

PERCEPTUAL DISCRIMINATION OF THAI TONES BY NAÏVE AND EXPERIENCED LEARNERS OF THAI

Ratree Wayland

University of Florida

<ratree@ufl.edu>

Susan Guion

University of Oregon

<guion@uoregon.edu>

0 Introduction

The present study was conducted to investigate the ability to discriminate the mid and low tone contrast 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, while subjects in the second group were experienced learners of Thai (EE). The variables under investigation were experience with Thai, discrimination of open vs. closed syllables, and the inter-stimulus-interval (ISI) of the presentation (500 vs. 1500 ms).

1 Methodology

Subjects: Sixteen native speakers of American English participated as experimental subjects and eight 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 U.S. The native English speaker subjects were divided into two sub-groups with eight subjects in each group: the naïve and the 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 the University of Oregon. The NT subjects were between the age of 23 and 28 years (*mean* = 24.5 years). The naïve English (NE) group were between the age of 21 and 47 years old (*mean* = 34 years old), and the experienced English (EE) group were between the age of 20 and 43 years of age (*mean* = 30 years old). Subjects in the NE group had no prior experience with Thai while those in the EE group have been studying Thai (*mean* = 2.5 years, range 1 to 5 years) and have lived in Thailand (*mean* = 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 1 below) of low and mid tone 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 contrasts (6-8) are closed syllables.

Table 1: *Minimal pairs used in the study*

Mid Tone		Low Tone	
1. [pi:]	‘year’	[pì:]	‘oboe’
2. [pa:]	‘to throw’	[pà:]	‘forest’
3. [k ^h a:]	‘to be stuck’	[k ^h à:]	‘galanga’
4. [t ^h a:y]	‘to guess’	[t ^h à:y]	‘to change’
5. [k ^h a:y]	‘to spit out’	[k ^h à:y]	‘a net’
6. [pa:n]	‘birthmark’	[pà:n]	‘sack’
7. [pan]	‘to share’	[pàn]	‘to pedal’
8. [ʔa:n]	‘saddle’	[ʔà:n]	‘to read’

These contrasts were produced in a Thai carrier phrase “[rau phû:t k^ham wâ: ...]”, “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 TC-DD8) and a head-mounted microphone (Shure, model SM 10A). The microphone was placed at a 45-degree angle approximately 13 mm from the mouth. The stimuli were later digitized using Cool Edit (Syntrillium Inc.) at 22.05 kHz, with a 16-bit quantization. Each target syllable was then excised out of the carrier phrase and saved as an individual file. All target syllables were normalized for peak intensity (50% of the scale).

Fundamental frequency at the beginning (F_0 onset) and at the end (F_0 offset) of the vowel of all target words were obtained using Pitchworks. These data were analyzed in a series of two-tailed paired-tests examining the difference between closed and open syllables. The results revealed that the mid and the low tones in open syllables were differentiated based on the F_0 onset [$t(15) = 2.14, p < .02$], while both F_0 onset [$t(8) = 2.31, p < .004$], and F_0 offset [$t(8) = 2.31, p < .01$] differentiated the mid and the low tones in closed syllable.

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 was tested by six ‘different’ trials, which consisted 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 [pi:]/[pì:] might consist of [pi:]-1, [pì:]-3, [pi:]-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 the 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 (i.e. different productions) of a single tone. Two catch trials testing each contrast consisted of three instances of one member of the pair (e.g., [pi:]-1, [pi:]-2, [pi:]-3), and the remaining two catch trials consisted of three instances of the other member of the pair (e.g., [pì:]-3, [pì:]-2, [pì:]-1).

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 (ITI) was always set at 1500 ms.

The subjects were tested individually in a quiet room in one session that lasted about 30-45 minutes using a PC. The 160 (8 pairs x 6 different trials + 8 pairs x 4 catch trials x 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 the word. They were told to push a button marked “1”, “2”, or “3” if the tone in one word differed from the tone in the other two words, but to click the fourth button, marked ‘none’, if they heard three words produced with the same tone. For example, the button marked “1” will be selected if they think the first word they heard was the one produced with a different tone from the second and the third words. All subjects were tested on both ISIs (500 and 1500 ms) and the order of presentation of the two tests was counter-balanced across subjects. To familiarize subjects with the stimuli and rate of presentations, a short practice session without feedback was provided. Moreover, in each block, the 80 experimental trials were preceded by five practice trials that were not analyzed.

2 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 catch trial. An A' value was then calculated¹ 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 .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’, then an A' score greater than .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 while 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 .94 and .96.

3 Results

Effect of ISI: 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 2.

As 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.

Table 2: Mean A' scores for all three groups for each ISI condition.

ISI	Group			
	NT	EE	NE	Mean
1500	.91 (.18)	.85 (.24)	.77 (.26)	.84 (.23)
500	.92 (.15)	.87 (.20)	.79 (.23)	.86 (.19)

These data were analyzed in a Group (3) x ISI (2) analysis of variance (ANOVA). This analysis yielded a significant main effect of Group, but only a marginal significant effect for ISI (Table 3).

Table 3: Results of Group x ISI ANOVA

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

There was no significant interaction between the two factors. A pair-wise comparison using the Tukey's method revealed that the NT speakers obtained significantly higher A' score than the NE only.

Effect of Syllable Type: Mean A' scores for closed and open syllables obtained for each group for both ISI conditions are shown below in Table 4. 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) x ISI (2) x Syllable Type (2) ANOVA. A significant main effect for both Group and Syllable Type was obtained (Table 5).

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.

Table 4: Mean A' scores for closed and open syllables for all three groups of subjects.

Type	Group			
	NT	EE	NE	Mean
closed	.97 (.06)	.97 (.05)	.87 (.15)	.94 (.26)
open	.88 (.20)	.79 (.26)	.72 (.27)	.80 (.24)

Table 5: Results of the Group x ISI x syllable type ANOVA

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