Brain and Language 180–182 (2018) 31–41

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Are lexical tones musical? Native language's influence on neural response to pitch in different domains

Ao Chen[a,b,d,*], Varghese Peter[b,e], Frank Wijnen[a], Hugo Schnack[a,c], Denis Burnham[b]

[a] *Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands*
[b] *MARCS Institute for Brain, Behaviour & Development, Western Sydney University, Sydney, Australia*
[c] *Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands*
[d] *School of Communication Science, Beijing Language and Culture University, Beijing, China*
[e] *Department of Linguistics, Macquarie University, North Ryde, NSW 2109, Australia*

ARTICLE INFO

Keywords:
Lexical tones
Musical pitch
Mismatch negativity
Cross-domain correlation

ABSTRACT

Language experience shapes musical and speech pitch processing. We investigated whether speaking a lexical tone language natively modulates neural processing of pitch in language and music as well as their correlation. We tested tone language (Mandarin Chinese), and non-tone language (Dutch) listeners in a passive oddball paradigm measuring mismatch negativity (MMN) for (i) Chinese lexical tones and (ii) three-note musical melodies with similar pitch contours. For lexical tones, Chinese listeners showed a later MMN peak than the non-tone language listeners, whereas for MMN amplitude there were no significant differences between groups. Dutch participants also showed a late discriminative negativity (LDN). In the music condition two MMNs, corresponding to the two notes that differed between the standard and the deviant were found for both groups, and an LDN were found for both the Dutch and the Chinese listeners. The music MMNs were significantly right lateralized. Importantly, significant correlations were found between the lexical tone and the music MMNs for the Dutch but not the Chinese participants. The results suggest that speaking a tone language natively does not necessarily *enhance* neural responses to pitch either in language or in music, but that it does *change the nature* of neural pitch processing: non-tone language speakers appear to perceive lexical tones as musical, whereas for tone language speakers, lexical tones and music may activate different neural networks. Neural resources seem to be assigned differently for the lexical tones and for musical melodies, presumably depending on the presence or absence of long-term phonological memory traces.

1. Introduction

Speech and music are two unique products of the human brain that serve communicative purposes, and are present across all cultures (Patel, 2008). The specificity of, or the commonality between these two domains has received much attention in cognitive neuroscience, and there is evidence of both association and dissociation between the two (Bidelman, Hutka, & Moreno, 2013; Krishnan, Gandour, & Bidelman, 2010; Nan, Huang, Wang, Liu, & Dong, 2016; Peretz & Coltheart, 2003; Wong & Perrachione, 2007). The aims of this study are to examine (1) the effect of tone vs non-tone language experience on neural discrimination of pitch in speech and music, and (2) cross domain correlation of the event related potentials (ERPs) that indicate neural pitch change detection in music and speech, and whether any such correlation is influenced by tone/non-tone language experience.

Both speech and music experience modulate auditory perception, and there is perceptual attunement to ambient input in both domains in development. For example, newborn infants are able to discriminate both native and non-native speech sounds (consonants, vowels and lexical tones), but from about 4 months onward sensitivity to non-native sounds deteriorates while sensitivity to native sounds is maintained or improves (Kuhl et al., 2006; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008; Werker & Tees, 1984). Once native phonetic categories are established, listeners perceive native phonemes categorically, whereas the non-native phonemes are perceived psycho-acoustically (Francis, Ciocca, & Ng, 2003; Hallé, Chang, & Best, 2004). Similar perceptual attunement has also been found in the music domain. Infants initially discriminate melodies from ambient and novel musical scales (Lynch, Eilers, Oller & Urbano, 1990), but this then becomes tuned to the structure of the ambient input (Lynch & Eilers, 1992; Trainor & Trehub, 1992) such that adults are less capable of discriminating non-

---

[*] Corresponding author at: Beijing Language and Culture University, No. 15 Xueyuan Road, 100083, Beijing, China.
E-mail address: irischen71@hotmail.com (A. Chen).

https://doi.org/10.1016/j.bandl.2018.04.006
Received 19 December 2017; Received in revised form 15 March 2018; Accepted 14 April 2018

native than native musical structures (Lynch et al., 1990; Schellenberg & Trehub, 1999; Trehub, Schellenberg, & Kamenetsky, 1999).

Over and above within-domain attunement, experience in one domain has been shown to enhance processing in the other. For example, speaking a tone language (i.e., a language in which pitch is used to distinguish lexical meaning) facilitates the discrimination of musical pitch (Bidelman et al., 2013; Chen, Liu, & Kager, 2016; Stevens, Keller, & Tyler, 2013). Conversely, musicianship enhances detection of lexical tone contrasts (Alexander, Wong, & Bradlow, 2005; Delogu, Lampis, & Belardinelli, 2006; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011), and musicians more easily learn to pair pitch patterns and word meaning, an ability similar to learning lexical tones (Wong & Perrachione, 2007). In a recent study, Chen, Liu et al. (2016) showed that being a native tone-language speaker not only facilitates music perception, but engenders a dissociation of the processing of musical and lexical tone. Chen et al. tested native tone language (Chinese Mandarin) and non-tone language (Dutch) listeners on discrimination of Mandarin lexical tones and musical phrases. As expected, Chinese listeners outperformed Dutch listeners in the music tasks. Importantly, significant positive correlations were found between the discrimination of lexical tones and musical phrases among the Dutch listeners whereas such correlations were absent in Chinese listeners. Based on these findings, Chen et al. proposed the “split hypothesis”, which holds that in perception, for tone language listeners, pitch variations that are used phonemically as lexical tones are split from other, non-lexical pitch variations, hence the absence of speech-music correlation among the Chinese listeners. For the Dutch listeners, as lexical tones do not play a phonemic role, tones and musical pitch are both perceived psychoacoustically in a unified process. Such cross-domain correlation has also been observed among native non-tonal language Turkish listeners (Chen, Roncaglia-Denissen, Roor, & Sadakata, 2016).

All these studies used behavioural methods. At the neural level, evidence has accumulated regarding plasticity induced by both language and music experience (Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011; Krishnan et al., 2010; van Zuijen, Sussman, Winkler, Näätänen, & Tervaniemi, 2005). A component of auditory event-related potentials (ERP), the mismatch negativity (MMN), is commonly used to examine neural detection of auditory change. MMN can be elicited using a passive oddball paradigm, in which listeners are presented with a stream of ‘standard’ sounds conforming to a certain regularity punctuated occasionally by ‘deviant’ sounds, dissimilar in some relevant dimension from the standards. If the brain detects the change from standard to deviant, then on the difference waveform obtained by subtracting the response to the standard from that to the deviant, the MMN is visible as a negative peak between 100 and 300 ms from deviant onset (Bishop, 2007; Näätänen, Paavilainen, Rinne, & Alho, 2007).

There is some discrepancy with respect to the effect of tone language experience on MMN responses to linguistic and non-linguistic pitch change. When non-speech stimuli (harmonic or pure tones) are closely matched to lexical tones (identical amplitude information, duration, and fundamental requency (F0)), native tone language listeners have been shown to exhibit comparable MMNs to lexical tones and non-speech analogues (Gu, Zhang, Hu, & Zhao, 2013; Xi, Zhang, Shu, Zhang, & Li, 2010), suggesting common neural mechanisms for processing both speech and non-speech pitch contours. Different results have been found comparing tone and non-tone language listeners’ MMN responses to pitch stimuli. Chandrasekaran, Krishnan and Gandour (2007a) found that Chinese listeners, compared to non-tone language listeners, showed larger MMNs in response to iterated ripple noises that capture the curvilinear characteristics of native lexical tones, but not for pitch contours represented by a linear rising slope which does not occur in real Mandarin Chinese speech. Chandrasekaran, Krishnan, and Gandour (2007b) found that whether Chinese listeners showed an enhanced MMN over English listeners depended on the acoustical salience of the tone contrasts. In Chandrasekaran, Krishnan, and Gandour (2009), non-speech homologues that were modelled on pitch differences within or between different Chinese tonal categories were presented to native Chinese speakers, native English musicians, and native English non-musicians. Regardless of the within- or between-category condition, the native Chinese listeners and the English musicians exhibited larger MMNs than the English non-musicians. In contrast, Kaan, Barkley, Bao and Wayland (2008) showed that although behaviorally English listeners discriminated Thai lexical tones less well than Chinese and Thai listeners, they showed a larger MMN compared to Chinese listeners. These inconsistent findings suggest that being a native speaker of a tone language modulates pitch MMN in a stimulus-specific way, and tone language speakers do not necessarily exhibit enhanced MMN compared to non-tone language speakers for lexical tones.

The above studies focused on the effect of language experience on neural responses to speech or non-speech pitch, but the non-speech stimuli lacked ecological validity and real-life function. As both language and music are universally used across human cultures to serve fundamental communicative purposes (Patel, 2008), it is crucial for understanding experience-dependent neural plasticity and the integration of high-level cognitive abilities that studies are conducted on the effect of language experience on neural processing of *music* – ecologically valid music stimuli, namely melodies that can occur in real life music. Therefore, in this study, as well as lexical tones, we used distinct three-note musical melodies with comparable pitch contours to the tones to investigate whether being a native speaker of a tone language or not modulates neural responses to lexical tones, and whether such modulation extends to the music domain.

How language experience may affect MMN lateralization is poorly understood. When presented with native lexical tones, MMN lateralization in tone language listeners differed across studies: some have found right lateralization (Luo et al., 2006; Ren, Yang, & Li, 2009; Xi et al., 2010); others no clear lateralization (Chandrasekaran, et al., 2007b); and yet others left lateralization (Gu et al., 2013). With regard to music, several studies report a frontal-central distribution of the MMN elicited by contour violation without clear lateralization (Trainor, McDonald, & Alain, 2002; Vuust, Brattico, Seppänen, Näätänen, & Tervaniemi, 2012). While scalp distribution of the MMN response does not necessarily reflect the location at which the MMN is generated, if the processing of lexical tones and musical pitch share neural resources, we would expect consistent manifestation of MMN lateralization across different conditions.

Besides MMN, several studies on pitch/tone discrimination report a negativity following MMN, likely to be the late discriminative negativity (LDN) (Cheour, Korpilahti, Martynova, & Lang, 2001). LDN is more frequently observed in children than adults, and tends to be more evident for speech than non-speech stimuli (Cheour et al., 2001). It has been suggested that, when presented with unfamiliar auditory stimuli, the LDN may reflect the transfer of the newly-encountered regularity into long term memory (Peter, Mcarthur, & Thompson, 2012; Zachau et al., 2005). Consistent with this hypothesis, Kaan et al. (2008) found an LDN in Chinese and English listeners presented with Thai tones, which decreased in amplitude after training, suggesting transfer of lexical tone information into long term memory. Whether LDN and MMN reflect the same change detection mechanism is still debated (Ceponiene et al., 2004; Čeponiene, Cheour, & Näätänen, 1998; Korpilahti, Krause, Holopainen, & Lang, 2001), and the cognitive function of LDN is not yet fully understood. Nevertheless, based on the research so far, we hypothesise that it is more likely to observe an LDN for lexical tones, if at all, in non-tone language speakers than in tone language speakers.

The aim of this study is to understand how language experience might shape neural responses to pitch change in different domains. We investigated Chinese and Dutch listeners’ MMN to Chinese lexical tones and to simple musical melodies. If there is a tone language benefit in pitch perception in general, we expect Chinese listeners to show more pronounced MMN than the Dutch listeners for both the music and lexical tone stimuli. In addition, based on the split hypothesis (Chen, Liu, et al., 2016; Chen, Roncaglia-Denissen et al. 2016), if the phonemic status of lexical tones for tone language listeners induces different neural responses than those to other non-phonemic pitch variations, we expect a dissociation between MMNs in the lexical tone condition and the music condition for the Chinese listeners. The non-tone Dutch listeners, on the other hand, should exhibit similar neural response in each condition, as both lexical tones and musical melodies should be processed in a psychoacoustic manner.

2. Methods and materials

2.1. Participants

Sixteen native Dutch language participants (NL hereafter, mean age 24.3 years, SD = 7.6 years, three males), and 16 native Mandarin Chinese language participants (CN hereafter, mean age 29.6 years, SD = 4.1 years, four males) were recruited for the experiment. All the participants are non-musicians, and none of the participants had more than three years of musical training. The Dutch participants were all raised and educated in the Netherlands, and seven of them had had music lessons (mean lesson duration = 1.6 years, SD = 0.8 years). None of the Dutch participants reported knowledge or experience with any tone language. The Dutch participants were either attending a university program or working at the time of the experiment. The Chinese participants were all raised in China with Mandarin as their first language and were tested in the Netherlands. They had finished at least high school education in China. They were attending a Dutch postgraduate degree program or exchange program at the time of experiment. Six of them had had music lessons (mean lesson duration = 1.5 years, SD = 1.9 years). Except for one Chinese participant who reported to play *Guqin* (a traditional Chinese string instrument) once a week, none of the participants were practicing music at the time of the experiment. A univariate ANOVA with language group being the independent variable did not find significant difference between the groups in terms of years of music training, F(1, 30) = 0.46, n.s. The Chinese participants were frequently exposed to Western music. All participants were right-handed, and none reported hearing or language impairment. All experimental methods used in the study were approved by the Ethics Committees for human research at Utrecht University (CHEN0108-ID-01-2015).

2.2. Stimuli

The Chinese rising tone (tone 2, T2) and dipping tone (tone 3, T3) were used as stimuli, each on the tone-bearing syllable /ma/. The speaker was a female native Chinese speaker, who was recorded speaking /ma2/ and /ma3/ in isolation in a sound-proof phonetic laboratory equipped with a DAT Tascam DA-40 recorder and a Sennheiser ME-64 microphone. The duration of the two tone contours (one token for each tone) were normalised to 667 ms. The pitch contours of naturally-produced /ma2/ and /ma3/ were extracted using Praat (Boersma & Weenink, 2011), and the pitch contours of the T2 were re-synthesized onto the original T3 syllable using the overlap-add method (Moulines & Laroche, 1995). In this way, the two lexical tones only differed in pitch and had identical segments and duration. The pitch of the initial 30 ms of both syllables was adjusted manually in Praat so that they began at the same F0 level. With respect to amplitude, both syllables were scaled to 70 dB in Praat. Five native Chinese speakers listened to the stimuli and were in agreement that all the stimuli sounded like natural speech.

The music stimuli were generated to match the pitch contours of the lexical tones. Piano tones F3 (175.61 Hz), F#3 (185.00 Hz), C#3 (138.59 Hz), G3 (196 Hz) and A#3 (233.08 Hz) were synthesised in Nyquist software (http://www.cs.cmu.edu/~music/music.software.html), which follows equal temperament tuning, with the frequency of middle A (A4) being the usual 440 Hz. The notes were generated with the default duration of one-16th of a note (250 ms) in Nyquist. Then F3, G3 and A#3 were concatenated to form a rising melody, and F3, C#3 and F#3 concatenated to form a dipping melody. In order to ensure continuity and naturalness, the whole melody was then adjusted in Praat using the overlap-add method to a duration of 667 ms, which resulted in slightly different durations for each note; 220 ms for the first, 226 ms for the second, and 221 ms for the third note. As can be seen in Fig. 1, the lexical tones and the musical melodies shared the same pitch onset and offset, and exhibited comparable pitch movement over the duration of the stimuli, but were each clearly lexical tones (continuous pitch contour realized by human voice) and musical melodies (discrete notes) respectively.

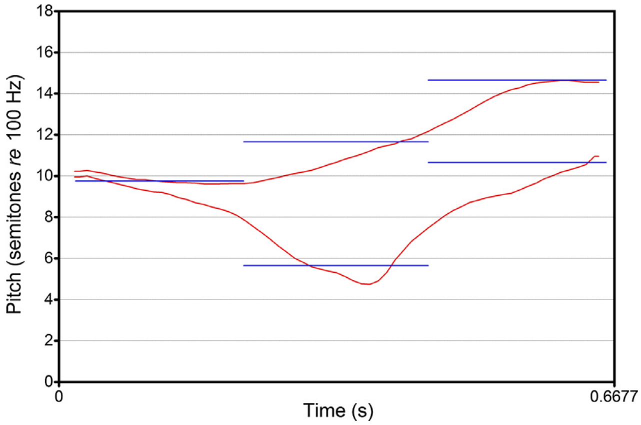

Fig. 1. Pitch contours of the musical melodies (blue) and the lexical tones (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Paradigm

MMNs was recorded in two blocks: a lexical tone block and a music block. Block order was counterbalanced across participants. In the tone condition, the standard was the T2 syllable, and deviant the T3 syllable; in the music condition, standard was the corresponding rising melody and deviant was the corresponding dipping melody.

Each block comprised 600 stimuli, of which 480 (80%) were standards and 120 (20%) deviants. Each block began with 10 repetitions of the standard; after which standards and deviants were presented in a pseudo-random order with the constraint that deviants were separated by at least two standards. The inter-stimulus interval (ISI) was randomly varied between 450 ms and 550 ms.

2.4. Procedure

The participants were tested in Utrecht, the Netherlands. The experiment was conducted in a sound attenuated room. During measurements, a participant-selected movie was played on a computer screen in silence with subtitles. The distance between the participant's eyes and the screen was ~1 m and the screen was placed in between two audio speakers which presented the experimental stimuli. EEG was recorded using a BioSemi system with 64 active Ag-AgCl electrodes from the scalp with the international 10–20 system layout with a sampling rate of 2048 Hz, and two additional active electrodes were placed at the right and left mastoid respectively. Horizontal eye movements (HEOG) were recorded by electrodes placed on the outer canthi of each eye. Vertical eye movements (VEOG) were recorded by electrodes vertically placed above and below the left eye. Electrode offset was adjusted to below 25 μV at the beginning of the experiment. The stimuli were presented at 70 dB SPL. For frequency MMN, as previous studies have shown no difference in MMN amplitude between the

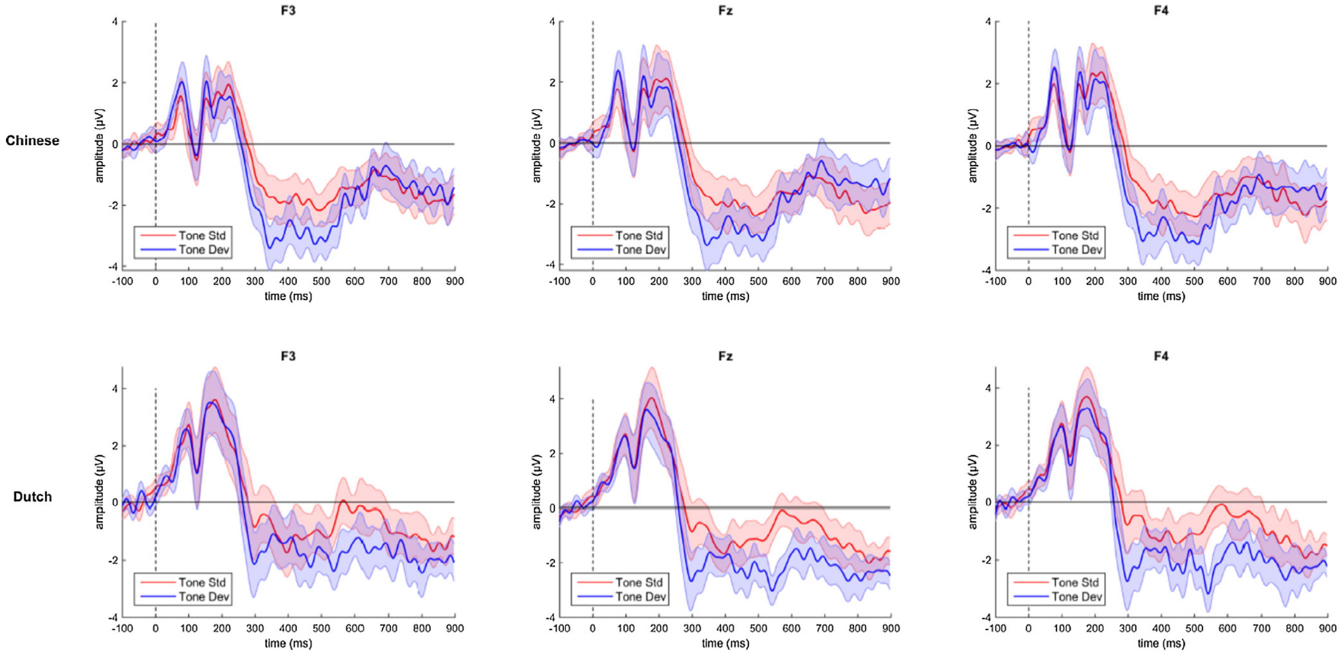

Fig. 2. Grand averaged response to the standard and deviant in the lexical tone conditions for Chinese and Dutch listeners. Shaded area indicates 95% confidence interval.

deviant-standard calculation and deviant-control calculation (i.e., deviant served as standard in a different control block) (Peter et al, 2012), and in order to ensure that the experiment could be finished within a reasonable time span, we chose not to run a separate block presenting only the deviant stimuli.

2.5. EEG analysis

All processing was conducted with the EEGLAB toolbox (Delorme & Makeig, 2004) (version 13.4.4b) in Matlab 2011b. The raw recordings were first down-sampled to 250 Hz, then filtered between 0.3 and 30 Hz. The continuous recordings were re-referenced to left and right mastoid, and segmented into 1000 ms epochs, starting at 100 ms before the onset (baseline) to 900 ms after the onset of the stimuli. Bad channels were visually identified and removed. Independent component analysis (ICA) was performed to remove ocular artifacts ('run_ica' function in EEGLAB). Independent components with known features of eye blinks and eye movements (based on activity power spectrum, scalp topography, and activity over trials) were identified visually for each participant and removed. Next, trials with amplitude greater than ± 100 μV were removed. After artifact reduction, values for the removed bad channels were interpolated. The artifact free trials were averaged to obtain the ERP waves for each participant. The mean number of accepted deviants in each condition of each participant group were: for the music condition, CN = 109 (SD = 9.7) and NL = 107 (SD = 12.3); for the lexical tone condition, CN = 108 (SD = 16.6) and NL = 113 (SD = 4.2). The difference waves were computed by subtracting the ERP to the standard stimulus from the ERP to the deviant stimulus. Individual waveforms were averaged to obtain the grand averaged waveform.

For MMN and LDN, to determine whether the responses to the standard and to the deviant differ significantly, point-by-point t-tests were conducted in a 21-time-point window surrounding the grand average peaks (10 points before and 10 points after), and if at least six consecutive time points showed a significant difference ($p < 0.05$, two tailed) between the response to the standard and the deviant, the mismatch peak was considered significant (Rinker et al., 2007). MMN and LDN peak latencies of each individual participant were identified as the latency of the most negative peak in a 40 ms window surrounding the grand average peaks (20 ms before and after) at Fz. MMN and LDN amplitude of each individual participant was calculated as the mean amplitude in the aforementioned 40 ms windows at F3, Fz, and F4.

The significance of the MMN and LDN responses were also tested using non-parametric cluster based mass permutation tests (Maris & Oostenveld, 2007) as implemented in Fieldtrip toolbox in Matlab (Oostenveld, Fries, Maris, & Schoffelen, 2011). This analysis was completely data driven and included all electrodes and all time points between 0 and 900 ms. First, a series of *t*-tests was computed at each electrode and at each time point. Then, clusters were formed over space by grouping electrodes (at least 2 adjacent electrodes) that had significant initial *t*-test results ($p < .05$) at the same time point. Clusters were formed over time by grouping adjacent time points that had significant t-values ($p < .05$). The sum of all t-values within each cluster provides a cluster-level t-score (mass t-score). A permutation approach was used to control for type I errors. For this the standard and deviant waveforms were randomly swapped and the *t*-tests were repeated 5000 times to generate a data driven null hypothesis distribution. The observed t-values from the first step was compared with the null-hypothesis distribution. The cluster was considered significant if the mass t-score fell in the top 2.5 or bottom 2.5 percentile of the distribution.

The cluster permutation statistics approach yields a conservative measure and there is a trade-off between sensitivity to local strong effects versus sustained smaller effects, which are diffused across scalp locations (Brusini et al., 2017). Therefore we used both the data-driven (cluster statistics) and literature-driven analyses with respect to MMN/LDN peaks.

3. Results and discussion

Results and discussion for each specific aspect, lexical tone, music and tone-music correlations are reported in turn, followed by a more general discussion and conclusions.

3.1. Lexical tone

3.1.1. Results

In line with many previous studies (e.g., Jacobsen & Schröger, 2003; Peter et al., 2012) three electrodes, F3, Fz, and F4 generated the largest

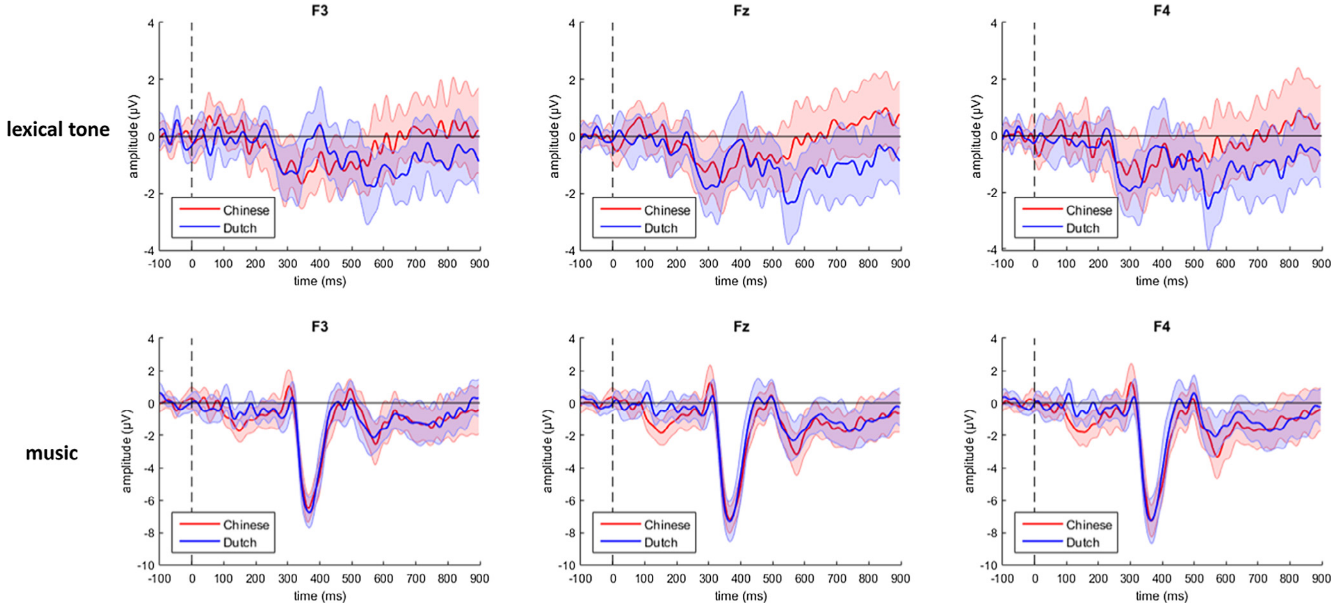

Fig. 3. Difference waves (deviant-standard) in the lexical tone and the music condition for the Chinese and the Dutch listeners. Shaded area indicates 95% confidence interval.

Table 1
Mean peak latency (ms) and mean amplitudes of MMN and LDN (μV) of the Chinese and the Dutch listeners in the lexical tone condition. Standard deviations are given in parentheses.

			CN	NL
MMN	latency at Fz		351 (8.50)	295 (13.99)
	amplitude	F3	−1.47 (1.57)	−1.23 (2.81)
		Fz	−1.42 (1.70)	−1.72 (2.53)
		F4	−1.35 (1.81)	−1.85 (2.60)
LDN	latency at Fz			550 (13.06)
	amplitude		F3	−1.59 (3.01)
			Fz	−2.19 (2.39)
			F4	−2.19 (3.02)

amplitude MMNs and were thus selected for statistical analysis. Fig. 2 plots grand average ERPs to the standard (T2) and deviant lexical tone (T3) for CN and NL, and Fig. 3 plots the difference waves in each condition. MMN peaks were identified in the 200–400 ms window at Fz after the stimulus onset for each language group. MMN peak latencies of the grand average of NL and CN are 292 ms and 344 ms respectively. As shown in Fig. 2, after the MMN, a second peak can be identified for the NL group on the difference wave, which is most likely the late discriminative negativity (LDN). For NL, the second peak was identified in the 400–600 ms window at Fz on the grand averaged response, with the peak latency being 548 ms. The mean latency and amplitude of MMN and LDN are given in Table 1. Based on the 21-point *t*-tests, the MMN peaks of both language groups, and the LDN of NL were significant. When a more conservative non-parametric cluster analysis was performed, the following time windows were found to be significant: for NL, 268–340 ms and 528–588 ms, corresponding to MMN and LDN; for CN, 328–500 ms, corresponding to MMN. Fig. 4 shows the topography of the MMN and LDN amplitudes at peak latencies for each language group.

For MMN latency, a univariate ANOVA was conducted with language group as the independent variable. A significant effect of language was found, $F(1, 30) = 192.32$, $p = 0.000$, partial $\eta^2 = 0.87$, where NL had a significantly earlier MMN peak than CN.

For MMN amplitude, a mixed-effect ANOVA was conducted with site (F3, Fz, or F4) as a within-subject variable and language (CN, NL) as a between-subject variable. No main effect of site, $F(2, 60) = 0.63$, $p = 0.54$ or language, $F(1, 30) = 0.06$, $p = 0.8$ was found, nor was the site by language interaction significant, $F(2, 60) = 1.22$, $p = 0.3$. For both groups, no significant MMN lateralization was found.

3.1.2. *Discussion*

Qualitatively, CN and NL exhibited similar brainwave morphology for the standard and the deviant. The two lexical tones began with the same pitch level, so it is possible that only when the F0 difference between the two tones reaches some threshold, the brain begins to respond to the F0 change. The later MMN latency for CN implies that a larger F0 difference was necessary for Chinese listeners to detect the tonal change. This is consistent with the notion that native tone-language listeners perceive lexical tones in a categorical manner such that there is insensitivity to differences between two pitch contours belonging to a single tone category, but sensitivity to a pitch difference that spans a boundary between two tones (Francis, et al., 2003; Hallé et al., 2004). As T2 and T3 share a similar pitch onset, and as the MMN emerges as a function of time, the Chinese listeners here may have been unresponsive to point-by-point T2-to-T3 F0 differences, responding to the deviant T3 as a possible member of the T2 lexical tone category until the pitch decrease in T3 becomes large enough to cross the T2/T3 boundary and trigger the perception of T3. Yet as the pitch contours of the lexical tones are continuous, it is difficult to pinpoint the exact time point where the difference between the standard and the deviant became noticeable for the brain. The non-native NL listeners presumably perceive the lexical tones psycho-acoustically and would thus be capable of responding to the within-category acoustic differences sooner, resulting in a shorter latency MMN. Alternatively, timing of the turning point (e.g., the location of minimum pitch) of the pitch contour and the magnitude of pitch change between the onset and the turning point have been found to be perceptually relevant for identification of T2/T3 for native speakers. Hence it might be the case that Chinese and Dutch listeners relied on different acoustic information for the detection of the change from T2 to T3, leading to different peak latencies of the MMN.

Unlike NL, CN did not show a clear LDN to their native lexical tones. While far from being well understood, it has been suggested that LDN might relate to the transformation of recurrent regularity to long term memory (Cheour et al., 2001; Peter et al., 2012; Zachau et al., 2005). Our findings are consistent with this notion: Here NL would have been responding to non-native pitch patterns, hence the clear LDN for this

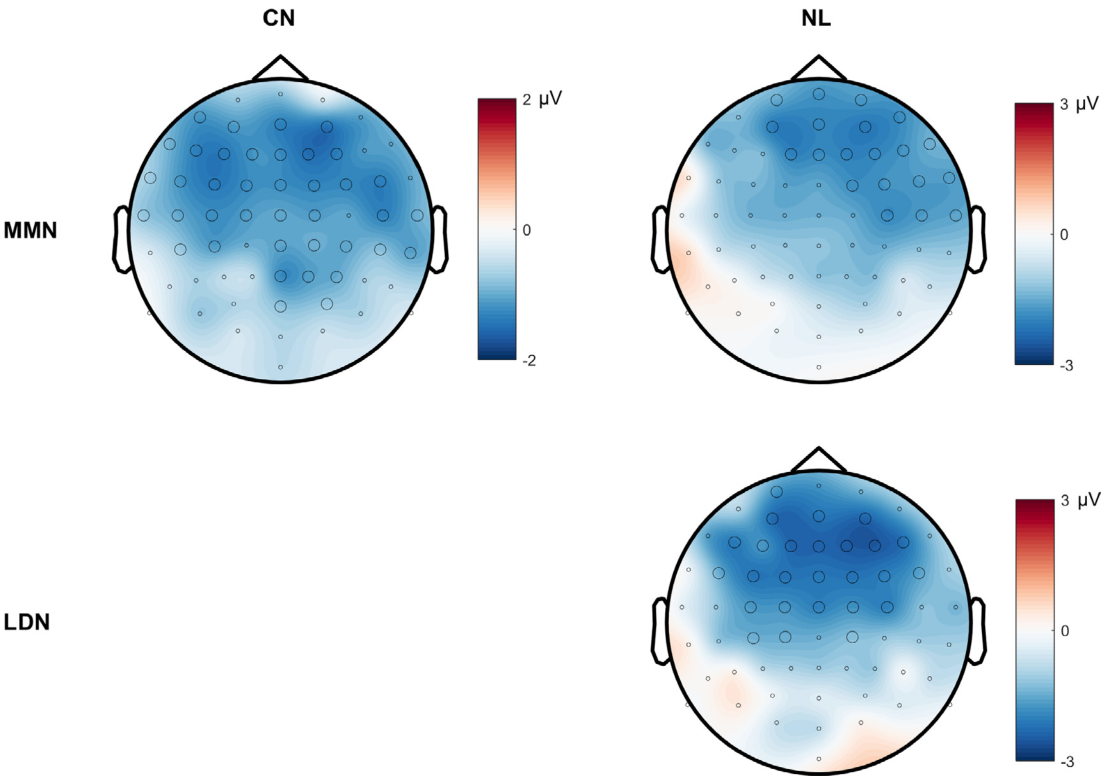

Fig. 4. Scalp distribution of MMN and LDN at corresponding peak latencies (MMN: CN = 344 ms, and NL = 292 ms; LDN: NL = 548 ms) in the lexical tone condition. The highlighted electrodes belong to a statistically significant cluster.

groups may indicate the incorporation of acoustic differences into memory. CN, on the other hand, were responding to their native lexical tones which would already be well represented in memory.

In accord with Chandrasekaran, et al. (2007a), Chandrasekaran, et al. (2007b), and Chandrasekaran et al. (2009), we did not find either left or right lateralization of MMN in the tone language, CN, listeners. A possible reason for this might be that only one token of both the standard and the deviant was used. Although the two lexical tones were clearly identified as native tones by Chinese judges, discrimination on the basis of *phonological* (*tonological*) categories may have been compromised by the lack of within-category variation. If variable standards and deviants had been used, phonological processing of the lexical tones might have been boosted, resulting in a more left lateralized MMN (but see Jacobsen, Schröger, & Alter, 2004; Shestakova et al., 2002). Further research on this is required.

These results demonstrate language-specific patterns when responding to lexical tones, and suggest that different neural mechanisms for change detection may be at play for tone versus non-tone language speakers.

3.2. Music stimuli

3.2.1. Results

Fig. 5 shows the grand averaged ERPs to the standard (F3-G3-A#3, rising) and the deviant melodies (F3-C#3-F#3, dipping) for CN and NL. The difference waves are shown in Fig. 3. As can be seen, for both groups, three peaks can be identified on the difference waves (Fz between 200 and 400 ms, 400–600 ms and 600–800 ms respectively) with the first two peaks being MMNs to the second and the third note, and the third possibly being an LDN. Table 2 provides the grand average peak latencies of CN and NL. Scalp distributions of the MMNs and the LDN are plotted in Fig. 6.

Significance of the MMN and LDN peaks was tested using point by point t-tests, as for the lexical tone condition, with a threshold of 6 consecutive points being significant within the 21 time points surrounding the grand average peak. For both CN and NL, all note2 MMNs, note3 MMNs and the LDN were significant. When the more strict non-parametric cluster analysis was performed, for the Dutch listeners, the time windows 328–374 ms, and 724–780 ms turned out to be significant, corresponding to MMN note2 and LDN, and the time window 544–596 ms was marginally significant ($p = 0.09$), corresponding to MMN note3; and for the Chinese listeners, the time windows 328–436 ms and 520–628 ms turned out to be significant, corresponding to note2 and note3 MMNs.

As in the lexical tone condition, individual MMN peak latencies and peak amplitudes were identified in the 40 ms windows surrounding the grand average peaks. Mean peak latencies and amplitudes of MMN and LDN for both the groups are shown in Table 3. First with respect to latency, for each peak, a univariate ANOVA was conducted with language group as the independent variable. Neither for note2 nor for note3, significant effect of language was found, $F(1, 30) = 0.45$, $p = 0.51$, $F(1, 30) = 1.38$, $p = 0.25$. For LDN, language showed a significant main effect, $F(1, 30) = 13.96$, $p = 0.01$, partial $\eta^2 = 0.32$, where NL had a significantly later peak than CN.

With regard to amplitude, three mixed-effect ANOVAs were conducted, one each for notes2 MMN, note3 MMN, and LDN with sites (F3, Fz, F4) as a within- and language group as between-subjects variable. For note2, the effect of sites was significant, $F(2, 60) = 5.73$, $p = 0.005$, partial $\eta^2 = 0.16$, with F3 having a smaller amplitude (less negative) than Fz. The factor language was not significant, $F(1, 30) = 0.12$, $p = 0.74$. There was no significant interaction between language and sites, $F(2, 60) = 0.14$, $p = 0.87$. For note3, site again was significant, $F(2, 60) = 3.98$, $p = 0.024$, partial $\eta^2 = 0.12$; F3 had a smaller amplitude than Fz, with no other significant pairwise comparison. Language was not significant, $F(1, 30) = 0.59$, $p = 0.45$, and the sites by language group interaction was marginally significant, $F(2, 60) = 3.19$, $p = 0.062$. For LDN, sites did not show significant main effect, $F(2, 60) = 1.79$, $p = 0.19$, and neither main effect of language, F

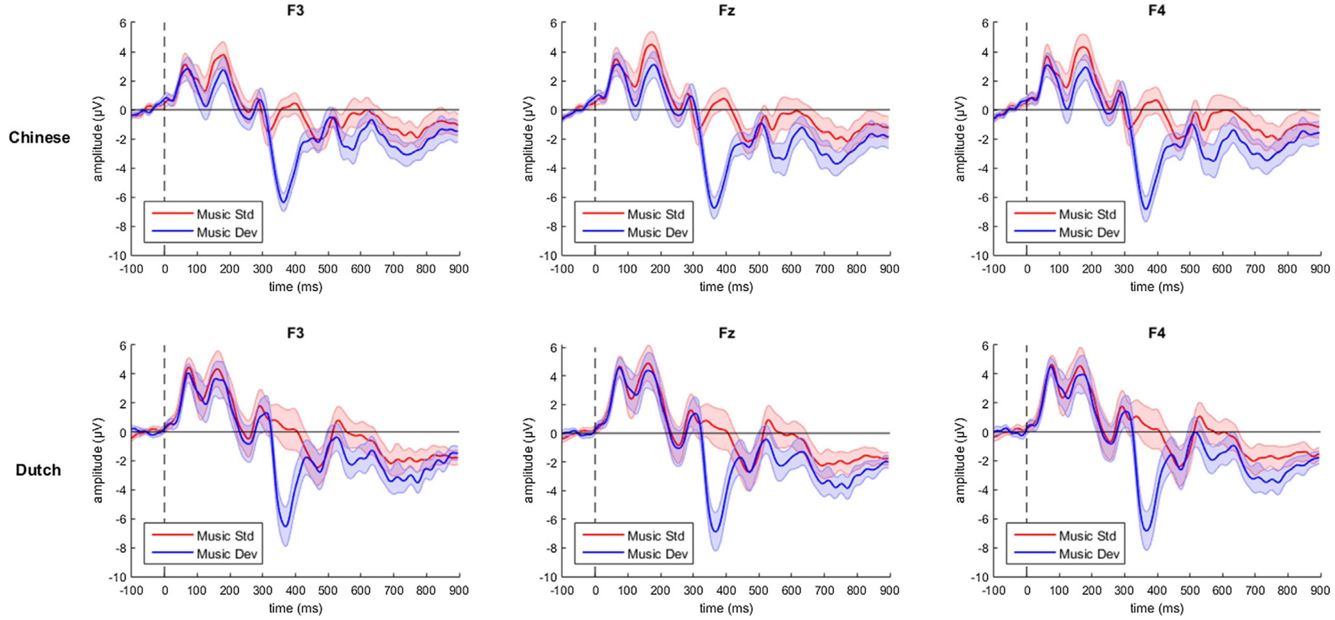

Fig. 5. Grand average response to the standard and the deviant in the music condition at F3, Fz, and F4 of the Chinese and the Dutch listeners. Shaded area indicates 95% confidence interval.

Table 2
Peak latencies (ms) on the grand average for note2 MMN, note3 MMN, and LDN at Fz of the three language groups in the music condition.

	Note2 MMN	Note3 MMN	LDN
CN	364	576	724
NL	364	564	740

$(1, 30) = 0.01$, $p = 0.93$, nor the language by sites interaction, $F(2, 60) = 0.38$, $p = 0.90$ was significant.

3.2.2. Discussion

As in the lexical tone condition, the two language groups exhibited similar ERPs. For both groups, the MMN to note2 is the most pronounced, with a much larger amplitude than the MMN to note3. There are two possible explanations for this. First, it is at note2 onset that the two melodies started to differ, and as this was the first change that listeners encountered, it may elicit a larger MMN. This interpretation is corroborated by the very consistent MMN peak latency between the two language groups. Note3, on the other hand, was the second change. It is likely that the neurons were not fully reset from the previous MMN by the time note3 occurred, leading to a less prominent MMN response to note3. Second, note2 changed the contour (i.e., note1 and note2 constituted a rise in the standard and a fall in the deviant) whereas note3 did not (note2 and note3 constituted a rise in both standard and deviant, with the standard and deviant differing in interval size of the rise). The contour change is likely to be more detectable than the interval change (e.g., Trehub, Bull, & Thorpe, 1984; Trehub & Hannon, 2006). These alternatives require further research for their resolution.

Tone language CN and non-tone language NL participants showed comparable MMN amplitudes. In this regard it should be noted that the change at note2 was highly perceptible, and the response would not be affected by any previous changes. Hence, it may be the case that NL and CN both performed at ceiling for MMN for note2. In other words, when a pitch change is salient, tone language listeners do not necessarily outperform non-tone language listeners. It is worth noticing that for the musical melodies, separate MMNs for individual notes are observed. Auditory input unfolds over time sequentially, and these results show that the brain is capable of registering each individual event in the auditory stream clearly and sequentially in a time-locked fashion. Clear onset of individual events helps the brain to segment the auditory stream.

With regard to LDN, given that the two melodies differed not only locally (i.e., at note2 and note3), but also globally (i.e., they exhibited different contours), the LDN in the music condition might reflect detection of global contour change. Due to the sequential nature of the auditory stream, a global pattern can only be established after the individual components have been presented. CN showed an earlier yet less robust LDN than NL, which may reflect a faster integration of pitch-based components into a global percept as a product of their tone language experience. LDN may have been boosted had more participants been tested. Investigation of the exact neural function of LDN requires further research.

3.3. Cross-condition correlation – Results

For both language groups, we investigated the correlation between the MMN peak latencies and amplitudes in the speech and music conditions. The LDN for the music condition was also included in the correlation analysis as it was found for both groups. For latency, none of the correlations were significant, possibly due to the consistency of the responses: MMN and LDN peak latencies showed limited variation, which may have made correlations, if any, difficult to discern.

Correlations for MMN amplitudes in the different conditions at Fz are shown in Table 4. Fig. 7 shows the scatter plots of the lexical tone MMN amplitudes versus music note2 MMN and music LDN amplitudes for CN and NL separately.

For NL, there was a positive correlation between music note2 MMN and lexical tone MMN, and between music LDN and lexical tone MMN. MMN amplitude reflects the perceived difference between the neural representations of the standard and the deviant sound. For these non-tone language listeners, the positive concordance between the degree of neural response to the lexical tone stimuli and the music stimuli, indicates the possible existence of a unified neural mechanism of change detection. CN, on the other hand, did not show such correlations, which suggests neural activation for lexical tones differs from that for musical pitch. In other words, it could be said that lexical tones are perceived as musical in non-tone NL speakers, but not in tone CN speakers. We tested the significance between correlations using Fisher's Z test. For note2,

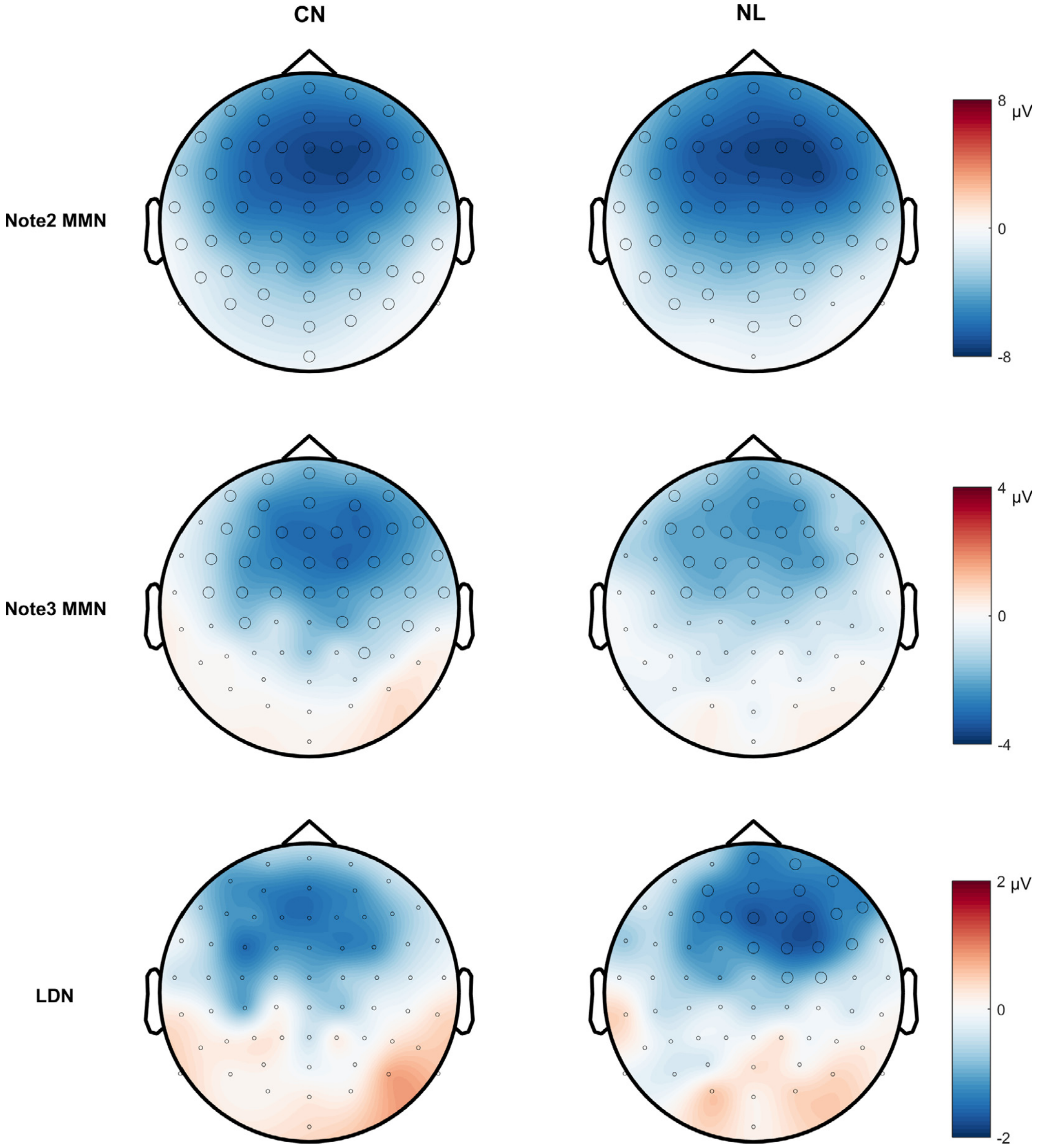

Fig. 6. Scalp distribution of MMN and LDN at corresponding peak latencies for each language group in the lexical tone condition. The highlighted electrodes belong to a statistically significant cluster.

Table 3
Mean peak latencies (ms) and mean amplitudes (μV) of MMN and LDN of the three language groups in the music condition. Standard deviations are given in parentheses.

			CN	NL
MMN note2	latency at Fz		366 (9.69)	368 (9.24)
	amplitude	F3	−5.91 (1.54)	−6.3 (2.75)
		Fz	−6.62 (2.15)	−6.82 (2.68)
		F4	−6.54 (2.20)	−6.78 (3.02)
MMN note3	latency at Fz		573 (12.24)	568 (13.05)
	amplitude	F3	−2.09 (1.67)	−1.92 (3.12)
		Fz	−2.81 (1.99)	−2.13 (2.77)
		F4	−2.90 (1.99)	−1.85 (2.58)
LDN	latency at Fz		722 (13.07)	740 (14.89)
	amplitude	F3	−1.33 (2.17)	−1.27 (2.10)
		Fz	−1.70 (2.28)	−1.58 (1.63)
		F4	−1.64 (1.80)	−1.64 (1.80)

Table 4
Pearson's r between MMN amplitudes in different conditions in for NL and CN.

		Music note2	Music note3	Music LDN
NL (Fz)	Lexical tone	0.50*	−0.10	0.51*
	Music note2		0.47#	0.68**
	Music note3			0.48#
CN (Fz)	Lexical tone	0.36	0.10	0.21
	Music note2		0.53*	0.47#
	Music note3			0.40

** Significance at 0.01 level.
* Significance at 0.05 level.
# Marginal significance (0.05 < p < 0.08).

the difference is not significant, z = 0.44, p = 0.33; for note3, the difference is not significant, z = −0.51, p = 0.31; for LDN, the difference is not significant, z = 0.89, p = 0.19. Seeing the potential type I error due to multiple comparisons, we further conducted permutation tests (Yoder, Blackford, Waller, & Kim, 2004), using the function *MPT.corr* function for R package https://www.psych.umn.edu/faculty/waller/

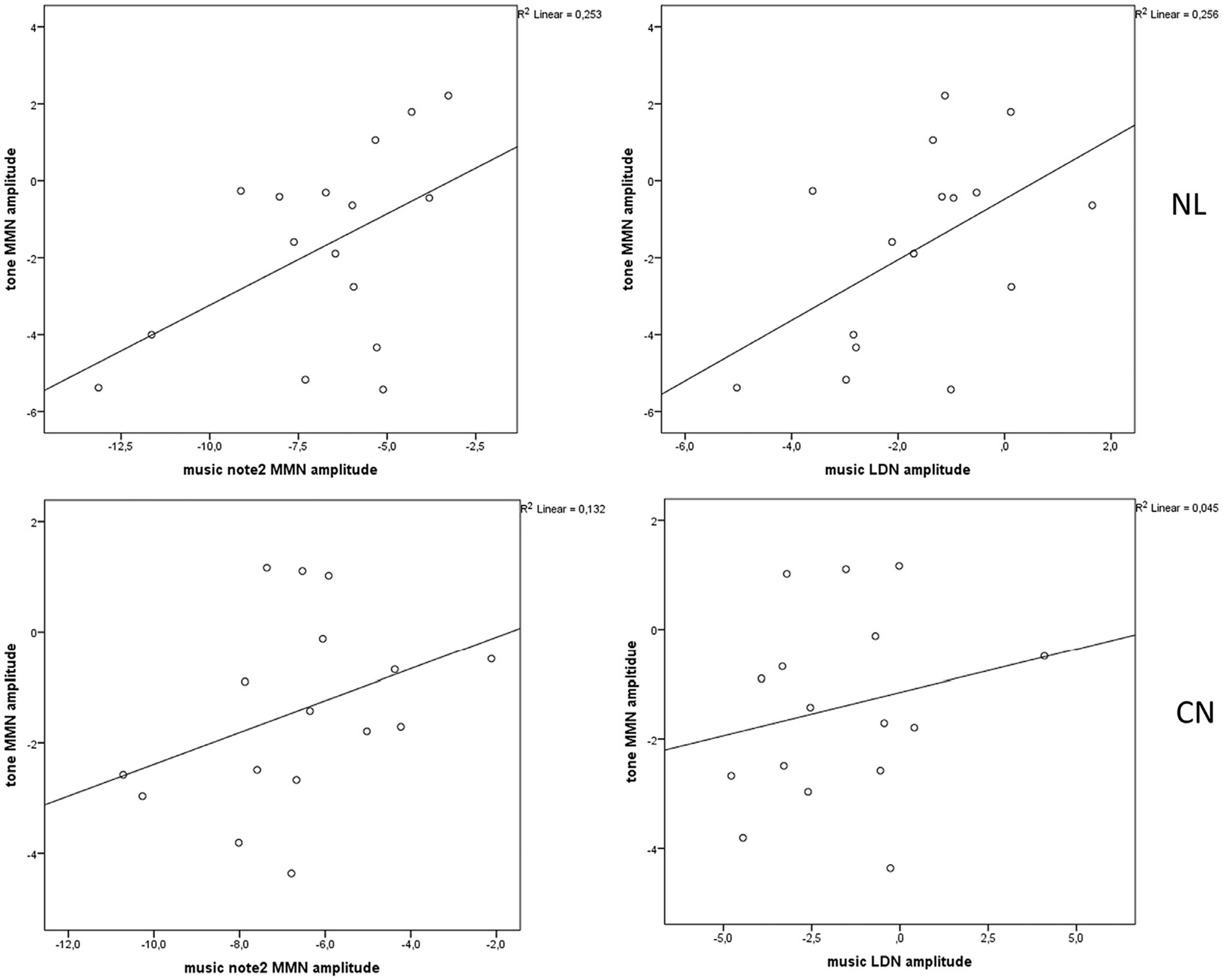

Fig. 7. Scatter plots of lexical tone MMN amplitude versus music note2 MMN and music LDN amplitude for CN and NL.

downloads/mpt/mptcorr.r. For the Dutch listeners, when a random 1000-time permutation was conducted, the lexical tone-note2, lexical tone-music LDN correlations turned out to be significant (one tail, positive correlation), with exact p values being 0.046 and 0.034. For the Chinese listeners, no significant correlation was found using the permutation tests. Admittedly, the difference in music-lexical tone correlation between the Chinese and Dutch listeners should be interpreted with caution. Nevertheless, the qualitative difference between the two language groups is unlikely to be artefacts. In Chinese, every syllable must carry a tone, and the native lexical tones are well stored in the long term memory of native speakers. The MMN to native phonemic difference is likely to reside on both short- and long-term memory traces, and exhibit different generators in comparison to non-phonemic differences. For native phonemes, not only traces in short-term memory were activated for the generation of MMN, but also recognition of phonological forms stored in long-term memory. Thus the phonemic function of the lexical tones for Chinese listeners may result in a different auditory neural network being used for lexical tones than for other non-speech (and even non-phonemic) pitch variations.

4. General discussion

To investigate how speaking a tone language natively influences neural detection of pitch change in language and music as well as their correlations, we tested Chinese and Dutch listeners' MMN in response to Chinese lexical tones and three-note musical melodies. We manipulated lexical tones and musical stimuli so that they were comparable in both pitch level and pitch contour, but despite these similarities, they capture specific features of language and music: the lexical tones are continuous pitch variations realized by human voices, and the musical melodies comprise discrete pitches without voices. With this in mind, it is notable that ERPs to lexical tones and to musical melodies exhibit very different waveforms, capturing the particular characteristics of the stimuli. Neural resources are possibly exploited domain generally when processing pitch, yet the brain also responds to the domain-specific features in music and language.

We did not find enhanced MMN for tone language speakers in either the lexical tone or the music condition. Thus it appears that when a pitch change is sufficiently salient, non-tone language listeners do not necessarily show attenuated neural detection compared to tone language listeners. With regard to the lexical tones, these results are consistent with the behavorial results in Chen, Liu, and Kager (2015), which showed that Dutch and Chinese listeners had comparable accuracy when discriminating monosyllabic Chinese lexical tones. Although it is difficult for non-native listeners to acquire lexical tones, such difficulties may not be due to lack of discrimination at the *acoustic* level. Rather, top-down linguistic processing, which for non-tone language speakers listening to what is obviously speech, albeit not speech in their native language, precludes incorporation of syllable-level pitch variations into stored representations. It would be useful for future studies to investigate whether the comparable MMN amplitude between native and non-native listeners is particular to T2-T3, which is the acoustically least salient contrast (Hume & Johnson, 2001), or whether it also holds for other more salient tonal contrasts. In addition, we did not reverse the standard and deviant in the current study, yet perceptual asymmetry in discriminating T2-T3 contrast has been observed in behavioral experiments among both native and non-native listeners, with such asymmetry being more pronounced among native listeners (Chen, Liu, et al., 2015). For future studies, it will be worth investigating whether such asymmetry, as well as any language experience induced difference, can be observed in the MMN response.

Chen, Liu, et al. (2016) found that when discriminating musical phrases in the Music Ear Test (MET, Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010), the native Chinese listeners outperformed the Dutch listeners. The current study, however, did not find more prounounced MMN among the Chinese listeners in the music condition. As the musical phrases in the MET consisted 3–8 tones and our musical stimuli were three-note miniature melodies, it is possible that when memory burden increases, tone language listeners start showing benefits in musical pitch processing. Yet whether such hypothesis holds for neural responses needs further investigation.

Nevertheless, our results show that being a native speaker of a tone language does modulate neural responses to both the native tones and musical stimuli. When responding to lexical tones Chinese listeners here showed a later MMN peak than Dutch listeners without an LDN. When listening to musical melodies, native speakers of Chinese showed an earlier LDN than the Dutch listeners. This pattern suggests that tone and non-tone language listeners' neural response to the lexical tones differed in terms of quality rather than quantity. The difference between note2 and lexical tone MMN peak latency is more evident among the Dutch listeners than among the Chinese listeners. The standard and deviant musical melodies differed from the second note, hence for both the Chinese and the Dutch listeners, it is impossible to show MMN earlier than the second note. The lexical tones, however, have continuous pitch contours, hence it is possible that the Dutch listeners detected the pitch change earlier than the time point corresponding to the onset of the second note in the musical condition, leading to an earlier MMN peak latency for lexical tones than music note2. The Chinese listeners showed a later MMN peak in the lexical tone condition than the Dutch listeners, hence there was comparable peak latency in the two conditions. Although Chinese listeners did not show enhancement in response to local pitch changes in musical melodies, they might discriminate the global contour difference more easily as shown by the earlier LDN peak. The enhanced representation of musical contour might be a carryover effect from their knowledge of the lexical tones. Yet the exact function of LDN should be further investigated.

Importantly, as predicted by the split hypothesis (Chen, Liu, et al., 2016; Chen, Peter, Burnham et al., 2016; Chen, Roncaglia-Denissen et al. 2016), MMN amplitudes between music and lexical tone are correlated for the Dutch, but not Chinese listeners. Elicitation of MMN depends on the short-term memory trace of the standard, which serves as a referent for detecting regularity violation (Näätänen et al., 2007). Beyond short-term memory, native phonological categories are well represented in long term memory, and language-specific long-term memory traces of phonological categories modulate MMN (Dehaene-Lambertz, 1997; Näätänen et al., 1997; Winkler et al., 1999, among others). When presented with phonological contrasts, the recognition of patterns stored in long-term memory may be activated (Näätänen et al., 1997). For the Dutch listeners, the top-down modulation from long term phonological memory is irrelevant for lexical tones, as this modulation is silent with respect to representation of F0 variation at the syllable level. Accordingly, for non-tone language listeners, lexical tones are processed in a similar manner to the musical stimuli. For the Chinese listeners, however, neural resources seem to be assigned differently for the lexical tones and for musical melodies, presumably depending on the presence or absence of long-term phonological memory traces. MMN is mostly generated by bilateral supratemporal cortices (Alho et al., 1998; Hari et al., 1984; Levänen, Ahonen, Hari, McEvoy, & Sams, 1996), and the lack of correlation between conditions in the Chinese listeners suggests that different cortical neural networks may be at play for the discrimination of phonologically contrastive versus musical pitch patterns. Whether such dissociation only occurs at the cortical level or is also evident at the subcortical level is worthy of further investigation. Previous studies haven found that Chinese amusics may also have difficulties with native lexical tone discrimination (Nan et al., 2016; Nan Sun & Peretz, 2010), which suggests some common pitch processing mechanisms across domains. Hence our results may also suggest that for tone language listeners neural resources are allocated differently for the processing of continuous pitch contours versus discrete pitch intervals. Further within-domain comparison is required to test such a hypothesis.

It should be acknowledged that we did not include conditions presenting only the deviant stimuli hence, although unlikely (Peter et al., 2012; Schröger & Wolff, 1996), it is still possible that the difference waves not only reflect neural change detection but also difference in neural refratoriness as the result of sensory adaptation (as the standards were presented more frequently, they have a greater refractory effect as compared to deviants). It would be useful for future studies to further compare MMN differences induced by language after ruling out any possible refractory effects. In addition, the musical stimul in the current study were tone triplets. Although they well captured the discreteness of musical pitch, they were miniatures rather than real musical phrases, so it would be useful for future study to investigate whether the brain is able to detect individual note changes when presented with melodies containing richer musical structure.

5. Statement of significance

Speech and music are two unique products of the human brain. This is the first study exploring whether language induced plasticity transfers to music. It also found the first evidence that language experience not only shapes responses within one domain, but also the correlation between neural responses in different domains.

Acknowledgement

This study was supported by Endeavor Research Fellowship funded by Australian Department of Education to the first author, with grant number ERF_PDR_113381_2013.The study was also supported by Discovery Project Grant funded by Australian Research Council to the last author, with grant number DP110105123. We also thank two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

Declarations of interest

None.

References

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. *Lisbon,* 397–400.
Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I. P., ... Ilmoniemi, R. J. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings. *Psychophysiology, 35*(2), 211–224.
Bidelman, G., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *Plos One, 8*, e60676.
Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychological Bulletin, 133*(4), 651–672.
Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (version 5.3.25).

Brusini, P., Dehaene-Lambertz, G., van Heugten, M., de Carvalho, A., Goffinet, F., Fiévet, A. C., & Christophe, A. (2017). Ambiguous function words do not prevent 18-month-olds from building accurate syntactic category expectations: An ERP study. *Neuropsychologia, 98*, 4–12. http://dx.doi.org/10.1016/j.neuropsychologia.2016.08.015.

Čeponiene, R., Cheour, M., & Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children: Evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 108*(4), 345–354.

Ceponiene, R., Lepistö, T., Soininen, M., Aronen, E., Alku, P., & Näätänen, R. (2004). Event-related potentials associated with sound discrimination versus novelty detection in children. *Psychophysiology, 41*(1), 130–141.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*, 148–156.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007a). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport, 18*(18), 1963–1967.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language, 108*(1), 1–9.

Chen, A., Roncaglia-Denissen, P. M., Roor, D., & Sadakata, M. (2016). Cross-domain pitch perception, the influence of native language. In Proceedings of 14th International Conference for Music Perception and Cognition, San Francisco, USA. pp. 364–364.

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience, 31*(6), 751–760. http://dx.doi.org/10.1080/23273798.2016.1156715.

Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of mandarin tone sandhi. *Language Sciences, 48*, 62–69. http://dx.doi.org/10.1016/j.langsci.2014.12.002.

Chen, A., Peter, V., & Burnham, D. (2016). Auditory ERP response to successive stimuli in infancy. *PeerJ, 4*, e1580. http://dx.doi.org/10.7717/peerj.1580.

Cheour, M., Korpilahti, P., Martynova, O., & Lang, A. (2001). Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and infants. *Audiology and Neurotology, 6*(1), 2–11.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport, 8*(4), 919–924.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing, 7*(3), 203–207.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.

Francis, A. L., Ciocca, V., & NG, K. B. Chit (2003). On the (non)categorical perception of lexical tones. *Attention, Perception, & Psychophysics, 65*, 1029–1044.

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology, 2*, 146.

Gu, F., Zhang, C., Hu, A., & Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in cantonese speakers as revealed by mismatch negativity. *NeuroImage, 83*, 637–645.

Hallé, P. A., Chang, Y., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395–421.

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., ... Sams, M. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neuroscience Letters, 50*(1–3), 127–132.

Hume, E., & Johnson, K. (2001). A model of the interplay of speech perception and phonology. In E. Hume, & K. Johnson (Eds.). *The role of speech perception in phonology* (pp. 3–26). Academic Press.

Jacobsen, T., & Schröger, E. (2003). Measuring duration mismatch negativity. *Clinical Neurophysiology, 114*(6), 1133–1143.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41*(4), 654–659.

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of thai, english and mandarin chinese: An event-related potentials training study. *BMC Neuroscience, 1148*, 113–122.

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language, 76*(3), 332–339.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics, 23*(1), 81–95.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 31*, 606–608.

Levänen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996). Deviant auditory stimuli activate human left and right auditory cortex differently. *Cerebral Cortex, 6*(2), 288–296.

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

Lynch, M., & Eilers, R. (1992). A study of perceptual development for musical tuning. *Perception and Psychophysics, 52*, 599–608.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science, 1*, 272–276.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience, 23*, 2701–2715.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. http://dx.doi.org/10.1016/j.jneumeth.2007.03.024.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241–265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367–1381.

Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication, 16*(2), 175–205.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., ... Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432–434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590.

Nan, Y., Huang, W., Wang, W., Liu, C., & Dong, Q. (2016). Subgroup differences in the lexical tone mismatch negativity (MMN) among mandarin speakers with congenital amusia. *Biological Psychology, 113*, 59–67.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain, 133*(9), 2635–2642. http://dx.doi.org/10.1093/brain/awq178.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*. http://dx.doi.org/10.1155/2011/156869.

Patel, A. D. (2008). *Music, Language, and the Brain.* NY: Oxford University Press.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience,* 688–691.

Peter, V., Mcarthur, G., & Thompson, W. F. (2012). Discrimination of stress in speech and music: A mismatch negativity (MMN) study. *Psychophysiology, 49*(12), 1590–1600.

Ren, G.-Q., Yang, Y., & Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience, 162*(1), 87–95.

Rinker, T., Kohls, G., Richter, C., Maas, V., Schulz, E., & Schecker, M. (2007). Abnormal frequency discrimination in children with SLI as indexed by mismatch negativity (MMN). *Neuroscience Letters, 413*(2), 99–104.

Schellenberg, E. G., & Trehub, S. E. (1999). Culture-general and culture-specific factors in the discrimination of melodies. *Journal of Experimental Child Psychology, 74*(2), 107–127.

Schröger, E., & Wolff, C. (1996). Mismatch response of the human brain to changes in sound location. *Neuroreport, 7*(18), 3005–3008.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., ... Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *Neuro Report,* 1813–1816.

Stevens, C. J., Keller, P. E., & Tyler, M. D. (2013). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music, 41*(1), 59–74.

Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience, 14*(3), 430–442.

Trainor, L. J., & Trehub, S. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 394–402.

Trehub, S., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development, 55*, 821–830.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*(1), 73–99.

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of music structure. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 965–975.

van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research, 23*(2–3), 270–276.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia, 50*(7), 1432–1443.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences, 20*(3), 188–196. http://dx.doi.org/10.1016/j.lindif.2010.02.004.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., ... Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7*(3), 357–369.

Wong, C. M. P., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*, 565–585.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170*(1), 223–231.

Yoder, P. J., Blackford, J. U., Waller, N. G., & Kim, G. (2004). Enhancing power while controlling family-wise error: An illustration of the issues using electrocortical studies. *Journal of Clinical and Experimental Neuropsychology, 26*(3), 320–331. http://dx.doi.org/10.1080/13803390490510040.

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: Early and late mismatch negativity to tone patterns. *Neuroreport, 16*(18), 2015–2019.