Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 36 (2008) 250–267
Journal of Phonetics
www.elsevier.com/locate/phonetics

Effects of two training procedures in cross-language perception of tones

Ratree P. Wayland*, Bin Li

Program in Linguistics, 4131 Turlington Hall, University of Florida, Gainesville, FL 32611-5454, USA

Received 18 August 2005; received in revised form 11 May 2007; accepted 1 June 2007

Abstract

This study evaluated two perceptual 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 inter-stimulus-interval (ISI) conditions (500 and 1500 ms). Participants received 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 listeners outperformed NE listeners both before and after training under both ISI conditions; (b) before training, NC listeners' discrimination was better under the longer ISI while NE listeners' performance was comparable across the two ISIs, but no ISI difference was observed for either group of listeners after training; (c) both NE and NC listeners' performances significantly improved after training, but the improvement was significantly greater among NE listeners under both ISI conditions, and (d), the amount of improvement was comparable across the two training procedures and across the two ISI conditions. These results suggest that both ID and SD training procedures were equally effective in improving NE and NC listeners' discrimination of the mid- vs. low-tone contrast in Thai and that prior experience with a tone language may prove advantageous in learning another tone language.

1. Introduction

Native-like production and perception in a second language (L2) is rarely achieved among L2 learners whose L2 learning does not begin until adulthood (Lenneberg, 1967). This finding led to the hypothesis that, similar to first language learning, L2 speech learning involves a timetable or a critical period (e.g., Lenneberg, 1967; Patkowski, 1989; Scovel, 1969). However, even after several decades of much research and debate, it remains inconclusive whether the critical period is conditioned by biological or environmental factors. It can be gleaned from previous research that the deterioration, slowing, or complete loss of some basic speech learning mechanism(s), inadequate L2 phonetic input (Flege, 1995a), and similarities and differences between the L1 and L2 phonological systems (Best, 1995) independently and/or collectively contribute to L2 speech perception and production difficulties among adult L2 learners. If the human speech learning mechanism is lost or reduced after the critical period, a reasonable question that one may ask is whether accurate production

*Corresponding author. Tel.: +1 352 392 0639x225; fax: +1 352 392 8480.
E-mail addresses: ratree@ufl.edu (R.P. Wayland), binli2@cityu.edu.hk (B. Li).

0095-4470/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2007.06.004

and perception of L2 speech sounds can be achieved among adult L2 learners. Thus, the overall goal of this study was to examine efficacy of short-term laboratory training on the perception of a non-native lexical tone contrast by adult listeners. Specifically, it investigated discrimination of the low- vs. mid-tone contrast in Thai by native American English (NE) and native Mandarin Chinese (NC) speakers.

Despite persuasive arguments and research findings in support of Lenneberg's Critical Period Hypothesis (CPH) for L2 acquisition (e.g., Johnson & Newport, 1989; Patkowski, 1989), a number of laboratory training studies have been conducted in recent years. Contrary to CPH, these laboratory training studies were conducted based on the assumption that the human perceptual system remains at least partially malleable over the life span. For example, a number of studies designed to train American English listeners to perceive non-native voicing contrast have been conducted. In 1982, Pisoni et al. (1982) reported that after a short period of laboratory training, monolingual speakers of American English were able to reliably label and discriminate voiceless aspirated, voiceless unaspirated and voiced stops differing in voice-onset time (VOT). McClaskey, Pisoni, and Carrell's (1983) study further reported that experience gained from discrimination training of VOT on one place of articulation (e.g., labial) could be transferred to another place of articulation (e.g., alveolar). Studies designed to examine the ability to identify the American English fricatives /θ-ð/ among native speakers of French (e.g., Jamie & Morosan, 1986) and the American English /l/ and /r/ among Japanese listeners (e.g., Bradlow, Pisoni, Yamada, & Tokhura, 1997; Lively, Logan, & Pisoni, 1993; Lively, Yamada, Tokhura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991) have also been conducted. In general, a significant improvement in the identification of non-native phones after training has been reported in all of these studies. More importantly, the improvement was found to extend to other non-native phones in novel contexts. The beneficial training effect has also been found to persist long after training (e.g., Bradlow et al., 1997; Flege, 1995b). These findings were interpreted as evidence in support of the view that the perceptual mechanisms used by adults in categorizing speech sounds can be easily modified with simple laboratory techniques in a short period of time. These findings also suggest that the human perceptual system remains at least somewhat malleable over the life span. This interpretation and suggestion are also consistent with the findings that adult L2 speech production and perception accuracy improve with experience (e.g., Flege, Bohn, & Jang, 1997; Flege, Takagi, & Mann, 1996; Yamada & Tokhura, 1992).

In contrast to segmental training studies, studies focusing on suprasegmental (e.g., stress/accent, intonation and lexical tones) learning remain scarce. Among these previous studies, only three are specifically designed to investigate the perception of non-native contrasts of lexical tones. A lexical tone contrast refers to a contrast in which variation in voice-pitch, an auditory impression of rates of vocal fold vibration (f_0), serves to differentiate word (lexical) meaning. In tone languages, differences in either average f_0 or f_0 contours over strings of otherwise identical phonemes distinguish different words in the lexicon from one another. In Mandarin Chinese, for example, the syllable /ma/ produced with four different pitch contours (i.e., 1 high-level, 2 high-rising, 3 low-falling, 4 high-falling will result in four different words: /ma1/ 'mother', /ma2/ 'hemp', /ma3/ 'horse', and /ma4/ 'scold'; Chao, 1948). On the other hand, pitch variation is not used to differentiate word meaning in an intonation language such as English. Instead, variation in pitch over an utterance can indicate whether it is a question or a statement. At word level, English exhibits lexical stress. That is, in a multi-syllabic word, one syllable is perceptually more prominent than neighboring syllables. The pitch of this 'stressed' syllable is typically higher than its neighboring 'unstressed' syllables.

Wayland and Guion (2004) investigated the ability to identify and discriminate the mid- vs. low-tone contrast in Thai by NE and NC listeners before and after auditory training under two inter-stimulus-interval (ISI) of presentation (500 ms vs. 1500 ms). Specifically, a categorial identification training procedure was used. In this training procedure, for each trial, listeners heard three Thai words produced by three different talkers. In 'change' trials, one of the three words was produced with a different tone from the other two, but all three words were produced with the same tone in 'no change' trials. The listener's task was to select the odd tone out if there was one, or to select 'none' when one was not detected. Correct feedback was given immediately after the response. The study found that the NC group outperformed the NE group in their ability to discriminate the two Thai tones under the 500-ms ISI condition before training and under both ISI conditions after training. However, a significant improvement in identification from the pre-test to the post-test was observed in the NC group under both ISI conditions, but not in the NE group. However, Wayland and Guion (2003) found that without laboratory training, American English listeners with prior experience (i.e., having studied

Thai and/or lived in Thailand) with Thai were better at discriminating the mid and low tones in Thai than those lacking the experience. Wang, Spence, Jongman, and Sereno (1999) reported a significant improvement in Mandarin Chinese tone identification among American L2 learners of Chinese after 2 weeks of tone identification training.

It is important to note that improvement observed after short-term laboratory training is believed to be due to modifications in selective attention and/or reweighting of multiple properties that define the non-native categories (Flege, 1995b). Disagreement exists, however, regarding the optimal method(s) for inducing such changes in the perception of speech segmental and suprasegmental properties (Flege, 1995b). Important methodological factors that may affect the outcome of laboratory training include training procedures (e.g., identification vs. discrimination) and the inter-stimulus-interval (ISI) used. The following section presents a review and discussion of studies on the effect of two training procedures, namely identification and discrimination, and the effect of ISI on speech perception.

2. Previous literature on effects of training procedures and ISI on speech perception

2.1. Identification vs. discrimination training

As mentioned, the effectiveness of short-term laboratory training on non-native phonetic contrasts among adult listeners has been reported in several studies (e.g., Bradlow et al., 1997; Flege, 1989; Flege & Wang, 1990; Jamie & Morosan, 1986; Lively et al.,1994; Logan et al., 1991; Polka, 1991; Pruitt, Strange, Polka, & Aquilar, 1990; Strange & Dittman, 1984; Wang et al., 1999; Wayland & Guion, 2003, 2004). Results of these studies showed a significant improvement in the ability to identify or discriminate non-native contrasts for most non-native subjects following training. The amount of improvement, however, varies considerably from study to study due to differences in the characteristics of the non-native subjects studied (i.e., naïve or experienced subjects) and the nature (i.e., synthetic or natural; single token or multiple-tokens from one or many talkers) as well as the type (i.e., vowel, consonants or lexical tones) of stimuli used. One important difference among these studies is the procedures used in the training (Strange 1992). Specifically, two training procedures, namely *identification* (ID) training and *same/different discrimination* (SD) training have been used. Commonly, stimuli used in training studies are multiple productions of minimal pairs that exemplify the phonetic contrast of interest (e.g., pick vs. pig for the /k/ vs. /g/ contrast).

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., /k/ or /g/). Feedback regarding the correct identity of the stimulus is given immediately after a response is received. Through the provision of feedback, subjects are expected to gradually learn the phonetic properties associated with each category. On the other hand, subjects who are administered an 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. Similar to the ID training, feedback is given once a response is received.

The advantages and disadvantages of the two training procedures have been a subject of debate. Logan et al. (1991) suggested that ID training might be more effective than SD discrimination training in that it encouraged subjects to rely more on phonetic codes stored in long-term memory than on rapidly fading sensory information in short-term memory. Similarly, Lively et al. (1994) assumed that the combination of ‘‘minimal uncertainty’’ of the two-alternative forced choice ID procedure and the provision of immediate feedback (p. 2076) 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. Jamie and Morosan (1986, 1989) suggested that SD 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. However, Polka (1992) argued that subjects who received ID training may learn to respond correctly by attending to any properties that might be used to differentiate the two non-native categories. Some of these properties may not be the ones used by native speakers.

The above disagreement over the advantages of ID and SD training stem partly from the existence of different types of SD discrimination procedures. According to Burnham, Earnshaw and Quinn (1987), the same process, perception of any psychoacoustic difference between the particular tokens, underlies the ID and SD discrimination procedures, especially among infants. However, these researchers left open the possibility that different neural or cognitive processes may be involved in ID and SD discrimination among adults. That is, unlike discrimination which only involves detection of perceptible differences among tokens, identification among adults may first involve the abstraction of common features that group sounds together and later the comparison of each token with this abstraction. Strange (1992), on the other hand, distinguished between the traditional SD (AX) discrimination and a 'categorial' SD discrimination. Unlike traditional AX discrimination, multiple tokens of each category are used in a categorially same/different discrimination; as such stimuli in each pair are always physically nonidentical. Thus, the subjects' task was not to decide whether one stimulus had been presented twice in succession (i.e., physically identical), but to determine if two different realizations of a single phonetic category had been presented (Flege, 1995b). It is argued that, unlike traditional AX discrimination, the presentation of multiple tokens of each category in a categorially same/different discrimination may encourage subjects to ignore within-category acoustic differences. Polka (1992), for example, maintained that a certain degree of perceptual constancy is required for successful performance in a categorial SD discrimination task since multiple tokens of each category are likely to encompass a wide range of acoustic variants.

Flege (1995b) directly compared the relative efficacy of the identification and categorial SD training procedures. Subjects in this study were two groups of Mandarin Chinese speakers who were trained to distinguish English /t/ and /d/ in word final position. One group of subjects received a two-alternative forced choice training and the other was administered a categorial SD training. The results obtained indicated that small but significant gains were observed in both groups. These gains were also evident in a delayed post-test 2 months after training. Moreover, for both groups, the effect of training generalized to words not included in the training. Interestingly, the magnitude of generalization did not differ significantly between the two groups. The author took these results as evidence against the view that identification training is superior to same/different training.

2.2. *Effect of inter-stimulus-interval (ISI) on perception*

ISI has been shown to affect speech perception among non-native listeners in previous studies (e.g., Burnham & Francis, 1997; Burnham, Kirkwood, Luksaneeyanawin, & Pansottee, 1992; Werker & Tees, 1983; Werker & Tees, 1984). For example, Werker and Tees (1984), reported that native English subjects who had no prior experience with Hindi or Thompson (a North American Indian language) had difficulty discriminating the Hindi voiceless, unaspirated retroflex vs. dental place of articulation (/ʈa- t̪a/) contrast and the Thompson glottalized velar vs. glottalized uvular (/k̀i-q̀i/) contrast at a 1500-ms ISI, but that they could discriminate them at a 500-ms ISI. According to these authors, the use of English native categories stored in long-term memory while performing the task was responsible for their failure at the longer ISI. Specifically, these authors proposed that a phonological mode of processing was activated during discrimination at the 1500-ms ISI. Since both members of the Hindi and Thompson contrasts mapped onto a single English phonological category (i.e., alveolar /t/ and velar /k/, respectively), the subjects showed poor discrimination at this longer ISI. In contrast, a phonetic mode of processing was involved in the 500-ms ISI condition. Since both members of the contrasts are phonetically distinct, the subjects were sensitive to these non-native contrasts at the shorter 500-ms ISI. A phonetic mode of processing is defined, by these authors, as a language-general mode of perception in which phones are discriminated based on their phonetic difference without any influence of the subject's prior linguistic experience. On the contrary, experience with the phonological system of a particular language, most likely the native language, is believed to constrain or enhance perception at the phonological level of processing (e.g., Burnham & Francis, 1997; Werker & Tees, 1983).

In addition to the phonetic and phonological modes of speech perception, the third mode of perception, namely the 'auditory' mode was proposed by Werker and Logan (1985). Similar to the phonetic mode, the auditory mode of processing is a language general mode of speech perception. However, unlike the phonetic mode, discrimination at the auditory level is based on acoustic variability between the individual non-native phones being discriminated rather than on (non-native) phonetically relevant categories. In other words, two acoustically different exemplars of a Hindi dental stop will elicit a 'different' response from native English speakers in a same/different discrimination task more frequently under the auditory mode of perception than under the phonetic mode of processing. However, the percentage of 'different' responses will increase under the phonetic mode of processing when two exemplars are drawn from two different Hindi categories (i.e., Hindi dental and retroflex stops). According to Werker and Logan (1985), auditory processing strategy was found under both 250- and 500-ms ISI conditions.

On the basis of Werker and Tees's (1983) proposal of two processing modes (i.e., phonetic and phonological), Burnham and Francis (1997) predicted that, when discriminating Thai lexical tone contrasts, the phonetic mode of processing would be activated among native (Australian) English speakers who had no prior exposure to Thai. Thus, their performance would be better in the 500-ms ISI condition than in the 1500-ms ISI condition. On the other hand, perception at the phonological level would be activated among the native Thai speakers. It was, therefore, predicted that they would perform better in the 1500-ms ISI condition than in the 500-ms ISI condition. However, these predictions were borne out for only some of the tone contrasts that were tested. Moreover, type of tonal contrast interacted significantly with processing levels. In particular, only the contour–contour (e.g., rising–falling) tone contrast exhibited beneficial effect of the short 500-ms ISI. Furthermore, the advantage of the 500-ms ISI was evident among both the native Australian English and the native Thai speakers. These results led the authors to speculate that the contour–contour contrast is acoustically more salient than either the level–level (e.g., mid-low, mid-high, low-high) or the level–contour (e.g., mid-rising, mid-falling, low-rising, low-rising) contrasts; and that this salient acoustic information was exploited in the phonetic mode of processing, thus a higher discrimination score in the 500-ms ISI condition.

However, alternative interpretations to the 500-ms ISI vs. 1500-ms ISI effect are available. First, using a lexical or shadowing tasks, phonological priming research has shown that presentation of a prime that shares some phonological similarity (i.e., shared initial 1, 2, or 3 phonemes) with a following target can affect a subject's response times in these tasks (see e.g., Hamburger & Slowiaczek, 1996 and references therein). More importantly, the ISIs used in these studies are typically around 500 ms. These findings, thus, suggest that stored lexical representation in the subject's long-term memory can be accessed within 500 ms. Therefore, an advantage offered to inexperienced subjects in discriminating non-native contrasts by a shorter ISI of 500 ms does not exclude their access to stored phonological information. Second, according to findings obtained from semantic priming studies, lexical access may take place in less than 200 ms (Sabol & DeRosa, 1976). Together, these results suggest that access to stored phonological category representations during a phonological discrimination task is possible within 500 ms. In other words, activation of both phonetic and phonological modes of perception is possible under an ISI of 500 ms or shorter.

Alternatively, when compared to the longer 1500-ms ISI, a weaker demands placed on a subject's working memory by a shorter 500-ms ISI may be responsible for its observed facilitory effect attested among non-native speakers. The reason that this working memory load effect is not observed for native speakers may be due to the fact that they are able to code the stimuli using categories from long-term memory. Thus, they are able to make a decision at a categorical level at both the shorter 500 ms and longer 1500-ms ISIs. In addition, a greater decision time offered by the longer ISI may explain why their discrimination is better in 1500-ms ISI than 500-ms ISI condition. In sum, the origin of ISI effects remains unclear, and a number of possible explanations can be offered to account for these effects. These include constraints on working memory and decision time or the activation of language-general (phonetic) vs. language-specific (phonological) modes of perception.

To our knowledge, an interaction between training procedure and ISI has yet to be explored. That is, a question remains as to whether the effectiveness of the two training procedures will remain comparable under different processing levels (auditory, phonetic or phonological), memory constraint or decision time. Since only one short ISI (250 ms) was used, it could be argued that in Flege's (1995b) study, relative degree of effectiveness of the two training procedures has been investigated under only one mode of processing (i.e., auditory or phonetic according to Werker and Logan (1985) discussed above), or when demand on listener's working memory is relatively low and decision time is relatively short. The current study is designed to fill this gap in the literature by including longer ISIs in its design, thus allowing for a comparison of the two training procedures at a different mode (i.e., phonetic or phonological level) of processing or when a greater load is placed on short-term memory.

3. The present study

This study was conducted to evaluate the relative efficacy of the ID and the categorial SD discrimination procedures in short-term laboratory training of a lexical tone contrast in Thai among native speakers of Chinese and native speakers of English. As mentioned, Chinese is a tone language but English is not. This study differed from Flege’s (1995b) in two important ways. First, unlike Flege (1995b), our study trained native Mandarin Chinese and native American English listeners to discriminate (Thai) tones. Second, while the inter-stimulus-interval (ISI) was fixed at 250 ms in Flege (1995b), two ISIs (500 and 1500 ms) were examined here. This study was guided by the following three questions: (1) Which training procedure would be more effective in training native Mandarin Chinese (NC) and native English (NE) listeners to perceive Thai tones?; (2) Will the performance of the participants vary as a function of ISI?; and (3) Would the effectiveness of the two training procedures vary as a function of ISI?.

Based on earlier work on lexical tone perception reviewed above, we hypothesize that NC listeners will perform better than NE listeners both before and after training. Additionally, we predict that the NE listeners’ performance will be superior under the shorter ISI than under the longer ISI. However, it is not clear if native experience with tones will facilitate the NC listeners’ performance under a shorter or a longer ISI processing condition. Additionally, since this study would be among the first to compare the effectiveness of the two training procedures in lexical tone perception, we are not certain if the same results as those found in Flege’s (1995b) on consonant perception would still hold for our study.

3.1. Thai and Mandarin tones

3.1.1. Thai tones

Thai, the national language of Thailand, has five phonemic tones: a low-tone ([k^{h}àː]’galangal, a kind of aromatic root often used in Thai cooking)’, a mid-tone ([k^{h}aː] ‘to be stuck or lodged in’), a high-tone ([k^{h}áː] ‘to engage in trade’), a falling tone ([k^{h}âː]’I, servant’), and a rising tone ([k^{h}ǎː]’leg’). Phonetically, these five tones may be characterized in terms of pitch contour, pitch height, and voice quality. These characteristics are schematically shown in Table 1 (after Hudak, 1987).

Level tones are referred to as ‘static’ tones and contour tones as ‘dynamic’ tones by Abramson (1978). According to Gandour (1979, 1983), the acoustic dimensions of average pitch level, pitch direction and pitch slope are the main perceptual cues to the discrimination of Thai lexical tones. Additionally, Abramson (1978) reported that although a different pitch level can be sufficiently used to identify static tones, successful identification of dynamic tones requires a detection of a rapid pitch movement. However, the finding that both native Thai and native Australian English adults who had no prior exposure to Thai found the dynamic contrast (rising vs. falling) to be the most easily discriminated, followed by the static vs. static (e.g., low vs. high) then static vs. dynamic (e.g., low vs. rising), suggested to Burnham and his colleagues that initial pitch level may have been the most salient perceptual cues for the listeners (Burnham et al., 1992). In addition, Abramson (1975, 1976) found that the low and mid tones are easily confused even for native speakers of Thai.

Table 1
Characteristics of Thai tones

Tone	Tone mark	Pitch contour	Pitch height	Voice quality
Mid	Unmarked	Level	Medium	Non-glottalized
Low	`	Level	Low	Non-glottalized
Falling	ˆ	Contour	High to low	Glottalized (creaky)
High	´	Level	High	Glottalized
Rising	ˇ	Contour	Low to high	Non-glottalized

The contrast between these two tones was also found to be the second most difficult to discriminate among naïve native Australian English speakers in Burnham and Francis, (1997) study.

3.1.2. Mandarin tones

Unlike Thai, Mandarin Chinese has only four phonemic tones. Tone 1 has been described as having a high-level pitch (/ma/ 'mother'), tone 2 has a high-rising pitch (/ma/ 'hemp'), Tone 3 has a low-falling pitch (/ma/ 'horse'), and tone 4 has a high-falling pitch (/ma/ 'scold') (Chao, 1948). It has been suggested that f_0 height and f_0 contour are the fundamental perceptual cues of Mandarin Tones, but that native Mandarin listeners appear to place more emphasis on the 'contour' dimension than the 'height' dimension (Gandour, 1984; Massaro, Cohen, & Tseng, 1985). Additionally, the f_0 turning point, namely the point at which the direction of the f_0 contour changes from falling to rising, has been found to differentiates Tone 2 from Tone 3 (Moore & Jongman, 1997; Shen & Lin, 1991). Furthermore, the perception of Tones 2 and 3 has been shown to shift with vowel duration (Gårding, Kratochvil, Svan tesson, & Zhang, 1986). That is, listeners reported hearing more Tone 3 when the vowel was systematically lengthened.

3.2. Method

3.2.1. Subjects

Subjects were 12 (5 male, 7 female) native speakers of Thai (NT), 30 (15 male, 15 female) native speakers of Mandarin Chinese (NC) from the Peoples Republic of China and 21 (5 male, 16 female) native speakers of American English (NE) with self-reported normal language and hearing abilities. Mean age for NT participants was 28 years (range 24–36), 28 years (26–36) for NC participants and 23 years (19–33) for NE participants. NT participants had been living in the US for an average of 27 months (8–84) and NC participants 23 months (4–80) at the time of testing. NC and NE participants had no prior experience with Thai. All participants were recruited from the student population of the University of Florida and were paid for their participation. After the pre-test (see Section 3.3 below), NC and NE, but not NT[1] participants were randomly divided into two subgroups before the training. One group received a two-alternative forced choice identification training (the ID group) while the other received a categorially same/different discrimination training (the SD group). There were 15 NC and 11 NE participants in the ID group, and 15 NC and 10 NE in the SD group. All subjects participated in all 3 phases (pre-training, training, and post-training) of the experiment which were completed in 4 consecutive days[2].

3.2.2. Stimuli

Stimuli were 5 minimal pairs (see Table 2) 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 to discriminate among all possible tone contrasts. When presented in isolation, these two tones are also known to be difficult for native Thai speakers to identify (Abramson, 1975, 1976). This tone contrast was thus chosen to avoid a ceiling effect, especially after training. The stimulus words were produced in a Thai carrier phrase [rau p^{h}ûːt k^{h}am wâː ___], 'we say the word ___'.

Each word was produced 3 times in random order by 5 male native Thai talkers. The recording took place in a quiet office setting using a high-quality DAT cassette recorder and a head-mounted microphone. The stimuli

[1] Since the main focus of the study was to compare the effects of training on lexical tone discrimination among non-native (but not among native speakers) NT participants' discrimination was only collected for comparison.

[2] Even though control participants were not included in the study, we have a good reason to believe that the perceptual gains shown after training reported among NC and NE in Section 4 was not due to the participants' test-taking experience. Specifically, we have recently collected control data for another study similarly designed to investigate the issue of 'talker normalization' in lexical tone perception. Unlike the current study, this study involves 1 day of perceptual training using (simpler) synthetic stimuli. Control participants did not participate in the training, but participated in the pre- and post-tests which took place on 2 consecutive days, thus there was only 1 day interval between the two tests. Both the perceptual Dprime scores and the reaction time data indicate that there was no improvement in the ability to discriminate lexical tones among the control subjects. Therefore, with a longer time interval between the pre- and the post-tests, and the more complex (multiple tokens produced by different talkers) stimuli employed in this current study, we are reasonably sure that improved performance among NC and NE participants was due largely to the training treatment, and not to the participants' test-retest experience.

Table 2
Minimal pairs used in the study

Mid-tone	Low-tone
1. [piː] 'year'	[pìː] 'an oboe'
2. [baː] 'nightclub'	[bàː] 'shoulder'
3. [k^{h}aː] 'to be stucked'	[k^{h}àː] 'galangal root'
4. [p^{h}aː] 'to accompany'	[p^{h}àː] 'to split (wood)'
5. [puː] 'a crab'	[pùː] '(paternal) grandfather'

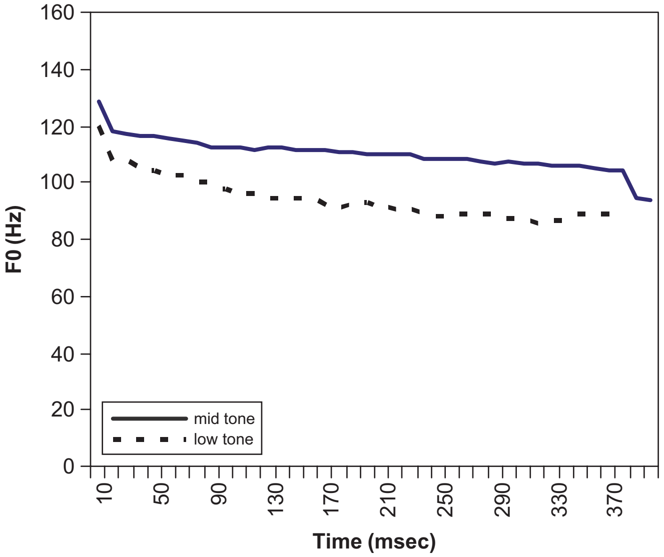

Fig. 1. Normalized f_0 tracks in Hz of a mid-tone [piː] 'year' and a low-tone [pìː]'an oboe' produced by talker 1.

were later digitized using a PC (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 [púː] 'grandfather' were produced by talkers 2 and 3, while the two productions of [puː] 'crab' were produced by talkers 1 and 5. The selection of the two productions was mainly based on their acoustic quality and naturalness as judged by the first author who is a native speaker of Thai. Based on these criteria, one token each of 'crab', 'oboe', 'shoulder', 'to split', 'stuck', and 'year' from talker 1; 'galangal', 'grandfather', and 'to split' from talker 2; 'nightclub', 'oboe', 'galangal', and 'grandfather' from talker 3; 'club', 'oboe', 'grandfather', 'stuck', and 'take' from talker 4; and 'crab', 'shoulder', 'take' and 'year' from talker 5. f_0 tracks of a mid-tone [piː] 'year' and a low-tone [pìː]'oboe' produced by talker 1 are shown in Fig. 1. In addition, the average f_0 for each stimulus word for each talker is given in Table 4 (included in Appendix A1). As shown in Table 4 and Fig. 1, average f_0 values of stimulus words produced with a low-tone were invariably lower than those produced with a mid-tone. f_0 contour of a mid-tone also differs from that of a low-tone. Fig. 2 shows f_0 tracks of a mid-tone [piː] 'year' produced by talker 5, and a mid-tone [baː] 'nightclub' and a low-tone [bàː] 'shoulder' produced by talker 3 and talker 5, respectively. As shown in Fig. 2, even though the average f_0 values of these two tones are comparable (107 Hz vs. 103 Hz) when produced by two different talkers, the low-tone contour shows a higher f_0 at onset followed immediately by a sharp f_0 decline. The mid-tone, on the other hand, exhibits a lower f_0 value at onset followed by a slow and gradual decline toward the end of the word. The same description applies to [puː] 'crab' and

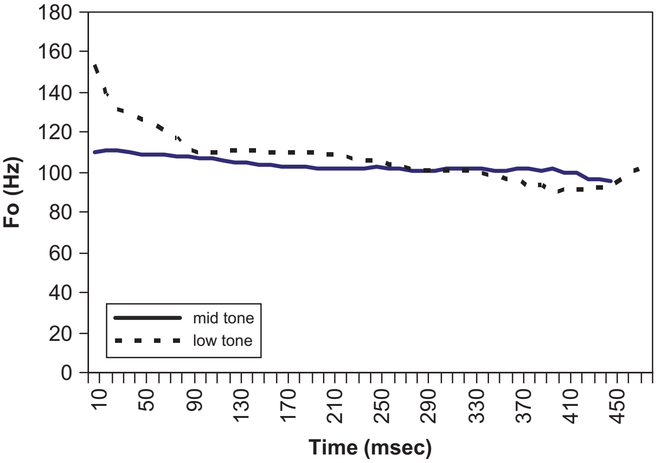

Fig. 2. Normalized f_0 tracks in Hz of a mid-tone [baː] 'nightclub' and a low-tone [bàː] 'shoulder' produced by talker 3 and talker 5 respectively.

[pùː] 'grandfather' produced by talker 1 and 2 whose average f_0 values are also comparable (118 Hz vs. 116 Hz).[3]

It is important to note, however, that, by comparison, the magnitude of the difference in both average pitch and pitch contour between the Thai mid and low tones just described, especially when they produced by different talkers, is much smaller than those of other contrasts. This is the reason why successful discrimination of this tone contrast has been proven challenging even among native Thai speakers.

3.3. Procedure

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., from a different talker) of either A or B. For example, a trial testing the contrast [puː]-[pùː] 'crab-grandfather' might consist of [puː]-1, [puː]-3, [pùː]-2 (where the number indicates different talkers). Both productions of each member of a contrast were distributed equally over the three positions resulting in 16 trials[4] for each contrast. After hearing all three stimuli, participants were asked to decide whether the tone of the second stimulus was 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 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 3000 ms.

3.3.1. Pre-test

The NT, NC and NE listeners were tested individually in a quiet room in one session that lasted approximately 45 min using a PC. The 80 (5 contrasting pairs × 16) trials for each ISI condition were randomly presented over headphones at a comfortable listening level. The listeners were told that each trial

[3] Similarity in f_0 values between these two stimuli may have been responsible for why a slightly higher number of Thai subjects committed errors in trials involving these two stimuli (i.e., between 5 to 6 out of 12 as opposed to fewer than 5 for trials involving other stimuli).

[4] If 1 and 2 are the two tokens of the low tone member of the contrast and 3 and 4 are those of the mid tone member, the 16 trials testing this contrast will be as follows: $A_1A_2B_1$, $A_2A_1B_1$, $A_1A_2B_2$, $A_2A_1B_2$, $B_1B_2A_1$, $B_1B_2A_2$, $B_2B_1A_1$, $B_2B_1A_2$, $B_1A_1A_2$, $B_1A_2A_1$, $B_2A_1A_2$, $B_2A_2A_1$, $A_1B_1B_2$, $A_1B_2B_1$, $A_2B_1B_2$, $A_2B_2B_1$.

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 practice session with 10 trials drawn from the 80 experimental trials was presented without feedback. In addition to the 10 practice trials, the 80 experimental trials (for each ISI) were preceded by five warm-up trials also drawn from the 80 experimental trials that were not analyzed.

3.3.2. Training

Following the pre-test, NC and NE, but not NT participants were asked to come back for the next two days for training sessions. Each session lasted 60 min. One group of NC ($n = 15$) and NE ($n = 11$) participants received a two-alternative forced choice identification (ID) training while the other ($n = 15$ for NC and $n = 10$ for NE, respectively) received a categorial same/different discrimination (SD) training. The stimuli used during training were identical to those used in the pre-test.

3.3.2.1. ID training. 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 system, these two tones were referred to as Tone A (mid-tone) and Tone B (low-tone) instead of Tones 1 and 2. Participants heard one stimulus per trial. Similar to the pre-test, the ITI was fixed at 3000 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 told that they could replay each trial as many times as they wished. 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 random. 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 s. 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 could complete this training as many times as they wished within the allotted time of 1 h. As it turned out, every participant was able to complete two sessions within 1 h.[5]

3.3.2.2. SD training. 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 s. However, the ITI remained at 3000 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), a same trial was followed by a different trial. That is, 'odd' number trials were same trials and 'even' number trials were different trials. Participants were asked to push a button marked 'same' for same trials and a button marked 'different' for different 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 s. Replay for a given trial was

[5] Since the training was self-paced (i.e., participants can replay trials as many times as they wish), the one-hour limit was imposed to make sure that the amount of time that participants exposed to the stimuli was comparable. As it turned out, none of the participants were too fast to complete more than two sessions or too slow to complete fewer than two sessions.

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. However, similar to ID training, every participant completed two training sessions in one hour.

3.3.3. *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.

3.4. *Data analysis*

Pre- and post-test data was statistically analyzed using SPSS 12.0 to test for the effects of L1 background (Thai, Chinese, and English), ISI (500 and 1500 ms) and training procedure (ID and SD). Since native Thai speakers did not participate in the training and the post-test phase of the experiment, pre-test scores were analyzed only for the effects of L1 background and ISI. For this analysis, a two-factor repeated measures ANOVA was first performed with L1 (Thai, Chinese, and English) as the between-subjects factor and ISI (500 and 1500 ms) as the within-subjects factor. A significant main effect of L1 was further explored through Bonferroni post-hoc pair-wise comparisons, and a significant interaction between L1 and ISI was further examined using a *t*-test.

To test for the effects of training and training procedure among native Chinese and native English speakers under both ISI conditions, a second analysis was performed using both pre- and post-test scores. For this analysis, a four-factor repeated ANOVA was performed with L1 (Chinese and English) and training Group (ID, SD) as the between-subjects factors, and ISI (500 and 1500 ms) and Test Time (pre- and post-test) as the within-subjects factors. A *t*-test was then used to further explore the significant interaction effects.

4. Results

4.1. *Pre-test*

Mean percent correct discrimination scores obtained during the pre-test by all three groups of participants[6] under each of the two ISI conditions (500 and 1500 ms) are presented in Fig. 3. As expected, NT listeners scored higher than both NC and NE listeners under the shorter 500-ms ISI. However, NT listeners' score was higher than that of NE listeners only under the longer 1500-ms ISI. Additionally, NC listeners scored higher than NE listeners under both ISI conditions. Furthermore, while NT listeners scored better under the shorter ISI condition, NC and NE listeners performed better under the longer ISI condition.

Percent correct discrimination score data reported in Fig. 3 was submitted to a repeated measures ANOVA with ISI (500 and 1500 ms) as the within-subject factor and L1 (Thai, Chinese, and English) as the between-subject factor. The alpha level was set at .05. The analysis yielded a significant main effect of L1 [$F(2,55) = 64.65$, $p < .001$], but not of ISI [$F(1,55) = .33$, $p = .57$]. Post hoc pair-wise comparison using the Bonferroni method revealed that NT listeners outperformed ($p < .04$) both NC (91% vs. 82%) and NE (91% vs. 58%) listeners ($p < .001$) under the 500-ISI condition. Interestingly, NT listeners performed significantly better than NE listeners only (86% vs. 61%) under the 1500-ISI condition ($p < .001$). More importantly, NC listeners outperformed ($p < .001$) NE listeners under both ISI conditions (82% vs. 58% for 500-ms ISI and 87% vs. 61% for 1500-ms ISI). ISI interacted significantly with L1 [$F(2,55) = 5.61$, $p < .008$] due to the fact that NT listeners performed significantly better under the shorter 500-ms ISI than under the longer

[6] To avoid entering spurious data (outliers) into the analyses, data from 3 NC listeners (2 from the ID group and 1 from the SD group) and 2 NE listeners (both from the SD group) were excluded, using the 2 standard deviations (SD) cut-off criterion (e.g., Flege, Munro, & McKay, 1995), from all analyses reported in the result section. One NC participant's pre-test score at 500-ms ISI was greater than 2 standard deviations (SD) below the group mean while the other two failed at the 1500-ms ISI condition. For the two NE listeners, one showed a perceptual gain that fell below 2 SD of the group mean under the 500-ms ISI while the other failed at the 1500-ms ISI condition. One NC listener showed a gain of .25% above 2 SD of the group mean under the 1500-ms ISI, but his score was included in the analysis because exclusion of his score did not change the outcome of the analyses. Scores of all remaining participants remained within 2 SD of the group mean.

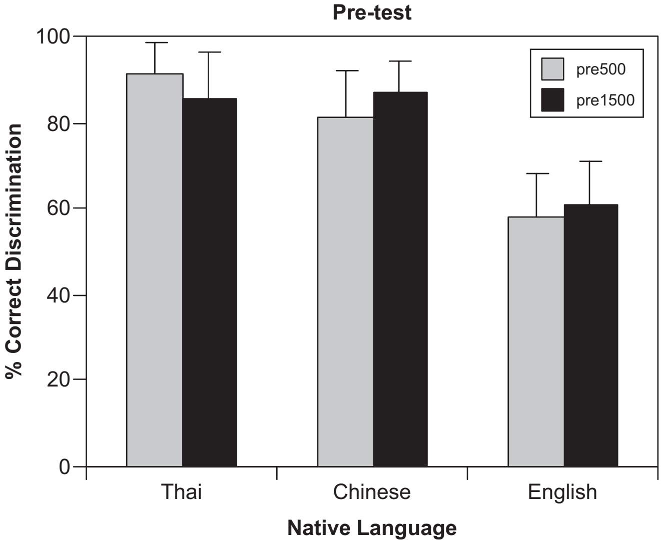

Fig. 3. Mean percent correct discrimination scores and standard deviations of the mid- vs. low-tone in Thai by native Thai, native Mandarin Chinese, and native American English listeners obtained under 500 and 1500-ms ISI.

1500-ms ISI condition (91% vs. 86%) [$t(11) = 3.22$, $p < .008$]. On the contrary, NC listeners scored better under the longer ISI than under the shorter ISI (87% vs. 82%) [$t(26) = -2.63$ $p < .01$]. The difference between the two ISI conditions did not reach significance among NE listeners (58% vs. 61%) [$t(18) = -1.19$, $p = .25$].

4.2. Effects of training and training procedures

Mean percent correct discrimination scores obtained before and after the training by the two training groups (ID and SD) under both ISI conditions among NC and NE listeners are presented in Fig. 4. Overall, NT performed at a higher level than NE under all conditions. This data was submitted to a four-factor repeated ANOVA with L1 (Chinese and English) and Training Procedure (ID and SD) as the between-subjects factors, and ISI (500 and 1500 ms) and Test Time (pre- and post-test) as the within-subjects factors. The analysis yielded a significant main effects of L1 [$F(1{,}42) = 82.73$, $p < .001$], ISI [$F(1{,}42) = 9.59$, $p < .003$], and Test Time [F (1,42) = 58. 46, p<001]. It is important to point out, however, that the main effect of ISI obtained was due mainly to the fact that NC listeners scored significantly higher under the longer 1500-ms ISI during the pre-test [$t(26) = -2.63$, $p < .01$]. Their scores under the two ISI conditions ceased to differ after the training [$t(26) = -1.59$, $p = .13$]. In contrast, NE listeners' performance remains comparable across the two ISI conditions both before [$t(18) = -1.19$, $p = .25$] and after training [$t(18) = -1.46$, $p = .16$].

Interestingly, the main effect of Training Procedure was not significant [F(1, 42) = . 13, $p = .72$] nor was the interaction between Training Procedure and Test Time [$F(1,\ 42) = 1.32$, $p = .26$] or ISI [$F(1,\ 42) = .56$, $p = .47$]. However, Test Time was found to interact significantly with L1. This was due to the fact that, averaged across both ISI conditions, perceptual gain after training was significantly greater (11.5% vs. 6.4%) among NE than NC listeners [$t(24) = -2.24$, $p < .03$]. No other significant interaction was found. These results indicated that overall NC listeners outperformed NE listeners, but NE exhibited a greater amount of perceptual gain after training than NC. Importantly, these results also suggested that both ID and SD training groups improved significantly and to the same extent after training. The improvement was also comparable under both ISI conditions.

In summary, the results obtained indicate that before training (a) NT listeners scored significantly higher than both NC and NE listeners under the shorter 500-ms ISI condition, but they (NT listeners) scored significantly higher than NE listeners only under the longer 1500-ms ISI, (b) NC listeners were better than NE

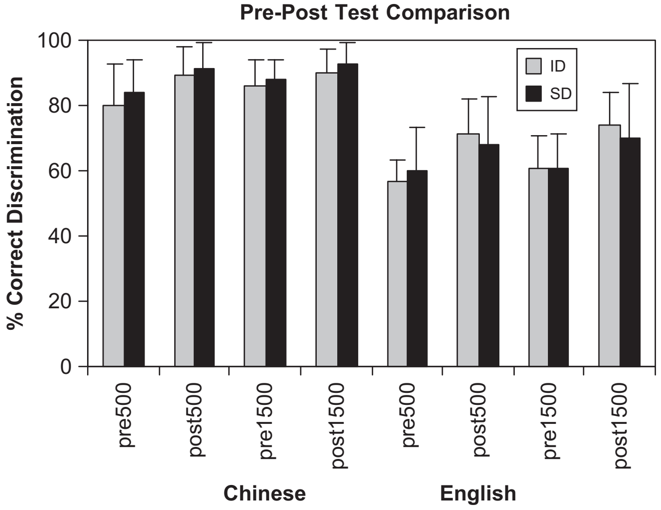

Fig. 4. Mean percent correct discrimination scores and standard deviations obtained by native Mandarin Chinese and native American English listeners in ID and SD training groups under 500 and 1500-ms ISI.

listeners under both ISI conditions, and (c) NT listeners performed significantly better under the shorter ISI but NC listeners did better under the longer ISI. NE listeners' performance, on the other hand, was comparable across the two ISIs. After training, the analyses showed that (a) both NC and NE listeners' performances significantly improved, but NE listeners' improvement was greater than that of NC listeners, (b) both NC and NE listeners performed equally well under the two ISI conditions, and (c) the amount of improvement among both NC and NE listeners was comparable across the two training groups and across the two ISI conditions.

5. Discussion

This study was designed to explore the effectiveness of two perceptual training procedures, namely the two-alternative forced choice identification and the categorial same/different discrimination, on the ability to discriminate the low- vs. mid-tone contrast in Thai under two different ISI conditions. Two groups of NC listeners and two groups of NE listeners who have never been exposed to Thai participated in the study. A group of native Thai listeners served as controls and only participated in the pre-test phase of the study while NC and NE listeners participated in all three phases (pre-test, training, and post-test). One group of NC and one-group NE listeners received the two-alternative forced choice identification training while the other received the categorial same/different training. The study sought to assess whether the two training procedures would improve the ability to discriminate the low- vs. mid-tone contrast in Thai to different degree. The study also asked whether the listeners' performance would differ between two training procedures and ISIs.

5.1. *Effects of training and training procedures*

The results indicated that NC and NE listeners were significantly better at discriminating the two Thai tones after training.[7] Specifically, an average gain between 7.7% (SD 500 ms) to 14.4 (ID 500 ms) was observed

[7] A post-hoc comparison between native Thai listeners' performance on the pre-1500 ISI and the native Chinese listeners' performance under the same ISI condition after the training (i.e., post-1500 ISI) using a t-test revealed that the NC's post-1500 ISI scores were significantly [$t(37) = 2.03$, $p < .05$) higher that that of NT's pre-1500 ISI scores. This result, however, should not be taken to mean that a short-term perceptual training is more effective than decades of native listening experiences. This is because it is possible that even the

Table 3
Mean and standard deviation (in parentheses) of percentage gains in discrimination after the training by each training group under each ISI conditions

NC				NE			
ID		SD		ID		SD	
500 ms	1500 ms	500 ms	1500 ms	500 ms	1500 ms	500 ms	1500 ms
8.8(10.0)	4.2(4.5)	7.5(8.4)	5.2(3.5)	14.4(11.1)	12.9(10.3)	7.7(15.7)	9.4(12.2)

among NE listeners (Table 3). On average, these gains were lower than those reported for previous tone training studies. For example, a gain of 21% was reported for the identification of Mandarin tones by native American English listeners in Wang et al. (1999). It is likely that this difference was due to a shorter period of training provided in the present study (1 h per day for 2 days vs. 40 min per day for 2 weeks). Moreover, participants in Wang et al.'s (1999) study were native American English learners of Mandarin Chinese while NC and NE participants in our study had no prior exposure to Thai tones prior to the study. Furthermore, as reported, NT listeners' performance was good, but not perfect suggesting that the Thai mid- vs. low-tone contrast was rather challenging to discriminate. This tone contrast also posed a relatively higher degree of perceptual difficulty for Mandarin listeners based on a pilot study.

Additionally, a perceptual gain between 4.2% (ID 1500 ms) to 8.8% (ID 500 ms) was realized among NC listeners in this study. However, since no previous data on the discrimination of lexical tones by native speakers of a different tone language is available for comparison, it is not possible at present to evaluate their performance.

On the other hand, the finding that NE listeners' discrimination improved to a significantly greater degree (under both ISI conditions) than NC listeners is of great interest. This result likely suggests a possible ceiling effect among NC listeners rather than relatively greater discrimination ability among NE listeners after training. That is, an initial lower level of discrimination ability among NE listeners during the pre-test afforded them a much greater room for improvement after the training. On the contrary, NC listeners whose initial performance was already at a high level quickly reached optimal level of performance (as suggested by the NT group's performance) afforded by the training. A further study with a relatively more challenging task (i.e., the categorial odd-ball identification task used in Wayland & Guion, 2004) with more tokens produced by different talkers for each tone may be needed to assess whether NC listeners will benefit from the training to the same extent as NE listeners.

More interestingly, it was found that the two training procedures (ID and SD) appeared to be equally effective in promoting the ability to discriminate the two Thai tones among both NC and NE listeners under both ISI conditions (i.e., 500 and 1500-ms ISI). That is, comparable perceptual gains were observed in both training groups under both ISI conditions after training. While a ceiling effect similar to the one described above may have been responsible for this finding among NC listeners, it is unlikely that such an explanation also applies to the same result obtained for NE listeners. This finding is in agreement with that of Flege (1995b) who found that both the ID and the SD training procedures were equally effective in training native speakers of Mandarin Chinese speakers to identify the English final voiced/voiceless stops. According to Flege, the ID training method is not superior to the same/different training in promoting a more robust phonological representation of speech sounds; rather the only advantage of the ID training seems to be a psychological one. Flege reported that participants who received identification training felt that they benefited more from the training than did participants who received categorial SD training.

(*footnote continued*)
native Thai listeners may benefit from the training. As mentioned, this low–mid tone contrast when presented in isolation was very difficult to discriminate among the Thai listeners. Further familiarity with the stimuli during training may also improve their performance on this difficult contrast. Second, multiple tokens produced by multiple speakers were used in this current study. Successful performance thus requires a certain level of abstraction or normalization across voice pitches, a process which, even for native listeners, may improve with training.

Additionally, the identification participants reported that they enjoyed the training more than the categorial SD participants did.

The result that ID and categorial SD discrimination are equally effective training procedures also supports the view that, like identification participants, participants in the categorial SD group are encouraged to ignore token-specific acoustic features and to focus instead on establishing 'perceptual constancy' or 'equivalence classes' based on phonetically relevant properties (Flege, 1995b; Polka, 1992; Strange, 1992). Additionally, this finding is in agreement with the observation that unlike traditional SD, categorial SD procedure encourages listeners to ignore within-category differences as well as to rely less on information in sensory memory (Strange, 1992).

It is possible however, that a subtle difference in relative efficacy between ID and categorial SD training still exists due to differences in their underlying processes. This possibility is suggested by the finding that overall, perceptual gains realized by participants in the ID group are numerically higher than those obtained by participants in the SD group. This was the case under the 500-ms ISI condition among NC listeners (8.8% vs. 7.5%) and under both ISI conditions (14% vs. 7.7% for 500-ms ISI and 12.9% vs. 9.4% for 1500-ms ISI) among NE listeners (Table 3). One might stipulate, for example, that identification training encourages listeners to pay more attention to a category's 'inclusionary' features, while both 'inclusionary' and 'exclusionary' features are simultaneously attended to in categorial SD training. That is, listeners who received ID training ask 'which category the token belongs to based on the acoustic features it exhibits' while those in the categorial SD training simultaneously ask 'why the two tokens should belong to the same or different phonetic categories.' The numerically higher discrimination scores obtained by participants in the ID training group after training suggest that focusing on 'inclusionary' or 'grouping' features may have been a more effective strategy in the long run. This hypothesis remains speculative and is not currently supported by Flege's (1995b) finding that no significant difference was found between the two training procedures after 10 days of training. Therefore, a longer period of training may be needed before the difference between the two training procedures materializes. Additionally, an analysis of reaction times data in future studies may yield further insights on the difference between the two training procedures.

5.2. *Effects of ISI*

During the pre-test, it was found that NT listeners performed significantly better under the shorter ISI condition than under the longer ISI condition. This finding is in disagreement with that of Burnham and Francis (1997). Specifically, these researchers found the advantage of the shorter 500-ms ISI, but only for the contour–contour (e.g., rising–falling) and not for the level–level (i.e., low–mid) tone contrast. As mentioned, the explanation provided by these researchers was that the contour–contour contrast is acoustically more salient than either the level–level (e.g., mid–low, mid–high, low–high) or the level–contour (e.g., mid–rising, mid–falling, low–rising, low–rising) contrasts and that this salient acoustic information was exploited in the phonetic mode of processing resulting in a higher discrimination score. Thus, the advantage of the 500-ms ISI on the mid–low-tone contrast found in this study is at odds with this explanation and suggests instead that small and subtle acoustic differences between these two level tones were better detected at the phonetic level of processing, at least among this group of native Thai listeners. It should be noted, however, that the stimuli used in the Burnham et al.'s study were produced by a single talker while the stimuli used in this study were produced by five talkers. As such, normalization of speakers' voice pitch was not something with which native Thai listeners in the Burnham et al.'s study had to contend. Moreover, the discrimination task employed in that study was the traditional AX same/different discrimination while a categorial AXB discrimination was used in the current study. It is likely that the categorial AXB task with tokens from multiple talkers employed in our current study placed a greater demand on listener's processing capacity than the AX discrimination task. Therefore, it may be reasonable to hypothesize that other factors (besides 'differences in processing levels') such as working memory load might have also been responsible for the current finding. Specifically, a lighter load placed on working memory by the shorter ISI may have facilitated the native Thai listeners' discrimination of multiple tokens of the low and the mid tones in Thai. Thus, the facilitative effects of the short ISI may be explained in terms of acoustic salience, 'phonetic' vs. 'phonological' levels of processing or working memory demand and other processing constraints.

In contrast to NT listeners, NC listeners benefited more from the longer ISI, at least during the pre-test. This result may suggest that, unlike NT listeners, NC listeners may have benefited more from a longer processing time afforded by the longer ISI, allowing them to perhaps compare the Thai tones to their native Mandarin tones. Their poorer performance under the shorter ISI suggests that they were not able to take advantage of the smaller load placed on working memory due to their lack of experience with f_0 variations across multiple tokens (i.e., produced by different talkers) of the low and mid tones in Thai. The fact that their performance improved and became comparable under the two ISI conditions after the training suggests that the training served to increase their sensitivity to the acoustic variances present in the multiple tokens of the two Thai tones. This increased phonetic sensitivity, in turn, allowed them to better discriminate the two Thai tones at a shorter time interval.

As for NE listeners, their lack of experience with tones at the phonetic level, as well as their lack of tonal representation at the phonological level, are likely the source of their poor performance under both ISI conditions during the pre-test. Their performance, while significantly improved, remained comparable under both ISIs after the training suggesting that the training increases their sensitivity to the phonetic characteristics of the two Thai tones and concurrently promotes the representation of the two Thai tones in their long-term memory.

5.3. Effects of L1 background

The finding that NC listeners performed significantly better than NE listeners both before and after training and under both ISI conditions is worth noting. This finding suggests that native experience with the tonal system of one tone language may subsequently facilitate the acquisition of another tone language. That is, the ability to perceive a change in average pitch or pitch contour that function to differentiate lexical meaning in one language, and the existence of tonal representations in long-term memory to which new tones can be compared and contrasted, may promote the learning of a different tone system. However, short-term laboratory auditory training, at least of the kind administered in this study, appears to increase sensitivity to the low- vs. mid-tone contrast in Thai to the same extent (but see the discussion on the effects of training procedures above) among both NC and NE listeners, who have no prior phonetic experience with tone and have no representation of tones in their long-term memory. Thus, while it may be advantageous for native speakers of a tone language to acquire another tone language, the ability of native speakers of non-tone languages to learn a tone language can be enhanced with experience afforded by a perceptual training procedure.

In conclusion, results obtained from this current study indicate that perception of lexical tone among non-native listeners can be improved with auditory training with either an ID or SD procedure and that native experience with a tone language positively affects the learning of a second tone language. However, due to a possible ceiling effect, a relatively more challenging perception task with many more tokens of the same tones produced by different talkers is suggested to further compare the extent to which native and non-native listeners benefit from the training. The generalizability of the effects of the training to novel stimuli should also be explored. Furthermore, different types of data including reaction times data should be collected to further assess relative degrees of the two training procedures.

Acknowledgments

We would like to thank all participants for their willing participation in the study. We also would like to express our gratitude to Professor Amanda Walley and two anonymous reviewers whose numerous constructive comments have significantly improved the quality of this manuscript. Many thanks also go to Marc Matthews and Russell Nekorchuk for their assistance in proof reading the manuscript.

Apendix A1

See Table 4 for the average f_0 of stimuli used in the study.

Table 4
Average f_0 of stimuli used in the study

Talker	Mid-tone	Average f_0 (Hz)	Low-tone	Average f_0 (Hz)
1	[puː] ‘a crab’	118	[pìː] ‘an oboe’	94
	[k^{h}aː] ‘to be stuck’	102	[bàː] ‘shoulder’	91
	[piː] ‘year’	110	[p^{h}àː] ‘split’	110
2			[k^{h}àː] ‘galangal root’	93
			[pùː] ‘grandfather’	116
			[p^{h}àː] ‘split’	117
3	[baː] ‘nightclub’	103	[pìː] ‘an oboe’	102
			[k^{h}àː] ‘galangal root’	86
			[pùː] ‘grandfather’	95
4	[baː] ‘nightclub’	113	[pìː] ‘an oboe’	107
	[k^{h}aː] ‘to be stuck’	116	[pùː] ‘grandfather’	107
	[p^{h}aː] ‘to accompany’	115		
5	[puː] ‘a crab’	130	[bàː] ‘shoulder’	107
	[p^{h}aː] ‘to accompany’	131		
	[piː] ‘year’	129		

References

Abramson, A. S. (1975). The tone of central Thai: Some perceptual experiments. In J. G. Harris, & J. R. Chamberlain (Eds.), *Studies in Tai linguistics in honor of William Gedney* (pp. 1–16). Bangkok: Central of English Language.

Abramson, A. S. (1976). *Thai tone as a reference system*. Haskins Laboratories Status Report on Speech Research 44, October–December (pp. 127–136).

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*, 319–325.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press, Inc.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tokhura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299–2310.

Burnham, D., Earnshaw, L. J., & Quinn, M. C. (1987). The development of the categorical identification of speech. In *Perceptual development in early infancy: Problem and issues* (pp. 237–275). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

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.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. (1992). Perception of central Thai tones and segments by Thai and Australian adults. In *Pan-Asiatic linguistics: Proceedings of the third international symposium on language and linguistics* (pp. 546–560). Bangkok: Chulalongkorn University Printing House.

Chao, Y. R. (1948). *Mandarin Primer*. Cambridge, MA: Harvard University Press.

Flege, J. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, *86*, 1684–1697.

Flege, J. (1995a). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 229–273). Baltimore: York Press, Inc.

Flege, J. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*, 425–442.

Flege, J., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Flege, J., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1161–1173.

Flege, J., & Wang, C. (1990). Native language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, *17*, 299–315.

Gårding, E., Kratochvil, P., Svan tesson, J. O., & Zhang, J. (1986). Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech*, *29*, 281–293.

Gandour, J. (1979). Perceptual dimensions of tone: Thai. In N. Liem (Ed.), *Southeast Asian linguistic studies, Vol. 3.: Pacific linguistics (Series C. No. 45)* (pp. 277–300). Canberra: Department of Linguistics, Australia National University.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 49–175.

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, *12*, 235–261.

Hamburger, M., & Slowiaczek, L. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin & Review*, *3*, 520–525.

Hudak, T. (1987). Thai. In B. Comrie (Ed.), *The world's major languages* (pp. 756–775). New York: Oxford Press.

Jamie, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /θ/-/ð/ contrast by francophones. *Perception & Psychophysics*, *40*, 205–215.

Jamie, D. G., & Morosan, D. E. (1989). Training new, non-native speech contrast: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology*, *43*, 88–96.

Johnson, J. S., & Newport, E. L. (1989). Critical periods in second language learning. The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Lenneberg, E. (1967). *Biological foundation of language*. New York: Wiley.

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. *Journal of the Acoustical Society of America*, *94*, 1242–1255.

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. *Journal of the Acoustical Society of America*, *96*, 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.

Massaro, D. W., Cohen, M. M., & Tseng, C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, *13*, 267–290.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, *34*, 323–330.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, *102*, 1864–1877.

Patkowski, M. (1989). Age and accent in second language: A reply to James Flege. *Applied Linguistics*, *11*, 73–89.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 297–314.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contribution. *Journal of the Acoustical Society of America*, *89*, 2961–2977.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, *52*, 37–52.

Pruitt, J., Strange, W., Polka, L., & Aquilar, M. (1990). Effects of category knowledge and syllable truncation during auditory training on American's discrimination of Hindi retroflex-dental contrast. *Journal of the Acoustical Society of America*, *87*, S72(A).

Sabol, M., & DeRosa, D. (1976). Semantic encoding of isolated words. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 58–68.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, *19*, 245–253.

Shen, X. S., & Lin, M. C. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, *34*, 145–156.

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.

Strange, W., & Dittman, S. (1984). Effects of discrimination training of the perception of /r/-/l/ by Japanese adults learning English. *Perception and Psychophysics*, *36*, 131–145.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tone. *Journal of the Acoustical Society of America*, *106*, 3649–3658.

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. *Applied Psycholinguistics*, *24*, 113–129.

Wayland, R., & Guion, S. (2004). Training native English and native Chinese speakers to perceive Thai tones. *Language Learning*, *54*(4), 681–712.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44.

Werker, J. F., & Tees, R. (1983). Developmental changes across children in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*, 278–286.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*, 1866–1878.

Yamada, R., & Tokhura, Y. (1992). Perception of American English /r/ and /l/ by native speakers of Japanese. In E. Tokhura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 155–174). Tokyo: Ohmsa.