Applied Psycholinguistics **24** (2003), 113–129
Printed in the United States of America
DOI: 10.1017.S0142716403000067

Perceptual discrimination of Thai tones by naive and experienced learners of Thai

RATREE WAYLAND
University of Florida at Gainesville

SUSAN GUION
University of Oregon

ADDRESS FOR CORRESPONDENCE
Dr. Ratree Wayland, Department of Linguistics, 4131E Turlington Hall, University of Florida, Gainesville, FL 32611-5454. E-mail: ratree@ufl.edu

ABSTRACT
This study investigated the ability to discriminate the middle and low tone contrasts in Thai by two groups of native English (NE) speakers and a control group of native Thai (NT) speakers. The first group was comprised of NE speakers who had no prior experience with Thai, whereas subjects in the second group were experienced learners of Thai (EE). The variables under investigation were experience with Thai, discrimination of open versus closed syllables, and the interstimulus interval (ISI) of the presentation (500 vs 1500 ms). The results obtained indicated that the NT group obtained higher discrimination scores than the NE or EE groups, the EE group obtained higher discrimination scores than the NE group, all three groups of subjects found open syllables to be more difficult to discriminate than closed syllables, and subjects in the EE group obtained higher discrimination scores for open syllables in the shorter than the longer ISI condition.

The majority of the world's languages use tone or pitch variations to convey semantic information at the lexical level (Goldsmith, 1994). Phonetic attributes of tones in various languages have also been described by linguists (e.g., Abramson, 1978; Gandour, 1979, 1983; Saravari & Imai, 1983). However, there have been relatively few studies on the acquisition of tones in the first language (L1) and second language (L2) acquisition literature. Studies in child language acquisition reported that tone-language infants use pitch to convey affect at around 8 months, which is earlier than their use of either pitch or segments to convey lexical contrast (e.g., Clumeck, 1980; Luksaneeyanawin, 1976). Moreover, the acquisition of tones appears to be relatively easier than that of segments and is completed at around the age of 23 months (e.g., Clumeck, 1980; Tuaycharoen, 1977, cited in Burnham & Francis, 1997). Fernald (1989, 1993) suggested that children may be more sensitive to pitch variation than adults.

Two types of tones, that is, dynamic (or contour) and static (or level) tones, have been identified by linguists (i.e., Abramson, 1978). In general, static tones are acquired earlier than dynamic tones. This is true for Mandarin (e.g., Li & Thomson, 1977), Cantonese (e.g., Tse, 1978), and Thai (Tuaycharoen, 1977, cited in Burnham & Francis, 1997). Furthermore, it has been suggested that, among dynamic tones, falling tones are easier to produce than rising tones. However, although this is true for children acquiring Mandarin (e.g., Li & Thompson), Tuaycharoen (1977) found that the rising tone was produced before the falling tone among Thai children.

With regard to the L2 literature, it has been shown that cross-language difference in suprasegmental features may have important repercussions for individuals who learn a L2. It has been shown, for example, that native speakers of a nontonal language (e.g., English) often have difficulty with the perception of tones in a tonal language like Chinese (e.g., Bluhme & Burr, 1971; Kiriloff, 1969; Wang & Spence, 1999). The perceptual difficulty experienced by these listeners may reflect the difference in the way tones are processed among native and nonnative listeners of a tonal language. Gandour and Harshman (1978), for example, found that listeners of two tonal languages (Thai and Yoruba) placed more emphasis on the linguistic tone dimensions, that is, the direction and slope of the fundamental frequency (F_0), whereas listeners of a nontonal language (American English) placed more emphasis on the nonlinguistic tone dimensions of average pitch and extreme endpoints. Similarly, Gandour (1983) found that, when compared to speakers of tonal languages, native speakers of English paid more attention to the average F_0 and less to the F_0 contour information in order to identify tones. Moreover, Lee and Nusbaum (1993) found that Mandarin Chinese listeners, but not English listeners, were slowed down by an irrelevant change in pitch level when making a segmental classification. This finding suggests that Mandarin listeners, but not English listeners, perceive pitch and segmental information in an integral manner. Moreover, in a dichotic perception experiment, Wang, Jongman, and Sereno (2001) found that Mandarin tones are predominantly processed in the left hemisphere by native Mandarin speakers, whereas they are processed bilaterally by American English speakers with no prior tone experience.

Previous research has also shown that the ability to perceive tones improves with experience. Auditory perception training, for example, has been shown to improve the perception of tones among nontonal language listeners. Wang and Spence (1999) showed that, after extensive auditory training, the ability to correctly identify Mandarin tones by native English speakers learning Mandarin increased from 69 to 90%. A corresponding neurophysiological change accompanying this improvement was recently observed (Wang et al., 2001). Thus, while native speakers of a nontonal language may experience difficulty in tone perception, the degree of difficulty may decline with experience.

The long-term goal of this study is to broaden our knowledge of the acquisition process of suprasegmental features in L2. As a specific goal, this study seeks to examine (a) the role of experience on the ability to discriminate tones and (b) the effect of different interstimulus intervals (ISIs) on tonal discrimination. To this end, two groups of adult native English speakers were tested on

Table 1. *Characteristics of Thai tones*

Tone	Tone Mark	Pitch Contour	Pitch Height	Voice Quality
Mid	Unmarked	Level	Medium	Nonglottalized
Low	ˋ	Level	Low	Nonglottalized
Falling	ˆ	Contour	High to low	Glottalized (creaky)
High	ˊ	Level	High	Glottalized
Rising	ˇ	Contour	Low to High	Nonglottalized

their ability to discriminate two tones of Thai. One group of subjects had prior experience with Thai and the other was naive. A group of native Thai speakers was also included as controls. The discrimination was performed under two different ISI conditions, 500 and 1500 ms.

THAI

Thai, or Siamese, is the national language of Thailand. It is spoken by approximately 60 million people. The dialect spoken in the capital city of Bangkok is considered the standard. However, outside Bangkok and the central plains, other languages of the Tai and Mon–Khmer families coexist with this standard dialect.

The phonemic inventory of standard Thai consists of 20 consonants, nine monophthongs, and three diphthongs. Each of the nine monophthongs may occur phonemically as long or short, for example short [a] and long [aː], as in [kʰàt] "to interrupt" and [kʰàːt] "to be torn." Phonetically, the long vowels are approximately twice as long as the short vowels (e.g., Hudak, 1990). All 21 vowel nuclei can occur with an initial consonant, a final consonant, or both initial and final consonants.

Each syllable in Thai carries a tone. There are five phonemic tones in Thai, namely a low tone ([kʰàː] *galanga*, a kind of aromatic root often used in Thai cooking), a middle tone ([kʰaː] "to be stuck or lodged in"), a high tone ([kʰáː] "to engage in trade"), a falling tone ([kʰâː] "I, servant"), and a rising tone ([kʰǎː] "leg"). These five tones may be characterized in terms of pitch contour, pitch height, and voice quality. The characteristics of the five tones are schematically presented in Table 1 (Hudak, 1990, p. 34). Based on their pitch contours, Abramson (1978) referred to level tones as static tones and contour tones as dynamic tones.

PREVIOUS RESEARCH

In an investigation of Thai and Australian English-speaking adults' tone and consonant perception, Burnham et al. (1992) found that Thai speakers were better able to discriminate tonal contrasts than the native Australian English speakers, who had had no prior experience with Thai. However, both groups of speakers found the dynamic contrast (rising vs. falling) to be the most easily discriminated. Burnham and Francis (1997) conducted another experiment investigating the ability to discriminate between pairs of Thai tones by native Thai and native English listeners. Four groups of subjects from each language background were included: adults, 8-year-olds, 6-year-olds, and 4-year-olds. All of the native English subjects had no prior experience with Thai. Following the work by Werker and Tees (1984), they also manipulated the ISI from 500 to 1500 ms in a same-different (AX) discrimination task. Werker and Tees (1984) had found that inexperienced subjects could not discriminate nonnative contrasts at the 1500-ms ISI but could discriminate them at the 500-ms ISI. They interpreted these finding to mean that at the 1500-ms ISI, subjects used native categories stored in long-term memory to perform the task. Because the nonnative contrast mapped onto a single native category, the subjects showed poor discrimination. In the 500-ms ISI condition, however, the subjects showed evidence of being sensitive to nonnative contrasts. The authors proposed that at the 1500-ms ISI, a "phonological mode" of processing was exhibited, whereas at a 500-ms ISI, a "phonetic mode" of processing took place. The phonetic mode was thought to be a language-general phonetic mode of perception in which phones are discriminated without any influence of linguistic experience. The phonological mode of perception was believed to be constrained or enhanced by experience with the phonological system of a particular language (e.g., Burham et al., 1992; Werker & Tees, 1983). Burnham and Francis (1997) predicted that, although native Thai speakers would perform better than native English speakers, the performance of both groups would vary according to the two ISIs. That is, the native English speakers would perform better with the ISI of 500 ms than with the ISI of 1500 ms because they would be using the phonetic mode, and the native Thai speakers would perform better with the ISI of 1500 ms than with that of 500 ms because they would be using the phonological mode. These predictions were borne out and were consistent with earlier findings (Burnham et al., 1992). There was, however, a significant interaction between type of tonal contrast and processing modes. The superior performance with the 500-ms ISI was evident only for the dynamic–dynamic (rising–falling) tone contrast. This was true for both the native English and native Thai speakers. Burnham and Francis (1997) suggested that the dynamic–dynamic contrast is acoustically more salient than either the static–static or static–dynamic contrasts and that this salient acoustic information was exploited in the phonetic processing mode, producing higher discrimination scores. These authors also found that reaction time for the 500-ms ISI was significantly shorter than that of the 1500 among 4-year-olds, a difference that was not found among older groups of children. The authors concluded that linguistic experience and/or age played a role in the shift from phonetic to phonological processing of tone. That is, tone perception becomes increasingly more phonologically based as one gains more experience with the ambient language. Irrespective of language background, tone perception may become more phonologically oriented over age.

It is possible to offer a different interpretation of the 500-ms ISI versus 1500-

ms ISI effect. Although it is clear that shorter ISIs offer inexperienced subjects an advantage in discriminating nonnative contrasts, it is not necessarily the case that a shorter ISI precludes access to stored phonological information for native speakers or experienced learners. Motivation for this proposal comes from results found in the priming literature. First, consider the effects of phonological priming. Presentation of a prime that shares some phonological similarity with a following target can affect response times on an experimental task such as lexical decision or shadowing (see, e.g., Hamburger & Slowiaczek, 1996, including the references). The ISIs used in such tasks are typically around 500 ms. Second, it has long been known from semantic priming studies that lexical access takes place in less that 200 ms (Sabol & DeRosa, 1976). Thus, a 500-ms ISI in a discrimination study would be plenty of time to access lexical representations, both phonological and semantic, in cases where the subjects possessed them. In other words, native speakers and perhaps learners of a language would be able to access stored representations of a word or phonological category within 500 ms. The facilitory effect found for nonnative speakers at a 500-ms ISI is probably due to a comparatively weaker demand placed on working memory. This effect is not found for native speakers because they are able to code the stimuli using categories from long-term memory. The finding that native speakers have better discrimination at 1500 ms for some contrasts could be due to greater decision time. Because native speakers are able to categorize the stimuli in terms of long-term memory representations, they are able to make a decision at a categorical level at both the 500- and 1500-ms ISIs. However, at the 1500-ms ISI they have a longer time to make the decision.

It is also important to note that an interaction has been found between the ISI and the type of stimuli under investigation. The categorical perception literature, for example, has revealed not only that vowels and consonants are perceived differently (e.g., Fry, Abramson, Eimas, & Liberman, 1962) but also that the ISI affects vowel and consonant perception differently. Pisoni (1973), for example, found that the delay interval (0–2 s) had relatively little effect for consonants but significantly impaired the within-category performance for vowels. On the basis of these and other related results, Studdert–Kennedy (1976) concluded that the difference between vowel and consonant perception lies in the nature of the acoustic cues that exist in these two types of sounds. That is, the transient nature of the acoustic cues for consonants induces a categorical, rather than a continuous, mode of perception. The results of Pisoni's (1973) study suggested that the categorical mode (as opposed to the continuous mode) of perception did not vary as a function of ISI. It remains to be seen, however, whether acoustic or auditory cues used in tone perception will induce a vowel or consonant-like perception and, more importantly, whether this will differ for native and nonnative speakers.

In conclusion, an ISI effect observed in the current study may be an effect of one or more of the following related factors: (a) short-term memory, (b) nature of stimulus cues, (c) categorical or continuous mode of perception, or (d) language-general (phonetic) versus language-specific (phonological) levels of processing.

THE PRESENT STUDY

The present study was conducted to compare and contrast the ability to discriminate a pair of Thai tone contrasts, namely the middle and low tones, by two groups of native American English speakers and one group of native Thai speakers. This study seeks to answer the questions of whether adult L2 learners of Thai whose native language lacks a tonal system would be able to discriminate the low and middle tones in Thai. The difference between these two tones has been found to be perceptually challenging even for native Thai speakers (Abramson, 1976; Burnham & Francis, 1997).

This study differed from that of Burnham and Francis (1997) in several respects. First, subjects in the Burnham and Francis study were native English speakers from Australia with no prior knowledge of Thai. Participants in this study included two groups of native American English speakers, one naive and the other experienced. Second, 10 tonal contrasts were examined by Burnham and Francis whereas only 1 contrast was examined in this study. Third, Burnham and Francis employed a same–different task with four possible pairings (AA, BB, AB, BA), whereas this study used a version of the ABX discrimination task. As will be described below, this task is designed to minimize response bias (guessing), which was not implemented by Burnham and Francis. This may be regarded as a significant improvement in the design of the present study. The stimuli in this study also differ from the Burnham and Francis (1997) study. They used productions of the syllable [ba:] produced with five different tones in Thai, resulting in three real words (middle, low, and falling tones) and two nonwords (high and rising tones). Stimuli in this study, on the other hand, are all real words and vary in their syllabic structure. The words used either ended in an open syllable (e.g., [a:] or [a:y]) or ended in a nasal stop (e.g., [a:n]). An acoustic analysis of the stimuli (presented in the method section) revealed that the middle and low tones were more distinct in the closed than the open syllables. Therefore, we predicted greater discrimination accuracy of the closed syllable stimuli than of the open syllable stimuli. Similar to Burnham and Francis (1997), this study employed two different ISIs (500 and 1500 ms). It is predicted that the shorter ISI will have a facilitory effect for the native English speakers because it puts less demand on working memory than the longer ISI. Because English does not have a tonal system, it is unlikely that naive English speakers will map the stimuli onto any type of tonal representation. The longer ISI might improve the discrimination of the native Thai subjects if it allows sufficiently more decision time. Keeping with our proposed view of the ISI effect, however, we expect the native Thai subjects to have sufficient time to access long-term stored phonological categories while performing the discrimination task at both ISIs. The interaction of ISI and contrast difficulty will also be discussed. Given the fact that the ABX task taxes working memory, we predicted that the facilitory effect of the shorter ISI for the nonnative speakers will be more pronounced in the more difficult contrasts. A comparison of differences and similarities in the performance of the inexperienced subjects versus the experienced subjects in this study may shed light on the role of experience in the acquisition suprasegmental features in general and the acquisition of tone in particular. Because

Table 2. *Minimal pairs used in the study*

Middle tone		Low tone	
1. [piː]	Year	[pìː]	Flute
2. [paː]	To throw	[pàː]	Forest
3. [kʰaː]	To be stuck, lodge in	[kʰàː]	Galanga
4. [tʰaːy]	To guess	[tʰàːy]	To change
5. [kʰaːy]	To spit out	[kʰàːy]	A net
6. [paːn]	Birthmark	[pàːn]	Jute
7. [pan]	To share	[pàn]	To pedal
8. [ʔaːn]	Saddle	[ʔàːn]	To read

subjects in this study are all adults who started learning Thai after puberty, the findings also allow a discussion on the role of age of learning in the acquisition of tone. The results of this study are also evaluated in light of earlier findings, especially those of Burnham and Francis (1997).

METHODOLOGY

Subjects

Sixteen native speakers of American English participated as experimental subjects and 8 native speakers of Thai participated as control subjects in the study. All native Thai speakers were from Bangkok and the native English speakers were originally from different regions in the United States (New York, 1; Washington, DC, 1; Maryland, 1; Michigan, 1; Ohio, 2; Illinois, 2; Wisconsin, 3; Idaho, 1; Oregon, 1; Washington, 1; Nebraska, 1; and California, 1). The native English speaker subjects were divided into two subgroups with eight subjects in each group, comprising the native and experienced groups. The native Thai (NT) subjects were recruited from the student population at the University of Florida at Gainesville and the native English speakers were mostly students from University of Oregon. The NT subjects were between the ages of 23 and 28 years (M = 24.5 years). The naive English (NE) group (7 females, 1 male) were between the ages of 21 and 47 years (M = 34 years), and the experienced English (EE) group were between the ages of 20 and 43 years (M = 30 years). Subjects in the NE group had no prior experience with Thai, whereas those in the EE group (3 females, 5 males) had been studying Thai (M = 2.5 years, range = 1–5 years) and had lived in Thailand (M = 4 years, range = 1–12 years). All subjects reported no prior history of speech or hearing impairment.

Stimuli

Stimuli were eight minimal pairs or contrasts (see Table 2) of low and middle tones of standard Thai produced by a 36-year-old female native speaker of Thai. Five out of eight contrasts (1–5) are open syllables and the remaining three

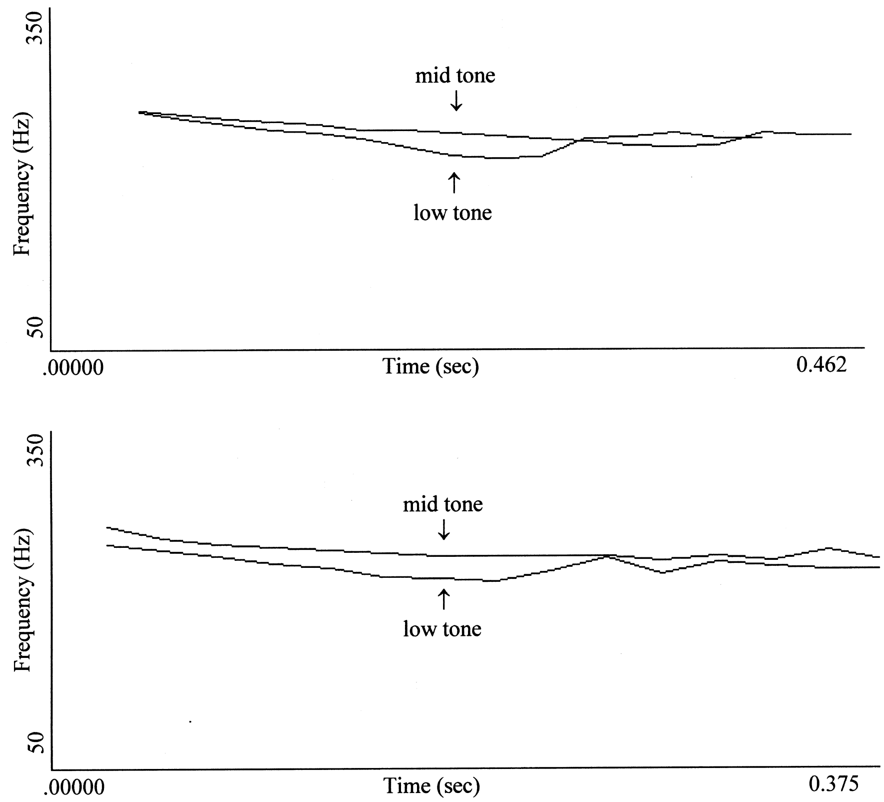

Figure 1. The fundamental frequency (F_0) traces of the standard Thai low and middle tones in open syllables /piː/–/pìː]/ (upper panel) and closed syllables /pan/–/pàn/ (lower panel).

contrasts (6–8) are closed syllables. These contrasts were produced in a Thai carrier phrase, "[rau phûːt k^{h}am wâː . . .]," 'we say the word' . . . Each contrast was produced three times in random order. The recording took place in a quiet office setting using a high-quality DAT cassette recorder (Sony TCDD8) and a head-mounted microphone (Shure, model SM 10A). The microphone was placed at a 45° angle approximately 13 mm from the mouth (as recommended by the manufacturer). The stimuli were later digitized using Cool Edit (Syntrillium Inc.) at 22.05 kHz, with a 16 bit quantization. Each target syllable was then excised out of the carrier phrase and saved as an individual file. All target syllables were normalized for peak intensity (50% of the scale). The fundamental frequency (F_0) traces of the middle and low tones for an open syllable pair and a closed syllable pair are shown in Figure 1. Acoustic measurements of F_0 mean, F_0 onset, F_0 offset, and the difference between F_0 onset and F_0 offset (F_0 onset – F_0 offset) are shown in Table 3. The measurements were made with the Pitchworks software using the autocorrelation method. The values shown

Table 3. *Acoustic measurements* (mean F_0, F_0 *onset*, F_0 *offset, and* F_0 *onset*–F_0 *offset*) *of stimuli used in the study*

Contrast	Word	Mean (Hz)	*SD*	F_0 Onset (Hz)	F_0 Offset (Hz)	F_0 Onset–Offset (Hz)
	[piː]	171	9	193	167	26
1.	[pìː]	161	13	188	142	46
	[paː]	167	7	185	170	15
2.	[pàː]	157	7	173	163	11
	[kʰaː]	172	11	201	165	37
3.	[kʰàː]	162	21	201	159	42
	[tʰaːy]	167	11	192	165	28
4.	[tʰàːy]	162	15	193	170	23
	[kʰaːy]	170	17	205	165	40
5.	[kʰàːy]	167	23	194	191	4
	[paːn]	165	6	180	162	18
6.	[pàːn]	153	8	169	147	21
	[pan]	170	8	191	166	25
7.	[pàn]	157	10	175	150	25
	[ʔaːn]	175	8	185	168	17
8.	[ʔàːn]	162	9	184	168	16

Note: The values are the mean of three tokens.

for each word are averaged across three tokens. These data were analyzed in a series of two-tailed paired *t* tests examining the difference between closed and open syllables. The results revealed that the middle and low tones in open syllables were differentiated based on the F_0 onset, t (15) = 2.14, $p < .02$, whereas both F_0 onset, t (8) = 2.31, $p < .004$, and F_0 offset, t (8) = 2.31, $p < .01$, differentiated the middle and low tones in closed syllables.

Procedure

The three productions of each word were used in constructing the test. The stimuli were presented in triads designed to test a single contrast. In any given triad, no two stimuli were exactly alike. Two instances of the same word were presented as two different productions.

Each of the eight contrasts were tested by six different trials, which consist of a single token of one word and two tokens of the other word with a different tone. For example, a trial testing the contrast [piː]–[pìː] might consist of [piː]-1, [pìː]-3, [piː]-2 (where the number indicates different productions). In the example given, the tone in the second stimulus is the odd item out because it contains a tone that differs from the first and third stimuli. The serial position of the odd item out was distributed equally over the three possible positions in the different trials.

Each tone contrast was also tested by four "catch" trials, which consisted of three physically different instances of a single tone. Two catch trials testing each contrast consisted of three instances of one member of the pair (e.g., [piː]-1, [piː]-2, [piː]-3), and the remaining two catch trials consisted of three instances of the other member of the pair (e.g., [pìː]-3, [pìː]-2, [pìː]-1).

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 was always set at 1500 ms.

The subjects were tested individually in a quiet room using a PC in one session that lasted about 30–45 min. The 160 (8 pairs × 6 different trials + 8 pairs × 4 catch trials × 2 ISIs) trials were randomly presented over headphones at a comfortable listening level. The subjects were told that each trial would be made up of three Thai words spoken by a female native Thai speaker and that they were to focus their attention on the tone or pitch level of each word. They were told to push a button marked 1, 2, or 3 if the tone in one word differed from the tone in the other two words, but to click the fourth button, marked none, if they heard three words produced with the same tone. All subjects were tested on both ISIs (500 and 1500 ms), and the order of presentation of the two tests was counterbalanced across subjects. To familiarize subjects with the stimuli and rate of presentations, a short practice session without feedback was provided. Moreover, the 80 experimental trials were preceded by five practice trials that were not analyzed.

DEPENDENT VARIABLE

The proportion of hits was determined for each contrast by determining how many times out of a maximum of six that the odd item out was correctly selected in the different trials. The proportion of false alarms was the number of times out of a maximum of four that an odd item out was incorrectly selected in a catch trial. An *A′* value was then calculated[1] for each of the eight contrast pairs for each subject to provide an estimate of phonetic sensitivity (see Snodgrass, Levy–Berger, & Haydon, 1985), taking into account the proportions of both hits and false alarms. An *A′* score of 0.5 would be obtained if the proportion of hits equaled that of false alarms. If the proportion of hits was greater than that of false alarms, an *A′* score greater than 0.5 would be obtained and vice versa when the proportion of hits was smaller than that of false alarms. Therefore, an *A′* score of 1 indicated perfect discrimination, whereas an *A′* score of 0.5 or less indicated a lack of phonetic sensitivity.

An inspection of the data suggested that the subjects understood and were able to perform the task. Except for two subjects in the NE groups, all subjects obtained a perfect score of 1.0 on at least one contrast. The highest scores for the two NE subjects were 0.94 and 0.96.

RESULTS

Effect of ISI

The *A′* scores were calculated for each subject in each ISI condition. The average *A′* scores for each group in each ISI condition are shown in Table 4. As

Table 4. *Mean A′ scores for all three groups for each ISI condition*

	Group			
ISI	NT	EE	NE	Mean
1500	0.91 (0.18)	0.85 (0.24)	0.77 (0.26)	0.84 (0.23)
500	0.92 (0.15)	0.87 (0.20)	0.79 (0.23)	0.86 (0.19)

Table 5. *Results of Group × ISI ANOVA*

Group	Tukey's (Group)	ISI
F (2, 89) = 6.96, $p < .001$	NT > NE, $p < .001$	F (1,189) = 3.62, $p < .059$

Table 6. *Mean A′ scores for closed and open syllables for all three groups of subjects*

	Group			
Type	NT	EE	NE	Mean
Closed	0.97 (0.06)	0.97 (0.05)	0.87 (0.15)	0.94 (0.26)
Open	0.88 (0.20)	0.79 (0.26)	0.72 (0.27)	0.80 (0.240)

expected, the native Thai subjects obtained higher A' scores, on the average, than both groups of native English speakers for both ISI conditions. The EE group also obtained higher scores than the NE group. These data were analyzed in a Group (3) × ISI (2) analysis of variance (ANOVA). This analysis yielded a significant main effect of group, but only a marginal significant effect of ISI (Table 5). There was no significant interaction between the two factors. A pair-wise comparison using the Tukey's method revealed that the NT speakers only obtained significantly higher A' scores than the NE speakers.

Effect of syllable type

Mean A' scores for closed and open syllables obtained for each group for both ISI conditions are shown in Table 6. As predicted, all three groups obtained higher A' scores for closed syllables than for open syllables. Moreover, native speakers of Thai obtained higher scores than the NE group on closed syllables and higher scores than both the EE and NE groups on open syllables.

These data were analyzed in a Group (3) × ISI (2) × Syllable Type (2) ANOVA. A significant main effect for both Group and Syllable Type was obtained (Table

Table 7. *Results of the Group × ISI × Syllable Type ANOVA*

Group	Tukey's (Group)	Syllable Type
F (2, 186) = 6.71, $p < .002$	NT, EE > NE, $p < .001$, .004	F (1,186) = 24.97, $p < .001$

Table 8. *Mean A′ scores for each group by each contrast*

Contrast	ISI (ms)	NT	EE	NE	Mean
1. [piː] [pìː]	1500	0.91 (0.07)	0.75 (0.25)	0.65 (0.25)	0.77 (0.19)
	500	0.87 (0.12)	0.86 (0.10)	0.62 (0.27)	0.78 (0.16)
2. [paː] [pàː]	1500	0.52 (0.23)	0.33 (0.16)	0.27 (0.09)	0.37 (0.16)
	500	0.60 (0.21)	0.45 (0.15)	0.40 (0.23)	0.48 (0.20)
3. [kʰaː] [kʰàː]	1500	0.97 (0.08)	0.90 (0.10)	0.86 (0.08)	0.91 (0.09)
	500	0.98 (0.03)	0.93 (0.07)	0.84 (0.11)	0.92 (0.11)
4. [tʰaːy] [tʰàːy]	1500	0.96 (0.06)	0.89 (0.13)	0.83 (0.16)	0.89 (0.12)
	500	0.95 (0.08)	0.83 (0.23)	0.80 (0.17)	0.86 (0.16)
5. [kʰaːy] [kʰàːy]	1500	0.99 (0.02)	0.99 (0.02)	0.96 (0.03)	0.98 (0.02)
	500	1.00 (0)	0.99 (0.03)	0.98 (0.03)	0.99 (0.02)
6. [paːn] [pàːn]	1500	0.96 (0.07)	0.96 (0.07)	0.87 (0.19)	0.93 (0.11)
	500	0.98 (0.03)	0.98 (0.03)	0.88 (0.06)	0.95 (0.04)
7. [pan] [pàn]	1500	0.99 (0.02)	0.99 (0.02)	0.80 (0.23)	0.93 (0.09)
	500	0.98 (0.03)	0.97 (0.05)	0.86 (0.11)	0.94 (0.06)
8. [ʔaːn] [ʔàːn]	1500	0.94 (0.08)	0.96 (0.08)	0.92 (0.08)	0.94 (0.08)
	500	0.97 (0.08)	0.99 (0.02)	0.92 (0.08)	0.96 (0.06)

Note: The standard deviations are in parentheses.

7). There was no significant interaction. A Tukey's pairwise comparison revealed that both the NT and EE groups obtained significantly higher A' scores than the NE group.

Effect of contrasts

Mean A' scores obtained by all three groups of subjects on each contrast in each ISI condition are shown in Table 8. These data showed that subjects in the NT group obtained relatively higher scores on all contrasts than the two native English speaker groups. The EE group also had higher scores than the NE group. Moreover, contrast 1 and, especially, contrast 2 obtained relatively lower scores than all other contrasts.

These data were analyzed in a Group (3) × Contrast (8) × ISI (2) ANOVA. A significant main effect of both the group and contrast factors was obtained (see Table 9). Tukey's pairwise comparisons for the group factor revealed that the NT group obtained significantly higher scores than the EE and NE groups. More interestingly, the Tukey's analysis also showed that the EE group obtained significantly higher scores than the NE group. Tukey's pairwise comparisons

Table 9. *Main effect of Group × Contrast × ISI*

Group	Tukey's (Group)	Contrast	Tukey's (Contrast)
F (2,169) = 23.54, $p > .001$	NT > EE, NE, $p < .02$, .001; EE > NE, $p < .001$	F (7, 168) = 64.269, $p < .001$	1 > 2 1, 2 <all other contrasts

Table 10. *Main and simple effects of Group × ISI on two difficult contrasts*

Main Effects					
Group			ISI		
F (2, 45) = 4.03, $p < .02$			F (1, 45) = 4.42, $p < .04$		
Simple Effects					
Group			ISI		
ISI 1500	ISI 500	Tukey's (Both ISI)	NT	EE	NE
F (2, 45) = 3.64, $p < .03$	F (2, 45) = 3.56, $p < .04$	NT > NE, $p < .03$	*ns*	F (1, 15) = 6.19, $p < .03$	*ns*

for contrast revealed that, whereas contrast 1 (*flute/year*) obtained a higher score than contrast 2 (*forest/throw*), both contrasts obtained significantly lower scores than all other contrasts (p values range from .03 to .001).

Effects of ISI for difficult contrasts

In this section, the prediction that the shorter ISI would have a more pronounced facilatory effect in the hardest (two) contrasts was tested. For this reason, a Group (3) × Contrast (2) × ISI (2) ANOVA was conducted (see Table 10). As expected, this analysis yielded a significant main effect of group as well as of ISI. The simple effect of group was found for both ISIs. Tukey's pairwise analysis revealed that for both ISIs, the NT group only obtained significantly higher scores than the NE group. More interestingly, the simple effect of ISI was found only for the EE group.

Summary of results

The results of the analyses presented here revealed that (a) when all contrasts were considered, there was no effect of ISI; (b) when data from only the most two difficult contrasts were considered, a significant effect of ISI was obtained.

This effect was significant, however, only for the EE group. For subjects in this group, higher scores were obtained for the shorter ISI. (c) All three groups of subjects were able to discriminate the tonal contrast in question better in the closed syllable condition than in the open syllable condition. Finally, (d) the NT group had higher discrimination scores than the NE and EE groups and (e) the EE group had higher scores than the NE group.

DISCUSSION AND CONCLUSION

This study was designed to explore the ability to discriminate between the low and the middle tones of Thai by two groups of native American English speakers, one naive and one experienced. Subjects in the naive group were those who have never been exposed to Thai, whereas those in the experienced groups had studied Thai and had lived in Thailand for varying amounts of time. The study sought to explore the question of whether experience with the target language resulted in improved ability to discriminate the tone contrast in question. The study also asked the question of whether the subjects' performance would differ as a function of ISI. The interaction between experience and ISI was also explored.

The results indicated that experience indeed did play a facilitative role in tonal discrimination. Subjects who were experienced with the target tones outperformed those who were not. This finding was expected and was consistent with results previously found in other studies on segmental features such as vowels and consonants (e.g., Best & Strange, 1992; Flege, Takagi, & Mann, 1996; Yamada & Tokhura, 1991). Moreover, the finding that there was no difference between the NT and the EE groups, at least in closed syllables, suggested that a nativelike discrimination can be achieved among adult L2 learners of a tone language.

The nonnative aspect of the EE subjects, however, emerged when the effect of ISI was explored. Similar to Burnham and Francis' (1997) finding on static versus static tone contrast, no effect of ISI was found for any of the three groups of subjects when all eight contrasts were considered. The ISI effect emerged only when the two most difficult contrasts were included in the analysis. This effect was found to be significant only for the experienced native English speakers. That is, higher scores were obtained for the shorter ISI (500 ms) for this group of subjects. As suggested earlier, the heavy demand on short-term memory posed by the oddity discrimination used in the study may have been responsible for this finding. In other words, information in short-term memory may dissipate with a longer ISI for the EE group. The fact that this limitation did not have an impact on the native Thai speakers' performance suggested that they were better able to hold in short-term memory the information necessary to differentiate the tones than the experienced native English speakers. This explanation implies that an interaction between the type of cues detected and the duration of their retention in short-term memory may exist. That is, given a difference in the kinds (and/or degree of salience) of acoustic cues, native and nonnative speakers may differ in their ability to code the stimuli in terms of tonal categories held in long-term memory. This may be due to more robust long-term memory representations in the NT listeners. This hypothesis, however, deserves further investigation.

As for the naive native English speakers, the absence of an ISI effect and their relative poor performance suggests that they suffered both from the inability to detect appropriate cues and the demand on short-term memory imposed by the discrimination task. The acoustic cues they extracted may not have been sufficiently salient to allow for a categorical processing at a shorter ISI, and a combination of a lack of tone representation in long-term memory and the decay of acoustic cues detected contributed to their failure at longer ISI.

Results of this study also revealed that all three groups of subjects obtained higher discrimination scores for closed syllables than for open syllables. Although this finding was expected based on acoustic measurements, it represented a new finding in cross-language tone perception. It should be emphasized, however, that this may not have been an effect of syllable type per se, but rather of acoustic salience. As shown earlier (see Methodology) whereas the middle and low tones in open syllables differed only in their F_0 onset, both F_0 onset and offset were significantly different in closed syllables.

In conclusion, results of this study suggest that experience plays a role in tonal discrimination and ISI affects the discrimination of the most difficult contrast only. These results should be further investigated using other types of discrimination tasks (i.e., AXB, AX), with other tone languages, and extended to other tone contrasts. Future research should also focus on the role of short-term memory, the role of L1 background, the degree of acoustic salience among tones, and the types of acoustic cues used in cross-language tone perception.

ACKNOWLEDGMENTS

The authors would like to thank Professor John Hartmann from Northern Illinois University at DeKalb, James Harnsberger for his assistance in finding some references cited in the study, Mohamed Al-Khairy for his technical assistance with the pitch traces shown in Figure 1, and J. J. Clark for her assistance in testing subjects. The authors also wish to thank two anonymous reviewers for their helpful comments.

NOTE

1. If the proportion of hits (H) equaled the proportion of false alarms (FA), then A' was set to 0.5. If H exceeded FA, then

$$A' = 0.5 + ((H - \text{FA}) * (1 + H - \text{FA}))/((4 * H) * (1 - \text{FA})).$$

However, if FA exceeded H, then

$$A' = 0.5 - ((\text{FA} - H) * (1 + \text{FA} - H))/((4 * \text{FA}) * (1 - H)).$$

REFERENCES

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

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

Best, C., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305–350.

Bluhme, H., & Burr, R. (1971). An audio–visual display of pitch for teaching Chinese tones. *Studies in Linguistics, 22,* 51–57.

Burnham, D., & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong* (pp. 29–46). 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.

Clumeck, H. (1980). The acquisition of tone. In G. Yeni–Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 257–275). New York: Academic Press.

Fernald, S. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60,* 1497–1510.

Fernald, S. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64,* 657–674.

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

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5,* 171–189.

Gandour, J. (1979). Perceptual dimension of tone: In G. I. Liem (Ed.), *Southeast Asian Linguistic Studies: Vol. 3. Pacific Linguistics (Series C, No. 45).* Canberra: Australia National University, Department of Linguistics.

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

Gandour, J., & Harshman, R. (1978). Cross-language difference in tone perception: A multidimensional scaling investigation. *Language and Speech, 21,* 1–33.

Goldsmith, M. (1994). Tone languages. In R. Asher & J. Simpson (Eds.), *The encyclopedia of language and linguistics* (Vol. 9, pp. 4626–4628). Oxford: Pergamon Press.

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

Hudak, T. (1990). Thai. In B. Comrie (Ed.), *Languages of east and south-east Asia.* Routledge: London.

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

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics, 53,* 157–165.

Li, C. & Thomson, S. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4,* 185–199.

Luksaneeyanawin, S. (1976). Some phonetic observations on the speech of Thai children. *Pasaa, 6,* 120–125.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253–260.

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

Saravari, C., & Imai, S. (1983). Perception of tone and short-long judgements of vowel variants of a Thai monosyllabic sound. *Journal of Phonetics, 11,* 231–242.

Snodgrass, J., Levy–Berger, G., & Haydon, M. (1985). *Human experimental psychology.* New York: Oxford University Press.

Studdert–Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 234–239). Springfield, IL: Thomas.

Tse, J. K. (1978). Tone acquisition in Cantonese: A longitudinal study. *Journal of Child Language, 5,* 191–204.

Tuaycharoen, P. (1977). *The phonetic and phonolgical development of a Thai baby: From early* *communicative interaction to speech*. Unpublished doctoral dissertation, University of London.

Wang, Y., Sereno, J., Jongman, A., & Hirsch, J. (2001). Cortical reorganization associated with the acquisition of Mandarin tones by American learners: An fMri study. In *Proceedings of the Sixth International Conference on Spoken Language Processing* (Vol. II, pp. 511–514). Beijing.

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

Werker, J., & Tees, R. (1983). Developmental changes across children in the perception of nonnative 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. (1991). Age affect on acquisition on non-native phonemes: Perception of English /r/ and /l/ by adult Japanese learners of English. In *Proceedings of the 12th International Congress of Phonetics Sciences.* Aix-en-Provence, France: Université de Provence, Service des Publications.