Effects of Native Language and Training on Speaker Normalization on Lexical Tone Perception: A Behavioral and ERP Study

Ratree Wayland[1], Edith Kaan[1], Mingzhen Bao[1] and Christopher Barkley[2]
[1]University of Florida and [2]University of California at San Diego

1. Introduction

Voice pitch or its acoustic correlate, fundamental frequency (F_0) serves different linguistic functions in 'lexical tone' languages such as Thai and Mandarin Chinese and 'non-lexical tone' languages like English and German. In lexical tone languages, a change in pitch differentiates word or lexical meaning. For example, in Thai, an otherwise identical phoneme string [k^ha:] produced with five different pitch level or pitch contour, namely 'mid', 'low', 'falling', 'high' and 'rising' will result in five different words: 'to be stuck', 'galangal root', 'to kill or servant', 'to engage in trade' and 'leg'. Previous behavioral studies (e.g., Wayland and Guion, 2003, 2004) have shown that perceptual discrimination of lexical tone contrasts are challenging for native speakers of 'non-tonal' languages. In addition, native speakers of a different tonal language have been found to be more successful at discriminating tonal contrasts from another tonal language than native speakers of a non-tonal language (Wayland and Guion, 2004).

A greater degree of perceptual difficulty observed among non-native speakers of a tonal language may have been due to the way they process lexical tones. For example, unlike speakers of a non-tonal language (English), speakers of a tonal language (Mandarin Chinese) seemed to perceive pitch and segmental information in an integral manner (Lee and Nausbaum, 1983).
In addition, of the five dimensions of F_0, namely average, direction, length, extreme endpoint, and slope, listeners of a tonal language (e.g, Thai and Yoruba) placed more emphasis on pitch direction and slope, while listeners of a non-tonal language (e.g, American English) placed more emphasis on average pitch and extreme endpoints (Gandour and Harshman, 1978). In addition, when compared to speakers of tonal languages, native speakers of English paid more attention to the average F_0 and less on the F_0 contour information in order to identify tones (Gandour, 1983).

However, a few studies (Wang et al, 2001; Wayland and Guion, 2004; Wayland and Li, in press), suggested that perceptual difficulty experienced by speakers of tonal and non-tonal languages can be overcome, to some extent, through experience afforded by a short-term laboratory training. For example, a 21% improvement in Mandarin Chinese tone identification was found among American English speakers learning Chinese after four weeks of training (Wang et al, 1999). Similarly, discrimination of the Thai mid vs. low tone contrast improved significantly among native English and native Chinese speakers after 2-5 days of training (Wayland & Guion, 2004; Wayland and Li, in press).

The aim of this current study was to investigate the effect of language background (tonal versus non-tonal native language) and training on the ability to perceive pitch contour among native speakers of a tonal language (e.g, Thai and Chinese) and native speakers of a non-tonal language (e.g, English). Unlike previous studies, both *pre-attentive* and *attentive* processing of lexical tones were investigated using electrophysiological and behavioral measures. Previous studies of lexical tone processing used either categorical discrimination or identification tasks. As these tasks involve conscious comparison or categorization, one could argue that performance in these behavioral tasks may have been affected by such factors as working memory load or attention. In the current study, in addition to attentive behavioral investigation, the pre-attentive discrimination of lexical tones and the effect of language background and training was investigated using Event-Related brain Potentials (ERPs).

ERPs can be recorded while the participant is presented with auditory stimuli, but is engaged in an unrelated task such as watching a movie. The ERP component that was of interest here is the mismatch negative (MMN). MMN is a frontal negative ERP that occurs at about 100-300 ms after stimulus onset. It is elicited to infrequent stimuli (oddball) that deviate from frequently presented (standard) stimuli in pitch, duration, voice onset time, or other acoustic or phonetic properties (Näätänen, 2001). Since this ERP component is elicited even while people are asleep or in coma, it is regarded as an index of pre-attentive processing of auditory differences. The MMN has been shown to increase in amplitude and, in some cases, to have a shorter peak latency as behavioral discrimination performance improves. The MMN is therefore a useful tool to study the (pre-attentive) processing and acquisition of non-native language contrasts (Näätänen et al, 1993; Kraus et al, 1995; Tremblay et al, 1997, 1998; Menning et al, 2002; Winkler et al, 1999; Peltola et al, 2003)

Thus far, only a few studies have applied the MMN to investigate the processing of lexical tones. For example, Chandrasekaran et al. (2007) investigated the effect of language background on lexical tone perception. It was found that MMN was elicited among both Mandarin Chinese and untrained English speakers to tone contrasts in Mandarin Chinese. However, only the Chinese participants showed a larger MMN to a distinction that was acoustically more salient. This pattern of results suggested that, to a certain degree, language background affects pre-attentive processing of lexical tones. The effects of both training and language background were investigated in Kaan et al. (2007). In this study, ERPs were recorded from native speakers of English, Mandarin Chinese and Thai while they were presented with three Thai tones, namely mid-falling, high-rising and low-falling, in an oddball paradigm. Only one token per tone was used. No difference between the groups was found before training. However, after a two-day perceptual training on the mid-level and low-falling tones, the English speakers showed an increase in MMN to untrained high-rising deviants, whereas the Chinese speakers showed a decrease of a later negativity in the same condition. This suggested that native speakers of tone and non-tone languages were sensitive to different aspects of the stimuli as a result of training. Interestingly, no effect of training was observed on the (trained) low-falling tone deviant. For this tone, all groups showed a large MMN before and after training. It is important to note, however, that behavioral performance at the start of the training was close to ceiling for all three subject groups. The differences found in the ERPs may, therefore, have not been indicative of improved perception of the tones. The ceiling performance may have been due to the use of only one token per tone category, as such normalization or abstraction of F_0 level or contour was not necessary for discrimination success. Due to both intra- and inter-talker variation in F_0, it has been shown that perception of lexical tones was more difficult, even among native speakers of a tonal language, when multiple tokens produced by multiple talkers were used (Wong and Diehl, 2003). Thus, to avoid ceiling effect, multiple tokens of three Thai tones were used (see Methods and Figure 1) in this current study. Successful categorial discrimination or identification of these stimuli would depend on the listener’s ability to normalize for F_0 slope and/or contour across the three tokens of each tonal category.

2. Methods

2.1. Stimuli

Nine stimuli were digitally created from one natural token of the Thai mid-level tone syllable [kha:] produced by a female speaker that had been digitized at 22,050 Hz sampling rate with a 16-bit amplitude resolution. Using the Praat speech analysis software, the duration of this original mid-level tone token was shortened from 610 ms to 450 ms. Its pitch contour was then manually modified to approximate the pitch contours of the natural tokens of the Thai low-falling and high-rising tones, also produced by the same female speaker. The entire F_0 contour of each of the three resulting stimuli (i.e, mid-level, low-falling, high rising) was then shifted down -15 Hz and -30 Hz to simulate three different talkers with three different pitch registers, thus yielding three tokens for each of the three tone categories, see Figure 1. All stimuli were normalized for RMS amplitude (98% of the scale). All 3 tokens of each tone category were then presented to two native Thai speakers (one male and one female) in isolation and were judged to be acceptable exemplars of each of the three tone categories.

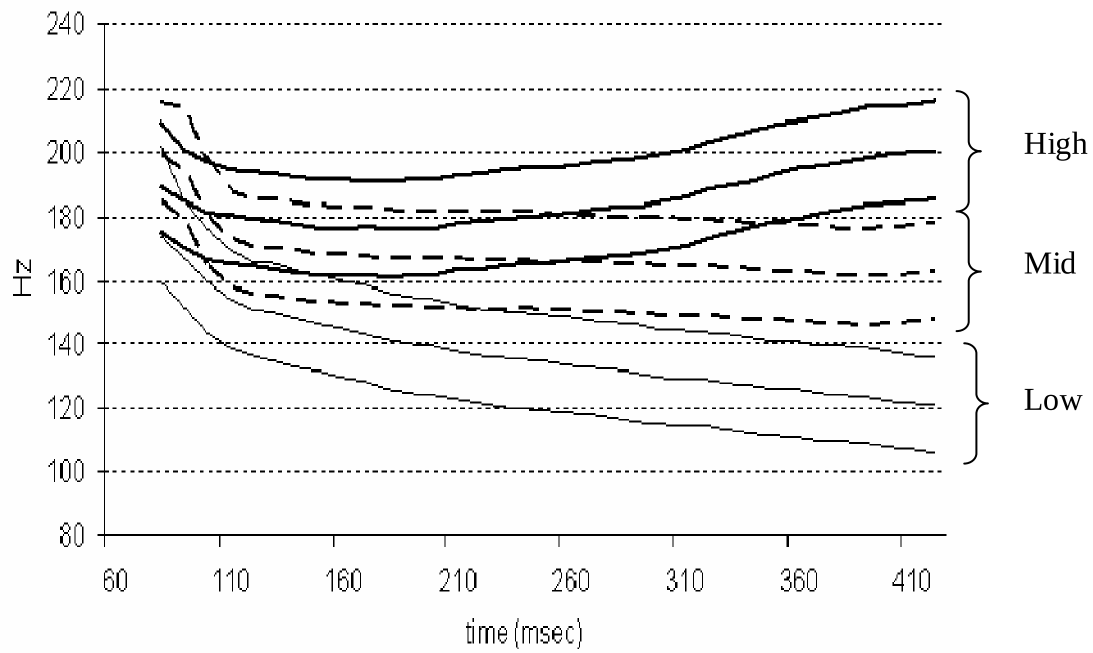

Figure 1: Pitch (F_0) tracks of all nine stimuli used in the study.

As shown in Figure 1, the main physical difference among the three tones is that of F_0 contour: relatively level for the mid-level tone, falling for the low-falling tone and rising for the high-rising tone. It should also be observed that besides F_0 onset values, pitch contours of the three tokens of the high-rising and the mid-level tones overlap considerably and do not become distinct until about 300 ms into the syllable. On the other hand, pitch contours of low-falling tone tokens differ from those of the mid-level and high-rising tone tokens at a relatively earlier point in the syllable. In addition, two low-falling tokens have a F_0 onset value that is lower than that of the mid-level tones, with one (the lowest) token of the low-falling starting lower than that of all other stimuli.

2.2. Subjects

Forty native speakers of American English (13 men), 32 native speakers of Mandarin Chinese (People's Republic of China) (11 men), and 11 native speakers of Thai (5 men), were recruited from the University of Florida community. Ten (2 men) of the 40 English speakers participated as control subjects. Informed consent was obtained from each participant according to the procedures of the University of Florida Internal Review Board. All participants were healthy young adults between the ages of 19-35. They were all right handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971), and none reported any history of neurological disease or language disorders. All passed a bilateral hearing screen (500 to 8,000 Hz measured at 25 dB HL). The American English speakers had no prior experience with a tone language and the native Chinese speakers had no prior experience with any other tone language. One Chinese speaker, however, did speak another Chinese dialect in addition to Mandarin Chinese. Participants were paid for participation. Behavioral data were obtained from all participants. However, ERP data (i.e., EEG recordings) were collected from only 12 English speakers (8 men), 12 Chinese speakers (6 men) and all 11 Thai speakers (5 men). Ten additional participants were run in the ERP study, but were discarded from analysis because of incomplete data sets.

2.3. Procedures

All participants were tested on four consecutive days. Stimuli were presented binaurally, one at a time over head phones at a comfortable hearing level (about 65dB). An ERP oddball task was conducted on Days 1 and 4 (for the 35 ERP participants); two categorization training sessions were each conducted on Days 2 and 3, with a behavioral discrimination task preceding (Day 2) and following (Day 3) the training. 10 English control participants only participated in the behavioral discrimination task without the categorization training.

2.3.1. Behavioural discrimination task

In the *behavioral discrimination task* (Days 2 and 3) the participant heard a sequence of three different stimuli A B C, separated by 575ms. A and B were always from the same tone category (either low-falling, high-rising or mid-level). The last stimulus, C, was either of the same or of a different category. The participant was asked to indicate whether the last stimulus was same or different from the first two stimuli by clicking a mouse button on the computer screen. As shown in Figure 1, since absolute values of F_0 onset and F_0 offset among the nine stimuli are similar or overlapped, successful performance on this task involves participant's ability to normalize for pitch slope or pitch contour. A total of 113 trials (108 experimental trials and 5 warm-up trials that were not analyzed) were presented. The response side for the 'same' and 'different' response button on the computer screen was counterbalanced among participants. If no response was given after 3 seconds, the next trial was started. Responses longer than 3 seconds (1.8-3.2% per session and language group) were treated as no-response errors. D' scores were calculated on the percentage of hits (correct 'different' response in case tone C was of a different type than A and

B) and false alarms (incorrect 'different' response when A, B and C were of the same category). Null responses were not included in D' score calculation.

2.3.2. Categorial identification training

In the *categorial identification training* sessions (Days 2 and 3), participants heard one stimulus per trial. They were asked to classify a token as being of tone A, B or C by clicking a box on the computer screen. During the introduction phase of the training, they heard the three tokens of tone A (low-falling), followed by the three tokens of tone B (mid-level), followed by the three tokens of tone C (high-rising). After this was repeated three times, the tokens were presented in random order for a total of 81 trials (each token presented 9 times) and identification accuracy was recorded. Participants were allowed to replay the stimulus. If an incorrect response was given, the correct button would blink for 5 sec. The inter-trial interval was 3 sec. One session lasted 30 minutes and was repeated on the same day after a short break. Data from one Chinese participant for the first training session on Day 2 were missing due to technical failure. Hence, this participant is omitted in all analyses involving this first session.

2.3.3. ERP oddball experiment

In the *ERP oddball task* (Days 1 and 4), the stimuli were presented in a continuous stream. Four stimulus blocks were presented, the order counterbalanced across participants: (1) mid-level presented as standard, high-rising as deviant; (2) high-rising as standard, mid-level as deviant; (3) mid-level as standard, low-falling as deviant; (4) low-falling as standard, mid-level as deviant. A total of 1200 stimuli were presented per block: 1080 of the standard category and 40 of each of the three deviant tokens (i.e, 10% deviants). The inter-stimulus (offset-to-onset) interval was randomized between 500-650 ms to prevent interference from regular biological rhythms on the waveforms. The order of the stimuli was pseudo-randomized such that two deviants were separated by at least two standards. The length of each block was 17 minutes. While the auditory stimuli were presented over headphones, participants watched a silent movie (Charlie Chaplin's 'The Gold Rush' or Buster Keaton's 'The General'). They were told that they would receive questions about the movie after each of blocks, and were instructed to ignore the sounds. A different movie fragment was played during each session. Each ERP session lasted about 2 hours, including set-up and debriefing.

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 EOG were recorded from 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 KOhm. The signal was sampled at a rate of 512 Hz, and was filtered off-line between .3 and 30 Hz. We only analyzed low-falling and high-rising stimuli. These were always presented with mid-level stimuli in the presentation blocks. Any differences between the ERPs to the low-falling and high-rising tones can therefore not be due to different alternate stimuli in the presentation blocks. Epochs were defined spanning –100 to 900 ms from the stimulus onset. EEG to low-falling and high-rising tone deviants were averaged separately. We also separately averaged the EEG to 120 low-falling and high-rising tones when these were used as standards. To avoid any potentially confounding effects from preceding deviant tones, we selected 120 standard stimuli that were preceded and followed by a standard stimulus. Trials with eye movements and other artifacts were rejected. The percentage of rejection was on average 28% per condition (SD 15%) in the Chinese group; 20% in the English group (SD 9%), and 26% (SD 13%) per each condition in the Thai group.

The *Mismatch Negativity* was analyzed using the F3 and F4 electrodes. These were electrodes where the MMN was largest on the lateral sites. First, difference waves (deviant minus standard) were calculated for the high-rising deviants minus standards, and low-falling deviants minus standards. Next, the most negative peak was found between 100 and 350 ms, and the mean amplitude for the windows spanning 100 ms centered around this peak was calculated for every channel, participant, tone type and session. Analyses were conducted on the mean difference in amplitude thus calculated.

2.4. Results

2.4.1. Behavioral discrimination

Figure 2 shows D' scores obtained from 30 experimental English speakers, 32 Chinese speakers, 11 Thai speakers and 10 control English speakers.

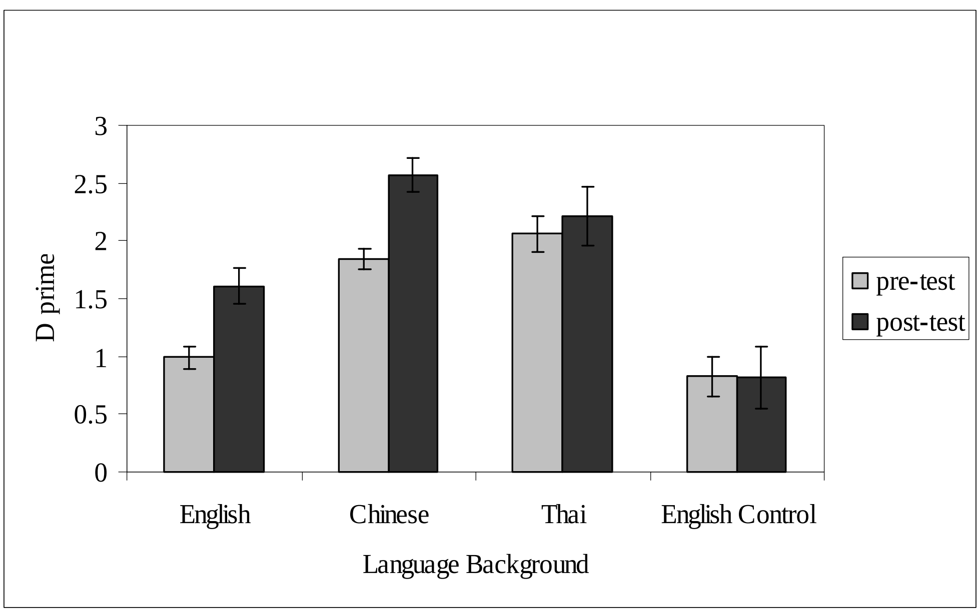

Figure 2: Mean D' scores and standard errors obtained from the behavioral discrimination task before and after categorial training among all participants.

Results of a repeated measures ANOVA revealed a significant main effect of Test Time [F(1, 79) = 18.128, p =.000] and Language Background [F(3,79)= 20.092, p.=000]. A significant interaction between the two factors [F (3, 79) =4.116, p =.009] was also obtained. Post-hoc pair-wise comparison showed that both the Chinese and Thai speakers outperformed the English speakers and the (English-speaking) control before training (p=.000). There was no difference between the English speakers and the control (p =1.0). After training, only the Chinese speakers outperformed the English speakers (p=.000). While numerically higher, the performance of the Thai speakers was no longer significantly better than that of the English speakers. In addition, while the Chinese and the English speakers showed a significant improvement after training [t(29) =-4.886, p=.000 for the Chinese speakers; t(31) =-5.688, p.=.000 for the English speakers] the Thai speakers and the control did not [t(10), -.667, p=.520 for the Thai speakers; t(9)=.090, p=.930 for the control]. Thus, unlike the Chinese and the English speakers, the Thai speakers did not benefit from the training. In addition, the fact that the improvement was observed only among the English-speaking group, but not among the control group suggested the effectiveness of the training.

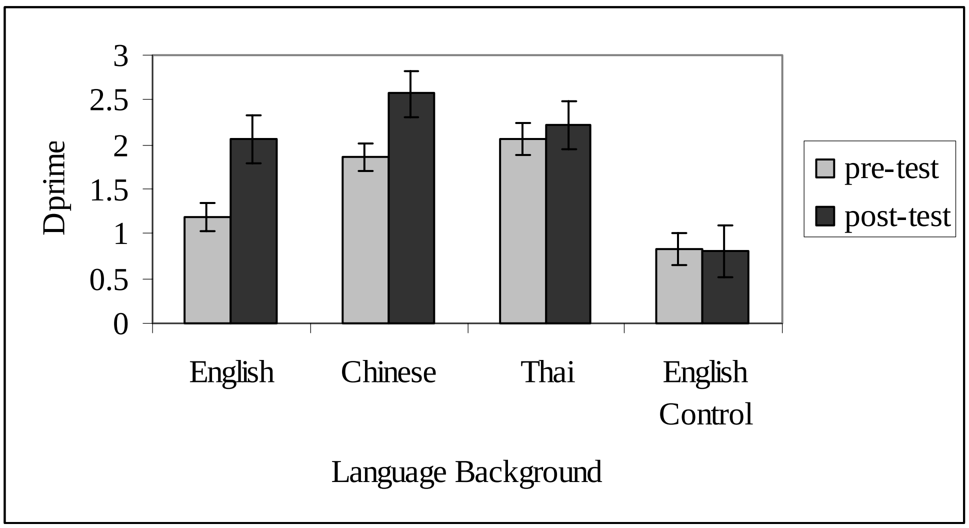

Figure 3: Mean D' scores and standard errors obtained from the behavioral discrimination task before and after categorization training among ERP experimental participants only. Mean D' score for control subjects is also included for comparison.

Behavioural discrimination scores before and after categorization training obtained from participants who participated in the ERP sessions are shown in Figure 3. Overall, performance of this subset of participants was similar to that of all subjects as a whole. Specifically, significant main effects of Language Background [F(3, 41)= 9.696, p=.000] and Test Time [F(1, 41) =17.153, p=.000] were found. In addition, a significant interaction between Language Background and Test Time [F(3, 41)=4.190, p=.011] was also evident. Again, this was due mainly to the fact that, unlike the English and Chinese speakers, the Thai speakers did not show a significant improvement after training. Consequently, while still numerically higher, the Thai (and surprisingly the Chinese) speakers' performances were no longer significantly higher than that of the English speakers after training. Furthermore, the English-speaking participants' performance also became significantly better than that of the control participants (p=.015) after training, suggesting that the improvement observed among the English-speaking participants was induced by the training, and not by a test-retest format.

2.4.2. Categorial identification (ID) training

Percentages of categorial identification errors obtained from all experimental participants are shown in Figure 4. A repeated measures ANOVA with Language Background as the between-subject factor and Session as within-subject factor was performed on the data. The analysis revealed a significant main effect of Language Background [F (2, 69) = 24.51, p =.000], Session [F(3,207)= 18.931, p=.000] and a near-significant interaction between Language Background and Session [F(6, 207)= 2.095, p=.056]. Post-hoc and follow-up tests suggested that, in general, the English speakers made more errors than the Chinese and the Thai speakers (p=.000) and the amount of errors committed by the Chinese and the Thai speakers was statistically comparable (p=.668). In addition, while number of errors committed during the first session was significantly higher than that of all other sessions among the English (p=.001) and the Chinese speakers (p=.05), the amount of errors made by the Thai speakers during the first session was nearly significantly more than those made during the last session, only (p=.068). These results indicated that the English and the Chinese speakers benefited more from the training that the Thai speakers.

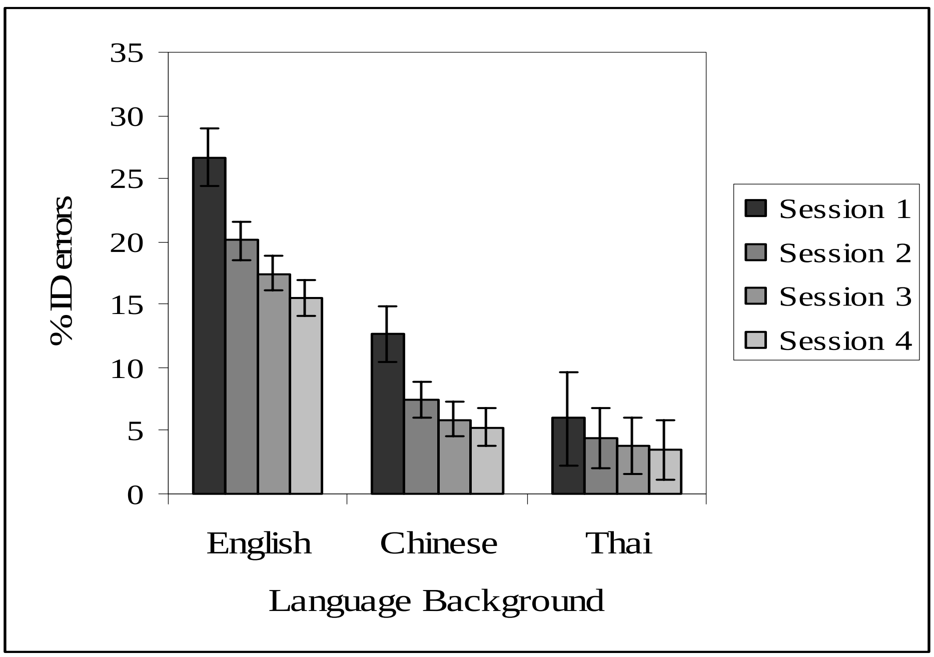

Figure 4: Percent ID errors across four training sessions obtained from all subjects

Identification errors exhibited across four training sessions by ERP participants are shown in Figure 5. These data were submitted to the same analyses as those obtained for all participants discussed above. The analyses yielded a significant main effects of Language Background [F(2, 31) = 11.035., p =.000], Session [F(3, 93) = 27.186, p=.000] as well as Language

Background and Session interaction [$F(6, 93)= 4.039$, $p =.001$]. Similar to the patterns of results found for all participants, post-hoc and follow-up tests revealed that, in general, the English speakers made significantly more errors than the Chinese and the Thai speakers and that the amount of errors committed by the Chinese and the Thai speakers were not significantly different. However, the main source of the significant interaction came from the fact that, while the errors committed by the English speakers during the first sessions were significantly higher than those of all later sessions, errors committed by the Chinese and the Thai speakers during the first session were nearly significantly fewer than those of the last session only ($p=.060$ for the Chinese and .068 for the Thai). This was somewhat different from the pattern of results obtained when data from all subjects were concerned. These results suggested that, among ERP participants, the English speakers appeared to benefit the most from the training.

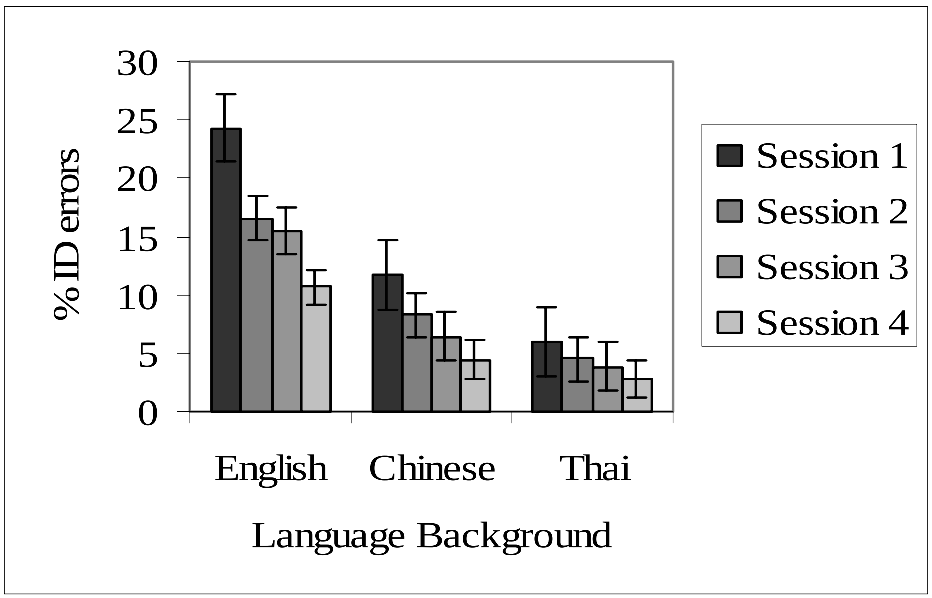

Figure 5: Percent ID errors among ERP subjects

2.4.3. ERPs to low-falling tones

ERPs to low-falling deviants minus standards at the F3 electrode for the English (solid line), Chinese (dotted line) and Thai (dot-dash) language groups are shown in Figure 6. Pre-training results are shown on the left panel and post-training results on the right panel. Negative is plotted up in this figure and in Figure 7. From this figure, a negative difference or MMN (arrow) starting at around 150 ms after stimulus onset (arrow) can clearly be seen for all three groups of participants. T-tests performed on the MMN amplitude at F3 and F4 versus a hypothetical zero showed that the MMN was significant before and after training in the English and Thai speakers [$ps <.004$], and after training in the Chinese speakers [before training: $p=.067$; after: $p<.001$].

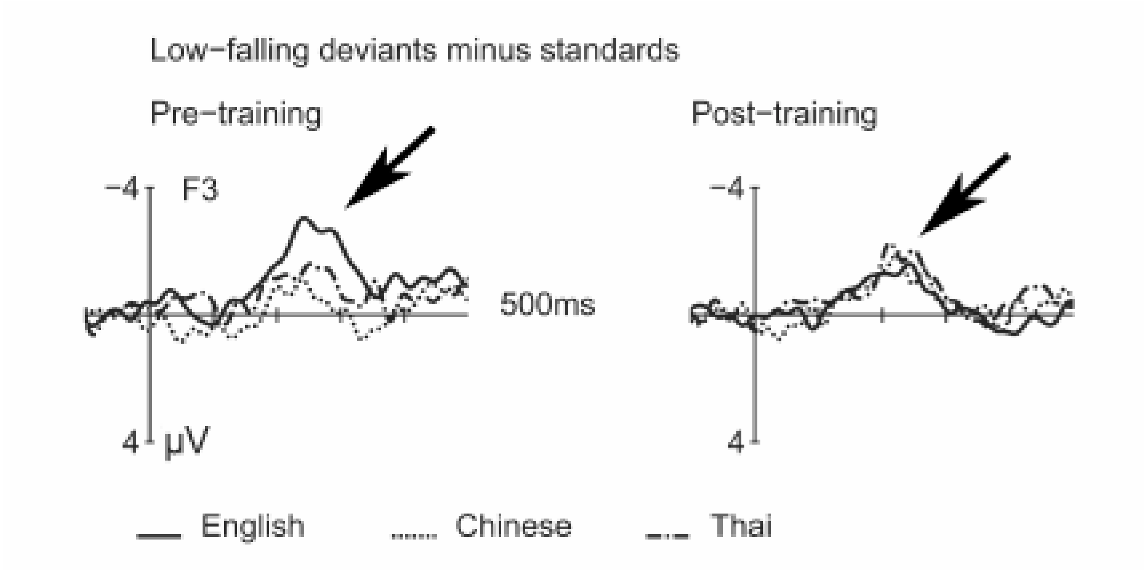

Figure 6: ERPs to low-falling deviants minus standards at the F3 electrode for the English (solid line), Chinese (dotted line) and Thai (dot-dash) language groups.

Training also appeared to weakly affect the differences in the MMN amplitude (deviant minus standard) among the language groups [Test Time x Language background: $F(2,32)= 2.69$, $p=.084$]: The groups weakly differed from each other before training [$F(1,32)=2.89$, $p=.07$], with the MMN being larger for the English group compared to the Chinese group [LSD post hoc comparison, $p=.025$], but not compared to the Thai group [$p =.12$]. After training, the groups did not differ [$ps >.68$].

2.4.4. ERPs to high-rising tones

ERPs to high-rising deviants minus standards are displayed in Figure 7. It is obvious that unlike ERPs to the low deviants minus standard, no MMN was present for any of the language group.

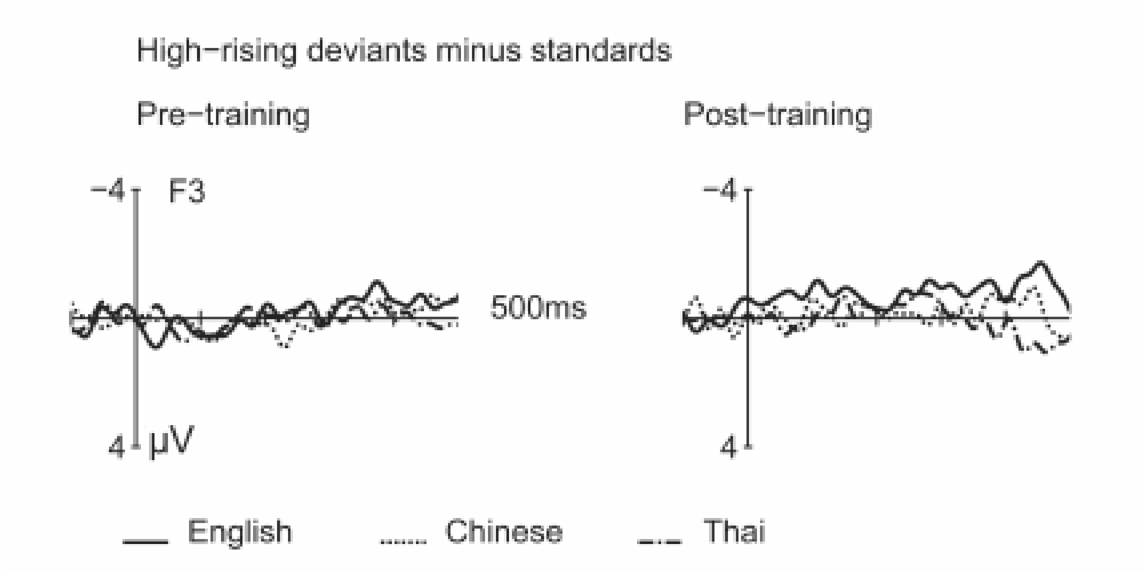

Figure 7: ERPs to high-rising deviants minus standards at the F3 electrode for the English (solid line), Chinese (dotted line) and Thai (dot-dash) groups. Left: pre-training results; Right: post-training results. Note the absence of a MMN.

Separate T-tests on the MMN amplitude at F3 and F4 versus a hypothetical zero showed no significant differences in any of the language groups before training [ps >.18]. After training, the MMN was only weakly present in the Thai group [Thai: p =.014; Chinese and English: ps >.062]. Interactions with Language Background and/or Test time were not significant.

2.5. Conclusion and discussion

This current study set out to investigate the effects of L1 background (English, Chinese) and training on the discrimination of Thai lexical tones, both at the attentive and pre-attentive level using behavioral and electrophysiological measures respectively. The stimuli used in the study were specifically constructed to allow for a comparison of participants' ability to abstract or normalize for the acoustic dimension of F_0 contour or F_0 slope. Specifically, three tokens of each of the three Thai tone categories (thus 9 stimuli) were used. The F_0 onset values of the stimuli largely overlap leaving F_0 contour or F_0 slope as the main physical difference among them (i.e., level versus falling versus rising). To successfully perform the categorial discrimination task employed, participants had to normalize for F_0 onset and offset and abstract the pitch contour or pitch slope across all three tokens of each tone category. Participants were also given a two-day categorial identification training. Discrimination D' scores obtained from all participants showed that the Thai and the Chinese speakers outperformed the English speakers before training, but that only the Chinese speakers outperformed the English speakers after training.

Data obtained from a subset of participants whose ERP data were also obtained mirrored those of all participants as a whole, except that neither the Chinese nor the Thai speakers outscored the English after training. Unlike the Thai group, however, the Chinese speakers did show a significant improvement after training.

Results of the categorial identification error rates showed that, as expected, the English group made a higher number of errors than the Chinese and the Thai groups. No difference between the Thai and the Chinese groups was found. In addition, it appeared that the English and the Chinese speakers benefited more from the training than the Thai speakers did, in that the amount of errors committed by the English and the Chinese speakers significantly decreased from the first to the last training session. While reduced, the error rate among the Thai speakers did not significantly differ from the first to the last training session.

Patterns of results obtained for the identification error rate among the ERP participants were similar, but not identical to those obtained when data from all participants were included. Specifically, it was found that the error rate among the Thai as well as the Chinese failed to decrease significantly from the first training session to the last.

ERP data suggested that all three groups of participants were sensitive to the physical difference between the mid-level and the low-falling tone at the pre-attentive level in that the MMN was elicited both before and after training to the low-falling tone deviant in all three groups of participants. The training seemed to have affected the MMN patterns observed. The group differed almost significantly before training with the MMN of the English being significantly larger than that of the Chinese, but no difference in MMN was observed across group after training.

Unlike ERPs to the low-falling tone deviant, no significant MMN was observed for the high-rising tone deviant among any group of participant before training. After the training, a weak MMN was observed only among the Thai. As noted earlier (section 2.1 Stimuli), F_0 onset values of all three tokens of the mid-falling and high-rising tones were overlapping, whereas the F_0 onset value of two tokens of the low-falling tone was much lower than that of the mid-level tones. Thus, a significantly larger MMN to the low-falling tone deviants observed among the English participants suggested, perhaps, that they were relatively more sensitive to this early F_0 onset difference than the Chinese (and the Thai) participants. The categorial identification may have increased their sensitivity to F_0 contour or slope, thus a reduction in MMN to the early F_0 onset difference. Similarly, a lack of a difference in F_0 onset values among the mid-falling and the high-rising tones may have been responsible for the absence of a significant MMN to the high-rising tone deviants among the English (and the Chinese) groups both before and after training.

In conclusion, behavioral data obtained suggested that, overall the Chinese (and the Thai) speakers outperformed the English speakers before the training. The English group, however, benefited the most from the training and their performance after the training, although still numerically lower, was no longer significantly different from the Thai group (and the Chinese group when only ERP participants were involved). However, it remains to be seen how long the training effect would last.

It is difficult at present to relate the behavioral data to the ERP data as they likely reflected processing at two different levels: attentive versus pre-attentive. As discussed above, the patterns of the MMN obtained for the low-falling and high-rising tones may have been an index of participants' sensitivity to physical properties (i.e., F_0 onset) of the signal. However, attention, working memory, and cross-language perceptual mapping of the tonal stimuli to existing L1 phonological system may have also influenced the behavioral data obtained.

2.5. References

Chandrasekaran, Bharath, Ananthanarayan Krishnan, and Jack Gandour (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Res* 1128: 148-156.

Gandour, Jack (1983). Tone perception in Far Eastern languages. *Journal of Phonetics* 11: 49-175.

Gandour, Jack and Richard Harshman (1978). Cross-language difference in tone perception: A multidimensional scaling investigation. *Lang Speech* 21: 1-33.

Kaan, Edith, Chris Barkley, Ratree Wayland and Mingzhen Bao (2007). Effects of Native Language and Training on Lexical Tone Perception: An ERP study. *Brain Res* 1148: 113-122.

Kraus, Nina, Therese McGee, Thomas Carrell, Cynthia King, Kelly Tremblay and Trent Nicol (1995). Central auditory system plasticity associated with speech discrimination training. *J Cogn Neurosci* 7: 25-32.

Lee, Lisa and Howard Nusbaum (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Percept Psychophys*, 53: 157- 165.

Menning, Hans, Satoshi Imaizumi, Pienie Zwitserlood and Christo Pantev (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, Risto (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, Risto, Erich Schröger, Sirel Karakas, Mari Tervaniemi and Petri Paavilainen. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport* 4: 503-506.

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

Peltola, Maija, Teija Kujala, Jyrki Tuomainen, Maria Ek, Olli Aaltonen, and Risto Näätänen (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neurosci Lett* 352: 25-28.

Tremblay, Kelly, Nina Kraus, Thomas Carrell and Therese McGee (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustic Society of America* 102: 3762-3773.

Tremblay, Kelly, Nina Kraus and Thomas McGee (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *Neuroreport* 9: 3557-3560.

Wang, Yue, Michelle Spence, Allard Jongman and Joan Sereno (1999) Training American listeners to perceive Mandarin tone. *Journal of the Acoustical Society of America*, 106: 3649-3658.

Wang, Yue, Joan Sereno, Allard Jongman and Joy Hirsch (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *J Cogn Neurosci* 15:1019-1027.

Wayland, Ratree and Bin Li (In press). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics.*

Wayland, Ratree and Guion (2004). Training native English and native Chinese speakers to perceive Thai tones. *Language Learning* 54: 681-712.

Winkler, István, Anne Lehtokoski, Paavo Alku, Martti Vainio, István Czigler, Valéria Csépe, Olli Aaltonen, Ilkka Raimo, Kimmo Alho, Heikki Lang, Antti Iivonen and Risto Näätänen (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research* 7: 357-369.

Wayland, Ratree and Susan Guion (2003). Perceptual Discrimination of Thai Tones by Naïve and Experienced Native English Speakers. *Applied Psycholinguistics* 24: 113-129.

Wong, Patrick and Randy Diehl. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tone. *Journal of Speech, Language, and Hearing Research* 46: 413-421.

Author contact information:

Ratree Wayland: ratree@ufl.edu