BRAIN RESEARCH 1148 (2007) 113–122

available at www.sciencedirect.com

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Effects of native language and training on lexical tone perception: An event-related potential study

Edith Kaan, Ratree Wayland, Mingzhen Bao, Christopher M. Barkley*

University of Florida, Linguistics, Box 115454, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Accepted 12 February 2007
Available online 21 February 2007

Keywords:
Lexical tone
Second language learning
Perceptual training
Mismatch negativity
Late negativity
Neural plasticity

ABSTRACT

Tone languages such as Thai use pitch differences to distinguish lexical meaning. Previous behavioral studies have reported that naïve listeners can discriminate among lexical tones, but that native language background affects performance. The present study uses ERPs to determine whether native speakers of a tone language (Mandarin Chinese) and of a non-tone language (English) differ in their pre-attentive discrimination among Thai lexical tones, and whether training has a different effect in these two language groups. EEGs were obtained from 10 native Mandarin Chinese speakers, 10 English and 10 Thai speakers in an oddball paradigm: The Thai syllable [k^ha:] pronounced with a high rising or low falling tone, was presented as an infrequent deviant amidst a standard mid level tone [k^ha:] syllable, while participants watched a silent movie. Next, the Chinese and English participants completed a 2-day perceptual identification training on the mid level and low falling tones, and returned for a post training EEG. The low falling tone deviant elicited a Mismatch Negativity (MMN) in all participant groups before and after training; the high rising deviant elicited no, or a smaller, MMN, which became larger after training only in the English group. The high rising deviant also elicited a later negativity (350–650 ms) versus the mid level standard, which decreased after training in the Chinese group. These results suggest that non-Thai speakers can pre-attentively discriminate among Thai tones, but are sensitive to different physical properties of the tones, depending on their native language. English speakers are more sensitive to early pitch differences, whereas native speakers of Mandarin Chinese are more sensitive to the (later) pitch contour.

1. Introduction

1.1. *Background*

Tone languages such as Thai use pitch differences to distinguish lexical meaning, that is the Thai syllable [k^ha:] pronounced with a low falling tone ("galanga, a kind of aromatic root"), means something completely different when it is pronounced with a tone that is low falling and then rising ("leg"), high falling ("I, servant"), high rising ("to do business in") or mid level ("to be lodged in"). Tone languages differ in their inventory of lexical tones. For instance, whereas Thai has five tones, Mandarin Chinese has four tones (high level, high rising, low dipping and high falling), and Yoruba three (low, mid, and high level). Because of these cross-linguistic differences, lexical tones provide a suitable means to investigate the effect of experience on speech perception, learning, and neural reorganization. Previous behavioral

* *Corresponding author*. Fax: +1 352 392 8480.
E-mail address: kaan@ufl.edu (E. Kaan).

0006-8993/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2007.02.019

studies have reported that naïve listeners can discriminate among lexical tones, but that native language background affects performance. The present study uses ERPs to determine whether (a) native language background affects pre-attentive discrimination among Thai lexical tones, and (b) whether training has a different effect in native speakers of a tone versus non-tone language.

Behavioral research has shown that native speakers of a non-tone language (e.g., English) often have difficulty perceiving tones in a tone language like Chinese (Bluhme and Burr, 1971; Kiriloff, 1969; Wang et al., 1999), whereas native experience with a tone language appears to facilitate the perception of tones in another tone language (e.g., Wayland and Guion, 2004; Wayland and Li, in press). This may be related to the difference in the way native and non-native speakers of tone languages process lexical tones. Gandour and Harshman (1978), for example, found that native speakers of a tone language (Thai and Yoruba) placed more emphasis on the linguistic tone dimensions, i.e., direction and slope of the fundamental frequency (F_0), while native speakers of a non-tone language (American English) placed more emphasis on the non-linguistic tone dimensions of average pitch and extreme endpoints. Similarly, Gandour (1983) found that, when compared to speakers of tone languages, native speakers of English paid more attention to the average F_0 and less to the F_0 contour information when identifying tones. Moreover, Lee and Nusbaum (1993) found that Mandarin Chinese, but not English listeners, were slowed down by an irrelevant change in pitch level when making a segmental classification. These findings suggest that Mandarin listeners, but not English listeners, perceive pitch and segmental information in an integral manner.

Speakers of tone and non-tone languages have also been found to employ different neural mechanisms in perceiving lexical tones. In a dichotic perception experiment, Wang et al. (2001) found that Mandarin tones were predominantly processed in the left hemisphere by native Mandarin speakers, but bilaterally by American English speakers with no prior tone experience. Similar results were found for Thai by Van Lancker and Fromkin (1973, 1978). Brain imaging studies report increased left hemisphere activation in native speakers processing lexical tones in their native language compared to non-native speakers (left frontal operculum, Gandour et al., 2000; left frontal, parietal and parietal–occipital areas, Klein et al., 2001). Non-native speakers show increased activation of the inferior frontal area in the right hemisphere (Klein et al., 2001), which has been shown to be more sensitive to non-linguistic pitch differences compared to the left (Zatorre et al., 2002).

The ability to perceive lexical tones has been shown to improve with experience. For example, Wang et al. (1999) showed that the ability of native English learners to correctly identify Mandarin tones increased from 69% before, to 90% after extensive auditory training. More recently, Wayland and Guion (2004) found that discrimination of Thai tones by native speakers of Chinese improved significantly after training. In addition, an improvement in tone identification has been found to be related to an improvement in tone production (Wang et al., 2003a). Training has also been shown to induce neurophysiological changes in native speakers of a non-tone language. English speakers who were trained to distinguish among lexical tones in Chinese showed a more extensive left hemisphere activation, and additional activation in the right inferior frontal gyrus after compared to before training, as measured by fMRI (Wang et al., 2003b).

It is important to note, however, that most of the existing data on tone perception have been obtained by having participants consciously compare or classify tone stimuli. This may have led to potential confounds due to attention, working memory or other factors. In the present study, we use ERPs to investigate how native speakers of Mandarin Chinese and native speakers of English process the distinctions between Thai lexical tones pre-attentively, that is, without paying conscious attention or performing an active discrimination task, and what effect training has on tone perception at this pre-attentive level.

One tool to investigate pre-attentive auditory discrimination is the Mismatch Negativity (MMN), or its magnetic equivalent, the MMNm or MMF (Mismatch field). The MMN is elicited to infrequent stimuli that are discriminated from frequent standard stimuli in pitch, duration, voice onset time (VOT) or other acoustic or phonetic properties. The MMN can be elicited even when participants are engaged in a different activity such as watching a movie or reading a book, or when they are asleep, which suggests that the MMN is an index of auditory discrimination at the pre-attentional level (see Näätänen, 2001, for an overview).

The MMN has been shown to increase in amplitude and, in some cases, to have a shorter peak latency as behavioral discrimination performance improves (Kraus et al., 1995; Menning et al., 2002; Näätänen et al., 1993; Tremblay et al., 1997, 1998). These changes in the MMN can even precede behavioral improvement during training (Tremblay et al., 1998). Moreover, an increased MMN can be observed for new contrasts that have not been trained on, that is, if the contrast is related to the contrast trained on and is relatively easy to discriminate (Gottselig et al., 2004; Tremblay et al., 1997).

Numerous studies have applied the mismatch negativity to study the adult perception and acquisition of non-native language contrasts (e.g., Kraus et al., 1995; Peltola et al., 2003; Tremblay et al., 1997, 1998; Winkler et al., 1999). For example, Tremblay et al. (1997) trained English speakers on the discrimination of a pre-voiced labial stop, which is not part of the English consonant inventory. ERPs after training showed an increased and longer lasting MMN. In addition, the effects of training were shown to carry over to the perception of an untrained pre-voiced alveolar stop (also not used in English). Using MEG, Menning et al. (2002) trained adult German speakers on durational contrasts of Japanese mora, showing an increase in discrimination performance as well as an increase in MMF amplitude and decrease in MMF latency after training. The MMN is therefore a fruitful tool to study perceptual discrimination at the pre-attentive level and the effects of training.

1.2. The present study

In the current study, we used an MMN oddball paradigm to investigate whether (a) there is a difference between native speakers of Thai, native speakers of a different tone language (Mandarin Chinese), and native speakers of a non-tone language (English) in the pre-attentive discrimination among Thai lexical tones, and (b) whether training has a different effect in the Chinese and English language groups. We presented the Thai syllable [k^ha:] pronounced with a high rising tone (High) and a low falling tone (Low) as infrequent deviants amidst a standard mid level tone (Mid).

We obtained ERPs from native speakers of Thai, Mandarin Chinese and American English, while they were being presented with the stimuli over headphones and were watching a silent movie, after they had been instructed that they would receive comprehension questions about the movie afterwards. The Chinese and English speakers subsequently completed a 2-day perceptual training task. Since the main focus of the study was to compare and contrast lexical tone processing and effects of training among Mandarin Chinese and American English speakers, the Thai group participated in only one training session. This was to obtain behavioral data to see to what extent the non-native listeners would differ from native Thai speakers in the identification of the tones. The perceptual training included the Mid and Low tones, but not the High tone. Previous studies have shown that the high rising tone is relatively easy to discriminate (e.g., Abramson, 1975). Not including this tone allowed us to investigate carry-over effects of training of the Low-Mid contrast to the easier High-Mid contrast. After training, the Chinese and English participants returned for a post-training ERP study, which used the same three stimuli as the pre-training ERP study.

Thai speakers were expected to show a relatively large MMN to both High and Low tone deviants versus the Mid standard tone, since all three tones are lexically distinctive in Thai. If Chinese and English speakers can (pre-attentively) discriminate among the tones, they would elicit a MMN to the deviant tones, which may be smaller than that of native Thai speakers (e.g., Näätänen et al., 1997; Peltola et al., 2003). With increased discrimination ability as a result of training, we expected the MMN to the High and Low tones to increase after the training sessions. Since behavioral studies suggest that speakers of a tone language are better at discriminating among tones in a foreign language than non-tone speakers do, we expected the MMN to the deviant tones be larger and earlier in the Mandarin Chinese speakers than in the English speakers, especially after training.

2. Results

2.1. *Question answering performance*

Participants performed well on the questions related to the movie fragments they saw during the EEG sessions [Thai: 69% correct; English: pre-training: 84%; post-training: 87%; Chinese: pre-training 80%, post-training: 86%]. This suggests that participants were paying attention to the movie.

2.2. *Behavioral perceptual training*

Percent correct identification and reaction time (of correctly identified trials that were not re-played by the participant—see Experimental procedures section) for the Low and Mid tones are given in Tables 1 and 2, respectively. Results are given for the first 25 trials of each tone of the first day of training, and the last 25 trials for each tone in the second day of training. Participants seldom replayed the trials (0.35% on average; 1.6% or less for the first and last 25 presentations per tone per group). The number of replays did not differ between the language groups, tones or training.

2.2.1. *Effects of language background (first trials of training)*

To test the effects of language background before any substantial training, we compared the data of the first 25 Mid tones and 25 Low tone trials of the first training session among the three language groups (Thai, Chinese and English). All participant groups were highly accurate in their identification of the two Thai tones at the beginning of the training. The groups did not differ in accuracy or response times [$Ps > 0.28$], and performance did not differ significantly between the Low and Mid tones [$Ps > 0.08$].

2.2.2. *Effects of training (Chinese and English only)*

To assess the behavioral effects of training in the non-native speakers, we compared performance on the first 25 Mid and Low tone trials, presented on the first day of training, with the last 25 Mid and Low trials presented on the second day, which was the last day, of training. Analysis of the identification accuracy data yielded a near-significant effect of Test time [$F(1, 18) = 4.16$, $P = 0.056$]. Identification of the Mid tone was more accurate than of the Low tone [$F(1, 18) = 6.82$, $P = 0.018$]. There were no differences between the Chinese and English groups [$Ps > 0.15$]. Analysis of the reaction time data showed that response times were significantly shorter after training [$F(1, 18) = 20.28$, $P < 0.001$]. There were no differences in reaction times between the tones [$F < 1$] or the two language groups [$Ps > 0.21$].

2.3. *Electrophysiological data*

ERPs to the High, Low and Mid tone, and difference waves (High deviant minus Mid standard; Low deviant minus Mid standard) are displayed in Figs. 1–3 for the Thai, Chinese and English groups, respectively.

All groups showed a MMN between 150 and 300 ms to the Low deviant versus the standard Mid tone, and no, or a much smaller MMN to the High deviant versus Mid standard tone. The MMN difference between the High deviant and Mid tone standard increased from pre-training to post-training in the English participants (see Fig. 3). The MMN was followed by a later negative difference, which decreased after training in the

Table 1 – Perceptual training task: identification accuracy

	Chinese		English		Thai
	First	Last	First	Last	First
Low	98.4 (2.0)	99.2 (1.6)	97.2 (5.3)	100 (0.0)	99.6 (1.3)
Mid	99.2 (1.7)	100 (0.0)	97.6 (4.3)	100 (0.0)	99.6 (1.3)

Percent correct identification and standard deviation (in parentheses) during the first and last 25 Low and Mid tone trials of training. The Thai participants only took part in one training session.

Table 2 – Perceptual training task: response latencies

	Chinese		English		Thai
	First	Last	First	Last	First
Low	472 (144)	380 (155)	538 (176)	396 (140)	570 (160)
Mid	457 (137)	367 (135)	563 (167)	382 (120)	570 (160)

Reaction times in ms, and standard deviation (in parentheses) to the Thai Mid and Low tones during the first and last 25 Low and Mid trials of training.

Chinese participants (see Fig. 2). These observations were confirmed by statistical analyses, which will be discussed separately for the early MMN and the later negative difference.

2.4. The MMN

2.4.1. Effects of native language (before training, all three language groups)

Statistical analyses on the MMN difference amplitude for the pre-training data showed a main effect of Tone: The MMN to the Low versus Mid tone was more negative (−2.38 μV) than to the High versus Mid tone (−0.75 μV), [$F(1,27)=30.98$, $P<0.001$]. There were no significant differences among the three language groups [Ps > 0.109]. The MMN peaked earlier for the High versus Mid tone (232 ms) than for the Low versus Mid tone (253 ms) [$F(1,27)=8.76$, $P=0.006$]. No latency differences were found between the groups [Ps > 0.262].

2.4.2. Effects of training (Chinese and English only)

Analysis of the pre- and post-training ERP data for the English and Chinese groups showed no main effect of Test time, or interaction of Test time with Tone [Ps > 0.37]. However, the Chinese and English participants did show a different effect of training for the two deviant tones [Tone by Test time by Group: $F(1,18)=5.36$, $P=0.033$]. Separate analyses on the MMN amplitude for the Low versus Mid tone, and High versus Mid tone showed that the two groups differed with respect to the effect of training for the High tone deviants [$F(1,18)=5.31$, $P=0.033$], but not the Low tone deviants [$F<1$]. English speakers showed an increase of the MMN for the High versus Mid tone after training [$F(1,9)=5.90$, $P=0.038$], whereas the Chinese showed no effect of training for the High versus Mid tone [$F(1,9)=1.23$, N.S.].[1] The MMN peak latency showed no significant effects of, or interactions with, the factor Test time [Ps > 0.247].

2.4.3. Statistical tests of the MMN difference (before and after training)

To test the robustness of the MMN effect within each of the three participant groups, the difference obtained by sub-

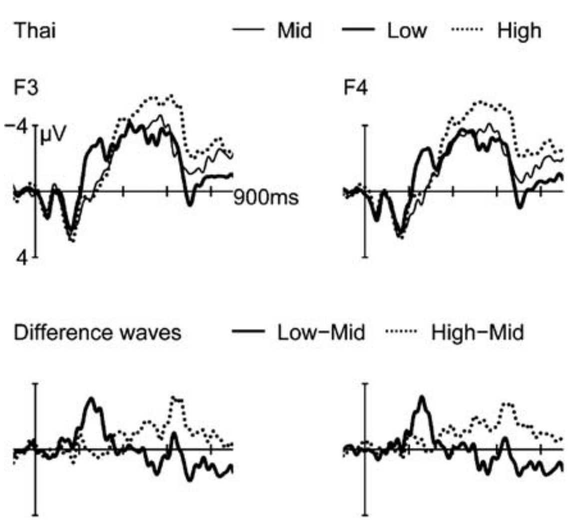

Fig. 1 – ERPs at the F3 (left column) and F4 electrodes (right column) for the Thai group. First row: ERPs. Thick solid line: Low tone deviant; Dotted line: High tone deviant; Thin solid line: Mid tone standard; Second row: Difference waves. Solid line: Low tone deviant minus Mid tone; Dotted line: High tone deviant minus Mid tone standard. In this and other graphs, the tick marks on the *x*-axis represent 200 ms. Negative is plotted upwards.

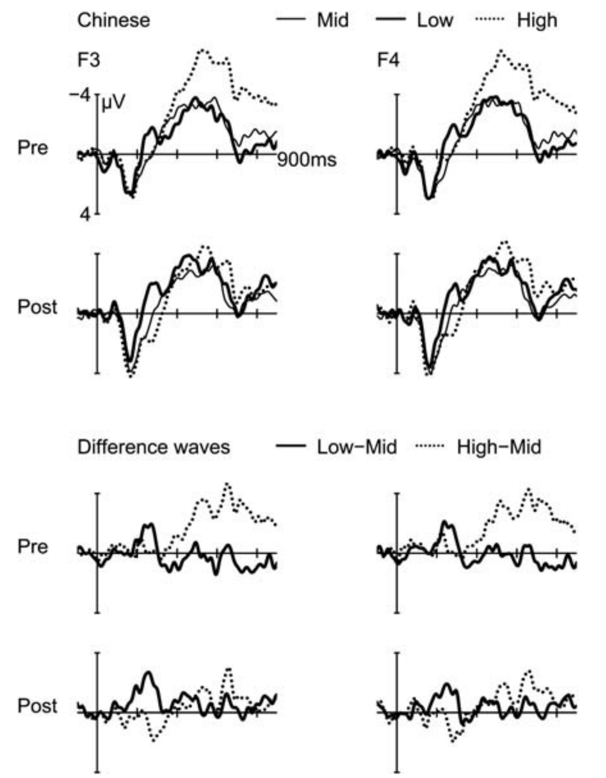

Fig. 2 – ERPs at the F3 (left column) and F4 electrodes (right column) for the Chinese group. First row: pre-training ERPs; Second row: post-training ERPs. Thick solid line: Low tone deviant; Dotted line: High tone deviant; Thin solid line: Mid tone standard; Third row: Pre-training difference waves; Last row: Post-training difference waves. Solid line: Low tone deviant minus Mid tone standard; Dotted line: High tone deviant minus Mid tone standard.

[1] The Test Time by Group interaction cannot be fully attributed to the ERPs to the Mid tone standard becoming more positive after training in the English compared to the Chinese: Analysis on the mean amplitude between 150 and 300 ms for the Mid tone showed no significant effect of Test time × Group [$F(1,18)=2.13$, $P=0.162$].

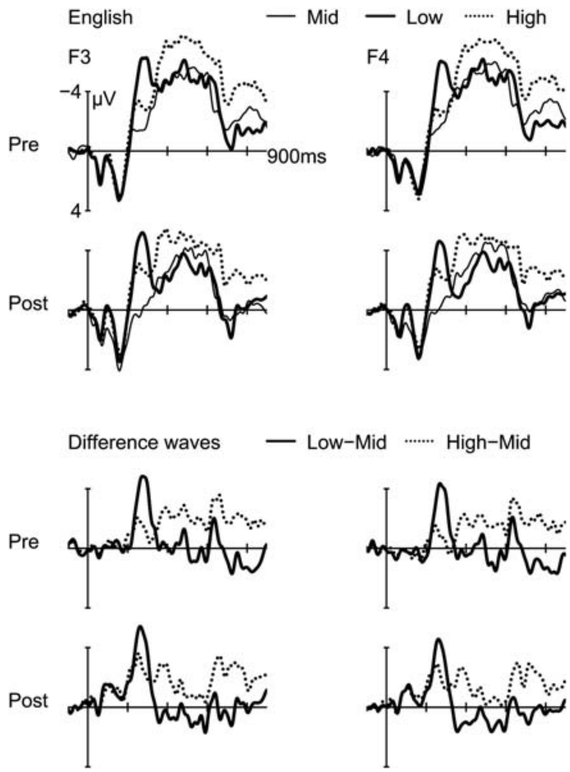

Fig. 3 – ERPs at the F3 (left column) and F4 electrodes (right column) for the English group. First row: pre-training ERPs; Second row: Post-training ERPs. Thick solid line: Low tone deviant; Dotted line: High tone deviant; Thin solid line: Mid tone standard; Third row: Pre-training difference waves; Last row: Post-training difference waves. Solid line: Low tone deviant minus Mid tone standard; Dotted line: High tone deviant minus Mid tone standard.

tracting the ERPs to the Mid standard tone from those to the deviants, was compared to a hypothetical zero. Since there were no hemispheric differences, we collapsed over the F3 and F4 electrodes. Before training, the MMN for the Low-deviant versus Mid-standard was significant in the English and the Thai participants, and approached significance in the Chinese, assuming an alpha of 0.005 to correct for the number of tests [Chinese: $t(9)=-3.34$, $P=0.009$; English: $t(9)=-4.51$, $P=0.001$; Thai: $t(9)=-4.63$, $P=0.001$]. After training, both the Chinese and the English participants showed a significant MMN for the Low tone versus Mid tone [Chinese: $t(9)=-5.80$, $P=0.000$; English: $t(9)=-5.46$; $P=0.000$]. The MMN to the High deviant versus Mid standard marginally differed from zero only in the English subjects before training [English $t(9)=-2.93$, $P=0.017$; Thai: $t(9)=-0.63$, NS; Chinese: $t(9)=-1.42$, NS]. After training, only the English showed a robust MMN for the High versus Mid tone [Chinese: $t(9)=0.30$, N.S.; English: $t(9)=-4.43$, $P=0.002$].

2.5. *The later negative difference*

2.5.1. *Effects of native language (before training, all three language groups)*

The late negative difference (350–650 ms) between the deviant and the standard was larger for High tones (−1.89 μV) compared to Low tones (0.164 μV) [$F(1,27)=56.98$; $P<0.001$]. The three groups did not differ [$Ps>0.44$].[2]

2.5.2. *Effects of training (Chinese and English only)*

Analyses on the Chinese and English data (pre and post-training) showed a significant interaction of Test time by Tone [$F(1,18)=5.02$, $P=0.038$]: the late negativity decreased after training for the High versus Mid tone, but not for the Low versus Mid tone [Effects of Test time for High tones: $F(1,18)=8.27$, $P=0.010$; Low Tones: $F(1,18)<1$, N.S.]. Furthermore, an interaction was found of Test time by Tone by Group [$F(1,18)=5.20$, $P=0.035$], suggesting that the Chinese and English differed in the effect of training on the tones. Only the Chinese showed a difference in the effect of training between the two tones [Test time by Tone: Chinese: $F(1,9)=10.97$, $P=0.009$; English: $F(1,9)<1$, NS]. In the Chinese data, the High tones showed a reduction of the negativity versus the Mid tones after compared to before training, which just failed to reach significance [$F(1,9)=4.59$, $P=0.061$]; whereas the negativity to the Low versus Mid tones was not different between pre and post-training [$F(1,9)<1$, NS]. The English group did not show any effect of Test time [$P=0.17$], suggesting that the Tone by Test time interaction reported above was mainly driven by the Chinese group.

2.5.3. *Statistical tests of the late negative difference (before and after training)*

To test the robustness of the late negativity within each of the three participant groups, we subtracted the ERPs to the Mid

[2] The late negativity showed some distributional differences with the early MMN in the present study, and hence cannot be considered a delayed MMN. The MMN had a frontal medial maximum, which was slightly stronger in the left than right hemisphere, whereas the late negativity had a broader distribution. The pre-training MMN amplitude for the Low tone minus Mid standard was compared with the pre-training late negativity amplitude for the High tone minus Mid standard, using the following electrode regions: Frontal: F3/4, F5/6, F7/8; Fronto-Central: FC3/4 FC5/6 FT7/8; Central: C3/4, C5/6, T7/8; Central-parietal: CP5/6; CP7/8, TP7/8 and Parietal: P3/4, P5/6, P7/8. A repeated measures ANOVA was conducted with, as (fixed) within-participant factors, Window (Low tone MMN/High tone late negativity), Hemisphere (left/right), Region (5 levels, frontal to parietal), and laterality (3 levels, from close to the midline to more lateral). Language group (3 levels) was included as a between-participant factor. The Greenhouse–Geisser correction (Greenhouse and Geisser, 1959) was applied when the degrees of freedom in the nominator were larger than one. Significant interactions of location and time window were found [Window × Laterality: $F(2,54)=7.19$, $P=0.012$; Window × Laterality × Hemisphere: $F(2,54)=4.11$, $P=0.039$; Windows × Region × Group: $F(8,108)=4.77$, $P=0.008$; Window × Laterality × Group: $F(4,54)=6.94$, $P=0.003$]. Separate analyses for each group showed a significant Window × Region interaction for the Thai group [$F(4,36)=12.98$, $P=0.001$], which remained significant after z-score correction (Kounios and Holcomb, 1994; McCarthy and Wood, 1985) [$F(4,36)=9.84$, $P=0.002$]. The MMN had a frontal maximum, whereas the late negativity was more equally distributed over the head. The English group showed a Windows × Laterality interaction [$F(2,18)=13.03$, $P=0.004$], which remained significant after z-score correction [$F(2,18)=5.422$, $P=0.014$]. The MMN was maximal towards the midline, whereas the late negativity was more equally distributed.

standard tone from that to the Low and High deviants, and compared the mean difference in the 350–650 ms interval to a hypothetical zero, collapsing over the F3 and F4 electrodes. The late negative difference for the High tone deviants versus Mid tone standards was significant in the Chinese participants, and approached significance in the English and Thai groups before training [Chinese: $t(9) = -6.58$, $P < 0.001$; English: $t(9) = -3.40$, $P = 0.008$; Thai: $t(9) = -2.34$; $P = 0.044$]. After training, the negative difference between the High deviant and Mid tone standard did not reach significance for either the Chinese [$t(9) = -2.28$, $P = 0.049$] or the English groups [$t(9) = -2.60$, $P = 0.029$], assuming an alpha of 0.005. The difference in late negativity amplitude to the Low tone versus the Mid tone was not significant in any group, neither before training [Chinese: $t(9) = 0.78$, N.S.; English $t(9) = -0.03$, N.S., Thai $t(9) = 0.47$, N.S.], nor after training [Chinese: $t(9) = -0.93$, N.S; English $t(9) = 1.24$, NS].

In sum, the Low deviant tone showed a MMN relative to the Mid standard tone, which was not affected by training. The High deviants showed no, or a smaller MMN compared to the Mid standards. This MMN increased in amplitude after training in the English group. In addition, the High deviants elicited a late negativity, which decreased in amplitude after training in the Chinese group.

3. Discussion

The aims of the present study were to investigate the effect of language background on the pre-attentive perception of Thai tones, and whether native speakers of Mandarin Chinese and native speakers of English differ in the effects of training. The prediction was that if Chinese and English speakers can pre-attentively detect differences among the tones, they would elicit a MMN to the deviant tones, which may be smaller than the MMN elicited in native Thai speakers. The MMN to deviant tones was expected to increase in amplitude and/or decrease in latency with increased discrimination ability as a result of training. In addition, based on previous behavioral studies, we expected the MMN to the deviant tones to be larger and earlier in the Mandarin Chinese speakers than in the English speakers.

We found that both Chinese and English groups performed close to ceiling on the behavioral training task, and elicited a MMN to the Low Thai tone deviant versus Mid tone standard, which, in contrast to expectation, did not differ from the Thai group and was not affected by training. The High tone deviant versus Mid tone standard showed no, or a smaller, MMN compared to the Low tone deviants. In addition, the High tone deviant elicited a later negativity versus the Mid tone standard. Again, these effects did not differ significantly between the Chinese, English and Thai speakers. Interestingly, the Chinese and English groups did differ with respect to carry-over effects of training: In the English, the MMN to the (untrained) High tone deviants versus Mid tone standards became larger after training, whereas the Chinese showed a decrease in the late negativity to the High tone deviants versus Mid tone standards after training.

Below we will discuss the present findings in light of our initial hypotheses. We will also discuss the differences between the Low and High tone deviants, and the differences between the language groups with respect to the effects of training.

3.1. *Effects of native language background in pre-attentional lexical tone perception*

The fact that the Chinese and English speakers did not differ from the Thai with respect to (1) the MMN to the Low tone deviant versus the Mid tone standard, (2) the late negativity for the High tone deviants versus Mid tone standards, and (3) behavioral discrimination even in the first stages of training, suggests that naïve listeners of both tone and non-tone background can detect differences among the Thai tones, both pre-attentively, as measured by the ERPs, and attentively as measured by the behavioral task. These results are apparently at odds with results from previous behavioral studies using Low and Mid tones. In these studies, naïve listeners, and especially English speakers, were shown to have difficulty discriminating among these two Thai tones (e.g., Wayland and Guion, 2003, 2004). The present results suggest that previous failure to behaviorally discriminate between these two tones in Thai may have been due to task demands. The stimuli used in the present study only consisted of a single token per tone type, whereas in previous behavioral studies multiple tokens (i.e., produced by multiple talkers) per tone type were used. Thus, unlike the current study, 'abstraction' or 'normalization' across tokens is required for successful discrimination in these earlier behavioral studies. This may have rendered categorization of the tones and perceiving the differences among tone types much harder in the behavioral studies compared to the present study. In addition, behavioral experiments employed an AXB (Wayland and Guion, 2003) or odd-man-out tasks (Wayland and Guion, 2004), rather than simple identification (but see Abramson, 1975), which may introduce confounds due to attention and working memory load. Future experiments manipulating the number of tokens and task demands need to be conducted to shed more light on this issue.

3.2. *Differences between Low and High tone deviants: MMN*

In contrast to the Low deviants, no, or a much smaller MMN was elicited for High tone deviants versus Mid tone standards in all three participant groups. This can be related to the smaller early pitch differences between the High and Mid compared to Low and Mid at the transition from the consonant to the vowel. This is shown in Fig. 4: Both High and Low deviant stimuli differed from the Mid tone standard stimulus between 90 and 140 ms from syllable onset, but this early difference was much larger for the Low tone. The MMN amplitude has been shown to become larger with increasing physical differences between the deviant and the standard stimulus (Näätänen, 1992; Sams et al., 1985), which would account for the MMN difference between the Low and High tones observed here.

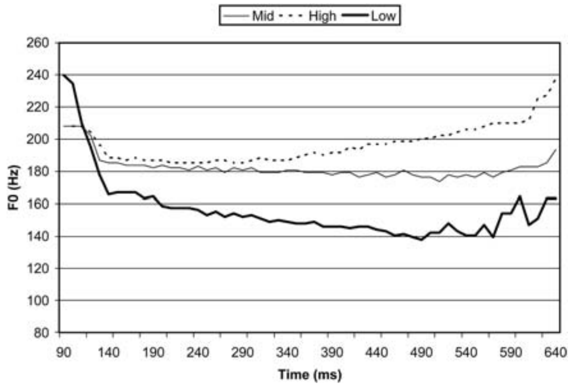

Fig. 4 – Normalized fundamental frequency (F_0) traces of the Thai tones used in the study. Dotted line: High rising tone; thick line: Low falling tone; thin line: Mid level tone.

3.3. *Differences between Low and High tone deviants: late negativity*

In contrast to the Low tone deviants, the High tone deviants elicited a larger late negativity relative to the Mid tone standards. It is unclear at present what this late negativity reflects. It is unlikely that it is a delayed MMN elicited by the late and slow differences in pitch slope between the High-rising and the Mid-level standard tone (see Fig. 4). First of all, this does not explain the occurrence of both MMN and late negativity to the High tone deviant in the English group. Second, the late negativity is present in all stimuli, and is a rather sustained response. Third, as discussed in footnote 2, the scalp distribution is different for the MMN and late negativity. Late negativities reported in the literature have been associated with cognitive, possibly pre-attentive processing of sound change (Ceponienè et al., 2004), or processing at a higher level of abstraction (Shafer et al., 2005; Zachau et al., 2005) including harmonic integration in music contexts (Koelsch et al., 2000; Loui et al., 2005). Alternatively, the late negativity may reflect reorienting of attention after involuntary attention to deviant stimuli (Schröger and Wolff, 1998; Shestakova et al., 2003). Under these interpretations, the High tone deviant may have involved more post-sensory processing or drawn more involuntary attention than the Low-falling deviants, perhaps due to its late and gradual change or easy discriminability (e.g., Abramson, 1975) from the Mid tone standards. However, no conclusive account can be provided at present.

3.4. *Effect of language background on training*

The three language groups did not show any significant differences with respect to the Low tone deviants before training. Since performance was already close to ceiling in the first block of training, training may not have altered the (pre-attentive) perception of the Low tone deviants. However, the ERPs to the (untrained) High tone deviant did differ as a result of training, and it is in this carry-over effect of training that the most interesting differences between the Chinese and English groups are seen: The English group showed an increase in MMN amplitude to the untrained High deviants versus Mid standards after training, whereas the Chinese showed a decrease in the late negativity to the High deviants versus Mid standard tones. These differences can be accounted for by behavioral differences in lexical tone perception reported in previous studies (e.g., Gandour and Harshman, 1978). In these studies, English speakers were shown to be more sensitive to pitch level than pitch contour, whereas native speakers of tone languages were more sensitive to pitch contour and pitch range. The present MMN data support this view. English speakers may have been better at sensing the subtle, early pitch level differences between the High and the Mid stimuli, leading to a more robust MMN for the High tone than in the other groups, even before training. Even though the training was on Mid and Low tones only, the training may have further heightened the English speakers' sensitivity to pre-attentively detect early pitch differences. As a result, the training effect carried over to the untrained tone (Gottselig et al., 2004; Tremblay et al., 1997), leading to a larger MMN for the High tone after training in this group. The Chinese, on the other hand, place more emphasis on pitch contour and pitch range, and hence ignored the subtle difference in pitch level, as judged by the absence of MMN for the High tones in this group. As a result, the Chinese did not profit much from training in detecting these early differences. A lack of a larger MMN after training suggests that they continue to ignore these early differences even after training. This is similar to the finding that training effects will not carry over to untrained difficult contrasts (Gottselig et al., 2004).

In contrast to the English participants, the Chinese showed a reduction of the late negativity to the High tone deviants versus Mid tone standards. In addition, this late negative difference was numerically larger in the Chinese group than in the English and Thai groups before training. If this late negativity reflects a shift in involuntary attention, as we suggested above, a potential and admittedly speculative explanation is the following. Before training, Chinese are more sensitive to (later) slope differences than the English. The Thai Mid-level and Low-falling tones can relatively easily be mapped onto tones 1 (high level) and tone 4 (high-falling) in Mandarin Chinese, respectively. However, the equivalent of the High-rising Thai tone is less obvious, since the closest Chinese equivalent, tone 2 (rising), rises earlier after onset than the Thai High-rising tone. The High-rising tone deviant would therefore draw involuntary attention of the Chinese participants and elicit a large late negativity, or involve more post-sensory processing. After training, the Chinese may have habituated, and less involuntary attention may have been drawn to High tone deviants, or less post-sensory processing resources may have been involved in processing these stimuli. Further research is obviously needed to clarify the nature of the late negativity and its reduction after training.

In sum, the present study shows that native speakers of tonal and non-tonal language can perceive foreign or non-native lexical tone contrasts at a pre-attentive level, and that perceptual training has different effects depending on language background. These results are consistent with previous behavioral finding that native and non-native speakers of tonal languages place emphasis on different phonetic dimensions of lexical tones.

4. Experimental procedures

4.1. Participants

Thirty-one participants were run: 10 native speakers of American English (7 men), 10 native speakers of Mandarin Chinese (6 men), and 11 native speakers of Thai (6 men). Informed consent was obtained from each participant according the procedures of the University of Florida Internal Review Board. All participants were healthy young adults, ages 20–35, with no history of neurological disease or language disorders as assessed by a self-report. All had a minimal bilateral hearing range of 500 to 8000 Hz measured at 25 dB HL. None of the Chinese speakers had any experience with a tone language other than Mandarin Chinese; two of the English participants had learned some Mandarin Chinese, but were not proficient as indicated by a self-report. Data from one male native Thai speaker were dropped from analysis due to too many artifacts.

The English and Chinese speakers participated in four sessions, one session per day on four consecutive days: Day 1: An ERP pretest; Day 2: the first training session; Day 3: the second training session; and Day 4: an ERP post test. The Thai speakers participated in only one ERP session, which was the same as the pretest ERP in the Chinese and English groups. In addition, the Thai participants completed only one training session, a few days after the EEG session. This was to obtain behavioral data to which the performance of the non-native speakers could be compared.

4.2. ERP pretest

4.2.1. Stimuli

Stimuli were instances of the Thai syllable [k^ha:], produced with three Thai lexical tones: a mid-level tone (Mid, standard), a high-rising (High, deviant), and a low-falling tone (Low, deviant). The mid-level and low-falling tones were chosen to parallel previous behavioral studies (Wayland and Guion, 2003, 2004). The low-falling and mid-level tones have been found to be hard to discriminate, even for Thai speakers (Abramson, 1975). The high-rising tone has been shown to be relatively easy to discriminate (Abramson, 1975), and was included for that reason. Although these Thai tones resemble lexical tones in Mandarin Chinese, the pitch contours are not identical. The stimuli were meaningful, actual words for Thai speakers, but not for the Chinese and English participants. Stimuli were produced by a female native speaker of Thai and digitized at 22050 Hz sampling rate with a 16-bit amplitude resolution. They were edited such that the three stimuli were comparable in voice onset time (93 ms for the High and Low tones, 89 for the Mid tone), and total duration of the stimulus (650 ms). All three stimuli were normalized for peak intensity (98% of the scale) and RMS amplitudes. The formant values (F1, F2) measured at the vowel steady state of the present stimuli indicate that the vowel /a:/ in all three tones is of the same quality (F1 equals 937 Hz, 1002 Hz and 1062 Hz; F2 equals 1686 Hz, 1601 Hz and 1635 Hz for the Mid, High and Low tone stimuli respectively). The pitch contour of the stimuli is shown in Fig. 4. Note that the three stimuli start to differ in pitch already at the transition between the [k^h]-stop and the vowel. The difference is larger for the Low versus Mid than for the High tone versus Mid tone. The Mid and the High tone converge in pitch from vowel onset, but later (approximately 200 ms into the vowel or 300 ms into the syllable) slowly diverge when the pitch of the Mid tone syllable gradually increases (see Supplementary Materials for the sound files and spectrograms of the stimuli).

4.2.2. Presentation

Stimuli were presented binaurally, one at a time over head phones. The physical volume was the same for all participants, and all participants reported it to be at a comfortable hearing level. The inter-stimulus interval (ISI) was randomized between 600 and 900 ms to prevent interference from regular biological rhythms on the waveforms. In total, 1200 stimuli were presented: 120 Low-deviants, 120 High-deviants (i.e., 10% of the stimulus set each), and 960 of the Mid tone standard stimulus. The order of the stimuli was pseudo randomized such that two deviants were separated by at least two standards. At least 120 of the standards tones were preceded and followed by a standard trial to serve as control for the deviant stimuli in the ERP analysis.

While the auditory stimuli were presented over headphones, participants watched a silent movie (Charlie Chaplin's 'The Gold Rush'). Participants were told that they would need to answer questions about the movie after the experiment, and were instructed to ignore the sounds. After the auditory stimulation, they received a form with 10 questions about the movie. The length of the EEG recording was 27 min. The total ERP test lasted about 1 h and 15 min, including set up and debriefing.

4.2.3. ERP recording

EEG was recorded from 39 Ag/AgCl scalp electrodes mounted in an elastic cap with active shielding (Easy-Cap, Falk Minow, Herrsching-Breitbrunn, Germany) combined with an ANT amplifier (ANT Software b.v., Enschede, The Netherlands). Electrode positions used were: Midline: Fz, FCz, Cz, CPz, Pz; Lateral left/right hemisphere: FP1/2, F7/8, F5/6, F3/4, FT7/8, FC5/6, FC3/4, T7/8, C5/6, C3/4, TP7/8, CP5/6, CP3/4, P7/8, P5/6, P3/4, O1/2. Horizontal and vertical EOGs were recorded from electrodes placed on the outer canthi, and below and above the right eye, respectively. Additional electrodes were placed on the right and left mastoids. The signal was acquired using the left mastoid as reference, but was arithmetically re-referenced off-line to the mean of the left and right mastoids. Electrode impedance was kept below 5 kΩ. The signal was sampled at a rate of 512 Hz, and was filtered off-line between 0.1 and 30 Hz.

4.3. Training sessions

After the first ERP session, native Chinese and English speakers returned on the next 2 days for two training sessions using the same Low and Mid tone stimuli as used in the ERP experiment. Thai participants participated in only 1 day of training. The distinction between the Mid and Low tone was expected to be difficult to perceive because previous research (e.g., Abramson, 1975) showed that, in isolation, these two tones are easily confused. No High tones were presented during training. First, a previous study has shown (e.g., Abramson, 1975) that non-native participants' performance on this tone is highly accurate; and second, not training the high-tone allowed us to investigate carry-over effects of training on untrained tones that are relatively easy to distinguish from standards (Gottselig et al., 2004).

Participants were told that they would hear one stimulus per trial with an inter-trial interval of 1.5 s. During the introduction phase of the training (20 trials), they would hear a production of Tone A (Low) alternated with a production of Tone B (Mid). In other words, odd number trials were Tone A (Low tone) and 'even' number trials were Tone B (Mid tone). They were asked to push a button marked 'A' on the computer screen for 'odd' trials and a button marked 'B' for 'even' trials. They were also told to focus on learning how these two Thai tones differ. A 'replay' prompt appeared if no response was given after 3 s and participants were allowed to replay each trial as many times as they wish. No specific instructions were given to the participants to sacrifice accuracy for speed. To avoid confusion and confounds, participants were not told what the actual meaning of the Thai words were, although we did tell them that the syllables were from Thai, and that in Thai, pitch differentiates between word meanings. If an incorrect response was given (e.g., the 'B' button was pushed for a Low tone trial), the correct button (i.e., 'A' button) would blink for a period of 5 s. A short break was offered at the end of this introduction phase. The procedure for the next phase of training was identical to that of the first phase except that the 'Low tone' and 'Mid tone' trials were now presented in random order in two blocks of 200 trials each (400 total). Participants were given a short break after each block. Participants' performance (i.e., percent correct identification and reaction time) during this phase was saved for further analysis. The training session on the second day of training was identical to that on the first. Both training sessions lasted approximately 30 min. This training method has proven successful in training non-native listeners to identify lexical tones in previous studies (Wayland and Guion, 2004; Wayland and Li, in press).

To obtain a baseline level to which to compare the effects of training, an analysis was conducted on the first 25 occurrences of each tone immediately following the introduction phase. Percentage correct and reaction time for correct responses were analyzed using a repeated measures ANOVA with language Group (English, Chinese and Thai) as a (fixed) between-participants factor, and Tone (Mid and Low) as (fixed) within-participants factors. Reaction times longer than 3 s were omitted from analysis, since this indicated that participants replayed the stimuli. To investigate the behavioral effects of training, accuracy and response time for the first 25 trials for each tone of the first block of training (as described above) were compared to that of the last 25 trials for each tone on the second day of training. In this case, the between-participants factor was language group (English, Chinese), and the within-participants factor Tone (Low, Mid), and Test time (first, last block of training).

4.4. *ERP post-training*

The Chinese and English speakers returned for a post training ERP session. The stimuli and procedure were similar to the ERP pretest, with the exception that a different fragment of the movie was shown and participants received different questions.

4.5. *ERP data analysis (adjust font size to header level)*

Epochs were defined spanning −100 to 900 ms from the stimulus onset. EEG to Low and High tones were averaged separately, relative to a 100 ms pre-stimulus baseline. We also separately averaged the EEG to 120 Mid tone standards, to which the ERPs to the deviant stimuli would be compared. We only included in this average standard Mid tones that were preceded and followed by a standard stimulus. This was to avoid potentially confounding effects of preceding and following deviant stimuli on the standard stimuli. Trials with eye movements and other artifacts were rejected (this was 25–27% per condition in the Chinese group; 21–28% in the English group, and 35% for each condition in the Thai group). Statistical analysis was conducted using the F3 and F4 electrodes to compare potential hemispheric differences. These were the lateral electrodes where the MMN was largest. First, difference waves (deviant minus standard) were calculated for the High minus Mid tone, and Low minus Mid tone conditions. Next, the most negative peak was found between 150 and 300 ms, and the mean amplitude for the windows spanning 100 ms centered around this peak was calculated for every channel, participant, tone and test time. Analyses were conducted on the mean amplitude thus calculated and on the peak latency. A later negativity was observed as well. Statistical analysis of this later negativity was conducted on the mean amplitude difference between 350 and 650 ms, based on visual inspection. To test for pre-training differences, we conducted a repeated measures ANOVA with Tone (High minus Mid, Low minus Mid), Hemisphere (F3/F4) as (fixed) within-participant factors, and Group (3 levels: Thai, Chinese and English) as a between-participant factor. To test the effect of training, we conducted a repeated measures ANOVA with Tone (High minus Mid, Low minus Mid), Hemisphere (F3/F4) and Test Time (pre or post training) as (fixed) within-participant factors, and Group (2 levels: Chinese and English) as a between-participant factor. In the latter analysis, we only report effects that included a main effect of, or interactions with, the factor Test Time. When a two or three way interaction was significant, separate repeated measurements ANOVAs were conducted to determine the source of the interaction.

The robustness of the MMN and late negativity was assessed for each group, tone and test session by comparing the difference (=deviant-standard) versus a hypothetical zero, using a *T*-test. Since there were no hemispheric differences of interest, F3 and F4 were collapsed. Alpha was set to 0.005 to correct for the number of tests.

Acknowledgments

This study was funded by a small research grant from the UF Asian Studies Program, awarded to RW. EK is currently supported by NIDCD grant number RO3 DC006160-01A2 and a UF Research Opportunity Seed Fund.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.brainres.2007.02.019.

REFERENCES

Abramson, A.S., 1975. The tones of central Thai: some perceptual experiments. In: Harris, J.G., Chamberlain, J.R. (Eds.), Studies in Tai Linguistics in Honor of William J. Gredney. Central Institute of English Language, Bangkok, pp. 1–16.

Bluhme, H., Burr, R., 1971. An audio-visual display of pitch for teaching Chinese tones. Stud. Linguist. 22, 51–57.

Ceponienè, 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, 130–141.

Gandour, J., 1983. Tone perception in Far Eastern languages. J. Phon. 11, 49–175.

Gandour, J., Harshman, R., 1978. Cross-language difference in tone perception: a multidimensional scaling investigation. Lang. Speech 21, 1–33.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., Hutchins, G.D., 2000. A crosslinguistic PET study of tone perception. J. Cogn. Neurosci. 12, 207–222.

Gottselig, J.M., Brandeis, D., Hofer-Tinguely, G., Borbély, A.A., Achermann, P., 2004. Human central auditory plasticity associated with tone sequence learning. Learn. Mem. 11, 162–171.

Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24, 95–112.

Kiriloff, C., 1969. On the auditory discrimination of tones in Mandarin. Phonetica 20, 63–69.

Klein, D., Zatorre, R.J., Milner, B., Zhao, V., 2001. A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. NeuroImage 13, 646–653.

Koelsch, S., Gunter, T., Friederici, A.D., 2000. Brain indices of music processing: “nonmusicians” are musical. J. Cogn. Neurosci. 12, 520–541.

Kounios, J., Holcomb, P.J., 1994. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. J. Exp. Psychol., Learn. Mem. Cogn. 20, 804–823.

Kraus, N., McGee, T., Carrell, T.D., King, C., Tremblay, K., Nicol, T., 1995. Central auditory system plasticity associated with speech discrimination training. J. Cogn. Neurosci. 7, 25–32.

Lee, L., Nusbaum, H.C., 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. Percept. Psychophys. 53, 157–165.

Loui, P., Grent 't-Jong, T., Torpey, D., Woldorff, M.G., 2005. Effects of attention on the neural processing of harmonic syntax in Western music. Cogn. Brain Res. 25, 678–687.

McCarthy, G., Wood, C.C., 1985. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr. Clin. Neurophysiol. 62, 203–208.

Menning, H., Imaizumi, S., Zwitserlood, P., Pantev, C., 2002. Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. Learn. Mem. 9, 253–267.

Näätänen, R., 1992. Attention and Brain Function. Lawrence Erlbaum Associates, Hillsdale, NJ.

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

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., Paavilainen, P., 1993. Development of a memory trace for a complex sound in the human brain. NeuroReport 4, 503–506.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Iimoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., Alkho, K., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385, 432–434.

Peltola, M.S., Bosch, L., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R., 2003. Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. Neurosci. Lett. 352, 25–28.

Sams, M., Paavilainen, P., Alho, K., Näätänen, R., 1985. Auditory frequency discrimination and event-related potentials. Electroencephalogr. Clin. Neurophysiol. 62, 437–448.

Schröger, E., Wolff, C., 1998. Attentional orienting and reorienting is indicated by human event-related brain potentials. NeuroReport 9, 3355–3358.

Shafer, V.L., Morr, M.L., Datta, H., Kurzberg, D., Schwartz, R.G., 2005. Neurophysiological indexes of speech processing deficits in children with specific language impairment. J. Cogn. Neurosci. 17, 1168–1180.

Shestakova, A., Huotilainen, M., Ceponienè, R., Cheour, M., 2003. Event-related potentials associated with second language learning in children. Clin. Neurophysiol. 114, 1507–1512.

Tremblay, K., Kraus, N., Carrell, T.D., McGee, T., 1997. Central auditory system plasticity: generalization to novel stimuli following listening training. J. Acoust. Soc. Am. 102, 3762–3773.

Tremblay, K., Kraus, N., McGee, T., 1998. The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. NeuroReport 9, 3557–3560.

Van Lancker, D., Fromkin, V.A., 1973. Hemispheric specialization for pitch and “tone”: evidence from Thai. J. Phon. 1, 101–109.

Van Lancker, D., Fromkin, V.A., 1978. Cerebral dominance for pitch contrasts in tone language speakers and in musically untrained and trained English speakers. J. Phon. 6, 19–23.

Wang, Y., Spence, M.M., Jongman, A., Sereno, J.A., 1999. Training American listeners to perceive Mandarin tone. J. Acoust. Soc. Am. 106, 3649–3658.

Wang, Y., Jongman, A., Sereno, J.A., 2001. Dichotic perception of Mandarin tones by Chinese and American listeners. Brain Lang. 78, 332–348.

Wang, Y., Jongman, A., Sereno, J.A., 2003a. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. J. Acoust. Soc. Am. 113, 1033–1043.

Wang, Y., Sereno, J.A., Jongman, A., Hirsch, J., 2003b. fMRI evidence for cortical modification during learning of Mandarin lexical tone. J. Cogn. Neurosci. 15, 1019–1027.

Wayland, R., Guion, S., 2003. Perceptual discrimination of Thai tones by naive and experienced learners of Thai. Appl. Psycholinguist. 24, 113–129.

Wayland, R., Guion, S., 2004. Training native English and native Chinese speakers to perceive Thai tones. Lang. Learn. 54, 681–712.

Wayland, R., Li, B. in press. Effects of two training procedures in cross-language perception of tones. J. Phon.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alkho, K., Lang, H., Iivonen, A., Näätänen, R., 1999. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. Cogn. Brain Res. 7, 357–369.

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Henninghausen, K., Schecker, M., 2005. Extracting rules: early and late mismatch negativity to tone patterns. NeuroReport 16, 2015–2019.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46.