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.

Ratree Wayland, John Hartmann & Paul Sidwell, eds. *SEALSXII: papers from the 12th meeting of the Southeast Asian Linguistics Society (2002)*. Canberra, Pacific Linguistics, 2007, pp.137-144.

Table 1: *Minimal pairs used in the study*

Mid Tone		Low Tone	
1. [piː]	'year'	[pi̖ː]	'oboe'
2. [paː]	'to throw'	[pa̖ː]	'forest'
3. [k^haː]	'to be stuck'	[k^ha̖ː]	'galanga'
4. [t^haːy]	'to guess'	[t^ha̖ːy]	'to change'
5. [k^haːy]	'to spit out'	[k^ha̖ːy]	'a net'
6. [paːn]	'birthmark'	[pa̖ːn]	'sack'
7. [pan]	'to share'	[pa̖n]	'to pedal'
8. [ʔaːn]	'saddle'	[ʔa̖ːn]	'to read'

These contrasts were produced in a Thai carrier phrase "[rau phû:t k^ham 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 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 [piː]/[pi̖ː] might consist of [piː]-1, [pi̖ː]-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 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., [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., [pi̖ː]-3, [pi̖ː]-2, [pi̖ː]-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 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 .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	
$F(1,186)= 24.97$, $p <.001$	

Table 6: *Mean A' scores for each group by each contrast. SDs are given in parentheses.*

Contrast	ISI	NT	EE	NE	MEAN
1.	1500	.91 (.07)	.75 (.25)	.65 (.25)	.77 (.19)
	500	.87 (.12)	.86 (.10)	.62 (.27)	.78 (.16)
2.	1500	.52 (.23)	.33 (.16)	.27 (.09)	.37 (.16)
	500	.60 (.21)	.45 (.15)	.40 (.23)	.48 (.20)
3.	1500	.97 (.08)	.90 (.10)	.86 (.08)	.91 (.09)
	500	.98 (03)	.93 (.07)	.84 (.11)	.92 (.11)
4.	1500	.96 (.06)	.89 (.13)	.83 (.16)	.89 (.12)
	500	.95 (.08)	0.83 (.23)	.80 (.17)	.86 (.16)
5.	1500	.99 (.02)	.99 (.02)	.96 (.03)	.98 (.02)
	500	1.00. (0)	0.99 (.03)	.98 (.03)	.99 (.02)
6.	1500	.96(.07)	.96 (.07)	.87 (.19)	.93 (.11)
	500	.98 (.03)	.98 (.03)	.88 (.06)	.95 (.04)
7.	1500	.99 (.02)	0.99 (.02)	.80 (.23)	.93 (.09)
	500	.98 (.03)	.97 (.05)	.86 (.11)	.94 (.06)
8.	1500	.94 (.08)	.96 (.08)	.92 (.08)	.94 (.08)
	500	.97 (.08)	.99 (.02)	.92 (.08)	.96 (.06)

Effect of Contrasts: Mean A' scores obtained by all three groups of subjects on each contrast in each ISI condition are shown in Table 6. 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) x Contrast (8) x ISI (2) ANOVA. A significant main effect of both the Group and Contrast factors were obtained (Table 7). Tukey's pairwise comparisons for the Group factor revealed that the NT group obtained a significantly higher score than both the EE and NE groups. More interestingly, the Tukey's analysis also showed that the EE group obtained a significantly higher score than the NE group.

Table 7: *Main effect of Group x contrast x ISI*

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

Tukey's pairwise comparisons for Contrast revealed that while 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 effect of ISI in the hardest (two) contrasts was tested. For this reason, a Group (3) x Contrast (2) x ISI (2) ANOVA was conducted. As expected, this analysis yielded a significant main effect of Group as well as of ISI (Table 8). The simple effect of Group was found for both ISIs. Tukey's pair-wise analysis revealed that for both ISIs, the NT group obtained significantly higher score than the NE group only. More interestingly, the simple effect of ISI was found only for the EE group.

Table 8: *Main and simple effects of Group x 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 1500	**ISI 500**	**Tukey's (both ISI)**
$F(2,45) = 3.64$, $p<.03$	$F(2,45) = 3.56$, $p<.04$	NT> NE, $p<.03$
ISI		
NT	**EE**	**NE**
ns	$F(1,15) = 6.19$, $p<.03$	*ns*

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.

4 Discussion and Conclusion

This study was designed to explore the ability to discriminate between the low and the mid tones of Thai by two groups of native American English speakers, one naïve and one experienced. Subjects in the naïve group were those who had never been exposed to Thai while those in the experienced groups had studied Thai and had lived in Thailand for varying amount of time. The study sought to explore the question of whether or not experience with the target language resulted in improved ability to discriminate the tone contrast in question. The study also asked the question of whether or not the subjects' performance differed as a function of ISI. The interaction between experience and ISI was also explored.

The results indicated that, indeed experience played a facilitative role in tonal discrimination. Subjects who were experienced with the target tones outperformed those who were not. This finding was expected and consistent with results previously found in other studies on segmental features such as vowels and consonants (e.g., Yamada & Tokhura, 1991; Best & Strange, 1992; Flege, Takagi & Mann, 1995). Moreover, the

finding that there was no difference between the NT and the EE, at least in closed syllables, suggested that a native-like discrimination can be achieved among adult L2 learners of a tone language.

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. We hypothesize that the heavy demand on short-term memory posed by the oddity discrimination task 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 non-native 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 naïve native English speakers, the absence of an ISI effect and their relative poor performance suggested 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 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. While 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 already mentioned (see Stimuli section) the mid and the 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 (a) experience plays a role in tonal discrimination, and (b) 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.

Notes

The authors would like to thank J.J. Clark from the University of Oregon for assistance in running subjects and Professor John Hartmann from Northern Illinois University at DeKalb, for his assistance in finding some references cited in the study.

[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-FA)*(1+H-FA))/((4*H)*(a-FA)). However, if FA exceeded H, then A' = 0.5-((FA-H)*(1+FA-H))/((4*FA)*(1-H)).

References

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

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.

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.

Snodgrass, J., Levy-Berger, G., and Haydon, M. (1985*). Human Experimental Psychology*. New York: Oxford University Press.

Yamada, R. & Tokhura, Y. (1991). Age affect on acquisition of non-native phonemes: perception of English /r/ and /l/by adult Japanese learners of English. In *Proceedings of the XIIth international congress of phonetics sciences, August 19-24, 1991*, Aix-en-provence, France.