Training English and Chinese Listeners to Perceive Thai Tones

Ratree P. Wayland
University of Florida, USA.
E-mail: ratree@ufl.edu

ABSTRACT

The ability to discriminate the mid and the low tone contrast in Thai by native English (NE) and native Chinese (NC) speakers were investigated before and after an auditory training. The variables under investigation were first language background and the inter-stimulus-interval (ISI) of the presentation (500 vs. 1500 ms). The NC group outperformed the NE group during both the pre- and the post-tests and in both ISI conditions. A significant improvement in discrimination from the pre-test to the post-test in the 1500 ms ISI condition was observed in the NC group, but not in the NE group. These results suggest that prior experience with the tone system in one tone language may be transferable to the perception of tone in another language.

1. INTRODUCTION

This study reports the findings of an experiment designed to answer the question of whether or not native speakers of a tone language will be faster, when compared to native speakers of a non-tone language, at acquiring tones from another tone language with which they have no prior experience. Given their native ability and experience in using fundamental frequency as well as other perceptual cues such as voice quality (i.e., breathy, creaky, glottalized), vowel duration etc. to differentiate among tones in their language, the answer should be affirmative, at least at the phonetic level of processing. However, tone languages may differ from one another both in the number of tones that exist in their phonological systems as well as in their use of phonetic features (e.g., F_0 height, F_0 range and F_0 contour) to categorize these tones. Thus, long-term representations of existing native tones may prove to be a source of interference at a higher level of processing (i.e., phonological).

With these two possible opposite predictions in mind, an auditory training experiment was conducted to test the ability to discriminate between the mid and the low tones in Thai by a group of native speakers of Chinese and a group of native English speakers. A comparison was made both before and after the training using two different Inter-Stimulus-Intervals (500 ms and 1500 ms). These two different ISIs have been claimed to trigger different levels of processing: a phonetic level void of native language (L1) interference for 500 ms ISI and a phonological level for 1500 ms ISI.

2. METHODS

2.1 Participants

Participants were six (3 male, 3 female) native speakers of Thai, six (2 male, 4 female) native speakers of English and five (2 male, 3 female) native speakers of Taiwanese and 1 female native speaker of Mandarin Chinese. All of the Taiwanese listeners are Taiwanese-Mandarin bilinguals. Results of previous studies [2] suggest that native Mandarin and Taiwanese-Mandarin bilingual speakers do not differ in the way they perceive and process tones. Thus, for the purpose of this study, the five native speakers of Taiwanese and the one native speaker of Mandarin Chinese will be grouped together and referred to as the native Chinese speakers group (NC). All of the participants were recruited from the undergraduate and graduate student population at the University of Florida at Gainesville. No known speech or hearing impairment was reported.

2.2 Stimuli

Stimuli were eight minimal pairs or contrasts (see table 1 below) of the low and the mid tones of standard Thai. Discrimination of these two tones is known to be difficult when presented in isolation even for NT listeners. This contrast was chosen in order to avoid a ceiling effect, especially among NC listeners after the auditory training. All stimuli were produced by five (3 female, 2 male) Thai speakers.

Mid Tone	Low Tone
1. [piː] ‘year’	[pìː] ‘oboe’
2. [paː] ‘to throw’	[pàː] ‘forest’
3. [kʰaː] ‘to be stuck’	[kʰàː] ‘galanga’
4. [tʰaːy] ‘to guess’	[tʰàːy] ‘to change’
5. [kʰaːy]‘to spit out’	[kʰàːy] ‘a net’
6. [paːn] ‘birthmark’	[pàːn] ‘juke’
7. [pan] ‘to share’	[pàn] ‘to pedal’
8. [ʔaːn] ‘saddle’	[ʔàːn] ‘to read’

Table 1 Minimal Pairs used in the study

These contrasts were produced in a Thai carrier phrase “[rau phûːt kʰam wâː …]”, “we say the word….” Each contrast was produced three times in random order. The recording took place in a quiet office setting using a high quality DAT cassette recorder (Sony TCD-D8) and a head-

ISBN 1-876346-48-5

mounted microphone (Shure, model SM 10A). The microphone was placed at a 45-degree angle approximately 13 mm (as recommended by the manufacturer) from the mouth. The stimuli were later digitized using Cool Edit (Syntrillium Inc.) at 22.05 kHz, with a 16 bit quantization. Each target syllable was then excised out of the carrier phrase and saved as an individual file. All target syllables were normalized for peak intensity (50% of the scale).

Each pair of stimuli was arranged into triads. Each of the eight contrasts was tested by ten 'change' trials, which consist of a single token of one tone and two tokens of the other tone. For example, a trial testing the contrast [pi:]/[pi̖:] might consist of [pi:]-1, [pi̖:]-3, [pi:]-2 (where the number indicates different speakers). In the example given, the tone in the second stimulus is the odd item out because it contains a tone that differs from the first and the third stimuli. The serial position of the odd item out was distributed equally over the three possible positions. Each tone contrast was also tested by five 'no change' trials, which consisted of three different instances (i.e., produced by different talkers) of the same tone. In four of the eight contrasts, two of the five no-change trials consisted of three different instances of the mid-tone member of the pair (e.g., [pi:]-1, [pi:]-2, [pi:]-3), and the remaining three no-change trials consisted of three different instances of the low-tone member of the pair (e.g., [pi̖:]-3, [pi̖:]-2, [pi̖:]-1). The reverse was true for the remaining four contrasts. Thus, overall there was an equal number of no-change trials for the low tone and the mid tones, i.e., 20 for each tone and 40 in total. To test the effect of ISI, two sets of stimuli were created. In one set, an interval between the three stimuli in each trial was set at 500 ms, and in the other at 1500 ms. However, the Inter-Trial-Interval (ITI) was always set at 1500 ms.

2.3. Procedures

2.3.1 Pre-test

The listeners were tested individually in a quiet room in one session that lasted about 45 minutes using a PC. The 120 (8 pairs x 10 change trials + 8 pairs x 5 no change trials) trials for each ISI condition were randomly presented over headphones at a comfortable listening level. The listeners were told that each trial will be made up of three Thai words spoken by three different native Thai speakers and that they were to focus their attention on the tone or pitch level of the word. They were told to push a button marked "1", "2", or "3" if the tone in one word differed from the tone in the other two words, but to click the fourth button, marked 'none', if they heard three words produced with the same tone. For example, the button marked "1" should be selected if they believe that the first word they heard was produced with a different tone than the second and the third word. Stimulus replay was allowed. However, once a button was selected, it could not be changed, and replay was disabled. All listeners were tested on both sets (ISI 500 and ISI 1500) of stimuli and the order of presentation of the two sets was counter-balanced across listeners. To familiarize listeners with the stimuli and rate of presentations, a short training session containing 6 change trials and 4 no change trials drawn from the 120 experimental trials were presented without feedback. Moreover, the 120 experimental trials (for each ISI) were preceded by five warm-up trials that were not analyzed.

2.3.2. Training

Following the pre-test, all participants were asked to come back for the next five days for auditory training sessions. Each session lasted 30 minutes. The stimuli and procedure for the training session were identical to those used in the pre-test. However, to avoid a bias toward either the ISI 500 or ISI 1500, the ISI chosen for the training session was 1 sec. Similar to the pre-test, replay was allowed. It was, however, disabled once a button was selected. Unlike the pre-test, if an incorrect button was selected, the correct button would blink for a period of five seconds. This was the only form of feedback provided to the participant. The majority of participants were able to complete the training exercise twice within the allotted 30 minutes. Since the training period (5 days) is relatively short and very little improvement was expected from one session to the next, we decided not to analyze data obtained from the training sessions, and to focus instead on a pre-test and post-test comparison.

2.3.3. Post-test

After the completion of the training sessions, participants were asked to come back the following day for a post-test. The stimuli and procedure for the post-test were identical to those of the pre-test.

3. RESULTS

3.1 Pre-test

Mean percentage of correct discrimination [(% correct identification for change trials (maximum of 10) + % correct rejection for no-change trials (maximum of 5)/2)] for all contrasts in each ISI for each group of listeners is shown in Table 2. A 25% [(1/4 (for change trials +1/4 for no change trials)/2] correct identification indicates a performance at chance level. It should be noted that the 62% and 60% correct discrimination scores for the native speakers of Thai are lower than those reported in previous studies. Using an AX discrimination task, Burnham et al. (1997), for example, obtained a mean correct identification[1] of 79.2% among adult native (Australian) English speakers they examined. However, a single talker was used in their study. Thus, it is possible that the use of multiple talkers and the categorial discrimination task, which places a relatively greater demand on working memory than the AX discrimination task, may have further reduced the discriminability of this difficult tone contrast.

The effect of ISI on mean percent correct for the pre-test scores was tested in a mix-designed 2-way Analysis of Variance (ANOVA) with Language (NT, NC, NE) as the between subject factor and ISI (500 vs. 1500 ms.) as the within subject factor. The results revealed a significant main effect of Language [F (2,141) = 59.45, p <.0001],

but not of ISI. The interaction between Language and ISI was not significant. These results suggested that while there was a significant difference in the performance across the language groups, the performance of each group was comparable in both ISI conditions. As expected, a Tukey's pair-wise comparison suggested that NT was significantly better than the two non-native groups (p<.0001) in their ability to discriminate the mid and low tone contrast in Thai. More interestingly, the Tukey's analysis also revealed that NC's discrimination ability was superior to the NE listeners (p <.03).

ISI 500 ms			ISI 1500 ms		
NT	NE	NC	NT	NE	NC
62 (14)	36 (13)	42 (13)	60 (13)	36 (15)	42 (16)

Table 2. Mean % correct identification and Standard Deviation (in parentheses) for pre-test in each ISI condition for all three groups of listeners.

3.2. Post-test

Percent correct discrimination scores for each ISI condition for NE and NC is shown in Table 3. As can be seen, NC listeners were able to correctly discriminate the difference between the low and the mid tones in Thai more often than NC. When these data were analyzed in a 2-way mix-designed ANOVA with Language (NC vs. NE) as the between subject factor and ISI (500 vs. 1500 ms) as the within subject factor, a significant main effect of both Language [F (1,94) = 27.92, p <.0001) and ISI [F (1,94) = 6.97, p<.009] was obtained. However, the interaction between the two factors did not reach significance [F(1,94)= 1.44, p >.23] Interestingly, a test of simple main effect revealed that while the NC group performed significantly better than NE in both ISI conditions, NC's performance was significantly better in the 1500 ms ISI than the 500 ms ISI condition (p<.007).

ISI 500 ms		ISI 1500 ms	
NE	NC	NE	NC
32 (15)	44 (15)	34 (16)	50 (13)

Table 3. Mean % correct discrimination and Standard Deviation (in parentheses) for post-test in each ISI condition.

3.3. Difference between pre- and post-tests

Table 4 and 5 show mean percent correct discrimination for the NE and NC groups for the pre-test and the post-test for each ISI condition. A series of one-way ANOVAs was performed to examine whether there was a significant improvement from the pre-test to the post-test for the NE and NC listeners. The analyses revealed that there was no significant change in the NE's performance from the pre-test to the post-test. This was true for both ISI conditions. On the other hand, a significant improvement was observed for the NC listeners in the ISI 1500 condition [F(1, 47) = 9.12, p <.002].

NE			
Pre500	Post500	Pre1500	Post1500
36 (13)	32 (15)	36 (15)	34 (16)

Table 4. Mean % correct discrimination and Standard Deviation (in parentheses) during the pre- and post-tests for each ISI condition for native English speakers.

NC			
Pre500	Post500	Pre1500	Post1500
42 (13)	44 (15)	42 (16)	50 (13)

Table 5. Mean % correct discrimination and Standard Deviation (in parentheses) during the pre- and post-tests for each ISI condition for native Chinese speakers.

4. SUMMARY OF RESULTS

The results of the analyses presented here revealed that (a) native Chinese listeners outperformed native English listeners in their ability to discriminate the difference between the mid and the low tone in Thai both before and after the auditory training, (b) there was no effect of ISI during the pre-test, however (c) there was an ISI effect during the post-test, but only for NC listeners. That is, the correct discrimination score for the NC listeners was significant better for the ISI 1500 condition than for the ISI 500 condition. Most interestingly, (d) a significant improvement in the performance before and after the auditory training was observed only among native Chinese listeners for the 1500 ISI condition. The NE group showed no improvement from the pre- to the post-test.

5. DISCUSSION

Results obtained from this study indicated that, prior to the perceptual training, the NC listeners were superior to the NE listeners at discriminating the mid and low tone contrast in Thai. This advantage was evident in both ISI conditions. If the view is taken that the two ISI conditions represent two different levels of processing, namely language-general (phonetic) vs. language-specific (phonological) levels of processing, then this result suggests that native experience with a tone language has heightened sensitivity in tone discrimination among the NC listeners at both levels. The ability to track F_0 movement, direction or change at word level required to discriminate among Chinese tones could have been transferred, thus accounting for the superior performance at the phonetic level (ISI 500 condition) of processing among NC listeners. Moreover, unlike NE listeners, NC listeners also have access to phonological representations of tones stored in their long-term memory. It is possible that native Chinese listeners attempted to map the low and the mid tones in Thai to one of the Chinese tones to aid their discrimination. While the mapping may not be perfect, as implied by the lower score than NT listeners, it still gives them an advantage over the native English listeners whose long-term memory does not include phonological representation of tones.

The finding that there was no ISI effect during the pre-test

ISBN 1-876346-48-5

in either group of listeners is worth noting. This result is in disagreement with that of Burnham et al. (1997) who found that there was an ISI effect among native Australian English listeners: unlike native Thai listeners, the subjects in this group performed better in the ISI 500 condition than in the ISI 1500 condition. If the view that an ISI effect is an effect of phonetic vs. phonological processing modes is taken (as in Burnham et al. 1997), then the lack of an ISI effect in the pre-test suggests that the same level of processing was used in both ISI conditions. It is not clear, however, whether it is the phonetic or the phonological level that is involved. Given the short period of time needed to access lexical representation in semantic priming studies [3], it is reasonable to assume that the phonological level of processing was involved. The question remains, however, as to why an ISI of 500 ms would condition greater discrimination ability in one group of non-native listeners (i.e., NE subjects in Burnham et al. 1997) but not in another group (namely NE subjects in this study).

An alternative explanation for the absence of an ISI effect in this study can be offered, however. This explanation leans toward the view that an ISI effect is an effect of short-term memory and acoustic salience. For native speakers of English, their poor performance at the ISI 500 condition may suggest that the acoustic cues they extracted may not have been sufficiently salient to take advantage of the relatively smaller load placed on the working memory by the shorter ISI. A combination of the lack of tone representation in long-term memory, the ineffective extraction of cues and the decay of acoustic cues detected contributed to their failure at longer ISI. For NC listeners, while better than NE, subtle acoustic or perceptual cues needed to differentiate between the two tones may have also been insufficient for them to perform well at the shorter ISI condition. Unlike NE, however, the lack of one-to-one mapping between the low and the mid tones in Thai and the Chinese tones may have been responsible for their lack of success in the ISI 1500 condition.

This explanation can also be extended to explain the most interesting finding of an ISI effect found among the Chinese listeners in the post-test. That is, after one week of auditory training, while there was no difference in their performance in the ISI 500 condition, native Chinese performed significantly better in the ISI 1500 condition. This result suggests that while the training did not improve their ability to discern phonetic difference between the low and the mid tones in Thai, it may have allowed them to notice that the Thai tones did not perfectly match any of the closest Chinese tones (the most likely candidates for similar tones being Mandarin Tone 2 and Tone 3 or Taiwanese mid-level and low-falling Tones). This, in turn, led them to rely less on the existing Chinese tone categories, and more on the new tone categories they were trying to form. However, their relatively lower performance in comparison to NT listeners suggests that a week of perceptual training may be too short a time for the formation of stable long-term representations for the two new tones. Failure to improve among NE listeners, on the other hand, suggests that after a week of training, they failed to both discern phonetic differences as well as to form long-term representation of the two Thai tones.

From this study, it is reasonable to conclude that a more extensive training would be required to train listeners from a non-tone language background to achieve the same level of performance as that of native speakers of a tone language. This is because they have to, not only acquire the new skill of detecting acoustic and auditory properties appropriate to differentiate among tones, a skill far more familiar to native listeners of a tone language, but also to form and incorporate tone representations in their long-term memory. On the other hand, native tone listeners can rely on their existing tone system to guide their perception of the new tones, depending on similarities or differences between the tone system of the native and that of the target language. This explanation, in turn, implies that native listeners of a non-tone language may benefit more, at least initially, from an auditory training in which a shorter ISI (i.e., 500 ms) is used, while native speakers of a tone language would benefit more from a perceptual training using a relatively longer ISI. With a shorter ISI, native speakers of a non-tone language have a better chance in detecting acoustic characteristics differences across tones (an important basis for future formation of tone representation) due to its lower demand on the working memory and possibly its allowance for a language general level of processing. On the other hand, native speakers of a tone language may benefit more from a comparison between the acoustic characteristics of the target tones and the long-term representation of the native tones. While this technique should prove effective for any tone contrasts to be trained, its effect may take longer to realize among pairs of tone contrasts that are acoustically and perceptually similar (such as the low and the mid tone in Thai used in this study). In summary, findings from this current study suggest that L1 background, ISI and training duration are important considerations in auditory perceptual training design. These considerations are of particular importance when such perceptual training is incorporated into a second language classroom where students' L1 background is diverse.

REFERENCES

[1] D. Burnham and E. Francis, "The role of Linguistic Experience in the Perception of Thai Tones," In A. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong*. Bangkok: Chulalongkorn University Press, 1997.

[2] J. Gandour, "Tone perception in Far Eastern languages," *Journal of Phonetics,* vol. 11, pp. 49-175, 1983.

[3] M. Sabol and D. DeRosa, "Semantic encoding of isolated words," *Journal of Experimental Psychology: Human Learning and Memory,* vol.2, pp. 58-68, 1976.