, et al. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. Brain Res. Cogn. Brain Res. 1999; 7:357–369. [PubMed: 9838192]

Xu Y. Contextual tonal variations in Mandarin. J. Phon. 1997; 25:61–83.

Yip, M. Tone. New York: Cambridge Univ. Press; 2003.

Ylinen S, et al. Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. Brain Res. 2006; 1072:175–185. [PubMed: 16426584]

Yuan L. Dictionary of Usage Frequency of Modern Chinese Words. Publishing House of Astronavigation, Beijing. 1990

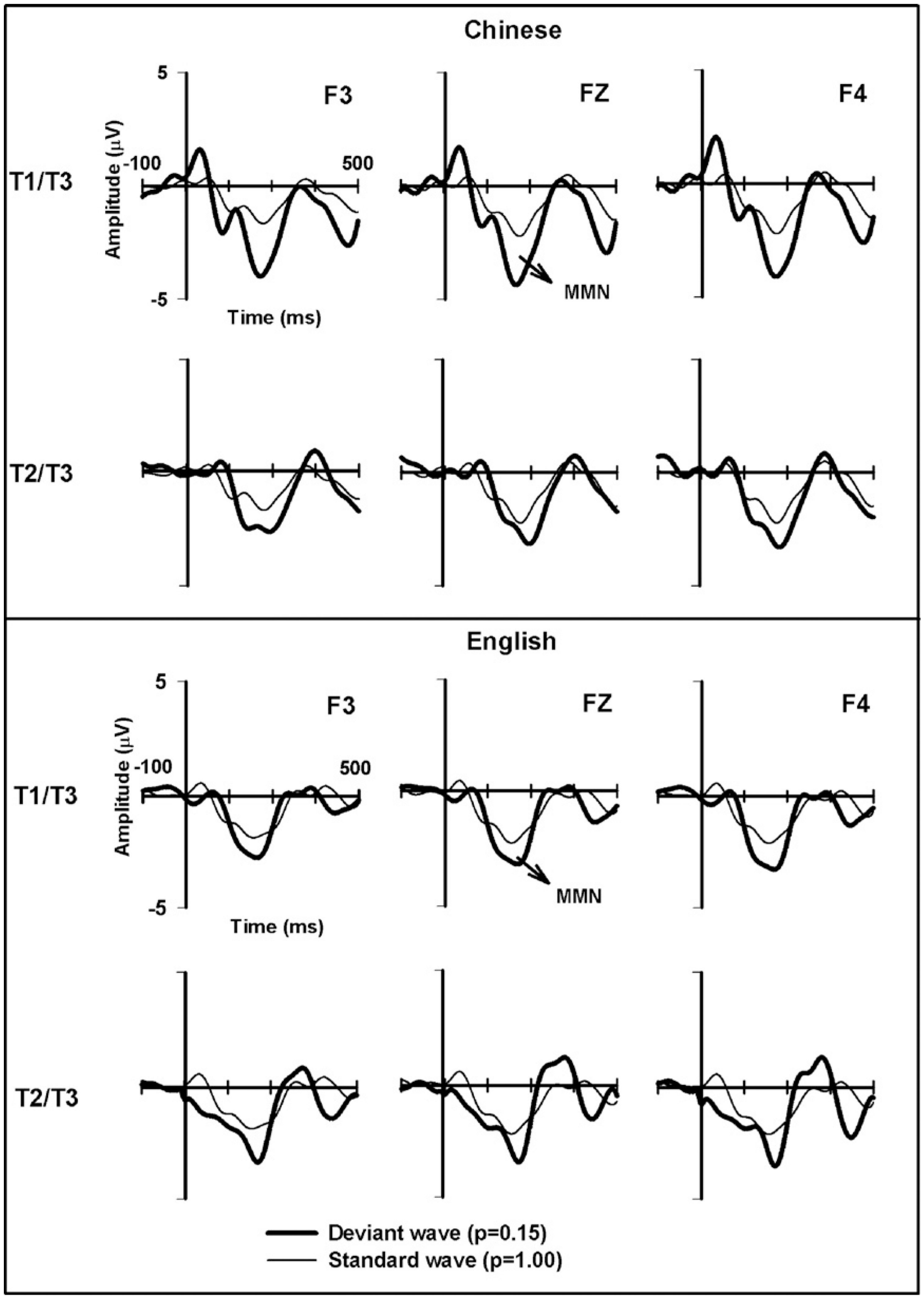

Fig. 1.

Grand-averaged standard (high probability, p = 1.00) and deviant waveforms (low probability, p = 0.15) obtained from the F3, Fz, and F4 electrode locations are displayed for the two language groups (Chinese, top panel; English, bottom panel) per experimental condition (high dissimilarity T1/T3, 1st row; low dissimilarity T2/T3, 2nd row). The standard waveforms were elicited by presenting T3 alone; the deviant waveforms were elicited by presenting T3 in the context of either T1 (T1/T3) or T2 (T2/T3). The T2/T3 condition peaks later than the T1/T3 condition for both groups. The ERPs obtained from the

Chinese group show a larger negativity for the T1/T3 condition relative to English speakers. Also, the Chinese subjects show a larger MMN response for the T1/T3 condition relative to the T2/T3 condition, a difference not seen in the English group.

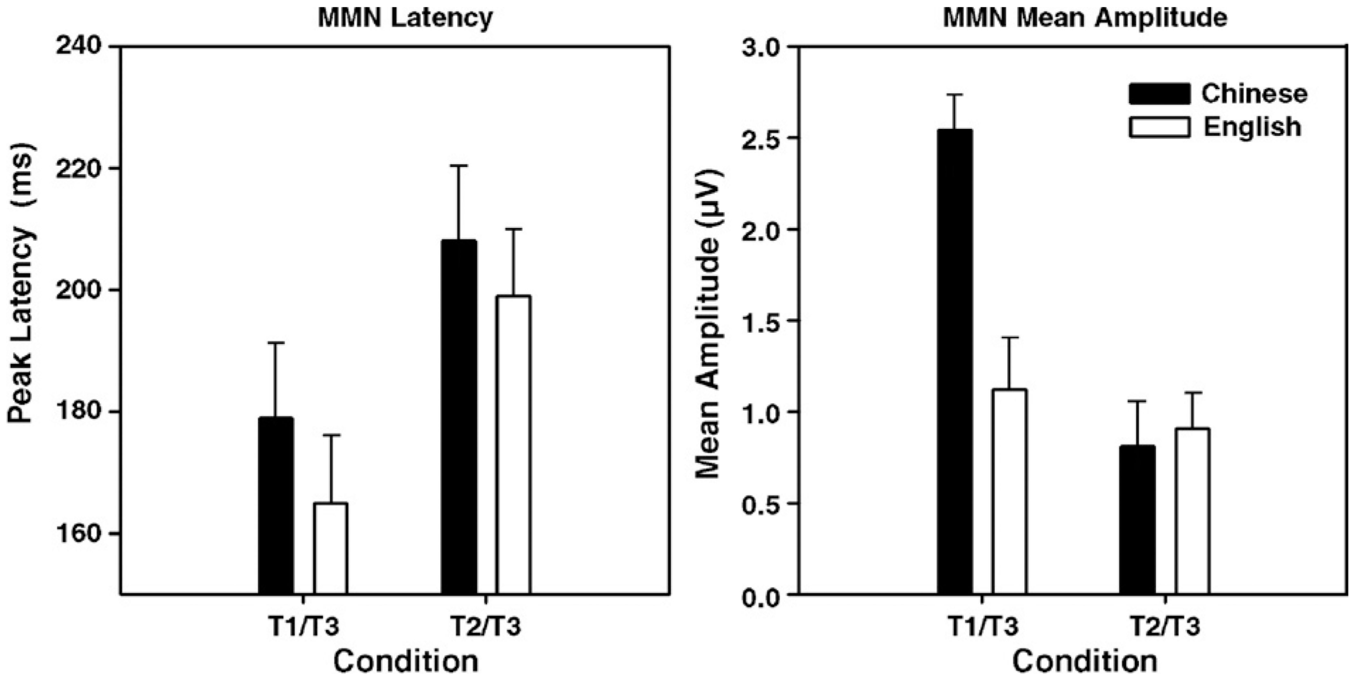

Fig. 2.

Mean peak latency (left panel) and amplitude (right panel) values are displayed for the two language groups (Chinese, English) per experimental condition (T1/T3, T2/T3) as measured from the Fz electrode location. For both groups, MMN reaches its peak negativity in the T2/T3 condition later in time than the T1/T3 condition. MMN amplitude is seen to be higher in the Chinese group relative to English in the T1/T3 condition only. Also, the MMN amplitude is larger in the Chinese group for the T1/T3 condition as compared to the T2/T3 condition. For English subjects, there is no difference between the two conditions.

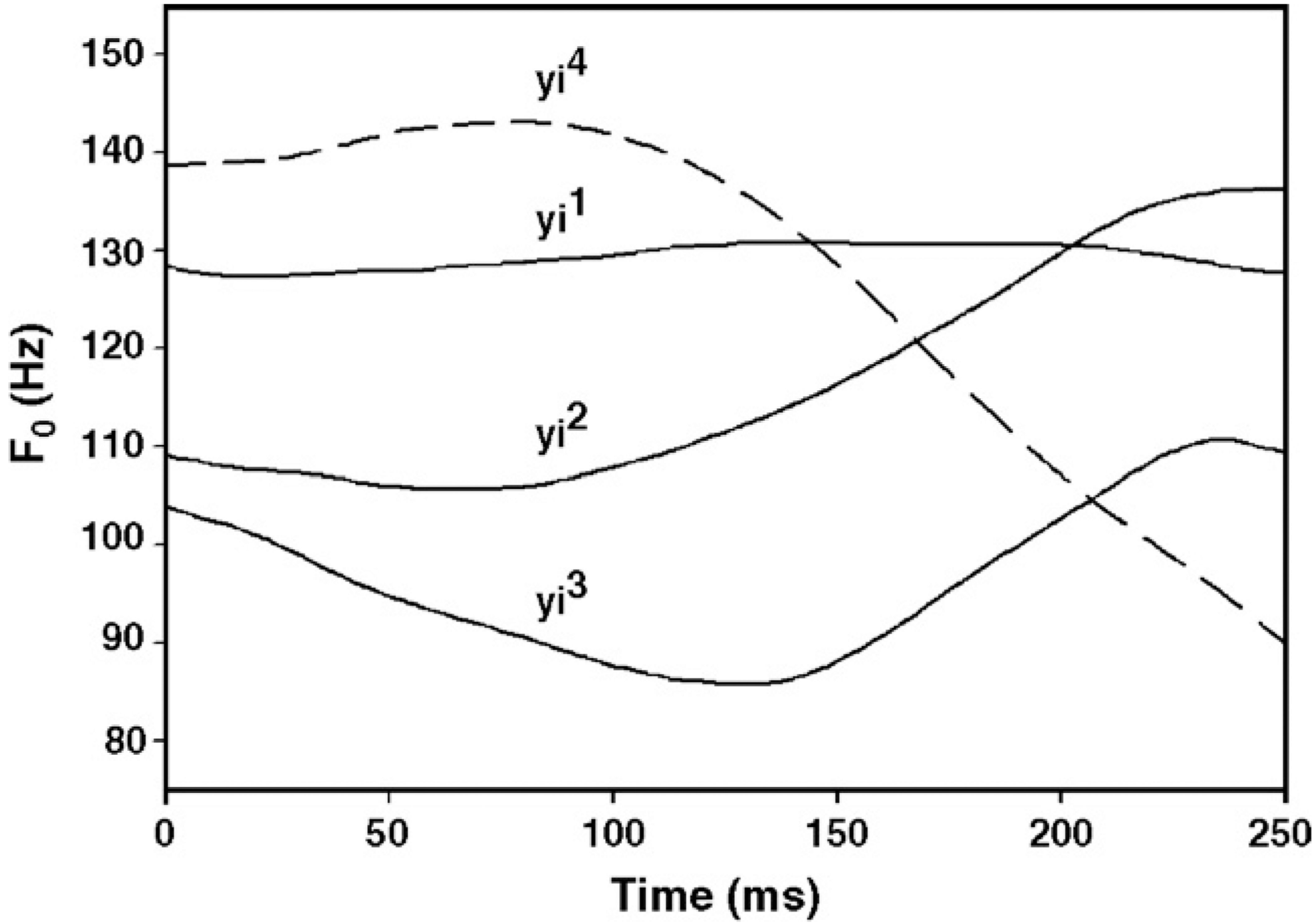

Fig. 3.

Average fundamental frequency contours of time-normalized Mandarin Chinese tonal stimuli adapted from Xu (1997). Superscript numbers (1–4) denote the four Mandarin tonal categories: yi^1 ‘clothing’; yi^2 ‘aunt’; yi^3 ‘chair’; yi^4 ‘easy’. Of those four tones, tones 1–3 (solid lines) were used in the ERP experiment. Two conditions were utilized; T1 was the standard in one condition (T1/T3), T2 in the other (T2/T3). In both conditions, T3 was presented as the deviant. Acoustically, T2 is closer to T3 than T1 in terms of F_0 onset, average F_0, and overall F_0 contour.

Author Manuscript
Author Manuscript
Author Manuscript

Table 1

Acoustic characteristics of experimental stimuli

Tone	F_0 parameters							
	Onset	Offset	Average	Slope [a] onset–offset	Slope [b] onset–TP	Slope [c] TP–offset	TP	ΔF_0
T1	129	128	129	0.00	0.02	−0.02	125	2
T2	109	136	117	0.11	−0.04	0.17	71	−4
T3	104	109	96	0.02	−0.13	0.19	133	−18

Onset, offset, and average F_0 values are expressed in Hertz (Hz). All slope values are expressed in Hz/ms. T1, T2, and T3 refer to the Mandarin high level, high rising, and low falling rising tones, respectively. TP, expressed in ms, refers to turning point, i.e., time at which the contour changed direction. F_0=voice fundamental frequency; ΔF_0=change in Hz from onset to turning point.

[a] Overall slope, measured from pitch onset to offset.

[b] Slope from the onset to TP. Since the level tone T1 has no clear turning point, slope was measured from onset to 125 ms (50% duration). Both T2 and T3 have negative slopes (i.e. falling F_0 contour).

[c] Slope from the TP to offset. Since the level tone T1 has no clear turning point, slope was measured from 125 ms (50% duration) to offset. Both T2 and T3 have positive slopes (i.e. rising F_0 contour).