TRAINING NATIVE SPEAKERS OF AMERICAN ENGLISH TO PERCEIVE THAI TONES USING HIGH STIMULUS VARIABILITY

By

DONRUETHAI LAPHASRADAKUL

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To my family, Khaodeedech and Laphasradakul

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Dr. Ratree Wayland, who has always been constantly supportive and encouraging, without her this study would not have been finished. My gratefulness also goes to the committee members, Dr. Caroline Wiltshire who always kindly helped me, ever since the days when I applied for the program, in 2004. Dr. Andrea Pham, Dr. Edit Kaan, and Dr. Jane Townsend have given me invaluable suggestions.

I would like to express my gratitude to the University of the Thai Chamber of Commerce under whose scholarship I studied at the Department of Linguistics, University of Florida, for five years. I would also like to express my appreciation to all of my friends. I always received help and support from them through the five-year pursuit of my study. My deepest gratitude goes to my family, Khaodeedech and Laphasradakul.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TRAINING NATIVE SPEAKERS OF AMERICAN ENGLISH TO PERCEIVE
THAI TONES USING HIGH STIMULUS VARIABILITY

By

Donruethai Laphasradakul

December 2010

Chair: Ratree Wayland
Major: Linguistics

The goal of this research was to investigate whether training naïve American English listeners to perceive Thai tones employing high-variability stimuli produced by multiple talkers in conversational speech would be more effective than training using low-variability, prototypical stimuli produced in citation forms. Due to time constraints, it was not realistic to conduct real time teaching sessions using the high stimulus variability in classrooms; therefore, the experiment was carried out in a laboratory setting.

Three groups (two experimental and one control) of native speakers of American English (NE) and one group of native Thai (NT) speakers participated in the experiment, which consisted of four phases: Pre-test, Training, Post-test and Retention test. There were 20 participants in each experimental group. In the Pre-test, all four groups of participants were asked to identify the low and mid tones in Thai. Immediately after the Pre-test, the two NE experimental groups underwent training. During training, participants in one experimental group, the Prototype group, were trained with prototypical stimuli produced in citation forms, while the other group, the High Variability group, was trained with high-variability stimuli produced by multiple speakers in a faster,

conversational speech style. Participants in the control group did not receive any training and only participated in the Pre-test and the Post-test, and NT speakers only participated in the Pre-test. The effectiveness of the two training methods was assessed immediately after training and again two weeks later. Besides accuracy rate and reaction time, the ability to make perceptual generalizations when presented with both old and new stimuli by new speakers was also evaluated.

As expected, NT speakers outperformed all three groups of NE speakers on the Pre-test, and there was no significant difference among the three non-native groups at that time. In addition, it was found that the training significantly improved tone identification accuracy among participants in both experimental groups, but there was no statistically significant difference in the two groups' amount of improvement. The effectiveness of the training was also retained two weeks after training. In addition, the training effects seemed to generalize to novel stimuli, both immediately after training and two weeks after training. However, a closer examination of the data suggested that the high-variability training paradigm may result in a more robust, longer-term representation of the two Thai tones than the low-variability training method. However, the high-variability effectiveness may not be immediate and takes longer to realize. Pedagogical values and implications of these findings are discussed. Shortcomings of the current study are noted and suggestions for further research are included.

CHAPTER 1
INTRODUCTION

In recent years, there has been a growing interest in teaching and learning Thai as either a second or foreign language. Thailand's economy is growing rapidly and because the country is considered a paradise for tourists, an increasing number of foreigners are coming to the country. These outsiders want to learn Thai to be able to communicate with Thai people and to become integrated into Thai society (Poomsan, 1995; Ponmanee, 2002a, 2002b). Nowadays, Thai is taught as a second language in Thailand and as a foreign language outside of the country. For instance, China, Korea, Japan, Vietnam, Malaysia, Indonesia, India, Australia, New Zealand, Canada, the United States, the United Kingdom, Sweden, and Germany, among others, all have programs for Thai as a foreign language (Department of Information, Ministry of Foreign Affairs, Bangkok, 2000). These programs play a major role in facilitating nonnative speakers' communication with Thai people. In order to accomplish this goal, it is crucial that Thai teachers in these programs implement optimal pedagogical methods that are effective in teaching the most difficult aspect of Thai which is Thai tones (Ponmanee, 2002a, 2002b; Wittayasakpan et al., 2002; Sittikesorn, 2005; Sathiansukon, 2005; Potibal, 2005; Udompan, 2005; Aroonmanakun, 2006). This research aimed to empirically examine if the common teaching method that has students perceive and produce Thai tone in citation forms such as tonal prototypes is effective in teaching Thai tones.

Previous studies (Abramson, 1979; Luksaneeyanawin, 1998; Arunreung, 1990) have shown that pitch contours of Thai tones vary greatly in conversational speech. Despite this fact, many Thai teachers still employ the traditional method in teaching Thai to non-native speakers. They usually present Thai tones in their prototypical, idealized citation forms and have L2 learners mimic his or her production. As a result, the learners become used to this idealized representation of Thai tones and use them in their production and perception. Producing and perceiving tones in fast, conversational speech where the idealized pitch contours are not retained, however, may prove difficult for them. This may be one of the reasons why perception and production of Thai tones remains challenging for non-native speakers even after years of trying to learn Thai as reported in a number of studies reviewed in the next chapter.

The traditional method of teaching Thai tones could be said to have been influenced by the Prototype theory and the abstractionist view of lexical representation. The Prototype theory maintains that humans categorize objects by matching them to the ‘prototype’ or the ‘ideal exemplar’ inside the category. This ideal exemplar is believed to contain the most representative features of the category (Kuhl, 1992). According to the abstractionist model, cognitive or mental categories such as words are represented in the lexicon in the form of normalized, abstract phonological representations (McClelland & Elman, 1986; Liberman & Mattingly, 1985; Norris, 1994). Speech input is abstracted from exact speech and compared to these prototypes or idealized representations by learners during processing. Consistent with this view, providing learners with only ideal exemplars of the five contrastive Thai tones in citation forms should facilitate their learning of these tones.

However, in recent years, new views on how humans categorize objects including mental objects known as the exemplar-based models have emerged in various domains of linguistic research including syntax, phonetics, phonology and language acquisition (Gahl & Yu, 2006). Central to the exemplar-based model is the idea that mental representations are comprised of memory traces of specific tokens. In other words, under the exemplar-based view, a category is comprised of a collection of memorized tokens, or ‘exemplars’. Each exemplar may, however, belong to many categories simultaneously. In an exemplar-based speech processing system, inputs are recognized and outputs are generated based on analogical evaluation across distinct memory traces of remembered tokens. Different exemplar-based models differ on how new tokens are assigned to relevant categories or integrated with pre-existing exemplars. All exemplar-based models assume, however, that the entire system of the category will shift slightly with each addition of a new experience. That is, as perceptual memories associated with a category accumulate and are incrementally updated, their distribution may shift.

Empirical evidence in support of the exemplar-based models came from different domains of linguistic research including studies in speech perception. Studies by Goldinger (1996) & Johnson (1997), for example, have shown that listeners relied on fine-grained phonetic details in word recognition suggesting that these phonetic details are coded in long term representation and are available during processing. Consistent with the exemplar-based models, speech categories comprised mainly of idealized exemplars should be different from the ones that consist of diverse and varied samples of the category, and that processing mechanisms and/or outputs from these two different category systems may vary. In this study, perceptual memories associated with two different category formation systems of Thai tones were compared and contrasted: input to one system comprised only of idealized exemplars of the category, while those of the other was phonetically diverse and varied. It was hypothesized that mental representations of Thai tones developed by the second type of input would be richer and thus more robust than those formed by the first system.

To test this hypothesis, a study designed to compare and contrast the effectiveness of two different training methods for lexical tone perception was conducted. In the first training method, the prototype method, participants were trained to perceive a contrast between the Thai mid and low tones using ideal exemplars of these two tones produced slowly and carefully in isolation by a female and a male native Thai speaker. On the other hand, participants administered the second training method were trained to discriminate between these two Thai tones using 'high-variability stimuli' such as stimuli produced in different phonetic contexts, and in a relatively faster, conversational style by multiple speakers, two female and two male native Thai speakers. The effectiveness of the two training methods was measured twice, once immediately after training and then again two weeks after training had ended. Effectiveness was measured by (1) perceptual accuracy rate (percentage of correct responses), (2) reaction times, and (3) ability to make generalizations when (a) old stimuli are produced by new speakers, (b) new stimuli produced by old speakers, and (c) new stimuli produced by new speakers. Participants were tested on two types of stimuli: prototypical such as low-variability stimuli produced in citation forms and high-variability stimuli produced in fast, conversational style by multi-talkers. Specific research questions are the following:

1. Which training method (prototype vs. high-variability) will result in a higher percentage of perceptual accuracy immediately after training and for which type of stimuli?

2. Which training method will produce faster reaction times in perception immediately after training and for which type of stimuli?

3. Which training method will lead to a greater ability to make generalizations immediately after training?

4. Which training method will be more effective (as measured by perceptual accuracy, reaction time, and ability to make generalizations) two weeks after training?

Based on the above research questions, it was hypothesized that:

1. Perception of the two Thai tones would improve after training among participants in both groups. However, since learning to categorize the two tones from highly-variable stimuli may take longer than learning from low-variable stimuli, a greater amount of improvement would be observed among participants in the Prototype training group than among those in the High Variability training group, particularly on low-variable, prototypical stimuli immediately after training. On the other hand, if high-variability training resulted in a more robust representation of the two tones, then participants in the High Variability training group should outperform those in the Prototype group on both types of stimuli two weeks after training.

2. Along the same line of reasoning provided in hypothesis 1 above, it was expected that reaction time would be shorter, particularly for low-variable stimuli among participants in the Prototype training group when tested immediately after training. However, two weeks after training, those in the High Variability training group were expected to exhibit shorter reaction times for all types (low and high-variable; new and old) of stimuli at two weeks after training.

3. A more robust representation of the two tones resulted from training with highly-variable stimuli which should also lead to better generalization ability. Thus, participants in the High Variability training group should be more accurate in their perception of the two Thai tones than those in the Prototype training group when tested with new stimuli two weeks after training.

4. Similarly, their reaction time to new stimuli should also become shorter when compared to those of participants in the Prototype training group.

The remainder of the dissertation is organized as follows. In Chapter 2, I give general background on the Thai language then present differences between Thai and English prosodic systems, issues of second language learning and cross-language speech learning, abstractionist vs. prototype theories of learning, and effects of training on SLA in general and on tone in particular. Chapter 3 presents the research methodology. Chapter 4 presents the results and statistical analysis on percent correct identification, and Chapter 5 presents the results and statistical analysis on reaction time. In the last chapter, Chapter 6, I focus on the implications of the results, the limitation of the present study and suggestions for future research.

CHAPTER 2
BACKGROUND

In this chapter, background information on the Thai tone system is first presented. This is followed by a brief overview of differences and similarities between Thai and English prosodic systems. Previous studies on Thai language acquisition by foreigners are also reviewed and discussed. Prototype and exemplar-based theories of memory and learning are briefly reviewed and three current speech perception models, namely Flege's Speech Perception Model (SLM), Best's Perceptual Assimilation Model (PAM) and Kuhl's Native Language Magnet effect (NLM) are discussed. Since the focus of the current study was on the effectiveness of laboratory training, results of previous training studies are also reviewed.

Thai Tone System

Thai is the national and official language of Thailand, a country in Southeast Asia. It is the mother tongue of the Thai people, the dominant ethnic group in Thailand. The Thai language is spoken in Thailand, Northern Malaysia, Cambodia, Southern Myanmar, and Laos. There are about 63 million total speakers (Pakasri, 2006). Thai is a member of the Tai group of the Tai-Kadai language family, the language family that is believed to have originated in what is now southern China. Four main dialects of Thai spoken in Thailand include Northern Thai, Central Thai, Northeastern Thai, and Southern Thai (Tingsapat, 1982). The Central Thai spoken in the capital city of Bangkok is considered the standard and official dialect of Thailand; it is used in the government school system.

The Thai phonological system consists of 21 consonants (see Table 2-1) and 21 vowel nuclei: nine monophthongs and three diphthongs (see Table 2-2). Phonemic contrastive length is found in Thai as each of the nine monophthongs may occur as long or short with different meanings, for example, short [a] and long [a:], as in [p̀ak] “to pierce” and [p̀a:k] “mouth”. Phonetically, long vowels are approximately twice as long as short vowels (Hudak, 1990). Thai is a tonal language, which means that the tone or pitch is used to determine the meaning of a word (Smith, 1995, 2005). In other words, besides consonants and vowels, differences in lexical tones or voice pitch are employed to distinguish the meaning of Thai words. Therefore, people learning Thai must be able to perceive and reproduce tones correctly in order to effectively communicate. Good pronunciation of consonants and vowels alone are not sufficient (Poomsan, 1995, Vongvipanond, 2000; Wittayasakpan et al., 2002; Wittayasakpan 2005). In his book, *‘Thai: A Complete Course for Beginners’* , David Smith not only informs the learners that Thai is a tonal language and that tones are very important in determining a Thai word’s meaning, but he also noted that “If tones make pronunciation in Thai seem more complex than in more familiar European languages, the learner will probably find Thai grammar considerably easier to absorb, for there are none of the complex verb tenses and noun endings which seem to dominate many people’s experience of language-learning” (Smith, 1995: 1).

In Thai there are five phonemic tones and four tone marks. The characteristics of Thai tones are presented in Table 2-3. Static tones are classified as having a level pitch, while dynamic, or moving tones, are identified as having a contour pitch (Satravaha, 2002). The fundamental frequency (F0) of the five tones in isolated monosyllabic words is shown in Figure 2-1 (adapted from Luksaneeyanawin, 1998: 377). It is these prototypes that are usually taught to Thai learners. Although the five-tone system in

Thai seems simple at first glance, their distributions are in fact rather complex. Not all syllables can bear all five ones which means that syllable structure plays a significant role in Thai tone assignment. According to Zsiga (2006), only CVV, CVS, and CVVS can bear all five tones. S stands for sonorant in Thai: /m, n, ŋ, w, j/ (sonorant means a sound that is produced without turbulent airflow in the vocal tract such as nasals and approximants). However, on CVO syllables, only high and low tones can occur, and on CVVO syllables, only falling and low tones are allowed. O stands for obstruent in Thai including /p, t, k/ (obstruent means a sound that is produced by obstructing airflow, causing increased air pressure in the vocal tract such as stops and fricatives). Zsiga also claimed that the tone-bearing unit (TBU) in Thai is the mora, not the syllable, and static tones are aligned to the rightmost mora, while contour tones such as falling and rising have one tone per mora.

Concerning variation of Thai tones, it has been observed that the phonetic realization of Thai tones varies due to several factors. The tonal context (tonal co-articulation between neighboring syllables) (Abramson, 1979) and sentence intonation (Luksaneeyanawin, 1998) are important factors influencing the pitch/fundamental frequency patterns of Thai tones. Additionally, Arunreung (1990) found that in Bangkok Thai, there was variation between generations of speakers which suggests that the language is undergoing change and that age is also a factor. Pitch levels and/or pitch contour of Thai tones have changed somewhat over time (Bradley 1911, Henderson 1949, 1982, Abramson 1962, and Arunrueng 1990) as shown in Table 2-5.

In this table, the numbers 1 through 5 represent the pitch level of the tones starting from 1 (low) to 5 (high). As shown in the table, Thai tones have changed gradually over time, particularly the contour tones. For example, the falling tone has changed from 31 (mid to low) in 1911 and 1928 to 343 (mid - high - mid) in 1990. Similarly, the pitch contour of the rising tone has changed from a linear rising tone (35 or 14) before the early 1900's to a falling-rising contour in the late 1990's. Pitch contours of static or level tones (mid, low and high) seem to have been alternately changed from a falling to a level pitch contour over time. However, it is uncertain to what extent the changes observed was due to differences in researchers' auditory impressions of the tones and thus their notations, particularly among earlier studies where supporting acoustic measurements were not available.

It is important to note that fluent Thai speech is usually fast and not every word is clearly produced. As shown below, Thai tones are not actually pronounced in complete forms but rather are generally spoken in reduced forms. While native Thai listeners are able to perceive the tones by relying on the context, L2 learners of Thai have difficulty perceiving these reduced tones. According to my preliminary study, tones produced in connected speech are generally shortened and do not show the complete forms of pitch contours. For example, the rising tone does not rise in natural speech as the prototype does and is therefore, similar to the mid and the low tones. Because of this variation, rising tone may be incorrectly perceived by L2 learners as either mid or low tone.

Figure 2-2 and Figure 2-3 show the pitch tracks of the mid tone and the low tone produced in conversational style by three speakers. As seen in the figures, the pitch levels of these tones vary from speaker to speaker. More importantly, there's a considerable amount of overlap between the pitch heights of the two tones such that a low tone produced by a high-pitched speaker is similar to a mid tone produced by a lowpitched speaker. Moreover, preceding or following tones affect the target tone. For example, as seen in Figure 2-3, the F0 of the low tone has a higher ending point than the low tone in citation form because the following tone is the high tone. According to previous studies, the mid and low tones are the most difficult pair to discriminate, even for native speakers of Thai, especially when they are perceived in isolation (Wayland & Guion 2004; Sathiansukon, 2005; Sittikesorn, 2005).

Differences between Thai and English Prosodic Systems

Prosody can be considered as the suprasegmental features which involve patterns in tone, tempo, and loudness that constitute the melody or rhythm or tune of a spoken language (Warren, 2000). The main difference between Thai and English lies in how voice pitch or an auditory impression of rate of vocal fold vibrations functions linguistically in these two languages. Thai is a tone language, or a language in which a difference in voice pitch is used to differentiate word meaning. It is important, therefore, to pronounce the tone associated with each word accurately in order to distinguish meanings (Kruatrachue, 1960). It has been estimated that as many as 60-70% of the world's languages are tonal (Goldsmith, 1994; Yip, 2002). Some examples of tone languages are Mandarin Chinese (885,000,000 speakers), Thai (60,000,000 speakers), Yoruba (20,000,000 speakers), and Swedish (9,000,000 speakers). In non-tonal languages such as English, on the other hand, variation in pitch level or pitch contour is used to express syntactical information (intonation) (Bolinger, 1949, 1962, 1978; Henderson, 1965; Pike, 1975).

Besides tones, other prosodic features or 'accentual system' in Thai including, accent, stress, intonation and rhythm have been examined by linguists such as Rudaravanija (1965), Hiranburana (1971), Luangthongkum (1978), Luksaneeyanawin

(1983), Potisuk et al. (1996). The effects of stress and intonation on Thai tones have also been investigated by Satravaha (2002). According to Luksaneeyanawin (1983), 'accent' refers to the potentiality of the syllable(s) in a word to be realized with 'stress'. 'Stress', on the other hand, refers to a subjective complex of some objective phonetic features such as a higher degree of respiratory effort, length, pitch, loudness etc. as compared with the unstressed syllable. However, the terms 'accent' and 'stress' are used interchangeably in studies on Thai accentual system.

Some Thai linguists (Luksaneeyanawin, 1983; Potisuk et al., 1990; Thubthong et al. 2000) have reported that Thai is a fixed accent language in which the accent or the stress is always assigned to the last syllable of the word. For polysyllabic words, the syllable-final position is also the position of primary stress. Concerning rhythmic categories, Thai has been considered a stress-timed language (Luangthongkum, 1977). In the original formulation, a stress-timed language refers to a language that each syllable has approximately equal duration whereas a syllable-timed language refers to a language that each foot has equal duration (Dauer, 1983). However, according to Grabe & Low (2002), although Thai's rhythmic characteristics were considered to have both features, Thai should pattern with the stress-timed languages because Thai has high Pairwise Variability Index (PVI) values. According to Low, Grabe, & Nolan (2001), the PVI is a phonetic measure of variability between vocalic and intervocalic speech durations and they predicted that stress-timed languages should have high PVI values.

In an analysis of intonation in Thai, Luksaneeyanawin (1983: 182) proposed that "the pitch contour of any syntactic unit is the realization of the integrated system of stress and intonation". She also investigated 'stylized intonation' which is related to phonetic characteristics such as pitch leveling and lengthening of duration. The common phonetic characteristics in stylized forms of speech such as chanting, recitation, calling, etc. are the following:

1. The pitch configuration of the tones is levelized: the static tones become level while the dynamic tones become a stepping narrow fall or rise.
2. The pitch range is narrower than normal intonation
3. In some extreme cases, such as in chanting and professional calling where the meaning of the words is unimportant, the dynamic tones are realized with level pitch
4. The extension of the syllable duration is very prominent
5. The degree of loudness is higher than normal speech, especially in callings

Luksaneeyanawin (1998), who investigated Thai lexical tone and sentential intonation, proposed that, to distinguish the meaning of the word, the system of tone in Thai is more important than the system of intonation since the latter just adds attitudinal and emotional meanings to the word. Unlike in Thai, pitch in English is used to distinguish 'lexical stress'. In other words, non-tone languages such as English do not depend on variation in pitch in order to distinguish the meaning of one word from another. In English, variation in pitch height at the word level is one of the cues used to make one syllable in a multi-syllabic word more prominent than others. At the syntactic level, differences in pitch are employed to signify a statement or a question. English has two types of stress: (1) lexical stress which refers to an accentuation of syllables within words, and (2) sentential stress, which refers to an accentuation of words within sentences (Cutler, 2005). Stress in English facilitates word recognition. Lexical stress is used to distinguish the word class of stress minimal-pair words such as pro’duce (verb) vs. ’produce (noun) (Wiltshire & Moon, 2003).

It is worth noting that there are varieties between English used in different countries such as British English, American English, Australian English, and Indian English. For instance, in an investigation of the phonetic properties of stress in Indian English (IE), it was found that the primary cue to lexical stress is increased duration. Moreover, unlike American English (AE), IE stress rules are conditioned by quantity or the number and size of the syllables in a word; for example, the main IE stress rule states that stress is placed on the rightmost heavy syllable, if there is one; if not, stress is placed on the first syllable of the word (a heavy syllable ends with a tense vowel or with a lax vowel followed by at least two consonants). The IE stress rules are not related to morphological category such as nouns or verbs like the rules in AE (Vijayakrishnan 1978 cited in Wiltshire & Moon, 2003). Also, in IE amplitude and durational differences between stressed and unstressed syllable are much smaller than those in AE. For example, on stressed syllable, IE speakers produce only a .15 dB average increase in volume, while the AE speakers produce a 3.27 dB increase on average. These differences may cause miscommunications between the two groups or slower processing for the AE speakers (Wiltshire & Moon, 2003).

Due in part to a difference in the linguistic function of pitch in tone and non-tone languages, native speakers of non-tonal languages often have difficulty acquiring languages with tonal systems such as Mandarin Chinese or Thai (Bluhme & Burr, 1971; Kiriloff, 1969; Wang et al., 1999; Potibul, 2005). It has been suggested that perceiving tonal distinctions is one of the most difficult challenges American English speakers face when they learn a tone language. (Wang 1999, 2001; Wayland & Guion 2003, 2004). Because AE speakers lack prior experiences using pitch differences to distinguish the meaning of words, they find it difficult to notice the pitch differences when they learn Thai tones.

Previous Studies on Thai Tone Acquisition

Even after non-native speakers of Thai have studied the language for a long period of time and are able to communicate in Thai, they may still have problems perceiving and pronouncing the tones (Ponmanee, 2002a, 2002b; Wittayasakpan, 2002; Sittikesorn, 2005; Sathiansukon, 2005; Potibal, 2005; Udompan, 2005; Aroonmanakun, 2006). Some studies claim that the difficulty may be due to the influence of the learner's native language (L1). Native knowledge of one prosodic system such as lexical tone or lexical stress may affect the acquisition of other systems. Wayland & Guion (2006) argue that adult learners can learn some aspects of a new prosodic system differing from their native system although they may acquire it imperfectly. According to Best (1995) and Flege (1995), similarities and differences between the L1 and L2 phonological systems can lead to perception and production difficulties among adult L2 learners.

Wayland et al. (1997, 2003, 2004, 2007) investigated the effects of native language on lexical tone perception. They mention that although naïve listeners demonstrate the ability to discriminate lexical tones, their performance is conditioned by their native language background. Wayland (1997) examined the contribution of both segmental and supra-segmental parameters to the perception of foreign accents in order to investigate whether the production of Thai consonants, vowels, and tones by native Thai speakers and native English speakers differ in terms of acoustic measurements and auditory evaluation by native Thai-speaking listeners. The results showed that adult native English listeners were almost successful in learning temporal aspects, such as voice-onset time and vowel duration, of certain Thai consonants and vowels. However, vowel quality and tones in Thai, that is, formant frequencies and fundamental frequency, still proved difficult for the native English speakers to master. Moreover, the production of Thai tones by adult native English learners was easily distinguished from native production by native Thai listeners, while some non-native production of tones was considered as ‘native-like’. The results also demonstrated that each Thai tone poses different degrees of difficulty for non-native speakers. Level tones seem to be more difficult than contour tones for native English learners to master; the rating scores for level tones were lower.

Moreover, it was demonstrated that first language background indeed affects the perception of Thai tones: native Chinese listeners had a greater ability of detecting and discriminating Thai tones (phonetic variations of the mid and low tones) than native English speakers both before and after training (Wayland & Guion, 2004). Based on the scores of the Pre-test and Post-test, the native Chinese listeners showed greater improvement than the native English listeners in both the shorter and longer inter-stimulus interval (ISI) conditions (500ms vs. 1500ms.). Thus, these results suggest that speakers of tonal languages such as Chinese have an advantage in L2 tonal perception because they can transfer their prior experience from their L1 tone system to another tone language like Thai, whereas speakers of non-tonal languages such as English are unable to transfer their experience with pitch successfully. In accordance with my preliminary study, many Thai teachers who teach Thai to speakers of both tonal languages like Chinese and non-tonal languages like American English have anecdotally found the above observation to be true.

A number of studies also support the findings that non-tone speakers have problems in learning another tone language. In previous studies of cross-language differences in supra-segmental features, the results suggested that the native speakers of a non-tone language such as English have problems perceiving or acquiring tones in tone languages such as Thai and Chinese (Burnham et al., 1992; Burnham & Francis, 1997; Wang, 1999, 2001). For instance, in a comparison between native Thai speakers and native Australian English speakers with no knowledge of Thai, the Thai speakers do better in the discrimination of tone contrasts (Burnham et al., 1992).

Further, it has been reported that even after two or three years of studying Thai at Chiang Mai University, Thailand, Korean students self-reported that the most difficult aspect in learning Thai was tone pronunciation (Ponmanee, 2002a). Potibal (2005) examined other Korean students who have been studying Thai for four years at Hankuk University of Foreign Studies, Seoul, Rebublic of Korean and similarly found that they had great difficulty pronouncing Thai tones. He noted that their errors stemmed from their inability to correctly produce pitch height for level tones and pitch contour for contour tones. Their production of contour tones was too short for the pitch contour to be fully and accurately realized. Korean students showed the influence of their mother tongue as they applied the Korean stress and intonation system to Thai. For instance, they pronounced words at the beginning of sentences incorrectly by changing a mid tone to a high tone because they usually placed stress there and this was consistent with Korean stress pattern (Potibal, 2005). In addition, conforming to Korean intonation

(Potibal, 2005), Koreans tend to produce high intonation at the middle of sentences in their L1 and they often do so when they speak Thai as well.

Udompan (2005) pointed out that Indonesian speakers also frequently have problems with tonal pronunciation due to the influence of their mother tongue. The significant differences between Thai and Bahasa Indonesian reside in the length of vowels and in tones. Neither of these features is phonemic in Bahasa Indonesia, but they have phonemic status in Thai. Due to the influence from their native language, Indonesian students do not pay enough attention to tones. Interestingly, it has been suggested that different tones pose different levels of difficulty for Indonesian students; they were able to accurately perceive and identify the contour tones better than the level tones. This is consistent with Wayland's (1997) report that contour tones may be easier than level tones for non-native listeners to perceive and perhaps produce. This is because a relatively more drastic change in fundamental frequency from high to low or from low to high found in contour tones may be more perceptually salient than a gradual linear change of F0 exhibited by level tones (Wayland, 1997). Udompan (2005) confirmed that changing the level of pitch facilitates Indonesian students' perception of tones. Danish students also have difficulty acquiring Thai tones. This was made evident by a study designed to investigate self- reported perception (using open-ended questionnaires and interviews) of some students' ability to produce Thai tones. The results confirm that, although these Danish students finished an Intensive 60-hour Thai Language Course, they were still highly concerned about their ability to perceive Thai tones correctly (Sathiansukon, 2005).

Effects of Experience in Speech Perception and Production Learning

Differences and similarities between the L1 and L2 sound systems, and experiences with the target language are important factors affecting L2 speech perception and production learning. Both phonemic and phonetic differences and similarities are considered the main sources in cross-language speech misperception and misproduction. While L1 and L2 phonemic system comparison known as 'Contrastive Analysis' was popular in the 1960's, later, a comparison at the phonetic level is believed to better predict or account for cross-language speech learning outcome. Current speech perception models such as the Speech Learning Model (SLM) (Flege, 1995), the Perceptual Assimilation Model (PAM) (Best, 1995), and the Native Language Magnet (NLM) (Iverson & Kuhl, 1995) take differences and similarities at both the phonemic and particularly at the phonetic level as main factors in the models.

The Speech Learning Model (SLM)

The purpose of the SLM is to account for success or failure in producing and perceiving phonetic segments in second language acquisition. The SLM focuses on difficulty of learning certain L2 speech sounds, age-related limits in L2 sound learning, and speech perception in terms of phonetic dissimilarity between L1 and L2. The core assumption of the SLM is that the ability to learn new speech sounds remains malleable across one's life span. The SLM proposes that the greater the dissimilarity between an L2 speech sound and a closest L1 sound is, the more likely the learners will notice the differences between the L1 and L2 sound. Moreover, the SLM claims that learning L2 sounds at an early age is more effective than learning as adult learners. This claim has been supported by a number of studies (Yamada, 1993, Flege & Munro et. al., 1995).

The Perceptual Assimilation Model (PAM)

In the domain of cross-language speech perception, the Perceptual Assimilation Model (PAM) has been one of the most influential models. Best and her colleagues proposed and developed the model to account for the process of speech perception of non-native sounds (Best, McRobert & Sithole, 1988; Best, 1995). The PAM viewed that different degrees of discriminability of pairs of non-native (L2) sounds is related to the perceived distance between articulatory gestures of a pair of L2 sounds and the L2 sound from L1 sound. Specifically, learners perceive L2 sound contrasts based on the degree of similarity as well as the discrepancy between the native and non-native sounds. According to PAM, there are three main perceptual assimilation types. Firstly, a two-category (TC) assimilation type where members of an L2 contrast assimilate to two different L1 sound categories and discrimination of this contrast is predicted to be relatively easy to discriminate. Secondly, a Category-goodness (CG) assimilation type: a pair of L2 sounds assimilate to the same L1 category, but is judged to differ in degrees of goodness of fit to the L1 sound. Discrimination is predicted to be relatively more difficult than the TC assimilation pattern. Thirdly, a single-category (SC) type: two L2 sounds assimilate to the same L1 category and both are judged to be equally deviant or similar to the L1 sound. Discrimination is predicted to be the most difficult. Therefore, according to their difficulty in discrimination, these three perceptual assimilation types can be ranked from easy to hard as follows: (TG)>(CG)>(SC).

Moreover, the PAM also formulated three other types of perceptual assimilation. The first one is the Uncategorized-Categorized (UC) type: one L2 sound is classified as one of the L1 sounds, but the other is not. It is predicted that discrimination for the (UC) is predicted to be high or good. The second one is the Uncategorized- Uncategorized

(UU) type: both L2 sounds cannot be classified into any L1 category and discrimination is predicted to range from fair to good. Finally, the Non-Assimilable (NA) type: both L2 sounds cannot be classified into any L1 category and are perceived as non-speech sounds. Discrimination is predicted to range from good to excellent depending on phonetic difference between the two foreign sounds. According to Best's Perceptual Assimilation Model (PAM) (Best, 1995), it is possible that American English speakers may perceive two lexical tones in Thai as the (UU) or the (NA) type. Discriminability of a pair of tone will thus depend on the phonetic distance between them.

The Native Language Magnet Effects (NLM)

This speech perception model focuses on infants' early stage of speech perception and the model was also proposed to account for adult listeners' speech perception development so that it is particular to their linguistic experience or their native language (Kuhl, 1991). The Native Language Magnet Effects (NLM) proposes that the native category goodness of speech sound is related to speech perception. According to NLM, an exceptional exemplar of phonetic categories, the prototype (P), was more difficult to be differentiated from its variants than a poor exemplar or the non-prototype (NP) because the prototype in a category assimilates the non-prototype members of the same category and shrinks the perceptual boundaries around it. This causes discrimination between prototype and non-prototype to be more difficult. This effect was referred to as the perceptual magnet effect (Kuhl, 1991, 1993; Kuhl et al., 1992, Iverson & Kuhl, 1996). However, the NLM does not offer specific prediction for learners who acquire a speech sound that does not fall into any category in their native language such as American English speakers who learn tones.

Exemplar-Based Models and Prototype-Based Models

When teaching Thai tones, most Thai teachers have long employed a traditional method consistent with the notion of an 'abstract' or 'prototype-based' learning model. According to the 'prototype-based' model, mental representation of a category is the prototype. Although different researchers defined a category's prototype differently (Barsalou, 1990; Nosofsky, 1987, 1992b; Posner & Keele, 1968; Reed 1972), the agreeable definition of 'prototype' is the exemplar with average values on all of the dimensions along a variety of the category's exemplars (Reed 1972; Rosch, Simpson, & Miller, 1976). A new stimulus is compared to the category's prototype and is accepted as a member of the category if it is sufficiently similar to the prototype. The prototypes of the five contrastive Thai tones in citation forms (as shown in Figure 2-1) are usually taught by Thai teachers to L2 learners. Unfortunately, as discussed earlier, after learning the Thai tones using this method, a substantial number of L2 learners of Thai still have difficulty in producing the tones correctly.

In contrast to the 'prototype' model, the exemplar theory holds that humans store exemplars they experience as whole, independent memory traces. Each category thus has no independent representation and instead is represented as collections of exemplars (Estes, 1986; Medin & Florian, 1992; Nosofsky, 1986; Reed, 1972). A new stimulus is compared to the category's exemplars and accepted as a member if it is similar enough to the exemplars (Estes, 1986; Nosofsky, 1986). As a result, all exemplars do not always share the same properties. Based on the different degree of similarity to the category's exemplars, some exemplars may be better or worse than the others. Therefore, instead of using unrealistic prototypes to teach Thai tones, using high-variability stimuli to teach tones would incorporate learning the tones in a variety of contexts. If nonnative learners of Thai learn to perceive high-variability stimuli, their mental representations of Thai tones would be richer and more robust, enabling them to perceive Thai tones correctly as they are realized in natural, native speech.

It has been demonstrated that high-variability stimuli positively affect speech perception learning (Mullennix & Pisoni, 1990; Jusczyk, Pisoni & Mullennix, 1992). For example, after learners were trained with high-variability stimuli produced by different speakers and at different speaking rates, their ability to perceive a new set of high-variability speech stimuli, such as produced by different speakers and in different phonetic contexts, significantly increased. Although their performance declined at the beginning of the training, it significantly improved later (Mullennix, 1989; Mullennix & Pisoni, 1990; Sommers et. al., 1994). Moreover, word-identification accuracy has also been shown to initially decrease under the high-variability training paradigm, but was found to later significantly facilitate listeners' sequential recall of words (Goldinger, Pisoni, & Logan, 1991). Similarly, Sommers & Barcroft (2007) found that, although variability in speaking-rate initially impedes word identification, it ultimately enhances L2 learners' knowledge of vocabulary. Variability of stimuli significantly improves participants' ability to perceive speech in the long term. Participants trained in this manner are also able to apply this ability to new speech sounds as well as to sounds produced by unfamiliar speakers (Lively et al., 1993, 1994). The ability to make generalizations to new talkers and new contexts is very important in learning Thai tones since there is an enormous amount of tonal variability in a real-world environment. It should be noted that the effectiveness of this training paradigm has been tested mainly on the learning of consonants and vowels. The goal of this study, on the other hand, is to explore if the high-variability training paradigm is more effective in improving L2 learners' ability to perceive lexical tones than the traditional method.

Training Studies (High-Variability Training Paradigm)

Numerous studies have shown that non-native language learners' perceptual system in both segmental and suprasegmental distinctions can be improved by perceptual/auditory training in a laboratory (Pisoni, Aslin, Perey, & Hennessy, 1982; Jamieson & Morosan, 1986; Logan, Lively, & Pisoni, 1991). For example, after American English listeners were trained in identifying non-native voicing contrasts, their performance significantly improved. The American English listeners were able to perceive voiceless aspirated, voiceless unaspirated, and voiced stops differing in voice onset time (VOT) after training (Pisoni, Aslin, Perey, & Hennessy (1982). Similarly, after auditory training, native speakers of French were able to discriminate between the two American English fricatives: /θ-ð/ (Jamieson & Morosan, 1986).

Besides perception of segments, perception of supra-segments such as tones has also been shown to improve with training. For example, it has been found that both perception and production of Mandarin tones by non-native speakers of Mandarin were significantly improved after the non-native speakers were trained with short perceptual tone training (Wang et al., 1999, 2003a). Moreover, Leather (1990) found that, in a group of Dutch speakers, production training on Mandarin tones had an effect on the learners' perception and perception training also improved their ability to produce the tones.

Based on these results, Wayland & Guion, (2004) conducted a study focusing on the transfer of prior experience with a tone system in a native language to experience with another tonal language. The findings suggested that the native speakers of

Chinese (NC) transferred their native ability in discriminating Chinese tones to their perception of Thai tones. Unlike the NC group, the native speakers of English (NE) lacked phonological representation of tones in their native long-term memory to help facilitate them in this process. Thus, even after the training, the NE speakers did not show significant improvement in perceiving Thai tones. Further, a comparison between naïve American English speakers of Thai and experienced American English learners of Thai confirms that experience with the target language (Thai) does facilitate these naïve American English listeners in acquiring tones (Wayland & Guion, 2003).

The above findings that perception of non-native segments and supra-segments improved with experience either in a laboratory or in natural setting challenged the strong version of the Critical Period Hypothesis (CPH) (Lenneberg, 1967; Scovel, 1969; Patkowski, 1989). According to CPH, plasticity of the human brain declines with age and language learning may be difficult or impossible beyond a critical time window which is believed to extend from birth to around puberty. Findings that adult L2 learners' perception of L2 sounds can be improved suggested, instead, that the human perceptual auditory system remains to some extent malleable throughout one's life span (Wayland & Guion, 2004).

This current study investigated Wayland & Guion (2004)'s suggestion that the non-native speakers of Thai might need more extensive training in order to reach the ability to perceive tones at the same level, or near the same level as native speakers. To facilitate the nonnative speakers' perception of Thai tones, an optimal training paradigm using high-variability stimuli produced by different speakers, both male and female, and with a fast speaking-rate, conversational style was employed. Furthermore, besides accuracy rate, the effectiveness of this training paradigm was also measured by learner's reaction times as well as their ability to generalize their perceptual ability to novel stimuli including new words produced by both new and old talkers.

Relationship between Speech Perception and Speech Production

This current research focused entirely on perception and excludes production. Nonetheless a brief theoretical overview of the relationship between speech perception and speech production is presented here. According to the Motor Theory of speech perception, there is a link between speech perception and speech production (Liberman, 1996; Liberman & Mattingly, 1985). Moreover, training in the perception domain can be transferred to better performance in the production domain. This notion is supported by a case study of Japanese learners who were trained to identify English /r/-/l / minimal pairs. These learners showed a close link between speech perception and production with training in one domain such as perception resulted in significant improvement in the other domain such as production (Akahane-yamada, Bradlow & Pisoni, 1996). The direct realist theory of speech perception (Fowler, 1986) supports the Motor Theory in terms of an inextricable link between speech perception and speech production.

However, according to the direct realist theory of speech perception, integrated event perception and action systems play a role in the perception-production link, not a specialized phonetic module as proposed by the Motor Theory. Moreover, according to the Motor Theory, speaker's intended neuromotor command of speech rather than its acoustic properties is the object of speech perception. On the other hand, besides taking acoustic and auditory patterns of speech as the object of speech perception, proponents of the acoustic-auditory theories of speech perception prescribe only an indirect link between speech production and perception (Steven & Blumstein, 1981; Diehl & Kluender, 1989). In this study, two Thai tones, mid and low, were selected due to L2 learners' observed difficulty differentiating these tones. Previous research has shown that these two tones are the most difficult to discriminate, even for native speakers of Thai, especially when they are perceived in isolation (Wayland & Guion 2004; Sathiansukon, 2005; Sittikesorn, 2005). The similarity between the low and mid tones is shown in Figure 2-4.

For this experiment, native speakers of American English were chosen as participants because their native language is non-tonal which makes it difficult for them to perceive Thai tones (Wayland & Guion 2003, 2004). According to Best (1995) and Flege (1995), similarities and differences between the phonological systems of Thai and American English may constitute significant factors causing tonal perception difficulties for American learners. Importantly, this research focuses on finding a better, more efficient method in teaching and learning Thai tones. Despite the fact that a number of studies have been conducted on teaching Thai to foreigners, none have exclusively focused on tones (Ponmanee, 2002a; Potibal, 2005; Sathiansukon, 2005; Sittikesorn, 2005). Even among those that emphasized phonology including consonants, vowels, and tones, tone is usually considered to be only a secondary aspect to investigate. Most importantly, the most effective method in teaching Thai tones has yet to be decided. Therefore, an experiment grounded in current linguistic theories should be conducted to find a more efficient solution. Equipped with a better teaching method, teachers of Thai would be able to help non-native learners to achieve their goal of learning Thai tones. A faster method would enable non-native learners learn Thai more quickly.

In sum, the above literature review suggested that due in part to a difference in the linguistic function of pitch in English and Thai and to the phonetic similarity between the Thai mid and low tones, native speakers of English would have difficulty perceiving the difference between these two tones. However, previous studies also suggested that the amount of perceptual difficulty they experience could be reduced with training. In addition, consistent with the exemplar-based theory of categorization and learning, a training paradigm using high-variability stimuli has been proven effective in improving non-native listener's sensitivity to non-native speech contrasts. The current study focused on the effectiveness of a high-variability perceptual training paradigm on the perception of the Thai mid and low tone contrast. Even though the study's focus was on tone perception, it was expected that improved perception, if observed, would result in improved production without further production training given the relationship between speech perception and production discussed above. Results of this currently study are of both theoretical and practical values. They would deepen our understanding of the mechanisms underlying the acquisition of a non-native prosodic system in general and of lexical tones in particular, and inform teachers, particularly of tone languages, of a perhaps more theoretically-sound and effective teaching method.

Table 2-1. Thai consonants

	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p p^h b		t t^h d			k k^h	ʔ
Nasal	m		n			ŋ	
Fricative		f	s				h
Affricate				tç $tç^h$			
Trill			r				
Approximant					j	w	
Lateral Approximant			l				

[Reprinted with permission from Abramson, A. S. & Tingsabadh, K. (1999), Thai final stops: Cross-language perception. *Phonetica, 56,* (page 118).]

Table 2-2. Thai vowels

	Place of articulation		
	front	central	Back
High	/i , i:/	/ɯ, ɯ:/	/u, u:/
Mid	/e, e:/	/ɤ, ɤ:/	/o, o:/
Low	/ɛ, ɛ:/	/a, a:/	/ɔ, ɔ:/
diphthongs	/ia, i:a/	/ɯ:a /	/ua, u:a/

[adapted from Jitapunkul, S., Maneenoi, E., Ahkuputra, V., & Luksaneeyanawin, S. (2003). Performance evaluation of phonotactic and contextual onset-rhyme models for speech recognition of Thai language. In *Eurospeech-2003*, (page 1841).]

Table 2-3. Characteristics of Thai tones

Tone	Tone mark	Pitch contour	Pitch height
Mid	Unmarked	Level	Medium
Low	ˋ	Level	Low
Falling	ˆ	Contour	High to low
High	ˊ	Level	High
Rising	ˇ	Contour	Low to high

[adapted from Hudak, T. (1990). *The world's major Languages,* Thai. In B. Comrie, (ed.), New York: Oxford University Press (page 760).]

Table 2-4. Thai tones

Tone	Word	Meaning
Mid	นา [na:]	'rice field'
Low	หน่า [nà:]	'apple custard'
Falling	หน้า [nâ:]	'face'
High	น้า [ná:]	'aunt'
Rising	หนา [nǎ:]	'thick'

Table 2-5. Comparison of phonetics of Thai tones (from 1911-1990)

Tone	Bradley 1911	Jones 1928	Henderson 1949	Abramson 1962	Henderson 1982	Arunrueng 1990
Mid	332	332	33	332	33	32
Low	221	22	22	21	21	21
High	341	341	343	343	34	34
Falling	31	31	51	451	52	343
Rising	35	14	14	14	14	213

[adapted from Arunrueng, A. (1990). Variation of the falling tone by age of speakers of Bangkok Thai. MA thesis. Chulalongkorn University (page 16).]

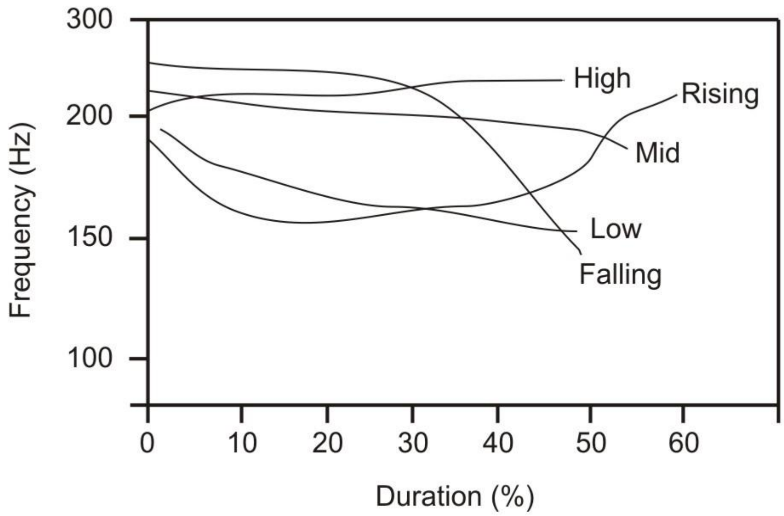

Figure 2-1. F0 contours of the five Thai tones

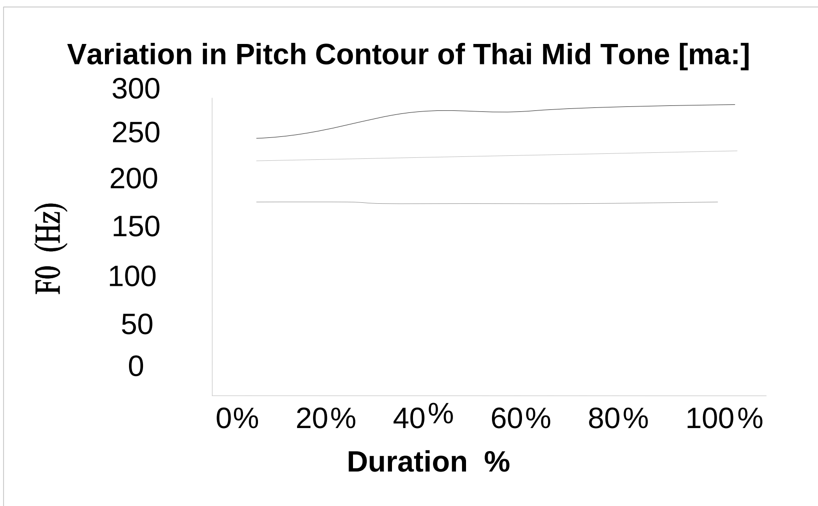

Figure 2-2. Thai mid tone syllable [ma:] produced in conversational style by 3 speakers

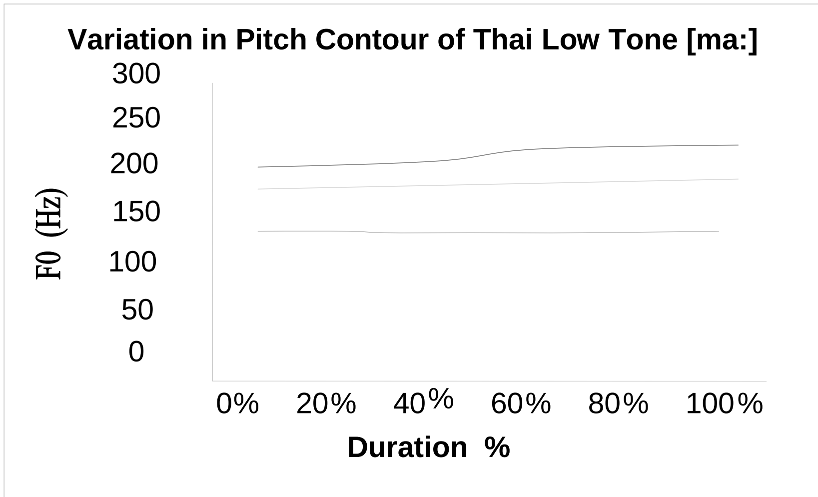

Figure 2-3. Thai low tone syllable [ma:] produced in conversational style by 3 speakers

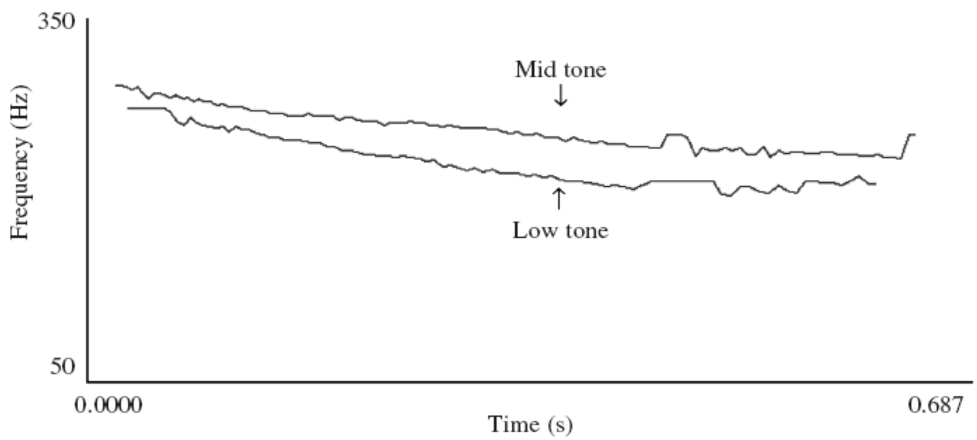

Figure 2-4. The fundamental frequency traces of the standard Thai low and mid tones. [reprinted with permission from Wayland, R. & Guion, S.G. (2004). Training English and Chinese listeners to perceive Thai tones: A Preliminary Report. Language Learning, 54 (4), (page 694).]

CHAPTER 3
METHODOLOGY

The overall design of this study involved testing three groups of native speakers of American English on their ability to perceive the distinction between two Thai tones: mid and low. None of the participants had prior experience with Thai or any other tonal language such as Chinese or Vietnamese, as determined by a language questionnaire. These participants were divided into three groups: two experimental groups and one control group. In addition, a control group of Thai speakers was tested. The study consisted of four phases: Pre-test, Training, Post-test, and Retention. During training, participants in one experimental group (the Prototype group) were trained with prototypical stimuli produced in citation forms while the other group (the High Variability group) was trained with high-variability stimuli produced by multiple speakers in a faster, conversational speech style. Participants in the control group did not receive any training and only participated in the Pre-test and the Post-test (American English control group); or only in the Pre-test (Thai control group). The effectiveness of the two training methods was assessed immediately after training and again two weeks after training. Besides accuracy rate and reaction time, the ability to make perceptual generalizations when presented with both old and new stimuli by new speakers was also evaluated.

Participants

Sixty naïve native speakers of American English: 27 men and 33 women, and 10 Thai native speakers (5 men, 5 women) were recruited from the student population of University of Florida. All were right-handed with ages ranging from 18 to 35 years old (mean = 24 years). All subjects were compensated for their participation at a rate of $ 8/hour, or if students participated for course credit, they received 1 credit/hour or up to a maximum of 2 credits for their participants. The English participants were randomly divided into three groups: two experimental groups and one control group. All three groups of participants were administered a Pre-test and a Post-test. However, only two experimental groups underwent training. The Thai Control group only did the Pre-test.

1. **The Prototype training group** (20 people: 8 men and 12 women) was trained to perceive the Thai mid and low tones with prototypical, low-variability stimuli produced in citation form by two Thai speakers (one male and one female) at a single slow, clear speaking-rate.

2. **The High Variability training group** (20 people: 9 men and 11 women) was trained with high-variability stimuli. Stimuli varied due to the use of multiple speakers: 4 speakers (2 men and 2 women), and speech occurred in a fast conversational style.

3. **The NE Control group** (20 people: 10 men and 10 women) was administered the Pre-test and, four days later, the Post-test, but did not undergo training.

4. **The Thai Control Group** (10 people: 5 men and 5 women) was administered the Pre-test, but did not undergo training.

Stimuli

Thirty minimal pairs (60 words) that contrasted the mid and low tones in standard Thai were used as stimuli in this study. The 60 words were real Thai words that were frequently used in the standard such as Bangkok variety of Thai. Additionally, in order to prevent an effect of syllable structures on tones, all stimuli were monosyllabic words containing the same or similar syllable structures. They were all ‘live’ syllables/words, that is, syllables that end with a long vowel or a sonorant final consonant: CVV, CVVS, or CVS (S are sonorants such as /m/ /n/ /ŋ/ /w/ and /j/). This study focused on live syllables/words, not dead or checked syllables/words such as those ending with a short vowel or a stop final consonant such as /p/, /t/, and /k/. Previous research has shown that subjects performed better on open syllables than on closed syllables (Wayland & Guion, 2003), perhaps due to the fact that the acoustic properties of low and mid tones in Thai are more salient in open syllables than in closed syllables. This is because F0 onset value is different in open syllables while both F0 onset and offset values are different in closed syllables (Wayland & Guion 2003). In addition, the mid and low tones were chosen for use in training due to their difficulty already discussed in Chapter 2.

The inter-stimulus interval was set at 1,500 ms to encourage participants' processing at the phonological level. According to Werker & Tees (1984) a phonological mode of processing occurred at the longer ISI (1500 ms) while a phonetic mode of processing occurred at the shorter ISI (500 ms). In addition, it is believed that prior linguistic experience affects processing at the phonological level, but not at the phonetic level of processing. The phonetic mode of processing is, therefore, viewed as a language-general mode of perception (Burnham et al. 1992; Werker & tees 1983). Evidence for these two levels of processing has been reported on lexical tone perception. In Burnham & Francis's (1997) for example, it was found that native Thai speakers discriminated Thai tones better under the longer ISI of 1500 ms while native Australian English speakers did better under the short, 500-ms ISI. The researchers asserted that this might have been due to the fact that native Thai speakers processed the tones at the phonological level and that the phonetic mode of processing was employed by native Australian English speakers. For further discussion on different modes of processing, see Wayland & Li (2008).

Stimuli for the Pre-Test

Two sets of stimuli were used in the Pre-test. The first set consisted of 10 minimal pairs of syllables produced with a mid or low tone in Thai. These 20 target words (as shown in Table 3-1) were produced in isolation by a Thai male (M1) and a Thai female speaker (F1) using a slow, clear, reading style. Each of the 20 target words was repeated 2 times generating a total of 80 tokens (20 words x 2 speakers x 2 repetitions). These 80 tokens were presented to participants twice, yielding 160 trials in total. In addition, the trials were presented in two blocks: 80 trials from the male speaker and 80 trials from the female speaker. The order of speakers (female-male and male-female) was randomized such that half of the participants were presented with stimuli produced by the male speakers first while the other half heard stimuli produced by the female first.

The second set of stimuli was the high-variability set. To create high stimulus variability, the same 10 target minimal pairs used to generate the first set were produced in various phonetic contexts. Specifically, they were produced at the end of the carrier sentences (see Appendix) and may be preceded by a low tone, a mid tone, or a high tone. As mentioned earlier in Chapter 2, preceding or following tones affect the target tone. For example, the F0 of the mid tone has a higher beginning point if the preceding tone is the high tone; the F0 has the lower beginning point if the preceding tone is the low tone. Thus, variations in the tones that precede the target tone produce desired variation in the pitch contour of that target tone. Therefore, each of the 20 target words (10 pairs) was produced in one sentential context preceded by various tones yielding 20 tokens in total. These 20 target tokens were produced by the same male and the same female speakers as the first set, but produced in a fast, conversational speed, thus generating 40 tokens (20 tokens x 2 speakers). The 40 tokens were presented four times for a total of 160 tokens. These 160 tokens were presented in two blocks: 80 tokens from the male speaker and 80 tokens from the female speaker with a random order of speakers (See Appendix).

Training Stimuli for the Prototype Training Group

Training stimuli for the Prototype training group consisted of 10 minimal pairs (shown in Table 3-2). These 20 target words (10 pairs) were produced by a Thai male (M2) and a Thai female speaker (F2*),* using a single speaking rate such as slow, careful reading pronunciation, generating a total of 40 tokens. Note that these two speakers were not the same speakers who produced the pre-test stimuli. These 40 tokens were presented to participants four times yielding 160 trials in total (see Appendix).

Training Stimuli for the High Variability Training Group

The same 20 target words (10 pairs) used in the Prototype training group were used in the High Variability training group. However, these 20 target words (10 pairs) were produced in a sentence final context similar to the high-variability stimuli set used in the Pre-test. To increase stimulus variability, the target words were pronounced by two male (M2, M3) and two female (F2, F3) speakers using fast, conversational speaking-rate and style**.** None of these speakers were the same as those who produced tones in the Pre-test materials. A total of 80 tokens (20 words x 4 speakers) were produced and presented two times. These 160 tokens were presented in two blocks: 80 tokens from the male speaker and 80 tokens from the female speaker (See Appendix).

Stimuli for the Post-Test

Stimuli for the post-test were identical to those used in the pre-test. There were two sets of stimuli. The first set consisted of 10 minimal pairs of syllables produced with a mid or low tone in Thai. The second set of stimuli was the high-variability set. The same 10 target minimal pairs used to generate the first set were produced in various phonetic contexts.

Stimuli for the Generalization Test

Four sets of stimuli were used in the Generalization test which included five old minimal pairs (previously used in the Pre-test) and 5 new (not used in the Pre-test) (see Table 3-3). These target words were produced by either old speakers such as M2, F2 speakers who produced training stimuli or new speakers such as M4, F4. Similar to Pre-test and Post-test stimuli, these stimuli were produced with different speaking styles: in isolation, slow, citation style; and in sentence final context at fast, conversational style. The Generalization test consisted of 80 tokens and 320 trials in total. Specific characteristics of the stimuli in each of the four sets were as follows.

> **Set 1: 20 tokens (10 words x 2 talkers) of 80 tokens,** old and new words produced by old talkers in sentence-final context at fast, conversational speaking rate.
>
> **Set 2: 20 tokens (10 words x 2 talkers) of 80 tokens,** old and new words produced by new talkers in sentence-final context at fast, conversational speaking rate.
>
> **Set 3: 20 tokens (10 words x 2 talkers) of 80 tokens,** old and new words produced by old talkers in isolation and in slow, citation form.
>
> **Set 4: 20 tokens (10 words x 2 talkers) of 80 tokens,** old and new words produced by new talkers in isolation and in slow, citation form.

In other words, set 1 and set 2 stimuli compared participants' perceptual generalizations to high-variability stimuli while set 3 and Set 4 tested their ability to generalize to low-variability stimuli.

Stimuli for Retention Tests

Two weeks after training, participants in the Prototype and the High Variability training groups were administered the same Post-test and the Generalization test again. To avoid confusion they will be referred to as Post-test vs. Retention test and

Generalization 1(or Gen 1) and Generalization 2 (or Gen 2) respectively. The stimuli and the format of these tests were otherwise identical.

Task

Tone identification task was used in this study. Two Thai tone, mid and low, were presented to participants. The participants were asked to indicate for each trial whether the tone they heard was a low or mid tone. Their responses were recorded and calculated as mean reaction time.

Procedure

Pre-Test

Before the experiment began, all participants were informed that they were participating in a speech perception experiment. Then they were asked to complete a language and education questionnaire. Participants were tested individually in a quiet, laboratory room. The Pre-test consisted of four blocks: two (one male speaker and one female speaker) with low-variability stimuli and two (one male speaker and one female speaker) with high-variability stimuli) and lasted approximately 30 minutes. Before taking the Pre-test, participants were presented with a short practice session with 10 trials in order to familiarize them with the stimuli and presentation rate (one stimulus per trial). They were told to focus on the difference between the mid and the low tone produced by Thai speakers. During the test, they were asked to identify the low and mid tone by selecting a button on the computer screen (a button marked ‘A’ for the mid tone and marked ‘B’ for the low tone). Half of the participants were tested with the low-variability stimuli first and the other half were tested with the high-variability stimuli first. The order of speakers (male speaker first/ female speaker first) was also counterbalanced across participants.

Training

After taking the Pre-test, participants in the High Variability training group and the Prototype training group were trained to identify the two Thai tones. The participants in the control group did not participate in any training. The participants heard one stimulus per trial and were asked to identify the tone category by selecting the corresponding button: A for the mid tone and B for the low tone. They could replay each trial as many times as they desired before making a decision. However, after the participants made a response, they could not replay the sounds. Feedback was given if an incorrect button was selected, the correct button would blink for 5 seconds. The two tones, the mid tone and the low tone, were randomly presented. Each participant completed four training sessions which lasted approximately 30 minutes each. The first session began on Day 1 after the Pre-test. The second and the third sessions took place on Day 2 and the last session took place on Day 3.

Post-Test

On Day 3 (after the fourth training session) participants in both training groups took the Post-test. To ascertain that the time lag between the Pre- and the Post-test was comparable for both the experimental (Prototype and High Variability Training) groups and the NE control group, participants in the control group were administered the Post-test on the fourth day after they took the Pre-test. Due to time limitations, we were not able to test them on the same day, such as Day 3, as those in the experimental groups. As mentioned earlier, the Post-test was identical to the Pre-test. This test thus tested the effectiveness of the training on stimuli heard on the Pre-test. It is important to note that participants were not trained on Pre-test stimuli. On day 4 (one day after the training and the immediate Pos-test), participants in the two training groups were administered the Gen 1 test designed to test which training method was more effective in terms of how well it generalizes to new stimuli.

Retention Test

Finally, two weeks after training, participants in the two training groups were asked to come back and took the Retention test (a delayed Post-test) and Gen 2 test. These two tests examined participants' retention of the effectiveness of the training on Post-test stimuli and on Gen 1 stimuli two weeks later. After training, percent correct identification of the two Thai tones for all subjects was calculated for statistical analyses. Based on previous research, it was expected that the higher degree of variability in the stimulus set used to train native speakers of American English in the High Variability training group would result in a more robust representation of the Thai low and mid tones in their long term memory. It was thus predicted that the American listeners in the High Variability training group would outperform those in the Prototype training Group on both the Post-test and the Generalization test. They were expected to gain the ability to distinguish the two Thai tones produced in new words and in both old and new words produced by new talkers.

Table 3-1. Minimal pairs of the pre-test and the post-test stimuli

	Mid-tone	Low-tone
1	[pa:] ‘to throw’	[pà:] ‘forest’
2	[pi:] ‘year’	[pì:] ‘an oboe’
3	[kha:] ‘to be stuck, lodge in’	[khà:] ‘galangal root’
4	[khu:] ‘a small canal’	[khù:] ‘to threaten’
5	[ku:] ‘I, me (ancient)’	[kù:] ‘to call from afar’
6	[thi:] ‘time, turn’	[thì:] ‘frequent’
7	[tha:ŋ] ‘way, path, road’	[thà:ŋ] ‘to spread apart, to widen’
8	[ph□:] ‘a large-raft boat’	[ph□̣:] ‘to expand, to stretch out’
9	[kɔ:] ‘group, pack, stalks’	[kɔ̀:] ‘to begin’
10	[taj] ‘kidney, wen’	[tàj] ‘to climb, to creep’

Table 3-2. Minimal pairs of the training stimuli

	Mid-tone	Low-tone
1	[phaj] ‘danger’	[phàj] ‘bamboo’
2	[kha:w] ‘fishy smell’	[khà:w] ‘news’
3	[pha:] ‘to lead, to bring along’	[phà:] ‘to split, to cut’
4	[pha:n] ‘troublesome’	[phà:n] ‘to pass’
5	[kha:j] ‘to spit out’	[khà:j] ‘ net, net-work’
6	[thaj] ‘freedom, independence’	[thàj] ‘to purchase back’
7	[kaw] ‘to scratch’	[kàw] ‘old, ancient’
8	[taw] ‘stove’	[tàw] ‘turtle’
9	[khom] ‘sharp point’	[khòm] ‘to press down, to suppress’
10	[kaj] ‘trigger’	[kàj] ‘chicken’

Table 3-3. Minimal pairs of the generalization (post-test) stimuli

	Mid-tone	Low-tone
1	[pha:] ‘to lead, to bring along’	[phà:] ‘to slit, to cut’
2	[pha:n] ‘troublesome’	[phà□n] ‘to pass’
3	[taw] ‘stove’	[tàw] ‘turtle’
4	[khom] ‘sharp point’	[khòm] ‘to press down, to suppress’
5	[thaj] ‘freedom, independence’	[thàj] ‘to purchase back’
6	[tha:j] ‘to predict, to guess’	[thà:j] ‘to excrete, to evacuate’
7	[pu:] ‘to spread out, to pave’	[pù:] ‘grandfather’
8	[tha:n] ‘to bear, to resist’	[thà:n] ‘charcoal’
9	[thɔ:] ‘to weave’	[thɔ̀:] ‘to punt, to pole’
10	[k□:] ‘you, he, she ’ (pro.)	[k□̣:] ‘old’

Table 3-4. Summary of procedure

Group	Day 1	Day 2	Day 3	Day 4	Two weeks later
High Variability	Pretest /training 1	Training 2, 3	Training 4 /post-test	Gen 1	Retention Gen 2
Prototype	Pre-test /training 1	Training 2, 3	Training 4 /post-test	Gen 1	Retention Gen 2
Control (NE)	Pre-test	-		Post-test	-
Control (Thai)	Pre-test	-	-	-	-

CHAPTER 4
PERCEPTION ACCURACY: PERCENT CORRECTION IDENTIFICATION

Percent correct identification of the two Thai tones obtained from both the Pre- and the Post-test was statistically analyzed using the Statistical Package for the Social Sciences (SPSS17) to test for the effects of Group (Prototype, High Variability, NE control, and Thai), Test Time (Pre-test and Post-test), and Stimulus Type (low-variability and high-variability). In this study, individuals whose scores fell more than two standard deviations (SD) below the group mean were considered outliers and were eliminated from the statistical analyses. Therefore, two outliers were removed and it was three percentages. No training data was obtained from the NE control group since they did not participate in the training phase. The native Thai speakers did not participate in either the training or the Post-test phase; therefore, only their Pre-test scores were analyzed and compared with those of the non-native groups.

The results will be reported starting with Pre-test and Post-test. The Pre-test data are compared among four groups: Prototype, High Variability, NE control and Thai groups; the Post-test data among three groups: Prototype, High Variability and NE control. The effects of two stimulus types are also examined: low-variability stimuli (produced in citation forms by two speakers, a male and a female speaker) and stimuli with high degree of variability (produced in sentence-final position by two male and two female speakers). Next, the training effects on participants’ perception will be reported by comparing the Pre-test with the Post-test among the three groups: Prototype, High Variability and NE control. Moreover, the results of the Generalization test comparing the two groups: Prototype and High Variability will be reported. In the last section, the results of the Retention-test (two weeks after the training) will be reported, by comparing the Pre-test to the Retention-test, and the Generalization test 1 to the Generalization test 2 between the two groups: Prototype and High Variability respectively.

Pre-Test

Mean percent correct identification of the two Thai tones, mid and low, for the four groups: Prototype (n=20), High Variability (n=20), NE control (n=20), and Thai groups (n=10), under two stimulus types: low-variability and high-variability and standard errors (SE), obtained during the Pre-test are reported in Table 4-1 and shown in Figure 4-1. Native Thai speakers earned the highest scores for both stimulus types (99.6% and 99.6%). Moreover, all groups' perception was more accurate for the low-variability stimulus than the high-variability stimulus condition (Prototype group mean = 77.9% vs. 73.5%, High Variability group mean = 80.0% vs. 77.7%, NE control group mean = 82.8% vs. 77.4%, Thai group mean = 99.6% vs. 99.6% and overall mean = 83.0% vs. 79.6%).

A repeated measures ANOVA with Group (Prototype, High Variability, NE control, and Thai) as the between-subjects factor, and Stimulus Type (low-variability and high-variability) as the within-subjects factor was performed on this data. As expected, the analysis yielded a significant main effect of Group [F (3, 66) = 6.90, p = .000]. Post-hoc pair-wise comparison using the Bonferroni method (p=.05) revealed that the Thai group outperformed all three non-native groups [means = 99.6%, 75.7%, 78.8%, 80.1%, for Thai, Prototype, High Variability, and NE control groups respectively, p <.01]. No significant difference among the non-native groups was found [F (2, 57) = .44, p = .646]. The main effect of Stimulus Type was also significant [F (1, 66) = 12.06, p = .001] confirming the above observation that participants performed significantly better on low-variability stimuli than on high-variability stimuli (mean=85.1% vs. 82.1%). No significant interaction between Group and Stimulus Type [*F* (1, 66) = 1.64, p = .187) was found. In sum, these results suggested that, prior to training, native Thai speakers outperformed all non-native speakers, all participants' performance was significantly better on low-variability than on high-variability stimuli, and performances of all 3 groups of non-native speakers were comparable.

Effects of Training

The effect of training was examined by comparing participants' performances on the four tests: Pre-test, Post-test, Retention-test (or delayed Post-test) and Generalization test. The Pre-test was compared with the Post-test (which were identical to those of the Pre-test) and Retention-test to evaluate the effect of training on the same words produced by the same speakers. Gen 1 and Gen 2 tests were also compared to assess if the effect of training is generalized to (1) new stimuli produced by new speakers, (2) new stimuli produced by old speakers, and (3) old stimuli produced by new speakers, as well as to investigate retention of the effects of the training.

Pre-Test to Post-Test Comparison

Mean percent correct identification for the three groups: Prototype (n=20), High Variability (n=20), and control groups (n=20), under two stimulus type conditions: low and high-variability, and two test times: pre-test and post-test, are reported in Table 4-2 and shown in Figure 4-2. Both experimental groups (Prototype and High Variability) performed better after the training under both stimulus type conditions: low-variability and high-variability. Concerning stimulus type, both experimental groups showed better performance for the low-variability stimuli than for the high-variability stimuli. Interestingly, the NE control group scored higher on the Post-test than on the Pre-test without training.

This data was submitted to a repeated measures ANOVA with Group (Prototype, High Variability, and NE control) as the between-subjects factor and Test Time (Pre- and Post-test) and Stimulus Type (low-variability and high-variability) as the within-subjects factors. The analysis yielded a significant main effect of Test Time [F (1, 57) = 46.02, p = .001] and Stimulus Type [F (1, 57) = 33.93, p = .001]; however, a significant main effect of Group was not found [F (2, 57) = 00, p = .999]. These results indicate that, when pooled across both types of stimuli, the participants' performance significantly improved after training (mean = 78.1% vs. 85.3%). Similarly, across both Test Times, the participants obtained significantly higher scores for low-variability stimuli than for high-variability stimuli (mean = 83.6% vs.79.9%). No significant interaction between Test Time and Stimulus Type was found [F (1, 57) = .51, p = .475]. Taken together, these results suggested that, immediately after training (Post-test), participants' perception of Thai tones significantly and equally improved. Moreover, their performance was better for the low-variability stimulus type than for the high-variability stimulus type.

Pre-Test, Post-Test and Retention-Test Comparison

Two weeks after training, participants were administered a Retention-test that was identical to the Post-test. Recall that the post-test was also identical to the pre-test. The administration of the retention-test allowed for an investigation of the effectiveness of the training two weeks after training. To test the participants' retention of the tone discrimination skills they had gained during the training, Pre-test, Post-test and Retention-test scores for the two experimental groups (the Prototype and the High Variability groups) were compared. Table 4-3, Table 4-4, Figure 4-3 and Figure 4-4 show mean percent correct identification of the two Thai tones, mid and low, for the two experimental groups: Prototype and High Variability, under two stimulus type conditions: low and high-variability, at three test times: Pre-test, Post-test and Retention-test.

As shown in the Table 4-3, Table 4-4, Figure 4-3, and Figure 4-4 above, both groups showed an improvement after training. It was also evident that the Prototype group did better than the High Variability group in the Post-test. However, as shown in the Retention-test scores, two weeks after training, the High Variability group's performance was better than that of the Prototype group: 97.2% vs. 93.3% for the low-variability stimuli and 94.0% vs. 88.0% for the high-variability stimuli. The Retention-test also showed that all groups' identification was more accurate under the low-variability stimulus type condition than the high-variability stimulus type condition (Prototype mean = 93.3% vs. 88.0%, High Variability mean = 97.2% vs. 94.0%, and overall mean = 95.2% vs. 91.0%).

A repeated measures ANOVA with Group (Prototype and High Variability) as the between-subjects factor, and Test Time (Pre-test, Post-test, and Retention-test), and Stimulus Type Condition (low-variability and high-variability) as the within-subjects factors revealed marginally significant main effect of Group [F (1, 18) = 4.20, p = .055]. This result suggested that, when pooled across three test times and two stimulus types, the High Variability training group outperformed the Prototype training group almost to a significant level (mean= 78.8% vs. 75.7%). Moreover, the analysis yielded a significant main effect of Test Time [means= 77.3%, 86.3%, and 93.1% for Pre-Post- and Retention tests respectively, [F (2, 18) = 21.76, p = .000]. Post-hoc pair-wise comparisons using the Bonferroni method suggested that the significant main effect of Test Time was largely due to the improvement from Pre-test to Post-test [p = .015], and from Pre-test to Retention-test [$p = .002$], but not from Post-test to Retention-test [$p = 1.000$]. Furthermore, a main effect of Stimulus Type was found [$F(1, 18) = 10.93$, $p = .004$] with the low –variability stimuli responded to more accurately compared to the high-variability stimuli (87.2 % vs. 83.9%). No significant interaction was found between Stimulus Type and Group [$F(1, 18) = .53$, $p = .474$] nor between Test Time and Group [$F(2, 17) = 2.04$, $p = .159$] nor between Stimulus Type and Test Time [$F(2, 17) = 1.73$, $p = .206$].

These results indicated that the training significantly improved the performance of participants in both experimental groups. The improvement was evident immediately after training and was retained two weeks after training. Interestingly, no additional amount of improvement was observed during the two weeks period between the Post-test and the Retention test. More importantly, overall, the High Variability training group performed almost significantly better than the Prototype group.

Effects of Training in the Generalization Test

Besides the Post-test and the Retention-test, the effectiveness of the training was also examined in the Generalization test administered immediately and two weeks after training. In this test, the effectiveness of the training on (1) new stimuli, and (2) new speakers was explored. It was hypothesized that the training affects to both new stimuli and new speakers.

Generalization Test 1 and Generalization Test 2 Comparisons

Generalization tests 1 and 2 were administered to examine if the effects of training generalized to new stimuli and new talkers. The two generalization tests were identical, except that Generalization test 1 (Gen1) was administered immediately after training and Generalization test 2 (Gen2) was administered two weeks after training. The

Generalization tests consisted of four sets of stimuli: (1) old words produced by old talkers, (2) new words produced by old talkers, (3) old words produced by new talkers, and (4) new words produced by new talkers. Mean percent correct identification scores from Gen 1 and Gen2 by the Prototype and the High Variability training groups are reported in Table 4-5 through Table 4-8, and Figure 4-5.

This data was submitted to a repeated measures ANOVA with Group (Prototype, and High Variability) as the between-subjects factor, and Test Time (Gen1, and Gen2), Stimulus Type (low-variability and high-variability), Talker (old and new), and Word (old and new) as the within-subjects factors. The analysis yielded a significant main effect of Talker [F (1, 18) = 7.27, p = .015], with performance being significantly better for stimuli produced by new talkers than by old talkers (mean= 85.0% vs. 82.9%); a main effect of Stimulus Type [F (1, 18) = 85.30% p = .000] confirming that significantly higher scores were obtained for low-variability stimuli than for high-variability stimuli (mean= 92.3% vs. 75.6%); a main effect of Word [F (1, 18) = 27.76, p = .000] with mean score for new words significantly higher than old words (mean = 85.7% vs. 82.3%); and a marginally significant main effect of Test Time [F (1, 18) = 3. 99, p = .061] with mean score of Gen1 higher than that of Gen 2 (mean score = 83.3% vs. 79.2%). However, no significant main effect of Group was found [F (1, 18) = 2.13, p = .161] even though the High Variability training group obtained higher score than the Prototype group (mean = 87.5% vs. 80.5%) overall.

These results suggested that the effectiveness of the training generalized to new talkers. However, the generalizability of the training effects appeared to decline over time (higher mean score for Gen1 than Gen2). High-variability stimuli were more difficult than low-variability stimuli. More interestingly, albeit statistically non-significant, the high-variability training paradigm appeared to result in a higher degree of generalizability than the prototype training method.

The analysis also yielded several 2- and 3-way significant interactions among factors. Significant interactions between Stimulus Type and Test Time [F (1, 18) = 11.30, p = .003], Stimulus Type and Talker [F (1, 18) = 8.59, p = .009], Test Time and Word [F (1, 18) = 5.28, p = .034], Talker and Word [F (1, 18) = 7.68, p = .013], Stimulus Type and Word [F (1, 18) = 10.09, p = .005], and Stimulus Type, Talker, and Word [F (1, 18) = 34.62, p = .000], were found. However, no significant interactions between Stimulus Type, Test Time, and Group [F (1, 18) = .00, p = 1.000], Stimulus Type, Talker, and Group [F (1, 18) = .34, p = .564], Test Time, Word, and Group [F (1, 18) = .21, p = .647], Talker, Word, and Group [F (1, 18) = .75, p = .398], Stimulus Type, Word, and Group [F (1, 18) = 2.72, p = .116].

Further analyses using one-way ANOVAs and T-tests were conducted to explore the significant interactions. Multiple comparisons were corrected using the Bonferroni method with the significant alpha level set at .05. The significant three-way interaction was explored first followed by 2-way interactions. Multiple t-tests conducted to explore the significant three-way interaction between Stimulus Type, Talker, and Word. First, the difference between the two stimulus types was compared. The analysis revealed that participants' performance on the low-variability stimuli was always significantly better than on the high-variability stimuli; old words produced by old talkers [Mean= 89.3 vs. 74.5, t (19) = 6.07, p = .000]; new words produced by old talkers [Mean=91.7 vs. 76.2, t (19) = 8.22, p = .000]; old words produced by new talkers [Mean = 93.9 vs. 71.4, *t* (19) = 10.54, p = .000; and new words produced by new talker [Mean = 94.3 vs. 80.5, *t* (19) = 10.82, p = .000].

When the difference between old and new talkers for each stimulus type (low and high-variability) and for each word type (old and new) were compared, it was found that the difference was significant for low-variability, old words [mean = 89.3 % vs. 93.9%, *t* (19) = -4.74, p = .000]; low-variability, new words [mean = 91.7% vs.94.3%, *t* (19) = -2.74, p = .013]; high-variability, new words [mean= 76.2% vs.80.5%, *t* (19) = -3.78, p = .001], but not for high-variability, old words [mean = 74.5% vs. 71.4%, *t* (19) = 2.29, p = .033]. These results, thus, suggested that overall participants performed significantly better on stimuli produced by new talkers than by old talkers. The only exception was found for high-variability, old words; their performance was better when these stimuli were produced by old talker, but the difference was not statistically significant.

T-test analyses comparing the difference between the two word types (old and new) for each stimulus type (low and high-variability) and for each talker (old and new), showed that participants performed significantly better on new words produced by new talker under the high-variability stimulus condition (mean = 80.5% vs. 71.4%) [*t* (19) = -5.89, p = .000], but not by old talker under the low-variability condition (mean = 91.7% vs. 89.3%) [*t* (19) = -2.40, p = .027], or old talker under the high-variability condition (mean = 76.2% vs. 74.5%) [*t* (19) = -1.44, p = .164] or new talker under the low-variability condition (mean= 94.3 vs. 93.9) [*t* (19) = .79, p = .437]. In other words, participants' performance on new words was superior to their performance on old words, but only when the new words were produced by new talkers under the high-variability stimulus condition such as conversational style.

A significant interaction between Stimulus Type and Test Time was further examined in one-way ANOVAs by comparing participants' performance on the two stimulus types for both Gen1 and Gen2. The analysis revealed that participants performed significantly better on low-variability stimuli than on high-variability stimuli immediately after training (Gen 1) [mean =92.5% vs. 74.1%, F (1, 39) = 88.68, p = .000] and also two weeks after training (Gen2) [mean= 92.1% vs. 77.1%, F (1, 39) = 15.16, p = .000]. The difference that was large in Gen1 than Gen2 was the source of this interaction.

Turning to the significant interaction between Test Time and Word, one-way ANOVAs revealed that participants' performance on Gen1 and Gen2 was comparable for both old [mean = 84.3% vs. 87.0%), F (1, 39) = .00, p = .994] and new words [mean = 84.3% vs. 87.0%), F (1, 39) = .53, p = .470]. However, performance on new words was significantly better than on old words for Gen2 [mean = 82.3% vs. 87.0%], t (19) = -8.02, p = .000], but not for Gen1 [mean=82.3% vs. 84.3%), t (19) = -1.90, p = .072]. Thus, the Test Time x Word significant interaction resulted mainly from the fact that, two weeks after training, participants' performance on new words was significantly better than their performance on old words.

In sum, analyses performed on the Gen1 and Gen2 tests which were designed to test the generalizability of the training effect revealed several interesting results. First, results obtained from the main repeated measures ANOVA analysis revealed that the effectiveness of the training generalized to new talkers. In addition, the generalizability of the training effects was, overall, more apparent immediately after training and seemed to have declined two weeks after training (higher score on Gen1 than onGen2).

Also, participants in the High Variability training group showed higher, but not statistically significant, degree of generalizability than those in the Prototype training group. Furthermore, participants found low-variability stimuli easier overall. Second, analyses conducted to examine interaction effects further revealed that a) participants preferred low-variability stimuli over high-variability stimuli regardless of whether they were old or new stimuli, produced by old or new talkers, b) participants' performance on new words was superior to their performance on old words, when these new words were produced by new talkers under the high-variability stimulus condition, c) participants performed significantly better on low-variability stimuli than on high-variability stimuli two weeks after training (Gen2), but not immediately after training (Gen1), and d) two weeks after training, participants' performance on new words was significantly better than on old words.

General Summary of Findings

In this chapter, statistical analyses performed on participants' identification accuracy data obtained from the Pre-, Post-, Retention and Generalization tests were reported. Results of the Pre-test showed that, as expected, native Thai speakers were superior to all three groups of non-native speakers, and that all groups performed significantly better on low-variability stimuli produced in citation forms than on fast, conversational, high-variability stimuli. After training, both training groups (Prototype and High Variability), and, surprisingly, the NE control group showed a significant and comparable amount of improvement (there was no significantly main effect of Group).

The effectiveness of the training remained significant two weeks after training for both training groups. More interestingly, when all three tests (Pre-, Post- and Retention) were considered, the High Variability training group outperformed the Prototype group at a near significant level. The effectiveness of the training was also found to extend to new words, particularly those produced by new talkers in conversational, high-variability style. Interestingly, the High Variability training group exhibited a higher degree of generalizability, overall (higher mean scores for Gen1 and Gen2) than the Prototype group, but the difference did not reach significance.

Table 4-1. Mean percent correct in mid and low tones identification in the Pre-test comparing between two stimulus types for each group

Group	Prototype		High Variability		NE Control		Thai	
Stimulus type	Low-vari	High-vari	Low-vari	High-vari	Low-vari	High-vari	Low-vari	High-vari
Mean	77.9	73.5	80.0	77.7	82.8	77.4	99.6	99.6
SD	14.5	15.5	15.0	14.8	16.5	17.2	0.4	0.5
SE	3.2	3.4	3.3	3.3	3.7	3.8	0.1	0.1

Table 4-2. Comparison of mean percent correct and SE between Pre-test and Post-test for low- and high-variability stimulus type conditions

	Prototype group				High Variability group				NE Control group			
	Low-vari		High-vari		Low-vari		High-vari		Low-vari		High-vari	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Mean	77.9	89.0	73.5	86.3	80.0	85.8	77.7	84.0	82.8	86.1	77.4	81.0
SD	14.5	14.0	15.5	12.7	15.0	15.0	14.8	15.1	16.5	18.2	17.2	16.8
SE	3.2	3.1	3.4	2.8	3.3	3.3	3.3	3.4	3.7	4.0	3.8	3.7

Table 4-3. Mean percent correct tone identification under low-variability stimulus type condition

	Prototype			High Variability		
	Pre-test	Post-test	Retention	Pre-test	Post-test	Retention
Mean	77.9	89.0	93.3	80.0	85.8	97.2
SD	14.5	14.0	9.6	15.0	15.0	2.4
SE	3.2	3.1	3.0	3.36	3.3	0.7

Table 4-4. Mean percent correct tone identification under high-variability stimulus type condition

	Prototype			High Variability		
	Pre-test	Post-test	Retention	Pre-test	Post-test	Retention
Mean	73.5	86.3	88.0	77.7	84.0	94.0
SD	15.5	12.7	12.5	14.8	15.1	4.1
SE	3.4	2.8	3.9	3.3	3.4	1.3

Table 4-5. Mean percent correct tone identification for Prototype group in the Generalization test 1 and 2 under low-variability stimulus type condition

	Prototype group low-vari condition							
	Gen1				Gen2			
	Old talker		New Talker		Old talker		New Talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	87.0	87.7	90.7	92.2	85.5	90.2	89.7	89.7
SD	16.0	18.7	15.3	14.0	19.5	14.8	18.1	17.1
SE	5.0	5.9	4.8	4.4	6.1	4.7	5.7	5.4

Table 4-6. Mean percent correct tone identification for Prototype group in the Generalization test 1 and 2 under high-variability stimulus type condition

	Prototype Group high-vari condition							
	Gen1				Gen2			
	Old talker		New Talker		Old talker		New Talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	69.8	69.0	70.1	73.1	72.5	75.0	67.8	77.9
SD	17.5	11.2	11.3	15.7	13.6	14.4	14.1	17.4
SE	5.5	3.5	3.5	4.9	4.3	4.5	4.4	5.5

Table 4-7. Mean percent correct tone identification for High Variability group in the Generalization test 1 and 2 under low-variability stimulus type condition

	High Variability Group low-vari condition							
	Gen1				Gen2			
	Old talker		New Talker		Old talker		New Talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	93.7	93.7	97.5	97.5	91.2	95.2	97.7	97.7
SD	5.0	9.8	2.6	3.1	5.1	5.9	2.7	4.6
SE	1.5	3.1	0.8	0.9	1.6	1.8	0.8	1.4

Table 4-8. Mean percent correct tone identification for High Variability group in the Generalization test 1 and 2 under high-variability stimulus type condition

	High Variability Group high-vari condition							
	Gen1				Gen2			
	Old talker		New Talker		Old talker		New Talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	75.8	77.5	73.8	84.0	79.8	83.3	74.0	87.0
SD	12.5	8.7	10.4	7.2	7.0	8.0	10.1	7.7
SE	3.9	2.7	3.3	2.2	2.2	2.5	3.2	2.4

Table 4-9. Means percent identification accuracy for factors that the main effect is significant collapsed over other variables

	Talker		Stims		Word		Test Time	
	Old	New	Low	High	Old	New	Gen-test1	Gen-test2
Mean	82.9	85.0	92.3	75.6	82.3	85.7	83.3	79.2

Table 4-10. Summary of significant results of the generalization tests

Factors tested	F value	P value
Talker	(1, 18) = 7.27	P = .015
Stimulus Type	(1, 18) = 85.30	P = .000
Word	(1, 18) = 27.75	P = .000
Test Time	(1, 18) = 3.99	P = .061 (marginally sig.)
Stimulus Type x Test Time	(1, 18) = 11.30	P= .003
Stimulus Type x Talker	(1, 18) = 8.59	P= .009
Test Time x Word	(1, 18) = 5.28	P= .034
Talker x Word	(1, 18) = 7.68	P= .013
Stimulus Type x Word	(1, 18) = 10.09	P= .005
Stimulus Type x Talker x Word	(1, 18) = 34.62	P= .000

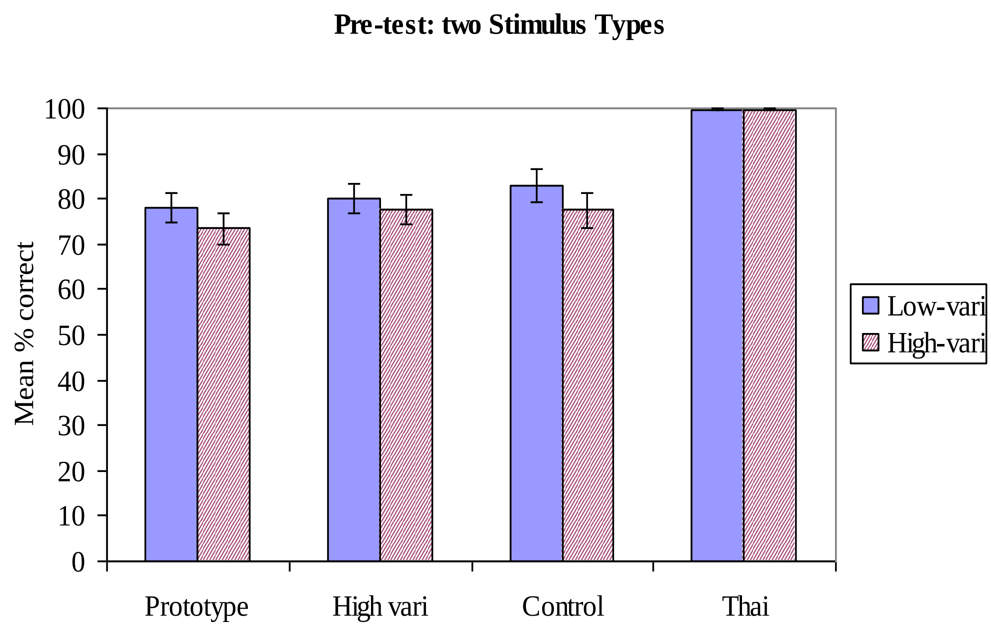

Figure 4-1. Comparison of mean percent correct and SE in the Pre-test between two stimulus types for each group

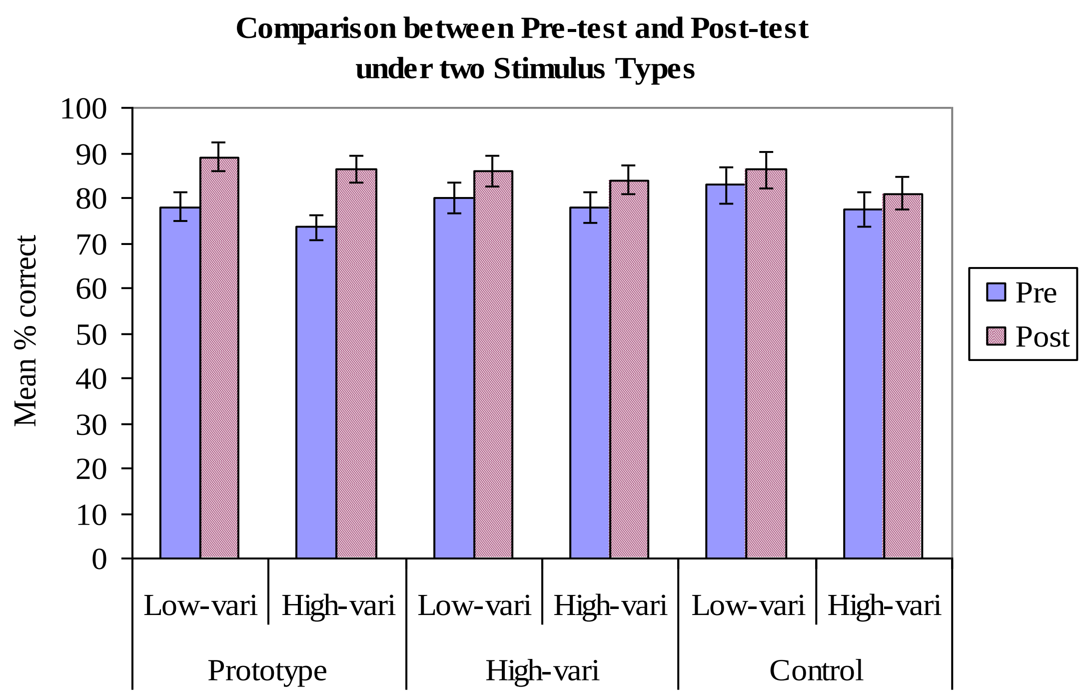

Figure 4-2. Comparison of mean percent correct and SE between the Pre-test and the Post-test under two stimulus type conditions

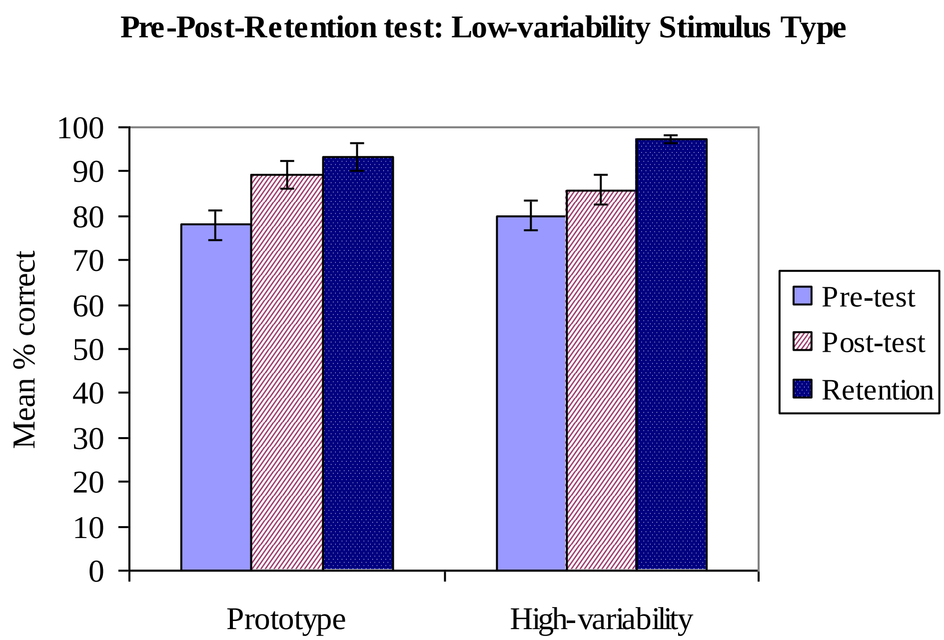

Figure 4-3. Mean percent correct tone identification and SE between the Pre-Post-Retention test under low-variability stimulus type condition

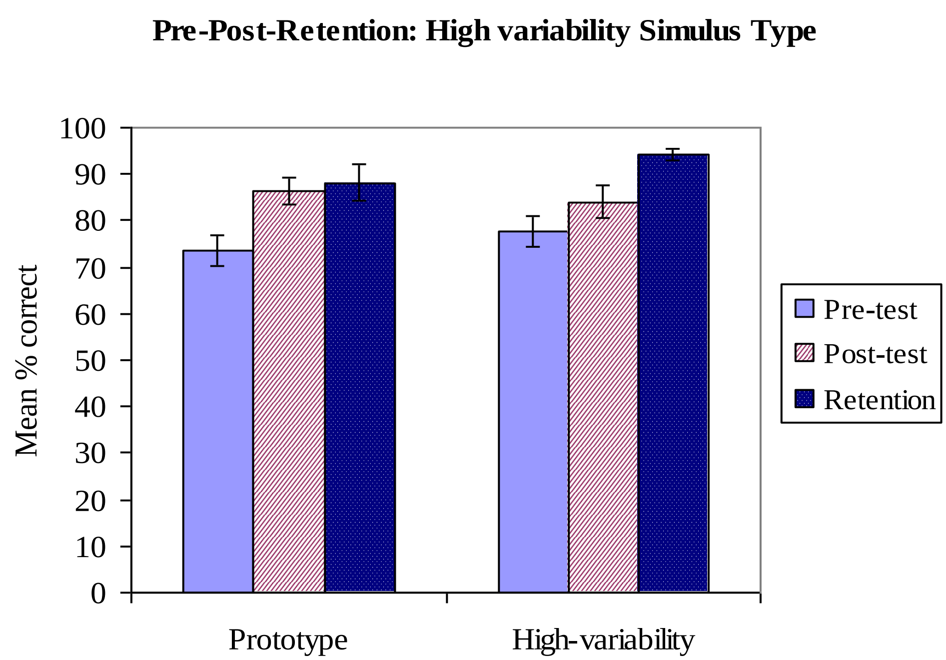

Figure 4-4. Mean percent correct tone identification and SE between the Pre-Post-Retention test under high-variability stimulus type condition

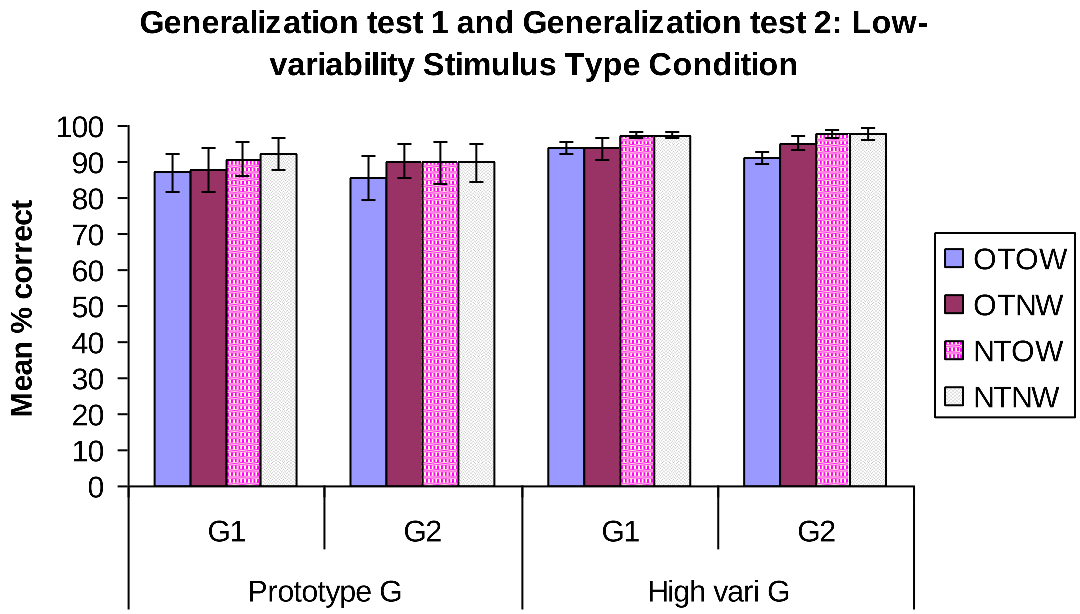

Figure 4-5. Mean percent correct tone identification and SE in the Generalization test 1 and 2 under low-variability stimulus type condition

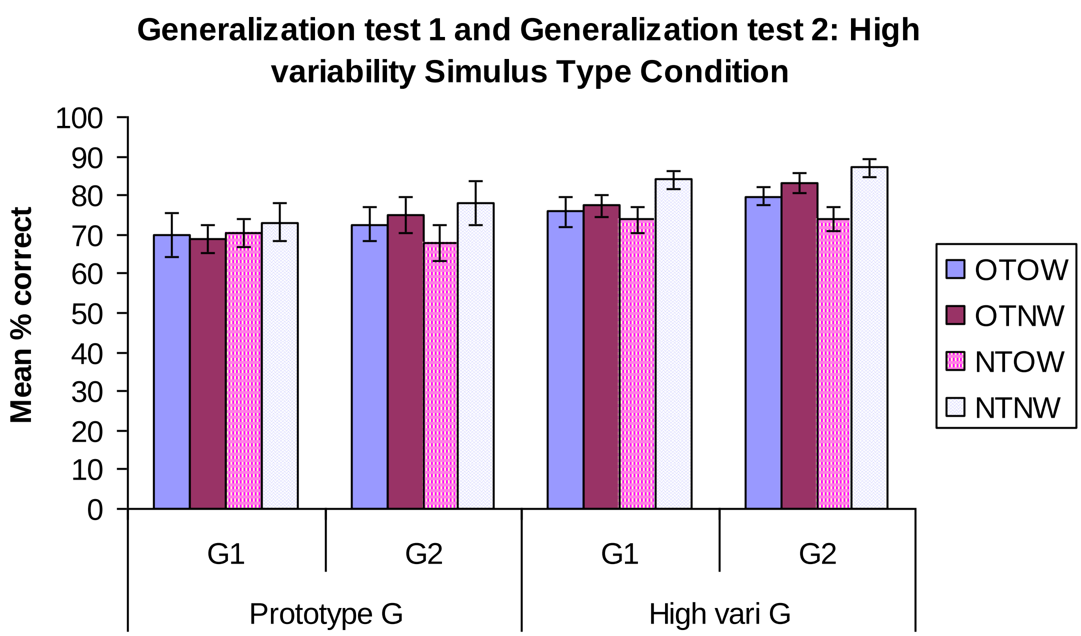

Figure 4-6. Mean percentage correct in tone identification and SE in the Generalization test 1 and 2 under high-variability stimulus type condition

CHAPTER 5
REACTION TIME

Besides mean percent correct identification reported in Chapter 4, participants' reaction time to correct responses were also collected and analyzed and the results will be reported in the same format as the mean percent correct identification. First of all, the results of reaction time to correct responses only were compared among four groups: Prototype, High Variability, NE control and Thai groups for Pre-test and among the three groups for Post-test: Prototype, High Variability, and NE Control. The effects of two stimulus types were also examined: low-variability stimuli (produced in citation forms by two speakers, a male and a female speaker) and stimuli with high degree of variability (produced in context by two male and two female speakers). Next, the training effects on participants' reaction time in perception will be reported by comparing the Pre-test with the Post-test among the three groups: Prototype, High Variability and NE Control. Moreover, reaction time data from the Generalization tests will be reported comparing between the two groups: Prototype and High Variability. In the last section, reaction time data from the Pre-Post and Retention tests and from the Generalization Test 1 and 2 will be reported.

Pre-Test

Reaction times (in milliseconds) in identification of the two Thai tones, mid and low, for the four groups: Prototype(n=20), High Variability (n=20), Control (n=20), and Thai groups(n=10), under two stimulus type conditions: low variability and high-variability obtained during the pre-test, are reported in Table 5-1 and Figure 5-1. Reaction time for the native Thai speakers and the NE control groups were faster than those of the Prototype and the High Variability training groups. However, for all groups, reaction time for high-variability stimuli was longer than for low-variability stimuli (mean = 675 msec. and 642 msec.).

A repeated measures ANOVA with Group (Prototype, High Variability, NE control, and Thai) as the between-subjects factor, and Stimulus Type (low and high-variability) as the within-subjects factor was performed on this data. The analysis yielded a non-significant main effect of Group (group mean averaged across all other variables: Prototype = 690 msec., High Variability = 710 msec., NE control= 615 msec., Thai = 620 msec.), F (3, 66) = 1.43, p = .240], but a significant main effect of Stimulus Type (mean for each stimulus averaged across all other variables: low-variability stimulus type = 642.5 msec., and high-variability = 675 msec.) F (3, 66) = 4.33, p = .041]. However, no interaction between Group and Stimulus Type [F (3, 66) = .05, p = .984] was found. These results suggested that, on the average, participants took longer to decide whether a mid or a low tone was heard when listening to high-variability stimuli.

Effects of Training

Effects of training on reaction times were also examined by comparing participants' performances on four tests: Pre-test, Post-test, Generalization test, and Retention-test. Pre-test was compared with Post-test and Retention-test to evaluate the effect of training on the same words produced by the same speakers. Gen 1 and Gen 2 tests were compared to assess if the effect of training is generalized to (1) new stimuli produced by new speakers, (2) new stimuli produced by old speakers, and (3) old stimuli produced by new speakers, as well as to investigate retention of the effects of the training.

Pre-Test to Post-Test Comparison

Reaction Times for the three Groups: Prototype (n=20), High Variability (n=20), and NE control groups (n=20), under two stimulus type conditions: low-variability and high-variability are reported in Table 5-2 and shown in Figure 5-2. Reaction time became faster after training for the Prototype and the NE control groups. The opposite, however, was true for the High Variability training group. Concerning stimulus type, all groups' reaction times were shorter for low-variability stimuli than for high-variability stimuli.

This data was submitted to a repeated measures ANOVA with Group (Prototype, High Variability, and NE control) as the between-subjects factor, and Stimulus Type (low-variability and high-variability) and Test Time (Pre- and Post-test) as the within-subjects factor. The analysis yielded a marginally significant main effect of Group [F (2, 57) = 3.01, p =.057], a significant effect of Stimulus Type [F (1, 57) = 22.32, p = .000], but a non-significant main effect of Test Time [F (1, 57) = .80, p =.374]. These results indicated that, on the average, reaction times for all three groups were comparable (mean for all three groups collapsed over both test times and both types of stimuli: Prototype = 670 msec., High Variability = 732 msec., control = 587 msec.), that they responded more quickly (640 msec. vs. 686 msec.) to low-variability stimuli than to high-variability stimuli, and that their response time did not improve significantly after training (mean for Pre-test = 673 msec. and Post-test = 653 msec.).

Pre-Test, Post-Test and Retention-Test Comparisons

Recall that the Post-test was identical to the Pre-test. The administration of the Retention-test allowed for an investigation of the effectiveness of the training on reaction time two weeks after training. To test for this effect, reaction times for Pre-test, Post-test and Retention-test for the two experimental groups (the Prototype and the High Variability groups) were compared. Table 5-3, Table 5-4, Figure 5-3 and Figure 5-4 show reaction time for the two experimental groups from the three tests: Pre-test, Post-test and Retention-test, under the low-variability and the high-variability stimulus conditions.

Overall, both groups' reaction time was faster under the low-variability stimulus type condition than the high-variability stimulus type condition (mean= 590 msec. vs. 680 msec.) for the Prototype group, and 540 msec. vs. 630 msec. for the High Variability group). In addition, the Prototype group's reaction time became shorter immediately after training (Post-test) while that of the High Variability group became longer. This was true for both types of stimuli (Prototype group mean = 620 msec. and 680 msec. and High Variability group mean = 720 msec. and 780 msec. for low- and high-variability stimuli type, respectively). However, two weeks after training (Retention-test), the opposite was true: reaction time for the High Variability group was faster than that of the Prototype group for both types of stimuli (mean = 540 msec. vs. 590 msec. and 630 msec. vs. 680 msec. for low- and high-variability stimuli respectively).

However, a repeated measures ANOVA with Group (Prototype and High Variability) as the between-subjects factor, and Test Time (Pre-test, Post-test, and Retention-test), and Stimulus Type (low-variability and high-variability) as the within-subjects factor revealed a significant main effect of only Stimulus Type [F (1, 18) = 16.26, p = .001], but not of Group [F (1, 18) = .00, p = .969], or Test time [F (2, 36) = .50, p = .608]. Furthermore, no significant interaction was found between Stimulus Type and Group [F (1, 18) = .25, p = .826] nor between Test Time and Group [F (2, 36) =

1.09, $p = .345$] nor between Stimulus Type and Test Time [F (2, 36) = 1.26, $p = .295$]. These results suggested that the training did not significantly improve participants' response time, neither immediately or two weeks after training, and that the high-variability training paradigm was not significantly more effective in reducing participants' reaction time than the prototype method.

Effects of Training in the Generalization Test

Besides the Post-test and the Retention-test, the effectiveness of the training on reaction time was also examined in the Generalization tests administered immediately and two-weeks after training. In these tests, the effectiveness of the training on reaction time on (1) new stimuli, and (2) new speakers was explored. It was anticipated that the training also affects on reaction time through both new stimuli and speakers.

Generalization Test 1 and Generalization Test 2 Comparisons

Effect of training in generalization to new stimuli was examined using two Generalization tests: Gen1 and Gen2. The Gen2 was identical to the Gen1 but it was administered two weeks after training. The Generalization tests consisted of four sets of stimuli: (1) old words produced by old talkers, (2) new words produced by old talkers, (3) old words produced by new talkers, and (4) new words produced by new talkers. Reaction time of the Gen1 and Gen2 by the Prototype and the High Variability groups are reported in Table 5-5 to Table 5-8, Figure 5-5, and Figure 5-6.

This data was submitted to a repeated measures ANOVA with Group (Prototype, and High Variability) as the between-subjects factor, and Test Time (Gen1, and Gen2), Stimulus Type (low-variability and high-variability), Talker (Old and New), and Word (Old and New) as the within-subjects factor. The analysis indicated that participants responded significantly more quickly to low-variability stimuli than to high-variability stimuli (596 msec. vs. 829 msec.), [main effect of Stimulus Type, *F* (1, 18) = 77.72, p =.000]. More interestingly, their reaction time became significantly faster two weeks after training (mean = 781 msec. vs. 644 msec.), [main effect of Test Time, *F* (1, 18) = 12.00, p = .003]. However, no significant main effect of Group (711 msec. vs. 714 msec.) [*F* (1, 18) = .25, p = .618], Talker (709 vs. 716) [*F* (1, 18) = .21, p = .646], and Word (710 msec vs. 716 msec) [*F* (1, 18) = 2.82, p = .110] was found suggesting that overall both groups' reaction times were comparable, that it took them the same amount of time to react to old and new talkers and to old and new words.

Interestingly, there were significant interactions between Stimulus Type and Word [*F* (1, 18) = 5.80, p = .027], and Talker and Word [*F* (1, 18) = 7.39, p = .014]. The significant interaction between Stimulus Type and Word was further examined by t-tests comparing the difference between the two Word types (old and new) for each Stimulus Type condition (low and high-variability). The analysis revealed that the difference between the old and new words was not significant under the low-variability stimulus type condition (584 msec. vs. 579 msec.) [*t* (19) = -57, p = .573], but it was significant under the high-variability one (809 vs. 836) [*t* (19) = -2,76 p = .012]. In other words, participants responded to old words significantly faster than to new words, but only when these words were produced in the high-variability stimulus condition.

The significant interaction between Talker and Word was further examined by comparing the difference between the two Word (old and New) for each Talker (old and New), the analysis revealed that the difference between the Old and New Word was not significant for the Old Talker (702 msec. vs. 694 msec.) [*t* (19) = .77, p = .450] was significantly different for the New Talker (691 msec. vs. 720 msec.) [*t* (19) =

-3.34, $p = .003$]. In other words, participants' reaction time was significantly slower for new words when they were produced by new talkers.

Summary of Findings

Statistical analyses performed on reaction time data revealed that, before training, all four groups of participants (Prototype, High Variability, NE control and Thai) took comparable amount of time to identify the low-level and mid-falling tones in Thai. However, they were significantly faster at identifying low-variability stimuli than high-variability stimuli. Interestingly, neither training method was found to significantly shorten the response time of any group or for any type (low- and high-variability) of (old) stimuli, either immediately or two weeks after training. Furthermore, even though there seemed to be an overall improvement in reaction time from immediately to two weeks after training on novel stimuli (shorter mean RT for Gen 1 than Gen 2), the training effects did not appear to generalize to new words or to new talker as it was found that reaction time for old words was significantly faster than for new words produced under high-variability condition, and that reaction time for new words produced by new talker was significantly longer than by old talker.

Summary of Accuracy and Reaction Time Data

When accuracy and reaction time data were considered together, the combined results were as follows. Firstly, results of the Pre-test showed that all groups of participants were significantly more accurate and faster at identifying low-variability than high-variability stimuli. The Pre-test results also showed that native Thai speakers were significantly more accurate, but not significantly faster than the non-native groups in identifying the two Thai tones. Secondly, immediately and two weeks after training, both training methods significantly improved tone identification accuracy of both training groups to a comparable degree. Their reaction times were also significantly shortened. When the Pre-, Post- and Retention-tests were considered together, it was found that the High Variability training group almost significantly outperformed the Prototype training group. A careful examination of Figure 4-3 and Figure 4-4 suggested that this result was likely due to the fact that the High Variability training group performed much better at the Retention-test than the Prototype group. Interestingly, their response time remained statistically comparable to the Prototype group suggesting that they have not sacrificed processing time for accuracy or vice versa. Thirdly, the effectiveness of both methods of training seemed to generalize to novel stimuli, particularly high-variability new words produced by new talkers at two weeks after training. That is, the analyses suggested that participants were more accurate at identifying the two Thai tones when they were new words produced in fast, conversational style by new talkers. Reaction time data suggested, however, that this effect may have come at the expense of a longer reaction time as it was found that reaction time for new words produced by new talker were significantly longer than for old words. These above results will be discussed in Chapter 6 along with tone language pedagogical implications, limitations of the study and future research.

Table 5-1. Reaction time in mid and low tones identification in the pre-test comparing between two stimulus type conditions for each group

Group	Prototype		High Variability		NE Control		Thai	
Stimulus type	Low-vari	High-vari	Low-vari	High-vari	Low-vari	High-vari	Low-vari	High-vari
Reaction time	670	710	700	720	600	630	600	640
SD	190	210	210	210	160	150	100	120
SE	40	40	40	40	40	40	60	60

Table 5-2. Comparison of reaction time and SE between Pre-test and Post-test for low- and high-variability stimulus type conditions

	Prototype group				High Variability group				NE Control group			
	Low-vari		High-vari		Low-vari		High-vari		Low-vari		High-vari	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
RT	670	620	710	680	710	720	720	780	600	520	630	600
SD	190	210	210	210	210	340	210	230	150	520	150	600
SE	40	60	40	50	40	60	40	50	40	60	40	50

Table 5-3. Reaction time in tone identification comparing between the Pre-Post-Retention test under low-variability stimulus type condition

	Prototype			High Variability		
	Pre-test	Post-test	Retention	Pre-test	Post-test	Retention
RT	670	620	590	710	720	540
SD	190	210	160	210	340	170
SE	40	60	50	40	60	50

Table 5-4. Reaction time in tone identification comparing between the Pre-Post-Retention test under high-variability stimulus type condition

	Prototype			High Variability		
	Pre-test	Post-test	Retention	Pre-test	Post-test	Retention
RT	710	680	680	720	780	630
SD	210	210	150	210	230	170
SE	40	50	50	40	50	50

Table 5-5. Reaction time in tone identification for Prototype group in the Generalization test 1 and 2 under low-variability stimulus type condition

	Prototype group Low-vari condition							
	Gen1				Gen2			
	Old Talker		New talker		Old Talker		New talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	620	620	620	600	560	560	600	580
SD	300	290	20	200	160	140	170	150
SE	70	70	50	40	50	40	50	50

Table 5-6. Reaction time in tone identification for Prototype group in the Generalization test 1 and 2 under high-variability stimulus type condition

	Prototype group High-vari condition							
	Gen1				Gen2			
	Old Talker		New talker		Old Talker		New talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	880	890	850	890	800	780	740	800
SD	270	290	290	300	210	220	230	240
SE	60	70	60	70	60	70	70	80

Table 5-7. Reaction time in tone identification for High Variability group in the Generalization test 1 and 2 under low-variability stimulus type condition

	High Variability group Low-vari condition							
	Gen1				Gen2			
	Old Talker		New talker		Old Talker		New talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	730	690	720	720	500	470	480	480
SD	270	260	320	270	100	150	130	130
SE	60	50	70	60	30	40	40	40

Table 5-8. Reaction time in tone identification for High Variability group in the Generalization test 1 and 2 under high-variability stimulus type condition

	High Variability group High-vari condition							
	Gen1				Gen2			
	Old talker		New talker		Old talker		New talker	
	OW	NW	OW	NW	OW	NW	OW	NW
Mean	880	910	910	980	730	730	740	760
SD	230	240	290	280	200	220	230	220
SE	50	50	70	60	60	70	70	70

Table 5-9. Reaction times means collapsed over other variables

	Talker		Stims		Word		Test Time		Group	
	Old	New	Low	High	Old	New	Gen-test1	Gen-test2	Prototype	High Variability
Mean	709	716	596	829	710	716	781	644	711	714

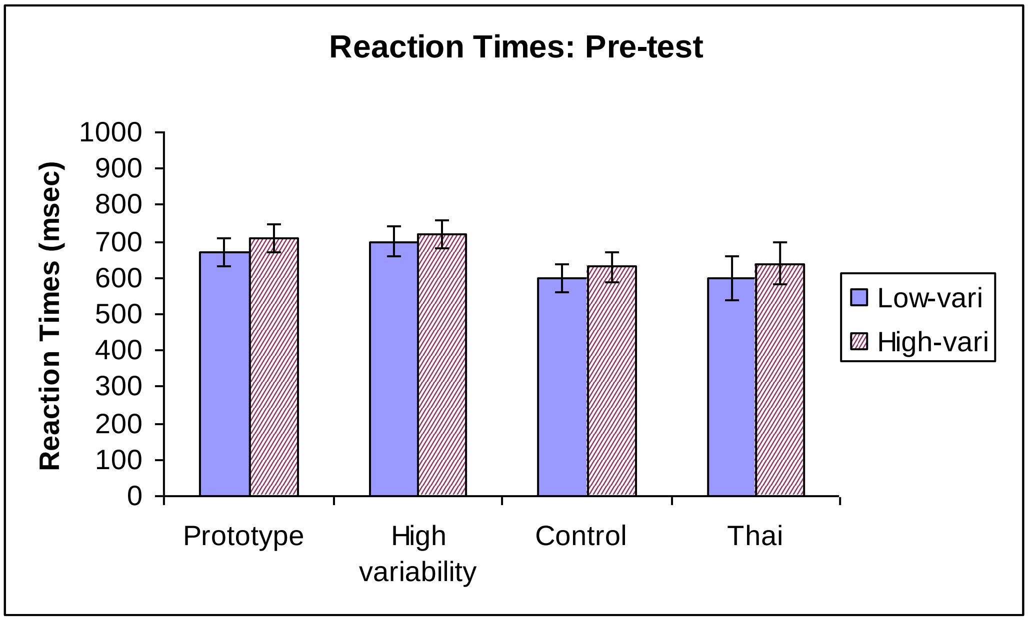

Figure 5-1. Reaction time in mid and low tones identification in the Pre-test comparing between the two stimulus type conditions and SE for each group

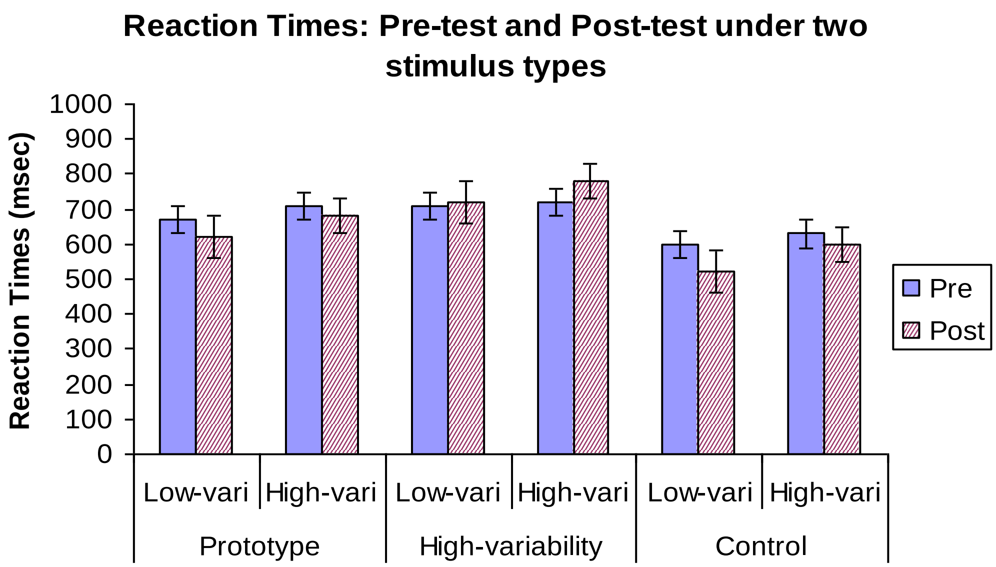

Figure 5-2. Comparison of reaction time and SE between Pre-test and Post-test under two stimulus type conditions

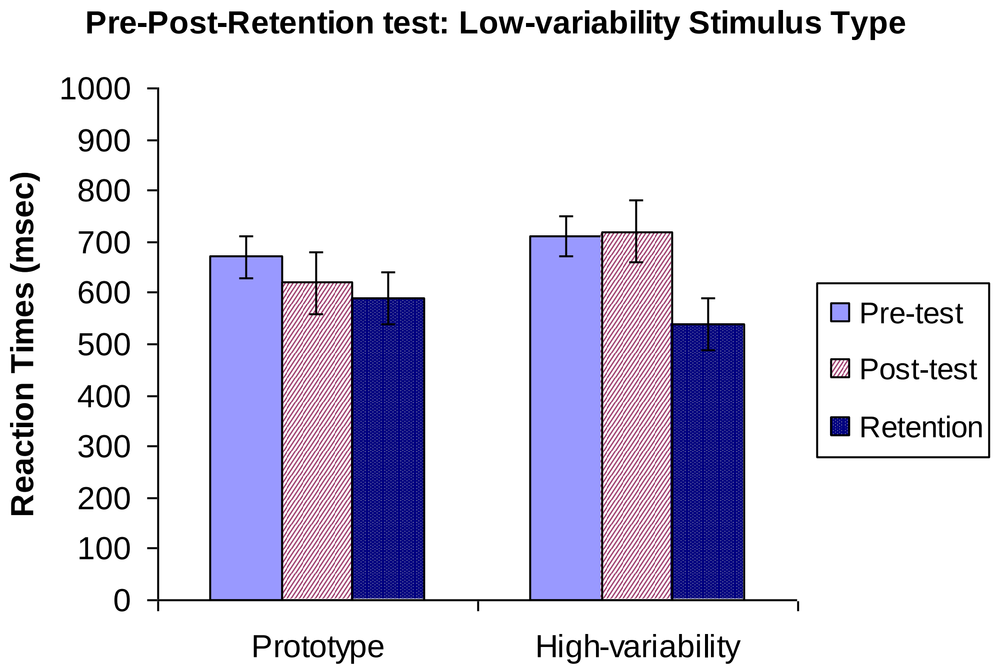

Figure 5-3. Reaction time in tone identification and SE comparing between the Pre-Post-Retention test under low-variability stimulus type condition

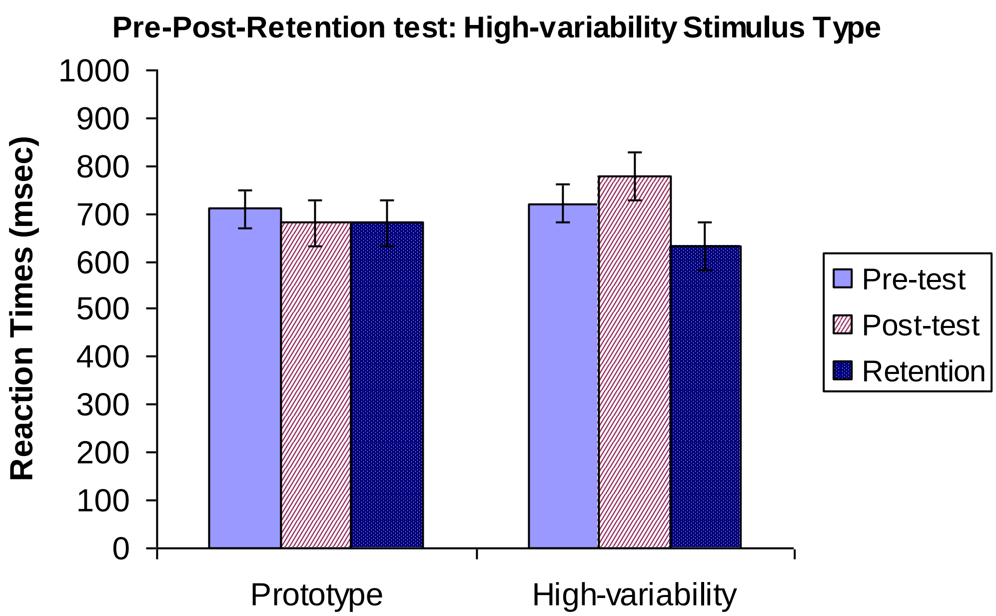

Figure 5-4. Reaction time in tone identification and SE comparing between the Pre-Post-Retention test under high-variability stimulus type condition

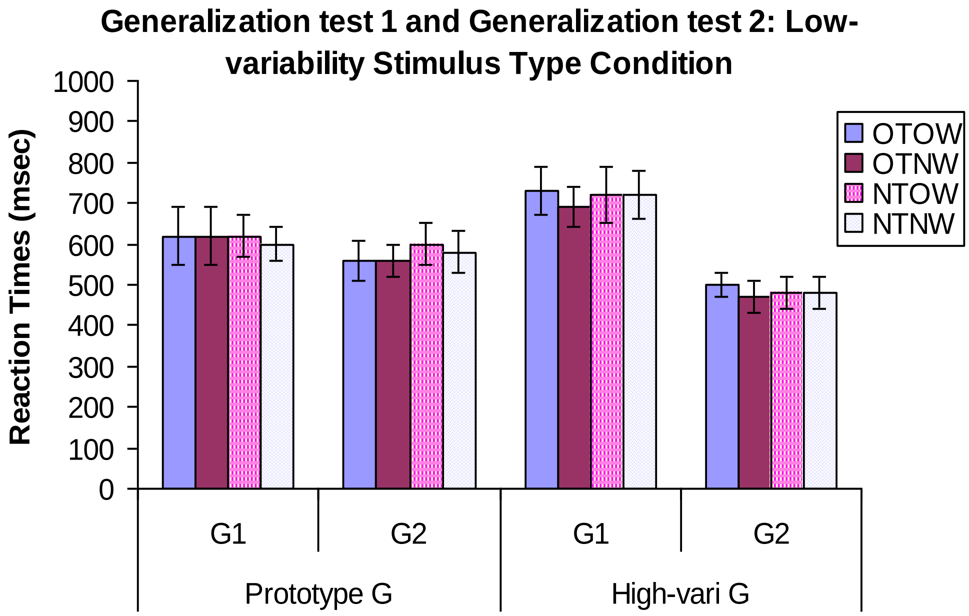

Figure 5-5. Reaction time in tone identification and SE in the Generalization test 1 and 2 under low-variability stimulus type condition

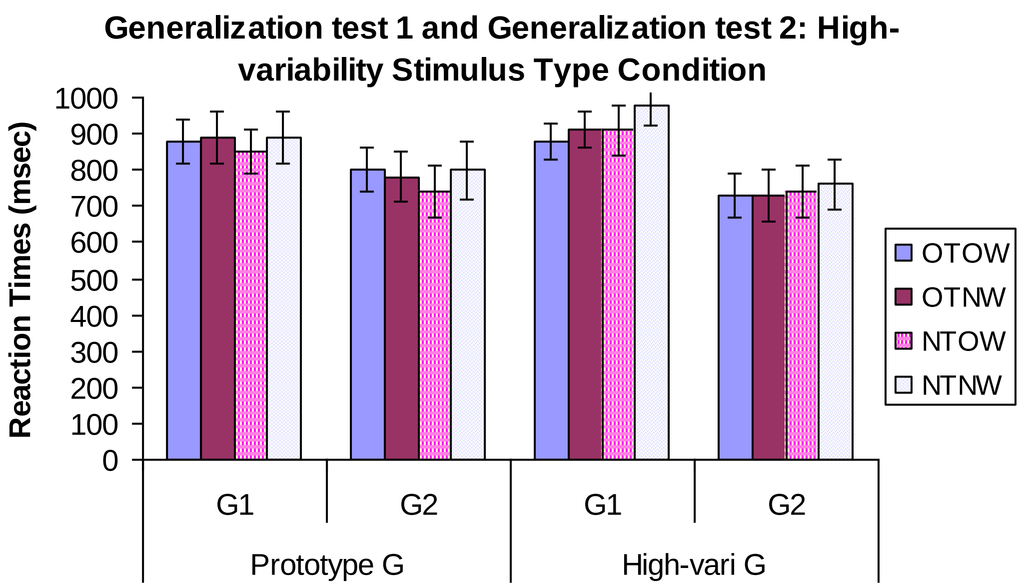

Figure 5-6. Reaction time in tone identification and SE in the Generalization test 1 and 2 under high-variability stimulus type condition

CHAPTER 6
DISCUSSION AND CONCLUSION

Relative efficacy of two methods used to train native speakers of English to identify the mid and the low tones in Thai was investigated. The two training methods differed mainly in the types of stimuli involved; the Prototype training paradigm used low-variability stimuli produced in isolation in a clear, citation style while the High Variability training method employed stimuli produced in a sentence-final context with a relatively faster, conversational style. The effectiveness of the training was measured immediately and two weeks after training from participants' identification accuracy rate as well as their reaction time (RT). Training effects on novel stimuli were also examined.

Pre-Test

The Pre-test results were compared among four groups: Prototype, High Variability, NE control, and Thai. First, as predicted, native Thai speakers were significantly more accurate at identifying the two Thai tones than all three groups of American English speakers under both stimulus type conditions: low-variability and high-variability. Second, all three groups of non-native speakers performed significantly better under the low-variability stimulus set. These results showed that the American groups did worse than the Thai speakers did, most likely because they had not had any experience with tones. As discussed in Chapter 1, non-native speakers of a tone language have difficulty when they learn tones, primarily because of their lack of prior experience with the target language (Piske, 2001).

RT data was consistent with the identification accuracy data and showed that RT was longer for high-variability stimuli than for low-variability stimuli for all groups of participants. These results suggested that additional variability in the stimuli resulted in an increase in processing load among both native and non-native speakers. The fact that native Thai speakers were not significantly faster than non-native speakers was somewhat surprising. No firm explanation can be provided at present. However, a relatively high identification accuracy (above 70%, chance = 50%) among non-native speakers suggested that the stimuli were easy for them to discriminate, which could explain why they responded as quickly as the native Thai speakers.

Pre-Test and Post-Test

Turning to the effect of training, comparing the Pre-test and Post-test performance among the three groups of American English listeners and the two stimulus types, the results showed that the training significantly improved the participants' perception of Thai tones, but the difference in improvement across the three groups was not significant. Interestingly, the increase in ability to discriminate the tones did not lead to an accompanying decrease in response time. Participants still thought about their responses for the same amount of time; they just achieved more accurate responses in the end. However, there was marginal evidence that the Prototype group reacted faster than the High Variability group, indicating that the performance of participants exposed to high-variability stimuli improved with training, but this may come at the cost of a higher processing load in comparison to low-variability stimuli. Overall, the pre-post training results suggested that training with simple laboratory techniques can improve or modify the adult human perceptual system that categorizes speech sounds in a short period of time, not only for consonant and vowel perceptions, as previous studies explored (Pisoni, Aslin, Perey, & Hennessy, 1982), but also for lexical tones. This evidence indicates that the human perceptual system is malleable and can be changed over one's life span (Best 1995; Flege, 1995).

It should be pointed out that the NE control group did not undergo training, yet they performed significantly better on the Post-test than on the Pre-test. However, unlike participants in the training groups, half of the participants in the NE control group were linguistics major students. Therefore, it is possible that their speech sound discrimination ability was superior to non-linguistics major students due either to their previous linguistic training or their linguistic inclination or propensity. Nonetheless, this result might lead one to think that simply exposing participants to target stimuli, with or without feedback, could lead to perception improvement. However, the fact that perceptual gain realized by participants in the NE control group was numerically lower than that of the two training groups suggested the additive benefit of the training or the provision of feedback. The statistically non-significant group difference observed was likely due to the duration of the training. In comparison to other studies, training duration employed in this study was relatively short. For example, the training lasted two weeks in Wang et al.'s (1999) study, and an average of 21% increase in Mandarin tone identification was observed. In this study, participants only underwent three days of training and perceptual gains of 11.99 % and 6.05 % for the Prototype and the High-Variability training groups, respectively, were recorded. A longer training period, such as two weeks or a month, may have yielded stronger and more obvious evidence of the training efficacy within and across groups.

Pre-Test, Post-Test and Retention Test

To investigate the longer-term effects of the training, a Retention test that was identical to the Pre-test and to the Post-test was administered two weeks after training. When all 3 tests are considered, what is interesting is that the difference between the two training groups reversed. Recall that the Prototype training group numerically outperformed the High Variability training group on the Post-test, which was administered immediately after training, but the High Variability training group numerically outperformed the Prototype training group on the Retention test two weeks later. In other words, perceptual gains realized among participants in the Prototype group were greater than those in the High Variability training groups immediately after training, but the opposite was true two weeks after training, leading to a nearly significant ($p=.055$) difference between the two groups. These results strongly suggested that long term representations of Thai tones formed under the high-variability training paradigm were more robust than those formed under the low-variability (prototype) training method. The fact that this result was found for both low- and high-variability stimuli lends further support to this claim.

Interestingly, it was also found that significant improvement in participants' accuracy scores was found from Pre- to Post-test and from Pre-test to the Retention test, but not from Post-test to the Retention test. This finding is interesting as it suggests that, similar to previous studies (Bradlow et al. 1999) the effectiveness of the training was retained at least two weeks after training. In addition, the fact that the High Variability group improved more between the Post-test and the Retention test than they did from the Pre-test to the Post-test suggests that the non-significant Post-test to Retention test difference was driven mainly by the data from the Prototype group. That is, while perceptual gains seemed to dissipate two weeks after training for the Prototype group, perceptual sensitivity seemed to continue to develop for the High Variability group.

Turning to reaction time, although the difference between the two training groups was not significant, a trend in the differential effects of training could be observed. Interestingly, the Prototype training group reacted numerically faster than the High Variability training group did for the Post-test whereas the High Variability training group reacted faster than the Prototype training group for the Retention test or two weeks after training in both stimulus type conditions. This result is consistent with mean percent correct identification, as the Prototype training group' s mean percent correct identification was higher than those in the High Variability training group for the Post-test, whereas the High Variability training group performed better than the Prototype training group on the Retention test two weeks after training.

In short, results of the combined Pre-, Post- and Retention tests suggested the superior long-term effects of the High-variability training paradigm. That is, contrary to the immediate, but quickly dissipating effects of the Prototype training paradigm, the High-variability training group endured a slow start, but had longer lasting improvement. This finding is consistent with findings of previous studies that performance after exposure to high-variability stimuli declined at the beginning of training and then significantly improved later (Mullennix, 1989; Mullennix & Pisoni, 1990; Sommers et. al., 1994).

Generalization Test 1 and Generalization Test 2

The results obtained from the Generalization tests illustrated the effects of training on learners' ability to generalize their knowledge to new words and new talkers. On both Generalization tests, the participants showed better performance on Generalization Test 1 than Generalization Test 2 under Low-variability Stimulus Type condition, New Talker and New word. The better performance on New Word stimuli might suggest that the participants were better at identifying new words because they matched the new words to the tonal representation, such as pitch contour, built/formed on old word during training. This finding provides evidence that the participants relied on the tonal representations they built/formed, not on each old word. This means that they were able to focus more on identifying the new stimuli without the conflict related to the old information of each old word. It may be easier for them to identify a new word than an old word, since for the new words they did not have the old information in their memory to be perfectly matched which may have confused the participants.

Concerning the result that the participants did better on New Talker than Old Talker stimuli, it is possible that the new talkers in the study produced the stimuli more clearly or more intelligible than the old talkers did. Other studies have shown that female talkers are more intelligible than male talkers (Bradlow et. al., 1996). However, the stimuli in this study were produced by both genders to avoid this effect. In interviews immediately after training, participants stated that the new talkers were more intelligible than the old talkers, and this may account for the result.

Turning to reaction time, the participants reacted faster in Generalization test 2 than on Generalization test 1. This finding supports the hypothesis that they needed time to form tonal representations in their memory, as they were able to react more quickly two weeks after training. Although a difference between the two experimental groups' reaction time for the Generalization test was not statistically significant, the High Variability training group showed numerically better performance than the Prototype training group did. This suggests that the High-variability training method may lead to a greater ability to make generalizations immediately after training. The ability to make generalizations is important to learning any phonetic aspect of any foreign language, particularly in learning Thai tones, given the enormous amount of tonal variability in a real-world environment. The High Variability training group seems to obtain more benefits in generalization from the higher degree of variability in the stimulus set they were exposed to during the training. This is consistent with previous studies stating the variability of stimuli facilitated participants' ability to perceive new speech sounds as well as to understand new speakers (Lively et al., 1993, 1994).

The initial research questions will now be addressed once more to illustrate the results.

1. Which training method (prototype vs. high-variability) will result in a higher percentage of perceptual accuracy immediately after training?

2. Which training method will produce faster reaction time in perception immediately after training?

3. Which training method will lead to a greater ability to make generalizations immediately after training?

4. Which training method will be more effective (as measured by perceptual accuracy, reaction time, and ability to make generalizations) 2 weeks after training?

For the first question, the results from all tests suggest that the Prototype training method did result in a higher percentage of perceptual accuracy immediately after training. Moreover, for the second question, the Prototype training group also reacted faster than the High Variability group at the Post-test administered immediately after training. However, for the third question, the High-variability training method led to a greater ability to make generalizations immediately after training. Most importantly, for the last question, the High-variability training method helped participants retain the Thai tones in the longer term, as shown by the Retention test administered two weeks after training. However, these are numerical data only; the differences by training group were not statistically significant.

In conclusion, this study investigated which method of teaching second language perception is more effective, the traditional Prototype approach or the more radical High-Variability approach. Unlike previous studies that focused on consonants and vowels, this study explored the effect of training on tone perception. Moreover, in this study, not only mean percent correct identification was examined, but also participants' reaction time and their ability to generalize to novel words and speakers. Tests were administered three times, before training, immediately after training and two weeks after training, the latter to investigate which training method is more effective for long-term gain. Importantly, the non-native speakers' performances significantly improved after High-Variability training, particularly two weeks after training, in both mean percent correct identification and reaction time. The results of this study suggest that the High-Variability training method has a longer and more enduring effect than the Prototype training method on tone perception for non-native speakers of tone languages.

The results support the Prototype theory in that providing learners with only ideal exemplar of the two Thai tones in citation forms does facilitate their learning of these tones immediately after training. However, the study also illustrated that, for long term gain, the results are consistent with the exemplar-based view. The entire system of the category shifted slightly with a new experience as seen from the result that the High-Variability training method has a longer and more enduring effect than the Prototype training method.

Applications to the Classroom

The results of this study provided useful information on effective training procedures on tone perception. Therefore, teachers who teach a tonal language such as Thai, Chinese, Vietnamese, and Yoruba, to L2 learners can employ the High-Variability training paradigm as a pedagogical tool in their teaching. Since long-term gain is obviously more important than immediate but temporary improvement, teachers should employ the High-Variability training paradigm. It would be more beneficial to get students to build a better long-term representation of tones. Alternatively, the teachers could combine the low- and high-variability training procedures, which may prove a more effective teaching method in their classrooms. For example, they may start their teaching with the Prototype training method providing the Prototype tones by themselves. After that, they may assign students to do activities to gain exposure to a variety of tones in real life, for instance, going out to Thai restaurants or travelling to tourist attractions to perceive tones from different talkers, different speaking styles, and speaking rates. Having been trained by a combination of these teaching procedures, the students should have more robust mental representations of tones in their memory.

Limitations of the Present Study

In this study, although the results showed that the participants performed better after training, some of their performances are numerically better but not statistically better. Therefore, further studies should use many more participants, more tokens of the same tones, more talkers, longer training, and a longer delay, perhaps of one to three months, before the Retention test. The most challenging task is designing an experiment to find out how much variability in each factor - word, talker, stimulus type condition, for instance - will provide optimal benefits in tone perception. Degree of variability should be considered very carefully to find a better solution. Moreover, to avoid a problem about the control group that might perform better than other participants due to their background; the participants should not be linguists or linguistics students although this may not always be possible.

Future Directions

Concerning the effectiveness of the training on generalization ability, the factors like Talker (New and Old) and Word (New and Old) should be expanded upon in further studies. For talker, it would be better if there were variations in speaking styles. For word, it would be better to employ higher-variability stimuli produced by multiple talkers in more sentential contexts, for example, the beginning, the middle, and the end of the sentences. However, a researcher needs to consider whether the stimuli can still be representative of a tone since a high degree of variability may cause excessive variation in a tone, and that tone may change and be realized as another tone. In addition, since this study did not test the generalization ability of the Control group, it is suggested that their generalization ability be further compared with the experimental groups to confirm the effectiveness of training.

To investigate the effect of training on working memory and long-term memory, the participants may be tested at multiple times, for instance, immediately after training and then two weeks, one month, two months, and three months after training. Then the results obtained from each test time will be compared to find significant improvement and the optimal period of time that is sufficient for learners to store high-variability information in their memory. In addition, it is suggested that other Thai tones or all five Thai tones be investigated and the results be compared. Moreover, concerning participants, comparing the ability of non-native speakers of tone languages like

American English and native speakers of tone languages like Chinese to perceive Thai tones using high-variability stimuli will generate better understanding of the effectiveness of training on tone perception for various listeners.

APPENDIX
STIMULI

1-20: Words in sentence final contexts of the Pre-test and the Post-test stimuli

1. พ่อบอกให้หยุดปา
[pʰɔ̂ː bɔ̀ːk hâj jùt paː]
Father told (Nid) to stop throwing.

2. ปู่เป็นคนรักป่า
[pùː pen kʰon rák pàː]
My grandfather loves the forest. My grandfather is a forest lover.

3. เธอรอฉันหลายปี
[tʰɤː rɔː tɕʰǎn lǎːj piː]
She has been waiting for me for years.

4. เขาชอบปี่
[kʰǎw tɕʰɔ̂ːp pìː]
He likes pipe.

5. อย่าให้เรื่องค้างคา
[jàː hâj rɯ̂aŋ kʰáŋ kʰaː]
Don’t let it be unresolved.

6. แม่กำลังปลูกข่า
[mɛ̂ː kamlaŋ plùːk kʰàː]
Mother is planting a galangal (plant).

7. ระวังจะเดินตกคู
[rawaŋ tɕàʔ dɤːn tòk kʰuː]
Don’t fall into the ditch!

8. เขาไม่กลัวแม้ว่าจะถูกขู่
[kʰǎw mâj klua mɛ́ː wâː tɕàʔ tʰùːk kʰùː]
He is not afraid even though he has been threatened.

9. ผู้หญิงไม่ค่อยพูดกู
[pʰûː yǐŋ mâj kʰɔ̂j pʰûːt kuː]
Women don’t like (to use) the word “*Koo*.”

10. สัตว์ป่าตะโกนกู่
[sàt pàː takoːn kùː]
Wild animals howl loudly.

11. เขาถูกตีสิบที
[k^{h}ǎw t^{h}ùːk tiː sìp t^{h}iː]
He is/was hit/bang/beaten 10 times.

12. คุณควรคิดให้ถ้วนถี่
[k^{h}un k^{h}uan k^{h}ít hâj t^{h}ûant^{h}ìː]
You should think carefully.

13. พ่อแม่ควรรักลูกให้ถูกทาง
[p^{h}ɔ̂ː mɛ̂ː k^{h}uan rák lûːk hâj t^{h}ùːk t^{h}aːŋ]
Parents should show/demonstrate their affection to their children in the right way.

14. เขาเดินขาถ่าง
[k^{h}ǎw dɤːn k^{h}ǎː t^{h}àːŋ]
He walks with legs wide apart

15. ฉันชอบนั่งกินข้าวที่แพ
[tɕ^{h}ǎn tɕ^{h}ɔ̂ːp nâŋ kin k^{h}âːw t^{h}îː p^{h}ɛː]
I like to have dinner on a raft.

16. ยายเป็นคนเอื้อเฟื้อเผื่อแผ่
[jaːj pen kon ʔɯ̂afɯ́ap^{h}ɯ̀ap^{h}ɛ̀ː]
My grandmother is a generous woman.

17. ที่หลังบ้านมีข่าสิบกอ
[t^{h}îː lǎŋ bâːn miː k^{h}àː sìp kɔː]
There are ten clumps of galangal in the backyard.

18. ปัญหานี้ใครก่อ
[panhǎː níː k^{h}raj kɔ̀ː]
Who cause this trouble?

19. ปู่ป่วยเป็นโรคไต
[pùː pùaj pen róːk taj]
Grandfather is sick with a kidney disease.

20. กระรอกตัวนี้ไม่ชอบไต่
[kràrɔ̂ːk tua níː mâj tɕ^{h}ɔ̂ːp tàj]
This squirrel doesn’t like scurrying.

21- 40: Words in sentence final contexts of the training stimuli

21. ขอให้ปราศจากโรคภัย

[kʰɔ̌ː hâj pràːtsatɕàːk rôːk pʰaj]
May you stay clear from illness.

22. เครื่องจักสานนี้ทำจากไผ่
[kʰrɯ̂aŋ tɕàk sǎːn níː tʰam tɕàːk pʰàj]
This handcraft is made of bamboo.

23. เธอไม่เคยมีราคีคาว
[tʰɤː mâj kʰɤːj miː raːkʰiː kʰaːw]
She has been virtuous.

24. ฉันเชื่อว่าเป็นการกุข่าว
[tɕʰǎn tɕʰɯ̂a wâː pen kaːn kùʔ kʰàːw]
I believe this is a rumor.

25. ลูกสาวคนนี้พ่อแม่หวังเป็นที่พึ่งพา
[lûːksǎːw kʰon níː pʰɔ̂ː mɛ̂ː wǎŋ pen tʰîː pʰɯ̂ŋpʰaː]
This daughter, the parents hope to rely on.

26. นิดกลัวฟ้าผ่า
[nít klua fâː pʰàː]
Nid fears thunder.

27. เธอจัดดอกไม้ใส่พาน
[tʰɤː tɕàt dɔ̀ːkmáj sàj pʰaːn]
She arranged the flowers into the tray.

28. ฉันสอบผ่าน
[tɕʰǎn sɔ̀ːp pʰàːn]
I passed the test.

29. ถ้าจับใบอ้อยจะรู้สึกคาย
[tʰâː tɕàp baj ʔɔ̂ːj tɕàʔ rúːsɯ̀k kʰaːj]
If you touch a sugar-cane leaf, your hand will feel irritated/rashed.

30. ปลาติดข่าย
[plaː tìt kʰàːj]
Fish was trapped in the fishing net.

31. ฉันรักประเทศไทย
[tɕʰǎn rák pràtʰêːt tʰaj]
I love Thailand.

32. เด็กถูกจับไปเรียกค่าไถ่
[dèk tʰùk tɕàp paj rîak kʰâː tʰàj]
A child was kidnapped (for ransom).

33. อย่าใช้เล็บเกา
[jàː tɕʰáj lép kaw]
Don’t scratch it with your finger nails.

34. เขาไม่ชอบขับรถเก่า
[kʰǎw mâj tɕʰɔ̂ːp kʰàp rót kàw]
He doesn’t like driving an old car.

35. แม่ทำความสะอาดเตา
[mɛ̂ː tʰam kʰwaːm saʔàːt taw]
My mother cleaned the stove.

36. เด็กบางคนไม่ชอบเต่า
[dèk baːŋ kʰon mâj tɕʰɔ̂p tàw]
Some kids don’t like turtles.

37. ความคิดเธอเฉียบคม
[kʰwaːm kʰít tʰɤː tɕʰìap kʰom]
Her mind is so sharp.

38. ไม่มีใครชอบถูกข่ม
[mâj miː kʰraj tɕʰɔ̂ːp tʰùːk kʰòm]
No one likes to be pressed.

39. นักยิงปืนกำลังลั่นไก
[nákjiŋpɯːn kamlaŋ lân kaj]
A sharpshooter is pulling the trigger.

40. พ่อครัวกำลังสับไก่
[pʰɔ̂ːkʰrua kamlaŋ sàp kàj]
The cook is chopping chicken.

41- 60: Words in sentence final contexts of the Generalization stimuli
(20tokens: 10 tokens of new words and 10 tokens from old words)

41. ลูกสาวคนนี้พ่อแม่หวังเป็นที่พึ่งพา
[lûːksǎːw kʰon níː pʰɔ̂ː mɛ̂ː wǎŋ pen tʰîː pʰɯ̂ŋpʰaː]
This daughter, the parents hope to rely on.

42. นิดกลัวฟ้า<u>ผ่า</u>
[nít klua fâː <u>pʰàː]</u>
Nid fears thunder.

43. เธอจัดดอกไม้ใส่<u>พาน</u>
[tʰɤː tɕàt dɔ̀ːkmáj sàj <u>pʰaːn</u>]
She arranged the flowers into the tray.

44. ฉันสอบ<u>ผ่าน</u>
[tɕʰǎn sɔ̀ːp <u>pʰàːn]</u>
I passed the test.

45. แม่ทำความสะอาด<u>เตา</u>
[mɛ̂ː tʰam kʰwaːm saʔàːt <u>taw</u>]
My mother cleaned the stove.

46. เด็กบางคนไม่ชอบ<u>เต่า</u>
[dèk baːŋ kʰon mâj tɕʰɔ̂p <u>tàw]</u>
Some kids don’t like turtles.

47. ความคิดเธอเฉียบ<u>คม</u>
[kʰwaːm kʰít tʰɤː tɕʰǐap <u>kʰom</u>]
Her mind is so sharp.

48. ไม่มีใครชอบถูก<u>ข่ม</u>
[mâj miː kʰraj tɕʰɔ̂ːp tʰùːk <u>kʰòm</u>]
No one likes to be pressed.

49. ฉันรักประเทศ<u>ไทย</u>
[tɕʰǎn rák pràtʰêːt <u>tʰaj</u>]
I love Thailand.

50. เด็กถูกจับไปเรียกค่า<u>ไถ่</u>
[dèk tʰùk tɕàp paj rîak kʰâː <u>tʰàj]</u>
A child was kidnapped (for ransom).

51. ถ้าไม่มั่นใจก็อย่า<u>ทาย</u>
[tʰâː mâj mân tɕaj kɔ̂ː jàː <u>tʰaːj</u>
If you are not sure, don’t guess

52. เด็กทารกกำลัง<u>ถ่าย</u>
[dèk tʰarók kamlaŋ <u>tʰàːj</u>]
Babies are defecating.

53. รีดเลือดกับปู
[rîːt lɯ̂at kàp puː]
To get blood from a crab. (English =Can’t get blood out of a stone)

54. เพื่อนบ้านก็รักปู่
[pʰɯ̂anbâːn kɔ̂ː rák pùː]
The neighbors also love my grandfather.

55. ทำข้อสอบเสร็จแล้วอย่าลืมตรวจทาน
[tʰam kʰɔ̂ːsɔ̀p sèt lɛ́ːw jàː lɯːm trùat tʰaːn]
When you are done with the test, don’t forget to go over all your answers.

56. ไฟฉายรุ่นนี้ไม่ต้องใช้ถ่าน
[fajtɕʰǎːj rûn níː mâj tɔ̂ŋ tɕʰáj tʰàːn]
This kind of flashlight does not require batteries.

57. ไหมแท้ที่แม่ทอ
[mǎj tʰɛ́ː tʰîː mɛ̂ː tʰɔː]
Real silk that my mother weaved.

58. เขาถือไม้ค้ำถ่อ
[kʰǎw tʰɯ̌a máːj kʰám tʰɔ̀ː]
He carries/holds a pole.

59. มันไม่เกี่ยวกับแก
[man mâj kìaw kàp kɛː]
It is not your business.

60. หลายคนบอกว่านิดแก่
[lǎːj kʰon bɔ̀ːk wâː nít kɛ̀ː]
Many people told Nid that she is old.

LIST OF REFERENCES

Abramson, A.S. (1962). *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments.* Columbia University.

Abramson, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. In T.L. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai And Mon-Khmer phonetics and phonology*: *Indigenous Languages of Thailand Research Project,* 1-9.

Aroonmanakun, V. (2006). *Using Language Data Base for learning and Teaching Thai.* Paper presented at the Conference on the Teaching Thai as a Foreign Language, Bangkok, Thailand.

Arunrueng, A. (1990). *Variation of the falling tone by age of speakers of Bangkok Thai*. MA thesis. Chulalongkorn University.

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

Bradley, C. B. (1911). Graphic analysis of the tone-accents of the Siamese Language. *JAOS*. 282-289.

Bradlow, A. R., Lively, S. E. Magnuson, J. S., Yamada,R. A. & Tohkura, Y. (1995). *Testing the importance of talker variability in non-native speech contrast training*. Paper presented at the 129th Meeting of the Acoustical Society of America, Washington, D.C.

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.

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

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.

Chun, D. (1982). *A contrastive study of the suprasegmental pitch in modern German, American English and Mandarin Chinese*. Ph.D. dissertation, University of California.

Clark, J., & Yallop, C. (1995). *An Introduction to Phonetics and Phonology*. Cambridge: Blackwell.

Clopper, C. G. & Pisoni, D. B. (2004). Effects of Talker Variability on Perceptual Learning of Dialects. *Language and Speech. 47(3).* 207-239.

Catford, J. C. (1989). *A Practical Introduction to Phonetic*. Oxford: Clarendon Press.

Flege, J. (1995). 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.

Gahl, S. & Yu, C. A. (2006). Introduction to the special issues on exemplar-based models in Linguistics. *The Linguistics Review*, 23, 213-216.

Goldinger, S. D., Pisoni, D. B. & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning Memory & Cognition, 17,* 151-162.

Goldinger, S. D (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166-1183.

Henderson, E., J.A. (1982). Tonogenesis: Some Recent Speculations on the Development of tones. *Transaction of the Phisological Society.*

Hudak, T. (1990). Thai. In B. Comrie, (ed.), *The world's major Languages* 756-775. New York: Oxford University Press.

Jamieson, D.G., & Morosan, D.E. (1989). Training New, Nonnative Speech Contrasts: A Comparison of the Prototype and Perceptual Fading Techniques. *Canadian Journal of Psychology*, *43(1), 88-96.*

Jitapunkul, S., Maneenoi, E., Ahkuputra, V., & Luksaneeyanawin, S. (2003). Performance evaluation of phonotactic and contextual onset-rhyme models for speech recognition of Thai language. In *Eurospeech-2003*, 1841-1844.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In *Talker variability in Speech Processing*, Johnson, K. & Mullenix, J. (eds.), 145-166. San Diego: Academic Press.

Kirk, K.I., Pisoni, D.B. & Miyamoto, R.C. (1997). Effects of stimulus variability on speech perception in hearing impaired listeners. *Journal of Speech, Language, and Hearing Research,* 40, 1395-140.

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

Kuhl, P. K. (1992). Psychoacoustics and speech perception: Internal Standards, Perceptual Anchors, and prototypes. In L. Werner & A. james (eds.): *Developmental Psychoacoustics.* American Psychological Association (pp. 293-332). Washington, DC: American Psychological Association.

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

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutchand English speakers. In J. Leather and A. James (eds.): *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech* (pp. 305-341). Amsterdam: University of Amsterdam.

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., Pisoni, D. B., 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 identifyEnglish /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874-886.

Luksaneeyanawin, S. (1993). Speech computing and speech technology in Thailand. In *Proceedings of the Symposium on Natural Language Processing,* 276-321. Bangkok, Thailand.

Luksaneeyanawin, S. (1998) .Intonation in Thai. In Hirst, D. and Di Christo, A. (Ed.), *Intonation Systems. A Survey of Twenty Languages*. Cambridge University Press, Cambridge.

Morén, B., Zsiga, E. (2006). The lexical and postlexical phonology of Thai tones. *Natural Language & Linguistic Theory, 24*, 113-178.

Nosofsky, M., R. (1988). On Exemplar-Based Exemplar Representations: Reply to Ennis (1988). *Journal of Experimental of Psychology, 117,* 412-414.

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

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29(2),* 191-215.

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.

Pakasri, Y., Wallapa, T. & Supak, M. (2007). *Thai in Daily Life*. Bangkok: Aempan.

Pisoni, D.B. (1997). Some Thoughts on "Normalization" in Speech Perception. In K. Johnson and J.W. Mullennix (Eds.). *Talker Variability in Speech Processing*. 9-32. San Diego: Academic Press.

Ponmanee, S. (2002a). *Foundations of teaching Thai as a foreign language*. Bangkok: Chulalongkorn University Press.

Ponmanee, S. (2002b). *Language and Teaching*. Bangkok: Klet Thai.

Potibal, P. (2005). *The Problems of Thai Pronunciation for Korean Students.* Paper presented at the International Conference on Teaching Asian Languages to Non-Native Speakers, Bangkok, Thailand, August 4-5.

Poomsan, B. B. (1995). *Thai for Beginners.* Bangkok: Paiboon Publishing.

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

Smith, D. (1995)*. Thai: A Complete Course for Beginners*. London: Hodder Headline Publishing.

Smith, D. (2005). *Teaching Thai to non-native speakers: some lessons from the past*. Paper presented at the International Conference on Teaching Asian Languages to Non-Native Speakers, Bangkok, Thailand, August 4-5.

Sathiansukon, S. (2005). *Situations and Problems Found in Intensive 60-hour ThaiLanguage Course: A Case Study of Danish Students Taking 10-Day and 15-Day Courses*. Paper presented at the International Conference on Teaching Asian Languages to Non-Native Speakers, Bangkok, Thailand, August 4-5.

Satravaha, N., Klinkhachorn, P. & Lass,N (2002). Tone classification of syllable-segmented Thai speech based on multilayer perception. *The Journal of the Acoustical Society of America.* 111, 2366-2366.

Sittikesorn, S. (2005). The relationship between the error of reading and the perception of Thai monosyllables: A case study of Japanese students who learn Thai as a foreign language. *Wannawitat, 5,* 168-185.

Sommers, M.S. & Barcroft, J. (2007). An integrated account of the effects of acoustic variability in first language and second language: Evidence from amplitude, fundamental frequency, and speaking rate variability. *Applied Psycholinguistics, 28,* 231-249.

Abramson, A. S. & Tingsabadh, K., (1999). Thai final stops: Cross-language perception. *Phonetica, 56,* 111-122.

Udompan, C. (2005). Problematic cases in pronouncing Thai sounds for learners. Paper presented at the International Conference on Teaching Asian Languages to Non-Native Speakers, Bangkok, Thailand, August 4-5.

Vongvipanond, P. (2000). *Survival Thai.* Bangkok: Chulalongkorn University Press.

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

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* II (pp. 511–514), Beijing.

Wang, Y., Jongman, A., & Sereno, J. A. (2003a). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America, 113,* 1033–1043.

Wang, Y., Sereno, J.A., Jongman, A., & Hirsch, J. (2003b). fMRI evidence for corticalmodification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience, 15(7),* 1-9.

Wayland, R. (1997) Non-native production of Thai: acoustic measurements and accentedness ratings, *Applied Linguistics, 18,* 345-373.

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

Wayland, R. & Guion, S.G. (2004). Training English and Chinese listeners to perceive Thai tones: A Preliminary Report. *Language Learning, 54 (4),* 681-712.

Wayland, R, Kaan E, Bao, M. & Barkley, CM (*2007* (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research, 1148*, 113-122.

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.

Wittayasakpan, S. (2002). *Bap rian pasaThai buengton nai boribot* Thai sueksa samrap chao tangchat. Bangkok: University Bureau.

Wittayasakpan, S. (2005). Teaching Thai as a Foreign Language*. Wannawitat,* 5, 215-261.

Yip, M. (2002). *Tone.* New York: Cambridge University Press.

BIOGRAPHICAL SKETCH

Donruethai Laphasradakul, daughter of Saner Khaodeedech and Benchaphorn Khaodeedech, received her Bachelor's and Master's degrees from the Thai department of Chulalongkorn University, Thailand. She completed her undergraduate education with a Top Scores Achievement Award in Thai Courses and Second Class honors. She was awarded a fellowship for her MA studies and then became a teaching assistant at the department. While studying for her MA, she started to work as a lecturer at the Department of Thai Language for Communication, the University of the Thai Chamber of Commerce (UTCC). She also won a full scholarship to the Ph.D. program in Linguistics at the University of Florida from UTCC and she will return to UTCC after graduation.

As a Thai language teacher who has enthusiastically been teaching Thai for eight years, she has always been fascinated by the use of Thai as a second language. The Seminar on Teaching Thai as a Foreign Language 2002, at Mahasarakham University, Thailand, inspired her chiefly to search for new aspects of Second Language Acquisition to strengthen her capability in teaching Thai to foreigners. Therefore, her aim is doing research on teaching Thai as a second language, focusing on Thai tones.