

## Effects of Two Training Procedures on Cross-Language Perception of Tones

Ratree Wayland & Bin Li

Program in Linguistics  
University of Florida, United States  
ratree@ufl.edu

### Abstract

This study evaluated two training procedures that might be used to increase native English (NE) and native Chinese (NC) listeners' ability to discriminate the mid vs. low tone contrast in Thai under two different inter-stimulus-interval (ISI) conditions (i.e., 500-ms and 1500-ms). Participants were assigned to receive training using either a two-alternative forced-choice identification (ID) procedure or a categorial same/different discrimination (SD) procedure. The results obtained indicated that (a) NC performed significantly better than NE both before and after the training under both ISI conditions, (b) both groups of listeners improved significantly after training and (c) both training procedures (ID and SD) were equally effective in promoting phonological representation of Thai tones among both groups of listeners.

### 1. Introduction

Several training studies have been conducted to evaluate the effectiveness of short-term laboratory training on nonnative phonetic contrasts among adult listeners. (e.g., [1], [2], [3]) All of these studies showed that a significant improvement in the ability to identify, or discriminate, nonnative contrasts was evident for most nonnative subjects. However, due perhaps to differences in the characteristics of the nonnative subjects studied, the nature as well as the range of stimuli used, a considerable amount of variation in the size of the improvement has been reported. As [4] pointed out, one important difference among these studies is the procedures used in the training. In particular, two training procedures, namely *identification* (ID) training and *same/different discrimination* (SD) training have been used. In ID training, subjects hear a single stimulus on each trial. Their task is to identify the stimulus in terms of two categories (e.g., /r/ or /l/). Feedback regarding the correct identity of the stimulus is given immediately after a response is received. Through the provision of the feedback, subjects are expected to gradually learn the phonetic properties associated with each category. On the other hand, subjects who are administered a SD training procedure hear two stimuli on each trial. Their task is to determine whether the two stimuli in each trial are instances of the same category or instances of two different categories.

Relative efficacy of the two training procedures has been a subject of debate. [1] suggested that identification training might be more effective than same/different discrimination training because it encouraged subjects to rely more on phonetic codes stored in long-term memory than on rapidly fading sensory information in short-term memory. [3] added that a two-alternative forced choice identification procedure promoted the formation of new and robust phonetic categories that are not adversely affected by acoustic variations irrelevant to the phonetic identity of the categories in question. These include variations induced by different speaking rates and/or

idiosyncratic characteristics of individual speakers. [5], [6] hypothesized that same/different training encouraged subjects to pay more attention to within-category acoustic differences than between-category acoustic properties. Consequently, the subjects may fail to recognize core acoustic information that defines the two categories. On the contrary, [7] argued that subjects who received identification training may learn to respond correctly by attending to any properties that might be used to differentiate the two nonnative categories. Some of these properties may not be the ones used by native speakers.

The effect of ISI on speech perception has also been documented in previous studies [8], [9]. [9] for example, found that inexperienced subjects could not discriminate non-native contrasts at the 1500-ms ISI, but that they could discriminate them at 500-ms ISI. These authors suggested that the subjects' failure at the longer ISI was due to their use of native categories stored in long-term memory while performing the task. Specifically, they proposed that the subjects showed poor discrimination at the 1500-ms ISI because a 'phonological mode' of processing was activated, and since the nonnative contrast mapped onto a single native phonological category, the subjects showed poor discrimination. On the other hand, the subjects showed evidence of being sensitive to non-native contrasts at the 500-ms ISI because a 'phonetic mode' of processing was involved. According to these authors a 'phonetic mode' of processing is a language-general mode of perception in which speech sounds are discriminated without any influence of linguistic experience. On the other hand, the 'phonological mode' of perception was believed to be constrained or enhanced by experience with the phonological system of a particular language [8], [10].

To our knowledge, a study designed to assess relative degree of effectiveness between the identification and categorial same/different training procedures under different ISI conditions has yet to be conducted. Such study may reveal relative efficacy of the two training procedures at the phonological and the phonetic levels of processing.

### 2. This Study

This study was conducted to evaluate relative efficacy of the identification and categorial same/different discrimination training procedures on the discrimination of Thai tones by NC and NE listeners under two ISIs (500-ms and 1500-ms) conditions. The investigation was guided by the following questions: (1) Would the ability to discriminate Thai tones by native Chinese and native English listeners improve after training; (2) Which training procedure, identification or categorial same/different discrimination, would be more effective; (3) Would the performance of the participants vary as a function of ISI; and (4) Would the effectiveness of the two training procedures vary as a function of ISI?

## 2.1. Methods

### 2.1.1. Subjects

Subjects were 12 (5male, 7female) native speakers of Thai (NT), 30 (14male, 16female) native speakers of Chinese (NC) and 22 (5male, 17female) native speakers of American English (NE) with self-reported normal language and hearing abilities. They were recruited from the student population at the University of Florida. Native Chinese and English speakers have no prior experience with Thai.

### 2.1.2. Stimuli

Stimuli were 5 minimal pairs of low and mid tones of standard Thai. Based on a pilot study administered on two native Mandarin listeners, this was the contrast proven to be the most perceptually challenging for Mandarin listeners among all possible Thai tone contrasts. When presented in isolation, these two tones were also known to be difficult for native Thai speakers to discriminate. This tone contrast was thus chosen to avoid a ceiling affect, especially after training. The stimulus words were produced in a Thai carrier phrase [rau phû:t k<sup>h</sup>am wâ:], meaning 'we say the word \_\_\_\_'. Five male native Thai talkers produced each word three times in random order. The recording took place in a quiet office setting using a high quality DAT cassette recorder and a head-mounted microphone. The stimuli were later redigitized using a Kay Elemetrics CSL station (25.0 kHz sampling rate and 16 bit quantization). Each target syllable was then excised from the carrier phrase and saved as an individual file. All target syllables were normalized for peak intensity (98% of the scale).

Two productions of each stimulus word were used and each production came from two different talkers. For example, the two productions of [pù:] 'grandfather' were produced by talkers 1 and 2, while the two productions of [pu:] 'crab' were produced by talkers 3 and 4.

## 2.2. Procedures

The two productions of each word were used in constructing the AXB discrimination test. In this test, stimuli were presented in triads. The first (A) and the last stimulus (B) was a production of each member of the contrasting pair, and the middle stimulus (X) was the other production (i.e., by a different talker) of either A or B. For example, a trial testing the contrast [pu:]-[pù:] 'crab-grandfather' might consist of [pu:]-3, [pu:]-4, [pù:] -2 (where the number indicates different talkers). Both productions of each member of a contrasting pair were distributed equally over the three positions resulting in 16 trials for each pair. After hearing all three stimuli, participants were asked to decide whether the tone of the second stimulus is the same as the tone of the first or the last stimulus by selecting a button marked 'first' or 'last' presented on the computer screen. Percent correct discrimination was calculated for statistical analyses.

To test the effect of the ISI, two versions of the test were created. In one version, the interval between the three stimuli in each trial was set at 500 ms, and in the other it was set at 1500 ms. However, the interval between each response and

the presentation of the next trial (the Inter-Trial-Interval or ITI) was always fixed at 3,000 ms.

**Pre-test** The listeners were tested individually in a quiet room in one session that lasted approximately 45 minutes using a PC. The 80 (5 contrasting pairs x 16) trials for each ISI condition were randomly presented over headphones at a comfortable listening level. The listeners were told that each trial would be made up of three stimuli spoken by three different native Thai speakers and that they were to focus their attention on the tone or pitch level of the stimuli. They were told to push a button marked "First" if the tone of the second stimulus was the same as the tone of the first stimulus and to click a button marked "Last" if it was the same as that of the third stimulus. All listeners were tested on both sets (500-ms ISI and 1500-ms ISI) of stimuli and the order of presentation of the two sets was counter-balanced across listeners within each group. To familiarize listeners with the stimuli and rate of presentations, a short training session with 10 trials drawn from the 80 experimental trials were presented without feedback. Moreover, the 80 experimental trials (for each ISI) were preceded by five warm-up trials that were not analyzed.

**Training** Following the pre-test, Chinese and English participants were asked to come back for the next two days for training sessions. Each session lasted 60 minutes. One group of Chinese (n=15) and English (n=11) participants received a two-alternative forced choice identification (ID) training while the other (n = 15 and 11 respectively) received a categorical same/different discrimination (SD) training. The stimuli used during training were identical to those used in the pre-test.

Participants in the ID training group were told that they would be trained to hear the difference between the two Thai tones used in the pre-test. To avoid any possible confusion with the Mandarin tone labeling system, the two Thai tones were referred to as Tone A (mid tone) and Tone B (low tone) instead of Tone 1 and Tone 2. Participants heard one stimulus per trial. Similar to the pre-test, the ITI was fixed at 3,000 ms. Participants were told that during the first phase of the training (20 trials), they would hear a production of Tone A (mid tone) alternated with a production of Tone B (low tone). In other words, odd number trials were Tone A (mid tone) and 'even' number trials were Tone B (low tone). They were asked to push a button marked 'A' on the computer screen for 'even' trials and a button marked 'B' for 'odd' trials. They were also told to focus on learning how these two Thai tones differ. They were allowed to replay each trial as many times as they wish. A short break was offered at the end of this phase.

During the second phase, (60 trials), the presentation of Tone A (mid tone) and Tone B (low tone) was randomized. For a given trial, participants were asked to decide whether they hear Tone A or Tone B by pushing the corresponding button on the computer screen. If their response was incorrect (i.e., button 'A' was pushed for low tone), the correct button (i.e., button 'B') would blink for a period of 5 sec. Similar to the first phase, replay was allowed before a response was made. However, replay was disabled once a response was given. Participants were told that they may complete this training as many times as they could within the allotted time of one hour.

On the average, each participant was able to complete two sessions within one hour.

Participants in the SD training group heard two stimuli per trial. To avoid any bias toward either the 500-ms ISI or the 1500-ms ISI, the ISI was fixed at 1 sec. However, the ITI remained at 3,000 ms. The two stimuli presented in a given trial were always produced by two different talkers (i.e., physically different). Stimuli in 'same' trials, were produced with the same tone (i.e., mid-mid or low-low) while those in 'different' trials, were produced with different tones (i.e., mid-low or low-mid). During the first phase of the training (20 trials), 'odd' numbered trials were 'same' trials and 'even' numbered trials were 'different' trials. Participants were asked to push a button marked 'same' for 'odd' trials and a button marked 'different' for 'even' trials. Similar to ID training, a trial replay was allowed. A short break was also offered after the end of this phase.

During the second phase of the training (60 trials), 'same' and 'different' trials were presented randomly. Participants were asked to decide whether they heard a 'same' or a 'different' trial by selecting the corresponding button on the computer screen. If an incorrect response was given (e.g., pushed 'same' button for a 'different' trial), the correct button (i.e., 'different' button) would blink for a period of 5 sec. Replay for a given trial was allowed before a button was selected. However, replay was disabled once a response was given. Participants were allowed to complete as many training sessions as they wish within one hour. On the average, participants completed two training sessions (each session included both phases of the training) in one hour.

**Post-test** After completion of the training sessions, participants were asked to come back the following day for a post-test. The stimuli as well as the procedures for the post-test were identical to those of the pre-test.

### 2.3. Results

#### 2.3.1 Pre-test

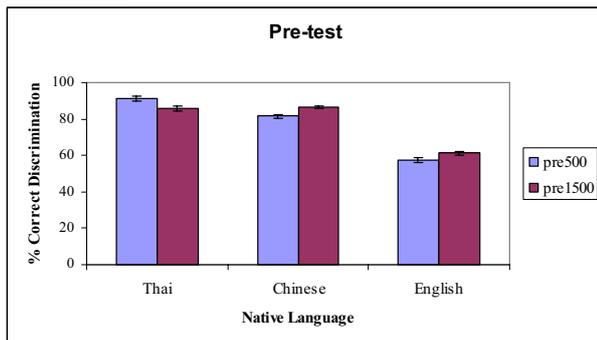


Figure 1. Percent correct discrimination during the pretest.

Figure 1 shows mean percent correct discrimination scores for the pretests of all three groups of participants under each of the ISI conditions (500 ms and 1500 ms). This data was submitted to a repeated measures ANOVA with ISI (500, 1500) as the within-subject factor and Language (Thai, Chinese, English) as the between-subject factor. The analysis yielded a significant main effects of Language [ $F(2, 60) =$

$62.02, p <.001$ ], but not of ISI [ $F(1,60) =.86, p = .36$ ]. The interaction between ISI and Language was significant [ $F(2, 60) = 5.37, p <.007$ ]. Bonferroni post-hoc tests revealed that NT listeners outperformed both NC ( $p <.04$ ) and NE listeners ( $p <.001$ ) under the 500-ISI condition, but they scored significantly better than NE listeners only under the 1500-ISI condition. Additionally, the NC listeners outperformed NE listeners under both ISI conditions ( $p <.001$ ). Moreover, a series of repeated measures one-way ANOVAs revealed that NT listeners performed significantly better under the shorter ISI ( $p <.005$ ) than under the longer ISI. On the contrary, NC scored better under the longer ISI than under the shorter ISI ( $p <.01$ ). The difference between the two ISI conditions did not reach significant among the English listeners ( $p =.10$ ).

#### 2.3.2. Effects of Training

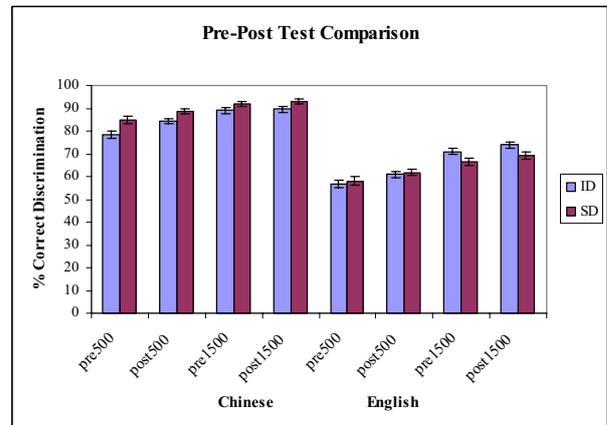


Figure 2. Percent correct discrimination before and after training.

Figure 2 shows mean percent correct discrimination scores of both training subgroups (ID vs. SD) among NC and NE listeners before and after the training under both ISI conditions. In general, both NC and NE listeners' discrimination of Thai tones improved after training. This is true for both the ID and the SD training groups and under both ISI conditions. However, their performance was better under the longer (1500-ms) ISI condition. For NC listeners, participants in the SD training group outperformed those in the ID group both before and after training under both ISI conditions. For NE listeners, on the other hand, while the SD training subgroup performed slightly better than those in the ID subgroup after the training under both ISI conditions.

This data was submitted to a mixed-design ANOVA with Training Group (ID, SD) as the between-subject factor and ISI (500-ms, 1500-ms) and Test Time (pre, post) as the within subject factor. For Chinese listeners, the analysis yielded a significant main effect of ISI [ $F(1,28) = 6.23, p <.02$ ] and Test Time [ $F(1,28) = 40.24, p <.001$ ], but not of Training Group [ $F(1,28) = 2.44, p =.13$ ]. The interaction between ISI and Test Time was also significant [ $F(1,28) =5.54, p <.03$ ]. Post-hoc repeated measures one-way ANOVAs revealed that their discrimination scores under the shorter ISI was significantly lower than those under the longer ISI before training ( $p <.01$ ).

However, the difference did not reach significance ( $p=.39$ ) after training. Additionally, the analyses indicated that their discrimination scores after the training were significantly higher than those before the training under both ISI conditions ( $p < .001$ ). Similarly, for native English listeners, the analyses yielded a significant main effects of ISI [ $F(1,19) = 5.49$ ,  $p < .03$ ] and Test Time [ $F(1,19) = 19.94$ ,  $p < .001$ ], but not of Training Group [ $F(1,19) = .19$ ,  $p = .666$ ]. However, there was no significant interaction among any of the factors.

### 3. Discussion

This study examined relative efficacy of the Identification and categorical same/different training procedures in promoting phonological representation of the low and the mid tones in Thai at both the phonetic and the phonological levels of processing. The results obtained suggested that the two training procedures were equally effective in improving the Chinese and the English listeners' ability to discriminate the two Thai tones at both levels of processing. This finding suggests that unlike traditional same/different AX discrimination task, the presentation of multiple tokens of each category in the categorical same/different discrimination may encourage subjects to ignore within-category acoustic differences. As Polka (1992) argued a certain degree of perceptual constancy is required for a successful performance in a categorical same/different discrimination task since multiple tokens of each category are likely to encompass a wide range of acoustic variants.

Equally interesting is the finding that NC outperformed NE both before and after training. This finding suggests that knowledge of the tonal system of one language facilitates the learning of a different tonal system.

### 4. Conclusions

Findings from this current study reveal that short-term laboratory training can improve adult listener's ability to discriminate nonnative suprasegmental features thus suggests that the human perceptual system remains at least partially malleable over the life span.

### 5. References

- [1] Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). "Training Japanese listeners to identify English /r/ and /l/: A first report", *J. Acoust. Soc. Amer.*, Vol 89, 874-886.
- [2] Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). "Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variability in learning new perceptual categories", *J. Acoust. Soc. Amer.*, Vol 94, 1242-1255.
- [3] Lively, S. E., Yamada, R. A., Tokhura, Y., & Yamada, T. (1994). "Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories", *J. Acoust. Soc. Amer.*, Vol 96, 2076-2087.
- [4] Strange, W. (1992). "Learning non-native phoneme contrasts: Interaction among subject, stimulus and task variables", in Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 197-219) Tokyo, Japan: Ohmsha.
- [5] Jamieson, D. G., & Morosan, D. E. (1986). "Training non-native speech contrasts in adults: Acquisition of the English /θ/-/ð/ contrast by francophones", *Perception & Psychophysics*, Vol 40, 205-215.
- [6] Jamieson, D. G., & Morosan, D. E. (1989). "Training new, nonnative speech contrast: A comparison of the prototype and perceptual fading techniques", *Canadian Journal of Psychology*, Vol 43, 88-96.
- [7] Polka, L. (1992). "Characterizing the influence of native language experience on adult speech perception", *Perception & Psychophysics*, Vol 52, 37-52.
- [8] Werker, J., & Tees, R. (1983). "Developmental changes across children in the perception of non-native speech sounds", *Canadian Journal of Psychology*, Vol, 37, 278-286.
- [9] Werker, J. F. & Tees, R. C. (1984). "Phonemic and phonetic factors in adult cross-language speech perception", *J. Acoust. Soc. Amer.*, Vol 75, 1866-1878
- [10] Burnham, D. & Francis, E. (1997). "The role of linguistic experience in the perception of Thai tones", in A. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong* (pp. 29-47). Bangkok: Chulalongkorn University Press.