

(Non-)categorical Perception of Mandarin Tones: A Comparison Between Speakers of
Tone and Non-tone Languages

by

Xianghua Wu

B.A., Beijing Language and Culture University, 1995

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

© Xianghua Wu, 2006
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other
means without the permission of the author.

(Non)categorical Perception of Mandarin Tones: A Comparison Between Speakers of
Tone and Non-tone Languages

by

Xianghua Wu

B.A., Beijing Language and Culture University, 1995

Supervisory Committee

Dr. Hua Lin, (Department of Linguistics)

Supervisor

Dr. Sonya Bird, (Department of Linguistics)

Departmental Member

Dr. John H. Esling, (Department of Linguistics)

Departmental Member

Dr. Richard King, (Department of Pacific and Asian Studies)

Outside Member

Dr. Daniel Bryant, (Department of Pacific and Asian Studies)

External examiner

Supervisory Committee

Dr. Hua Lin, (Department of Linguistics)

Supervisor

Dr. Sonya Bird, (Department of Linguistics)

Departmental Member

Dr. John H. Esling, (Department of Linguistics)

Departmental Member

Dr. Richard King, (Department of Pacific and Asian Studies)

Outside Member

Dr. Daniel Bryant, (Department of Pacific and Asian Studies)

External examiner

ABSTRACT

This thesis examines the factors that affect categorical perception of Mandarin tones by English, Thai and Mandarin speakers. A pre-study and a main study were conducted. The pre-study was carried out to investigate the correspondence between Mandarin and Thai tones. The main study compared the categorization of two Mandarin tone contrasts, one between the high level tone (T1) and falling tone (T4), and the other between the rising tone (T2) and falling-rising tone (T3). The results suggest that linguistic experience, natural auditory sensitivity and perceptual training all affect Mandarin tone categorization, among which, the effect of linguistic experience is most noticeable. This study also demonstrates that Best et al.'s (1988, 1994, 1995) Perceptual Assimilation Model (PAM)

is useful in characterizing not only segmental assimilation patterns, as illustrated in many previous studies, but also suprasegmental categorization patterns.

TABLE OF CONTENTS

ABSTRACT.....	iii
TABLE OF CONTENTS.....	v
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
ACKNOWLEDGEMENTS.....	x
DEDICATION.....	xi
Chapter One: INTRODUCTION.....	1
1.1 Purpose of this study.....	1
1.2 Research questions.....	3
1.3 Outline.....	4
Chapter Two: LITERATURE REVIEW.....	6
2.1 Categorical perception.....	7
2.2 Categorical perception of segments.....	8
2.2.1 Language experience effects on perception.....	8
2.2.2 Language-independent effects on perception.....	11
2.3 Categorical perception of tones.....	14
2.3.1 Tone languages.....	14
2.3.2 Perception of tone continua.....	16
2.3.2.1 Cross-language studies.....	16
2.3.2.2 L1 studies.....	21
2.3.3 Other cross-language studies on tone perception.....	23
2.4 Summary.....	27
Chapter Three: STUDY I: CORRESPONDENCE BETWEEN MANDARIN AND THAI TONES.....	29
3.1 Introduction.....	29
3.2 Description of Mandarin and Thai tone systems.....	29

3.2.1 Mandarin tone.....	29
3.2.2 Thai tone.....	31
3.3 Method.....	34
3.3.1 Subjects.....	34
3.3.2 Stimuli.....	34
3.3.3 Procedure.....	40
3.4 Results.....	41
3.4.1 Correspondence of tokens on the T1-T4 continuum to Thai tones.....	41
3.4.2 Correspondence of tokens on the T2-T3 continuum to Thai tones.....	43
3.5 Discussion.....	45
Chapter Four: STUDY II: IDENTIFICATION AND DISCRIMINATION OF TONE CONTINUA.....	48
4.1 Introduction.....	48
4.2 Method.....	49
4.2.1 Subjects.....	49
4.2.2 Stimuli.....	50
4.2.2.1 Stimuli in categorical task.....	50
4.2.2.2 Tone training stimuli.....	52
4.2.3 Procedure.....	53
4.2.3.1 Identification test.....	53
4.2.3.2 Discrimination test.....	54
4.2.3.3 Perceptual tone training.....	55
4.3 Analyses and results.....	57
4.3.1 Identification tests.....	58
4.3.1.1 Probit.....	58
4.3.1.2 Pooled results.....	59
4.3.1.3 Comparison between trained and untrained subjects.....	63
4.3.2 Discrimination tests.....	65
4.3.2.1 Pooled results.....	65
4.3.2.2 Comparison between trained and untrained subjects.....	74
4.4 Summary and discussions.....	78
Chapter Five: GENERAL DISCUSSIONS.....	82
5.1 Categorical perception of tone.....	83
5.1.1 Presence of auditory sensibility.....	84
5.1.2 Acquired F_0 pattern.....	86
5.1.3 Learned acoustic experience.....	92
5.1.4 Summary.....	94

5.2 Perceptual assimilation model (PAM).....	96
5.2.1 Perceptual assimilation patterns.....	96
5.2.2 Assimilation patterns of Thai listeners.....	97
5.2.3 Assimilation pattern of English listeners.....	99
5.2.4 Summary.....	101
Chapter Six: CONCLUSION.....	102
6.1 Summary of the findings.....	102
6.2 Limitations and further studies	106
6.3 Concluding remarks	108
REFERENCES.....	110
APPENDIX I.....	118
APPENDIX II.....	118
APPENDIX III.....	120

LIST OF TABLES

Table 3-1.	F ₀ values (in Hz) of the four Mandarin tones produced by a male and a female speaker (Wu, 1986, adapted from Wang, 2002)	30
Table 3-2.	F ₀ values (in Hz) of the five Thai tones produced by male and female speakers (Gandour & Potisuk 1991, Burnham & Francis, 1997)	31
Table 3-3.	Comparison of Mandarin and Thai tones (Chao, 1948; Grandour, 1978, Ladefoged, 2001, 237-238).....	32
Table 3-4.	Predicted counterparts of Mandarin tones in the Thai tone system.....	33
Table 3-5.	F ₀ pattern and duration of selected “best” tones.....	36
Table 3-6.	F ₀ range and duration of the manipulated tones.....	38
Table 3-7.	The first and last points and turning points (bold) along tone contours as well as occurrence time in the two continua: (A) in T1-T4, (B) in T2-T3.....	39
Table 3-8.	Number of responses and average similarity ranking between T1 and T4 and the Thai tones.....	43
Table 3-9.	Number of responses and average similarity ranking between T2 and T3 and the Thai tones.....	45
Table 4-1.	Intercept and slope for Mandarin, Thai, and English participants’ data.....	59
Table 4-2.	Intercepts and slopes for trained and untrained Mandarin, Thai, and English groups.....	63
Table 4-3.	Mean correct discrimination levels (%) for Mandarin, Thai, and English groups on the AX discrimination task.....	66
Table 4-4.	Mean correct discrimination levels (%) for the English group and the musically trained English subgroups on the AX discrimination task.....	67
Table 4-5.	Mean correct discrimination levels (%) for trained and untrained Mandarin, Thai, and English listeners.....	74
Table 5-1.	Comparison of the occurrence of the turning point (Occurrence) and F ₀ between the tonal onset and the turning point (F ₀) of T2 and T3 in Shen & Lin (1990) and the present study.....	89

Table 5-2. The first and last points and turning points (bold) along tone contours as well as the occurrence time in the T1-T4 continuum.....	91
Table 5-3. Intercepts and slopes for English listeners with and without musical training.....	93

LIST OF FIGURES

Fig. 2-1. Classic patterns of categorical perception (Figure 1) and non-categorical perception (Figure 2) (adapted from Strange, 1995).....	8
Fig. 3-1. Intensity and pitch envelopes of t1 and t8: (A) in T1 (left) and T4 (right), (B) in T2 (left) and T3 (right).....	37
Fig. 3-2. Eight tokens (t1, t2, t3, etc.) for the syllable <i>shi</i> in the two continua: (A) in T1-T4, (B) in T2-T3.....	40
Fig. 3-3. Correspondence of Mandarin tonal variants on the T1-T4 continuum to Thai tones.....	42
Fig. 3-4. Correspondence of Mandarin tonal variants on the T2-T3 continuum to Thai tones.....	44
Fig. 4-1. Identification curves of the Mandarin, Thai, and English groups: (A) on the T1-T4 continuum; and (B) on the T2-T3 continuum.....	59
Fig. 4-2. Two-step discrimination curves for the pooled English group and subgroup of participants with and without musical training: (A) on the T1-T4 continuum and (B) on the T2-T3 continuum.....	68
Fig. 4-3. Two-step discrimination curves for the Mandarin, Thai, and English groups in the two continua: (A) in T1-T4 continuum, (B) in T2-T3.....	70
Fig. 4-4. Discrimination curves for trained and untrained subjects in Mandarin (Panel A, M refers to Mandarin), English (Panel B, E refers to English), and Thai (Panel C, T refers to Thai) groups: (A) on the T1-T4 continuum; (B) on the T2-T3 continuum.....	76
Fig. 5-1. Identification curves of the Mandarin, Thai, and English groups on the T1-T4 continuum.....	91

ACKNOWLEDGEMENTS

I would like to thank all the participants in my two studies, who contributed their invaluable time to this research. Without their generosity, the completion of this study would have been impossible.

I would like to thank my supervisory committee members. To my supervisor, Dr. Hua Lin, I would like to extend my deepest gratitude for her guidance and stimulating encouragement throughout my Master's program. I am also extremely thankful to Dr. Sonya Bird and Dr. John Esling for their help, especially in the field of phonetics. Thanks also to my outside member, Dr. Richard King and my external examiner, Dr. Daniel Bryant, for their valuable comments.

I would like to thank Dr. Alexander Francis, Dr. Pierre Hallé, Professor Jackson Gandour, Dr. Yue Wang, Dr. Yi Xu, Dr. Mary Lesperance and Mr. David Popham, who shared with me their expertise in linguistics and statistics.

I would like to thank my Thai consultants from Bangkok and Victoria: Niramis, Huiwen, Chev, Thip and Mrs. Dearden, for giving me so much help.

I would like to thank all the people in the Linguistics Department. My special thanks to Dr. Leslie Saxon for being my listener and mentor. I am also very thankful to Allison, Qian, Jun, Yuko, Kyoko, Claire, Carmen, Janet, Aliko, Ryan, Rachel and Shadiya for always being supportive.

DEDICATION

To my parents

Chapter One

INTRODUCTION

1.1 Purpose of this study

It is well known that linguistic experience significantly influences speech perception. Listeners tend to perform better in perceiving segmental or suprasegmental features that exist in their native languages (Best and Strange, 1992; Kuhl & Iverson, 1995; Hallé et al., 2004; Lee et al., 1996; Gandour, 1983, 1984; Burnham, 1997 etc.). It is also believed that listeners perceive native phonemic contrasts as different categories, while they may perceive non-native phonetic contrasts as a single category (Lieberman, 1957; Liberman et al., 1967; Hallé et al., 2004). In addition, as revealed by Best and Strange (1992), Wang et al. (1999), Wayland and Guion (2003), and Iverson et al. (2005), speech perception is a developmental process: compared to inexperienced listeners, experienced listeners' perception of non-native phones is closer to that of native speakers; by means of learning or training, the experienced listeners have had significant exposure to certain non-native languages or acoustical features.

However, listeners' linguistic experience does not fully account for the perception of non-native contrasts. Some non-native phonetic features can be perceived very well, as

seen in English adult and child listeners' perception of Zulu click consonants (Best et al., 1988). Previous studies also indicate that listeners' perception of some native vowel or tone contrasts is non-categorical (Stevens et al., 1969; Abramson, 1977); that is, listeners are not able to classify these native contrasts into different categories. Moreover, although perceptual training has been found to improve tone language speakers' discrimination of tone changes, it also impedes them from discerning differences between native tonal categories (Wang, 1976). In light of such findings, researchers have proposed that the presence of natural auditory sensitivity and other psychophysical factors mediate the influence of linguistic experience (Holt et al., 2004; Francis et al., 2003; Stevens et al., 1969; Beddor & Gottfried, 1995).

Given these previous research findings, it is meaningful to integrate the influence of linguistic experience and psychophysical factors into one empirical study and examine how perceptual training affects speech perception. Very few studies have investigated how linguistic experience and other factors interact in the perception of lexical tones (Wang, 1976). The present study attempts to fill this gap by comparing the tone perception of native speakers from three different language backgrounds: two tone languages, Mandarin and Thai, and one non-tone language, English. The main study, preceded by a pre-study, is carried out within the framework of categorical perception, which includes an AXB identification experiment as well as an AX discrimination experiment. Two synthesized tone continua with two of the four Mandarin tones as the endpoint tones in each continuum are used as stimuli. In order to find out whether tone

perception (categorization) is also a developmental process, the effect of natural tone training is also examined by comparing the performance of trained and untrained listeners from each language group in the main study. The pre-study aims to determine the correspondence between the Mandarin and Thai tones

1.2 Research questions

The goal of this study is to investigate factors that affect tone perception (categorization). According to previous studies, these factors include language background, innate auditory sensitivity, and learning or training experience. Musical background also affects tone perception, as revealed by Gottfried et al. (2000, 2004) and Alexandra et al. (2005) (see Chapter 2). The research questions focus on three areas: (1) the role of language background in tone perception (categorization), particularly the role of tone language experience; (2) the relationship between tone contrasts and regions of natural auditory sensitivity; and (3) the difference between the performance of trained and untrained subjects. The impact of musical experience will be discussed only as necessary; given that this study does not strictly control this variable. The research questions are:

- Is the perception of any native tone contrast categorical for Mandarin listeners?
- Given their exposure to tone in their first language (L1), do Thai listeners have an advantage in categorizing Mandarin tones compared to English listeners?
- Does the lack of tone distinctions in English render English listeners “deaf” to Mandarin tones?

- Does tone training affect Mandarin listeners' categorization of native tone contrasts?
- To what extent does tone training affect Thai and English listeners' categorization of Mandarin tones?
- Is auditory sensitivity salient between some tone contrasts but not between others? If so, does this factor affect listeners' categorization?
- Closely related to the previous question, does the presence of auditory sensitivity have different effects on the categorization of Mandarin tones for native versus non-native speakers, for tone language speakers versus non-tone language speakers, and for Mandarin versus Thai speakers?
- Are listeners with a musical background better at perceiving (categorizing) Mandarin tones than listeners without this background?

Two studies, the pre-study (Study I) and the main study (Study II) are carried out in order to answer the above questions. Study I is an experiment in which Thai speakers map Mandarin tones to their native Thai tones. The correspondence between the Mandarin and Thai tones is expected to help account for the performance of the Thai listeners in Study II, which is designed to address the above-mentioned research questions.

1.3 Outline

This thesis is organized as follows: Chapter 1 introduces the purpose of the study

and the research questions. Chapter 2 provides background information on previous studies of speech perception, tone perception in particular. Chapters 3 and 4 report, respectively, the results of Study I on the correspondence between Mandarin and Thai tones as perceived by Thai speakers, and the results of Study II on the identification and discrimination of tone continua. Chapter 5 discusses the findings of Studies I and II. Chapter 6 outlines the conclusion of this thesis; identifies the limitations of this study; and provides suggestions for future research.

Chapter Two

LITERATURE REVIEW

Categorical perception has interested researchers since the 1950s (Liberman et al., 1957). Within this topic, one of the main questions is whether or not language experience is a necessary condition for the categorical perception of speech (Diehl et al., 2004). This chapter reviews previous studies representing two different points of view on this issue. The first section of this chapter provides a general introduction to the topic of categorical perception. The second section focuses on categorical perception of segments. The third section focuses on categorical perception of tone; also included in this section is a description of cross-linguistic studies on the effect of linguistic experience on tone perception, many of which have been conducted outside the framework of categorical perception. The last section summarizes the list of remaining issues, which are addressed in the current study.

2.1 Categorical perception

Categorical perception, as defined by Gandour (1978: 59), “refers to the phenomenon whereby small steps along an acoustic continuum will produce perceivable differences when they occur between phonetic categories, but not when they occur within a phonetic category” (see also Liberman, 1957; Liberman et al., 1967). Categorical perception contrasts with continuous (non-categorical) perception, in which there is no significant between-category discrimination.

In a typical categorical perception experiment, a series of synthetic phones varying in an acoustic parameter and ranging perceptually across two or more base phones are presented to listeners for phonemic labeling or identification, and for discrimination of pairs of stimuli located near each other in the series (Diehl et al., 2004: 155). For perception to be considered to be categorical, two patterns must occur: (1) the identification curve should show a steep slope at a category boundary; (2) the discrimination level should be close to chance for stimulus pairs within a phonemic category, but nearly perfect for stimulus pairs that straddle a category boundary (see Fig. 2-1 (Figure 1, Strange, 1995)). If one or neither of these perceptual patterns is seen, continuous perception is said to occur (see Fig. 2-1 (Figure 2, Strange, 1995))

Fig. 2-1. Classic patterns of categorical perception (Figure 1) and non-categorical perception (Figure 2) (adapted from Strange, 1995).

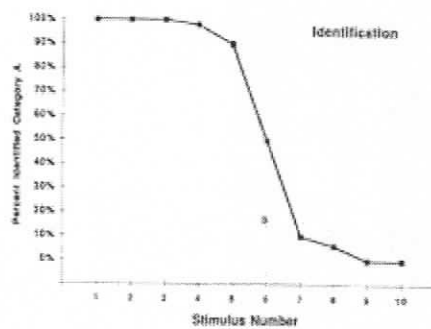


Fig.1 Categorical perception of consonant contrasts. Identification function (above) shows phoneme boundary (50%) crossover at #6; relative discrimination function (below) shows poor within-category discrimination.

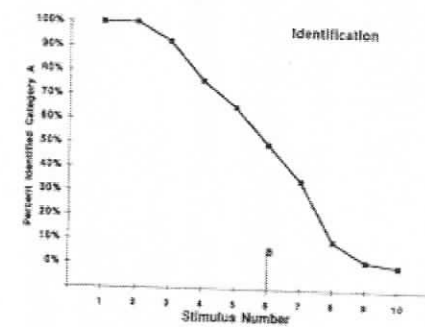


Fig. 2 Continuous perception of (steady-state) vowel contrasts. Identification function (above) shows phoneme boundary (50%) crossover at #6; relative discrimination function (below) shows good within-category discrimination.

2.2 Categorical perception of segments

2.2.1 Language experience effects on perception

Early studies on categorical perception were mainly based on the discrimination and identification of voice distinctions in syllable-initial stop consonants, for example, /ba/ versus /pa/, /da/ versus /ta/ and /ga/ versus /ka/. Lisker and Abramson (1964) identified a key phonetic correlate of voice contrasts—voice onset time (VOT), that is, the interval

between the release of the articulatory occlusion and the onset of voicing.

Cross-linguistically, initial stops tend to occur in one of three ranges of VOT values: long negative VOTs (voicing onset leads the articulatory release by 50ms or more); short positive VOTs (voicing lags behind the release by no more than 20ms); and long positive VOTs (voicing onset lags behind the release by more than 25ms). Lisker and Abramson (1970) and Abramson and Lisker (1970) next examined VOT perception among native speakers of English, Spanish, and Thai; Spanish and English have two-way VOT contrasts while Thai has three-way contrasts. All three language groups showed clear evidence of categorical perception. However, the locations of phoneme boundaries and the associated peaks in discriminability varied among the groups, reflecting differences in the way each language realizes voice distinctions. The English and Spanish groups categorized the stimuli as either voiced or voiceless, with a boundary between categories at +20 to +40 VOT for English and at -5 VOT for Spanish. Thai speakers, however, categorized the stimuli into three categories (voiced, voiceless unaspirated, and voiceless aspirated), with boundaries at about -20 VOT and +40 VOT, respectively. These results suggested that categorical perception of VOT arises from language experience, with listeners becoming more sensitive to phonetic differences that play a functional role in their language and/or less sensitive to differences that do not.

In another study of categorical perception, Miyawaki et al. (1975) explored the /r/ and /l/ distinction to test the effect of linguistic experience on the perception of a cue that is known to be effective in distinguishing between /r/ and /l/ in English. Twenty-one

Japanese and thirty-nine American adults were recruited for a test on the discrimination of a set of synthetic speech-like stimuli. The 13 stimuli in this set varied in the initial stationary frequency of the third formant (F3) and its subsequent transition into the vowel over a range sufficient to produce the perception of /ra/ and /la/ for American subjects and to produce /ra/ for Japanese subjects (/r/ is not in phonemic contrast to /l/ in Japanese). For Americans, the discrimination of the speech stimuli was nearly categorical, i.e. comparison pairs which were identified as different phonemes were discriminated with high frequency, while pairs which were identified as the same phonemes were discriminated relatively poorly. In comparison, discrimination of these synthesized stimuli by Japanese subjects was only slightly better than chance for all comparison pairs; thus the perception of the English /r/ and /l/ contrast was non-categorical for the Japanese subjects. These results suggest that linguistic experience accounts for the categorical perception of the /r/ and /l/ contrast among the English listeners.

In their study on English approximants, Best and Strange (1992) examined how non-native phones are perceptually assimilated into native phonetic categories. They tested the identification and discrimination of three synthetic American English approximant contrasts, /w-j/, /w-r/, and /r-l/ in syllable-initial position by American English speakers and Japanese learners of English. The English approximants differ more or less with respect to their phonemic status in Japanese, as well as in the phonetic details of the most similar Japanese phonemes. Realizations of /j/ are quite similar in the two languages; the English /w/ is realized with lip-rounding whereas in Japanese, /w/ is

produced with spread lips, similar to the back unrounded Japanese vowel /ɯ/ (Vance, 1987); American English /r/ is a retroflex or palato-alveolar central approximant while Japanese is usually an alveolar tap /ɾ/ rather than an approximant; English /l/ is an alveolar lateral approximant while Japanese does not employ a distinct /l/ phoneme (Vance, 1987). Three 10-step continua of /wak/-/jak/, /wak/-/rak/, and /rak/-/lak/, were created by interpolating between the first base phone and the tenth base phone on F1, F2, and F3 onset frequencies. For American subjects, discrimination and identification of the three continua exhibited categorical perception. For Japanese subjects, the identification and discrimination of the /w-j/ and /w-r/ contrasts also showed categorical perception, while the perception of the /r-l/ contrast was non-categorical. These results further support the idea of language-specific influences in the perception of English approximant contrasts. Moreover, this study also evaluated differences between subgroups of the Japanese subjects, who had two different levels of English experience. Japanese subjects with intensive English conversation experience showed identification and discrimination patterns that were more similar to the Americans' than those with little English experience.

2.2.2 Language-independent effects on perception

Stevens et al. (1969) examined the discrimination and identification of synthetic vowels by speakers of two languages, English and Swedish. Two series of vowels were generated, each of which varied in F1/F2/F3 frequency over a range that encompassed

three vowel categories. One series contrasted three front unrounded vowels that are phonologically distinct in both languages. For the Swedish listeners, the three vowels corresponded to /i/, /e/, and /ä/, but for the English listeners, they corresponded to /i/, /ɪ/, and /ɛ/; the other series /i/, /y/, and /ʌ/ included front vowels that are phonemic in Swedish, but not in English. The results indicated that the ability of the subjects to discriminate between the vowels is relatively independent of their linguistic experience: Swedish and American English subjects exhibited similar performance in the discrimination tests, though they had somewhat different results in the identification tests. Both groups showed continuous perception along the entire range of both continua. These results contrasted with the cross-language studies of consonant contrasts in that the discrimination of vowels by American English and Swedish listeners showed no language-specific influence, suggesting that some natural perceptual boundaries exist between vowel categories.

Later studies found that language experience is not a necessary condition for the categorical discrimination of VOT stimuli. Eimas et al. (1971) presented the synthetic 6-step VOT continuum between the voiced and voiceless stop consonants /b/ and /p/, as prepared by Lisker and Abramson, to 1- and 4- month old infants from an English-speaking environment. The six stimuli had VOT values of -10, 0, +20, +40, +60, and +80msec. This study found that these infants were able to discriminate differences in VOT for stimulus pairs that straddle the English /ba/-/pa/ boundary, but they were not able to discriminate equivalent VOT differences when the stimuli were from within the

same English category. As pointed out by the authors, one of the implications of this study is that the means by which the categorical perception of speech is accomplished may well be part of the biological makeup of the organism; and moreover, that these means must be operative at an unexpectedly early age. Further supporting this view, Lasky et al. (1975) found that infants raised in a Spanish-speaking environment can discriminate differences in VOT if the stimuli are located around either the Spanish or the English voice boundary, but otherwise show no evidence of discriminative ability.

In line with the evidence presented above, Kuhl and Miller (1975) have proposed that there are natural boundaries in auditory perceptual space, across which discrimination is enhanced. In this study, four chinchillas were first trained to respond differently to /t/ and /d/ consonant-vowel syllables produced by four English native speakers in three vowel contexts: /a/, /u/, and /i/. This training then generalized to novel instances, including the synthesized syllables /da/ and /ta/ (VOT of 0 and 80 ms, respectively). In a subsequent experiment, nine synthetic stimuli with VOT between 0 and +80ms were presented for identification. The form of the labeling functions and the “phonetic boundaries” for chinchillas and English-speaking adults were similar. To test the generality of the agreement, Kuhl and Miller (1977) presented the chinchillas with synthetic stimuli that had bilabial (/b-p/) and velar (/g-k/) places of articulation. No significant distinctions between the English speakers and chinchillas on the values of phonetic boundaries were obtained. These boundaries or “discontinuities,” as suggested by Holt et al. (2004) tend to coincide with regions of natural auditory sensitivity, thereby

enhancing the distinctiveness, intelligibility, and learnability of linguistic contrasts. This theoretical position places into question the influence of linguistic experience in the categorical perception of VOT.

The results of the aforementioned studies conflict in terms of the influence of linguistic experience on the categorical perception of segments. Segmental perception seems to be language-specific in some cases, but not in others. More research is needed to determine exactly which components of speech lead to categorical perception.

2.3 Categorical perception of tone

Up to this point, the discussion has focused on categorical perception of segments. In this section, tone perception will be discussed.

2.3.1 Tone languages

Languages of the world can be classified into tone or non-tone languages, depending on whether or not they use pitch to distinguish word meanings. Tone languages, which employ pitch at the lexical level, are widely spoken around the world; examples of tone languages are Mandarin Chinese and Thai in Asia, Yoruba in West Africa, Swedish and Norwegian in Europe, and Navajo and Dogrib in America (Gandour, 1978; Krauss, 1979). In previous studies, Mandarin, Thai, and Cantonese have been the most studied tone languages; in studies of tone perception, these languages are often compared to English, a non-tone language.

Pike (1948) categorized tones into two types, level and contour. This division reveals differences between tones from the perspective of fundamental frequency (F_0), the primary feature of tone (Lin, 1988; Abramson, 1978; Howe, 1976). In the perception of a level tone, no rise or fall in pitch can be perceived. By contrast, contour tones contain a rise or a fall or a combination of rises and falls. For instance, Mandarin has one level tone and three contour tones. Following Chao's 5-pitch level scale, the four Mandarin tones are represented as 55 for high level tone (T1), 35 for rising tone (T2), 214 for falling-rising tone (T3), and 51 for falling tone (T4). Using this same system, the five Thai tones may be represented as 32 for mid tone, 21 for low tone, 51 for falling tone, 45 for high tone, and 214 for rising tone (Ladefoged, 2001; see also Chapter 3 for a comparison of the Mandarin and Thai tone systems). Compared to the Mandarin and Thai tone systems, Cantonese has more tonal patterns. The six Cantonese tones are 55 for high level tone, 25 for high rising tone, 33 for mid level tone, 21 for low falling tone, 23 for low rising tone, and 22 for low level tone (adapted from Francis et al., 2003). However, F_0 height and direction (Gandour, 1983) are not the only features of tone; other acoustical correlates of tones include duration and amplitude (see section 3.2.1 in Chapter 3). The following sections review previous studies on tone perception by native and non-native speakers, both from within and outside the framework of categorical perception.

2.3.2 Perception of tone continua

2.3.2.1 Cross-language studies

Although very few studies have examined categorical perception of tone, existing studies have nonetheless produced interesting findings on the factors that influence tone perception. This section reviews studies on categorical perception of tone, including research on the effects of linguistic experience and perceptual training. Studies by Hallé et al. (2004) and Wang (1976) have highlighted differences in tone perception by native speakers of tone versus non-tone languages. Stager and Downs (1993) have argued that native tone experience decreases listeners' sensitivity to absolute frequency differences that correspond to within-category differences in tone. Zue (1976) and Blicher et al.'s (1990) study, however, found similar patterns for tone and non-tone language speakers in the identification and discrimination of Mandarin tones.

Hallé et al. (2004) tested native Taiwanese Mandarin and French listeners on three tone continua derived from natural Mandarin utterances within carrier sentences. Classic assessments of identification and discrimination of each tone continuum were conducted with both groups of listeners. In one experiment, Taiwanese listeners labeled the tones of target syllables within the context of carrier sentences and discriminated tones presented as single syllables. In a second experiment, both French and Taiwanese listeners completed an AXB identification task based on the presentation of single isolated syllables. Finally, in a third experiment, French listeners performed an AXB discrimination task. Overall, the results indicated that Taiwanese listeners' perception of

tones is quasi-categorical, whereas that of French listeners is psychophysically based. In other words, French listeners perceived the three Mandarin tone continua non-categorically. Nevertheless, French listeners showed considerable sensitivity to tone contour differences, though to a lesser extent than Taiwanese listeners. Thus, the findings suggest that despite the lack of lexical tone contrasts in the French language, French listeners are not absolutely “deaf” to tonal variations.

Wang (1976) reported the results of an experiment on the categorical perception of Mandarin tones. Eleven synthetic tonal variants from rising tone (T2) to high level tone (T1), ranging from 105 Hz to 135 Hz in steps of 3 Hz, were superimposed on the vowel /i/. The stimuli were presented to speakers of Mandarin and English in a two-step ABX discrimination test and in a labeling identification test. The Mandarin speakers varied in the degree to which they were influenced by their native language. The untrained subjects who had never participated in any psychophysical experiments showed a strong identification boundary between rising and level tones at about the mid-point of the continuum, with a corresponding peak in discrimination accuracy. The native speakers who had previous experience with psychophysical experiments showed comparably heightened discrimination accuracy at the center of the continuum, but their discrimination accuracy remained relatively high toward the level end of the continuum; that is, their within-category discrimination was high. The highly trained speakers (the experimenters) showed no peak in discrimination across the category boundary, but still showed a high degree of discrimination accuracy toward the level end of the continuum

in a manner similar to that exhibited by native speakers of English unaccustomed to hearing Mandarin. In other words, both highly trained native Mandarin speakers and English speakers responded only to psychophysical differences between all eleven tonal variants.

Stagray and Downs (1993) hypothesized that native speakers of tone languages would have poorer discriminatory sensitivity than non-tone language speakers because the former group had learned to categorize sounds of similar frequency together to facilitate their perception of tone phonemes. During their experiment, Mandarin and English listeners were asked to judge whether tonal variants at increments within the frequency range of a level tone phoneme category sounded the same or different in pitch than standard tones at 1000 Hz and 125 Hz. Psychometric functions at both frequencies indicated that, as compared to the English listeners, the Mandarin listeners' difference limens for frequency¹ (DLFs) were significantly larger, and their psychometric functions were poorly shaped. These results suggested greater perceptual confusion among tones of similar frequency for Mandarin speakers. According to the authors, Mandarin listeners exhibited poorer differential sensitivity for frequency because they perceived standard and variable tones as being within the same pitch range of a learned, level tone phoneme category. In other words, they had learned to categorize sounds of similar frequency

¹ Difference limen (DL) refers to the minimal increment in a stimulus needed to produce a just-noticeable difference in sensation. Difference limen for frequency (DLF) is the just-noticeable difference in the frequency of a signal or the smallest detectable difference between two frequencies (Web link: <http://e-viataal.org/DEFG.html>). The larger the DLF, the poorer the listeners' sensitivity to frequency is.

together to enhance their perception of tones. These findings suggested that listeners can differentially tune their auditory systems to certain physical properties of a sound as a function of their linguistic experience.

By contrast with the studies described above, two cross-linguistic studies investigating the categorization of Mandarin rising tone (T2) and falling rising tone (T3) conducted by Blicher et al. (1990) and Zue (1976) (see also Gandour, 1978; Shen & Lin, 1990 and Blicher et al., 1990) concluded that the categorization of Mandarin T2 and T3 is largely independent of specific linguistic experience.

Blicher et al. (1990) tested the hypothesis that talkers use length differences to enhance the perceptual contrast between Mandarin T2 and T3, with syllables carrying T2 tending to be shorter than those carrying T3. In their experiment, a short stimulus series (350ms) and a long stimulus series (450ms) was synthesized for each of the syllables /bi/, /ba/, and /bu/ and presented to native Mandarin and English speakers. Each series included 11 stimuli that varied incrementally in F_0 contour from a T2 exemplar to a T3 exemplar, both of which were Mandarin morphemes. Subjects (both Mandarin- and English-speaking) were first trained with feedback to assign the short and long series-endpoint stimuli to two categories based on F_0 contour alone. Next, subjects were asked to identify the entire stimulus series (both short and long) on the basis of the two training categories. For both Mandarin- and English-speaking subjects, longer syllable duration reliably shifted the labeling boundary toward more T3 responses. Therefore, syllable lengthening appears to auditorily enhance F_0 cues for the T3 category. The

comparable performance of Mandarin- and English-speaking subjects suggests that linguistic experience had little effect on the categorization of Mandarin tones.

Zue (1976) researched the categorical perception of Mandarin tones using an acoustic continuum consisting of nine tonal variants intended to span the continuum between T2 and T3. All nine tonal variants, superimposed on the synthetic syllable /pao/, displayed a linear rising fundamental frequency contour (100 to 160 Hz). The tokens differed in the duration of the beginning steady-state portion (0-400msec, 50msec steps) up to the point of the rise. With this stimulus continuum, both Mandarin and English subjects displayed a sharp category boundary at tonal variant 5, and a peak in discrimination at the category boundary. This result suggests that both groups of subjects simply divided the psychophysical continuum into a “predominantly rising” group of stimuli (tone variants 1-4) and a “predominantly level” group of stimuli (tonal variants 6-9). Accordingly, this study concluded that linguistic experience is not a necessary condition for the categorical perception of tone.

In this section, some studies show tone categorization is language-specific while other studies do not. Some studies argue that linguistic influence emerges only if the tone falls into native F_0 ranges. The results of some studies indicate that training impedes listeners' categorization of tone contrasts. Furthermore, studies on the perception of Mandarin T2 and T3 suggest that the duration of tones and the occurrence of the turning point where a tone moves from one to the other direction are crucial indices for listeners to differentiate the two tones, but are independent of language experience.

2.3.2.2 L1 studies

The question of language experience is controversial not only in cross-linguistic studies of tone perception, but also in perceptual studies of native speakers of tone languages. Some research shows that native speakers of tone languages do not perceive all tonal contrasts categorically, thus suggesting that language experience has only a weak effect on tone perception, as argued by Abramson (1977) and Francis et al. (2003). Other research demonstrates that specific perceptual cues based on linguistic experience affect the categorical perception of tone (Shen and Lin, 1990).

Abramson (1977) tested 33 native Thai speakers with 16 synthetic tone variants in level F_0 trajectories, ranging from 92 Hz to 152 Hz, in steps of 4 Hz superimposed on the syllable /k²aa/. These 16 stimuli were intended to span the range of high, mid, and low level tones. The results of discrimination experiments showed that the discrimination peaks were not located in the vicinity of the identification boundaries; that is, the perception of level tones by native Thai speakers was not categorical.

Francis et al. (2003) conducted an elaborate study on the categorical perception of lexical tones in Cantonese. Six experiments were conducted in this study. The first three were designed to investigate the perception of level and contour tone continua derived from Cantonese, including a low- to mid- to high-level continuum; a high-level to low-rising to high-rising continuum; and a low-falling to low-rising to high-rising continuum. Based on the results of these experiments, three further experiments were conducted to investigate the effects of external context on the categorical perception of tones. Each continuum included 10 tonal variants imposed on the target syllable /ji/. Three groups of native Cantonese speakers participated in a labeling identification task and an AX (same or different) discrimination task; each group was tested on their perception of a given continuum. The first group, which was assigned to perceive the level tone contour, showed no

peak in discrimination accuracy at the location of the identification boundary or at any other point along the discrimination continuum. The second group, which was assigned to perceive the high-level to high-rising to low-rising continuum, showed obvious sensitivity at a peak across a pair of tonal variants that span a category boundary, as determined by an identification task. The third group, which was also assigned to perceive a contour tone continuum (from falling to low-rising to high-rising), was more sensitive to the falling to low-rising contour transition than to the low-rising to high-rising contour. In the second of three experiments, stimuli were embedded in a sentential context. In this context, listeners showed sharper category boundaries with level tones than in the citation context, although there was still no evidence for the hypothesis that level tones are perceived categorically. The results of this study suggest that the categorization of lexical tones is influenced mainly by psychophysical factors, i.e. the auditory sensitivity that coincides with the category boundary (Klatt, 1973; Holt et al., 2004). Language experience thus failed to account for tone categorization.

Shen and Lin (1990) investigated the perceptual cues that native speakers employ to discriminate Mandarin T2 and T3. Two fundamental frequency continua were superimposed on the Mandarin syllable /wu/: on the first continuum, F_0 went from 190 to 160 to 250 Hz; and on the second continuum, F_0 went from 190 to 175 to 265 Hz. These endpoints approximated acoustic measurements for T2 and T3 in real speech. The difference between the two continua lay in the extent of the initial fall, which was 30 Hz for the first continuum and 15 Hz for the second continuum. The results of this experiment demonstrated that the distinction between T2 and T3 is cued by the timing of the turning point, which is correlated with the degree of the initial fall. In a second experiment, 10 minimum pairs of T2 and T3 spoken by four native speakers were presented to Mandarin listeners for labeling. Acoustic analysis of the labeled tokens indicated that

mistakes occurred when the correlation between the timing of the turning point and the degree of the initial fall was violated. In other words, native speakers categorized T2 and T3 following the threshold of the degree of the initial fall and the timing of the turning point as a function of their language experience.

The studies described above were all conducted within the framework of categorical perception: listeners were presented with tone series for discrimination and/or identification. The conflicting results reported illustrate the controversy over the role that language experience plays in categorizing tone contrasts. To provide a wider background for the present study, in the following section, some related cross-linguistic studies of tone perception conducted outside the framework of categorical perception are discussed.

2.3.3 Other cross-language studies on tone perception

It is well known among linguists that the majority of the world's languages are tone languages (Fromkin, 1978). Unlike non-tone languages, tone languages use pitch change to differentiate meanings at the lexical level. With regard to the influence of language experience, one question related to L2 tone perception emerges: compared to speakers of non-tone languages, do speakers of a tone language perform better or worse in discriminating and identifying the tone system of a different tone language? Gandour (1983) investigated the perceptual dimensions of tone and the effect of linguistic experience on a listener's perception of tone. Wayland and Guion (2004) concluded that L1 tone experience is transferable to the perception of an L2 tone system; Lee et al. (1996) found better performance for one tone language group, but not for another one, relative to listeners in a non-tone language group. Finally, Wang et al. (2004) examined speakers of two different tone languages, bilingual speakers of tone and non-tone languages and speakers of a

non-tone language and demonstrated the role of linguistic experience in the hemispheric processing of lexical tone. These studies are reviewed in more detail below.

Gandour (1983) examined the perception of tones by English speakers and speakers of four Asian tone languages (Mandarin, Cantonese, Taiwanese, and Thai) to identify the perceptual dimensions that speakers of tone and non-tone languages employ when perceiving tones. The stimuli consisted of 19 different fundamental frequency trajectories (five level, four rising, four falling, three falling-rising, three rising-falling) superimposed on a monosyllable /wa/. The multidimensional scaling model INSAL (individual differences scaling) was applied to assess the importance of the perceptual dimensions of F_0 height and direction. Gandour found that the perceptual weight of these two dimensions was related to the listeners' linguistic experience. In particular, the dimension of direction was more important to speakers of the tone languages than to speakers of English; and more important to Thai listeners than to Chinese listeners. For English listeners, more importance was attached to F_0 height than to the pitch direction of the stimuli. These results suggested that listeners' tone perception strategies depend to some extent on the linguistic function of pitch in their native languages.

Wayland and Guion (2004) investigated the benefit of native tone experience in the perception of another tone system by presenting Thai tone stimuli to native English and native Chinese listeners. Results of the perception tests for the two groups were examined before and after a short training period in relation to two variables: first, language background and second, the inter-stimulus interval (ISI) of the presentation (500ms versus 1500ms). The native Chinese group outperformed the native English group in its ability to discriminate two Thai tones: mid tone and low tone under the ISI 500ms condition before training; and under both ISI conditions after training. A significant improvement in identification from the pretest to the posttest was observed

in the native Chinese group under both ISI conditions, but not in the native English group. These results suggested that prior experience with the tone system of one language may be transferable to the perception of tone in another language.

Lee et al. (1996) investigated this same question by comparing tone perception among three groups of subjects: speakers of two tone languages, Cantonese and Mandarin, and speakers of a non-tone language, English. All the subjects were presented with natural tones in Cantonese and Mandarin, superimposed on word and non-word syllables in two different experiments. The interval between the presentation of stimuli, and the level of interference during intervals, were manipulated. The level of interference increased in three conditions: the immediate condition, the delay condition, and the counting condition. The first two conditions differ in the setting of the interstimulus interval. For the intermediate condition, the interstimulus interval was close to zero, while for the "delay" condition, the ISI was set to 5 seconds. For the "counting" condition, the ISI was also 5 seconds. In contrast to the "delay" condition, during this 5-s interval, subjects were required to count backward by repeatedly subtracting 3 from a two-digit number that appeared on the center of the computer screen. The results of AX (same/different) discrimination tests showed that Cantonese speakers were better than Mandarin and English speakers at discriminating Cantonese tones; and better than English speakers at discriminating Mandarin tones. However, although Mandarin speakers were better than Cantonese and English speakers at discriminating Mandarin tones, they were no better than English speakers at discriminating Cantonese tones. The authors concluded that speakers of tone languages have an advantage over speakers of non-tone languages in perceiving lexical tone systems. They also argued that "this advantage came not simply from the lexical content of tones, but also from familiarity with the tones themselves regardless of whether or not the tones had corresponding lexical entries" (p.539). Thus, listeners

seemed to acquire general abilities of tone discrimination. The authors also suggested that the advantage might be neutralized when perceiving a more complicated tone system.

Relative to the two studies discussed above, Wang et al. (2004) provide a different account of the advantage that native speakers of a certain tone languages may have in tone perception, but they also suggest that the advantage does not extend to speakers of another tone language. Their investigation compared hemisphere lateralization of Mandarin tones for four groups of listeners: native Mandarin listeners, English-Mandarin bilinguals, Norwegian listeners with L1 tone experience, and American English listeners with no tone experience. Tone pairs were dichotically presented and listeners identified which tone they heard in each ear. For the Mandarin listeners, 57% of the tonal errors occurred in the left ear, indicating a right-ear (left-hemisphere) advantage. The English-Mandarin bilinguals exhibited native-like patterns, with 56% left-ear errors. However, no ear advantage was found for the Norwegian or American listeners (48% and 47% left-ear errors, respectively), a pattern which had also occurred when native speakers of Thai were tested on hummed tones and when non-native speakers were musically trained (Van Lancker & Fromkin, 1973, 1978). These results indicated left-hemisphere dominance for the perception of Mandarin tone by native and proficient bilingual listeners, but no evidence of lateralization among non-native listeners, regardless of their familiarity with lexical tone. Thus, to some extent, this study cast doubt on the advantage that speakers of tone languages have when perceiving other tone systems. Results of the studies reviewed in this section indicate discrepancy on the influence of a tone language background on the perception of a non-native tone system. It would be interesting to continue the investigation of the effect of tone language experience within the framework of categorical perception.

Tone perception is also believed to be affected by a listener's musical background. Gottfried et al. (2000, 2004) found that English listeners with advanced musical training outperformed those with less or no musical training, both in discriminating and labeling Mandarin tones. Alexander et al. (2005) proposed that musical experience was transferable to the perception of lexical tones. Their findings are similar to those exhibited in Gottfried et al., as presented above.

2.4 Summary

As demonstrated above, previous research is inconclusive on the question of whether categorical perception at the segmental and suprasegmental levels is dependent on language experience. The conflicting results of the studies discussed above suggest that categorical perception is essentially a biological process as opposed to a developmental process. If categorical perception is a biological process, natural auditory mechanisms produce speech categorization for native and non-native listeners, and even for non-human listeners. However, if categorical perception is a developmental process, acquired L1 experience for native speakers and learned L2 experience for non-native speakers play a crucial role in categorizing L1 or L2 phonetic features and patterns.

With regards to tone categorization, as noted above, the effect of perceptual cues and L1 tone experience have produced conflicting results. With respect to perceptual cues, despite the fact that tone categorization studies have widely employed stimuli with equalized length, it has been argued that the duration of tone plays an important role in categorizing tones, especially the rising (T2) and falling rising tone (T3) in Mandarin. With respect to L1 tone experience, although some researchers agree that L1 transfer occurs when tone stimuli are discriminated by native listeners of a different tone language, other researchers have proposed the opposite position: tone experience

in one tone language does not significantly affect the discrimination of tones in another phonological tone system. Beyond these two issues, other questions relevant to tone categorization include the role of perceptual training, prior musical experience, and the role of the timing of the turning point in Mandarin T2 and T3 (a perceptual cue). With respect to the latter issues, further research is required to determine: (1) whether perceptual training improves or impedes tone categorization; (2) whether prior musical experience improves the categorization of synthesized tones, along the lines seen with natural tones; (3) whether the role of the timing of the turning point is as critical as suggested in Shen and Lin (1990, see above). The present study is designed to address these issues.

Chapter Three

STUDY I: CORRESPONDENCE BETWEEN MANDARIN TONES AND THAI TONES

3.1 Introduction

Study I is an experimental study that examines the correspondence between Mandarin and Thai tones as perceived by native Thai speakers. Some Mandarin tones correspond to Thai tones in terms of their F_0 contours and height. For instance, Mandarin T4 moves from a high to a low level in the same way as the Thai falling tone (both would be characterized as 51 in Chao's description). This correspondence may shed light on the shared tonal experience of Thai and Chinese speakers and is useful in the generation of predictions and/or hypothesis as to how Thai speakers categorize Mandarin tones—a major question that is dealt with in Study II, the main study of this thesis (see Chapter four).

3.2 Description of Mandarin and Thai tone systems

3.2.1 Mandarin tone

Lexical tones are classified as level and contour (static and dynamic tones respectively, in Abramson, 1978) according to their fundamental frequency (F_0). With level tones, only F_0 height

differences can be perceived. By contrast, in contour tones, a rise or a fall or a combination of rises and falls is perceivable. In Mandarin, each syllable carries a tone that serves to differentiate word meanings. There are four distinctive tones in Mandarin: a high-level tone (T1), a rising tone (T2), a falling-rising tone (T3), and a falling tone (T4). Some tone researchers argue that T3 is virtually a low-level rising tone (Lin, 1998, 2001). Following Chao's 5-pitch level scale (Chao, 1948, 1968), level and contour tones in Mandarin are depicted as 55 (T1), 35 (T2), 214 (T3), and 51 (T4), with "5" representing the highest pitch and "1" representing the lowest pitch. One tonal variation is T3 in a non-final position and not followed by another T3; in this context, T3 is realized as 21 (see Hallé et al., 2004).

The three most important perceptual cues for tone are fundamental frequency (F_0), amplitude, and duration. Nevertheless, tones are characterized primarily by F_0 contours and F_0 height (Abramson, 1974), since more than 90% of a tone's cues are conveyed by fundamental frequency (Lin, 1988; Abramson, 1974). Table 3-1 shows the F_0 values associated with the four Mandarin tones as produced by a male and a female native speaker.

Table 3-1. F_0 values (in Hz) of the four Mandarin tones produced by a male and a female speaker (Wu, 1986, adapted from Wang, 2002).

Tone	Gender	F_0 pattern		
		Beginning point	Turning point	Ending point
Tone 1 (High level)	Male	190	-	177
	Female	307	-	305
Tone 2 (Rising)	Male	124	-	178
	Female	222	-	318
Tone 3 (Falling-rising)	Male	124	68	141
	Female	221	165	242
Tone 4 (Falling)	Male	223	-	80
	Female	351	-	166

3.2.2 Thai tone

Compared to Mandarin, Thai has a wider variety of tones. Traditionally, Thai tones were identified into high, mid, low, falling, and rising tone (Gandour, 1978). According to Abramson (1978), the former three and the latter two were classified as static and dynamic tones, respectively. Gandour and Potisuk (1991) compared the F_0 values of the five Thai tones for old and young male Thai native speakers. The average F_0 of the five tones produced by the “young” male group is given in Table 3-2. The F_0 value of the five Thai tones produced by a female Thai native speaker was measured by Burnham et al. (1997, 40). The mean F_0 is also given in Table 3-2 (the tone description follows the traditional system employed in the above two studies).

Table 3-2. F_0 values (in Hz) of the five Thai tones produced by male and female speakers (Gandour & Potisuk, 1991; Burnham & Francis, 1997).

Tone	Gender	F_0 pattern	
		Beginning point	Ending point
Mid tone	Male (young)	107	102
	Female	198.4	182.8
Low tone	Male (young)	114	92
	Female	173.8	147
Falling tone	Male (young)	138	94
	Female	223	147.6
High Tone	Male (young)	121	122
	Female	188.6	211.8
Rising Tone	Male (young)	110	134
	Female	160.2	184.8

It may not be warranted to propose that Thai tones are generally lower than Mandarin tones, based solely on a comparison of Tables 3-1 and 3-2 (the data were not collected from the same speaker). However, it is still reasonable to assume that Mandarin and Thai tones with the same pitch patterns may not be realized with comparable F_0 values. In other words, the Thai tones may be produced with a lower F_0 than the corresponding Mandarin tones.

Based on the traditional descriptions, Ladefoged (2001) presented a more detailed descriptive system to reflect the phonetic features of the five Thai tones. He applied Chao's 5-pitch level scale to the Thai tones. For instance, Thai falling and mid tones were depicted as 51 and 32, respectively. Thai rising tone was described as a low-falling-rising tone (214) in Ladefoged's system. Chao's and Ladefoged's systems differ in that, with the exception of T1 (high-level tone) Chao did not mark the starting F_0 height for the Mandarin tones, whereas Ladefoged labeled the starting F_0 height for both Mandarin and Thai tones¹ (see Table 3-3).

Table 3-3. Comparison of Mandarin and Thai tones (Chao, 1948; Gandour, 1978; Ladefoged, 2001, 237-238).

Pitch	Mandarin tone		Thai tone		
	Tone number	Chao's description	Tone number	Ladefoged's description	Traditional description (Gandour)
55	1	High level			
35	2	Rising			
214	3	Falling-rising	4	Low falling-rising	Rising
51	4	Falling	2	High falling	Falling
32			0	Mid falling	Mid
21			1	Low falling	Low
45			3	High rising	High

As shown in Table 3-3, there are two contour tones common to Thai and Mandarin: falling tone (51) and falling-rising tone (214), while the Mandarin high-level tone and rising tones have no counterparts, or at any rate no identical counterparts, in Thai. However, it is likely that Thai listeners would still assimilate these two non-native tones to two native tones in their phonology. As pointed out in a series of L1 and L2 acquisition studies, within the Mandarin tone system, T2 and T3 are the most similar (Huang, 2002; Shen & Lin, 1990). Therefore, it is very likely that Thai

¹ My question focuses on Ladefoged's description of Mandarin T2 (35). Since, as Ladefoged acknowledges (Ladefoged 2001, p. 237), pitch 3 on a 5-pitch level scale represents mid-level height, it may be improper to label Mandarin T2 as a high rising tone (see p. 237). The latter would be a better description for Thai tone 3 (45).

speakers would draw a correspondence between Mandarin T2 (MT2) to Thai falling-rising tone [TT4 (Thai Tone 4)]. Compared to MT2 (Mandarin T2), it is more difficult to predict which Thai tone MT1 (Mandarin T1) would correspond to. In terms of F_0 direction, none of the Thai tones have a flat contour. But if F_0 height is taken into consideration, TT3 (high-rising tone) is the closest to MT1. However, as shown in Hallé et al. (2004), Wang et al. (1999), and Xu (1997), in real speech, MT1 bears a falling contour with a shallow slope (also see Fig. 3-2 in this study). In addition, as pointed out by Gandour (1983), Thai speakers attached more importance to F_0 direction than to F_0 height when perceiving tones. Therefore, it is unlikely that MT1 would correspond to TT3 (high-rising tone), given the contrasting F_0 directions of these two tones. It is also unlikely MT1 would be assimilated to TT2 (falling tone) because Mandarin has an identical tone in its system. Therefore, TT0 (mid-falling tone) is the tone most similar to MT1 because it has a falling contour with a shallow slope and its F_0 is higher than TT1 (low-falling tone), although the latter also has a falling contour with a shallow slope. My predictions for the correspondences between the Mandarin and Thai tones are summarized in Table 3-4.

Table 3-4. Predicted counterparts of Mandarin tones in Thai tone system.

Pitch	Mandarin tone		Corresponding Thai tone			
	Tone number	Chao's description	Pitch	Tone number	Ladefoged's description	Traditional description (Gandour)
55	1	High level	32	0	Mid falling	Mid
35	2	Rising	214	4	Low falling-rising	Rising
214	3	Falling-rising	214	4	Low falling-rising	Rising
51	4	Falling	51	2	High falling	Falling

3.3 Method

3.3.1 Subjects

Fifty-one native Thai speakers participated in this study in a language lab at Thammasat University in Bangkok. The participants, who ranged in age from 18-24 years, were university students in Bangkok who had never been exposed to Mandarin or to any other tone languages. None of the participants reported any speech or hearing impairments. They participated voluntarily in this study and received no financial compensation. A trilingual speaker of Thai, Mandarin, and English administered the test on-site.

3.3.2 Stimuli

The stimuli were derived from four naturally produced Mandarin tones superimposed on the target syllable *shi* [ʃə]. *Shi* has the least marked (CV) structure (Major et al., 1996) and can be combined with all four Mandarin tones to form frequently occurring Mandarin words; the latter are considered to be more native in production than words that occur less frequently, as proposed by Levi and Winters (2005). Therefore, it is reasonable to assume that native listeners' performance will be better when they are presented with high-frequency words (Hallé et al., 2004; Wang, 2001). *Shi* was also adopted as a target syllable in previous studies, such as Wang (2002).

Four commonly used Mandarin words: 师 (*shi*1, teacher), 十 (*shi*2, ten), 使 (*shi*3, to let), and 是 (*shi*4, to be) were recorded by a 26-year-old, Beijing-born, female native speaker of Mandarin in a soundproof booth at the Phonetics Laboratory in the Department of Linguistics at the University of Victoria. Each word was recorded five times in a random order.

The 20 recordings (4 syllables × 5 times) were then played to three native Mandarin-speaking judges, two of whom were female graduate students in Linguistics (one is the author); the third

judge was a male graduate student who taught Mandarin pronunciation for more than a year at the University of Victoria. The three judges first labeled the tones for each word and then ranked the naturalness of the tone on each token on a scale of 1 to 4, with “4” indicating the highest degree of naturalness. The three judges’ labeling and rankings were in agreement for over 97% of the recorded words.

The four “best” tones were then selected from those ranked as most natural using the sound-editing software Praat (Boersma & Weenink, 2006). Two perceptual cues,² F_0 contour and duration, were applied to the choice of the best tones. The amplitude cue was not considered here because it appeared consistent and individual for each tone type. Two criteria were considered in selecting the best tones. First, the selected tones needed to clearly exhibit the characteristic F_0 contour of each tone, that is, high-level for Tone 1 (T1), mid-rising for Tone 2 (T2), low falling-rising for Tone 3 (T3), and high-falling for Tone 4 (T4). Nevertheless, some variables may emerge with the naturally produced tone, especially for T1 and T2, whose F_0 contours usually exhibit a slight falling contour (see Table. 3-1) and a falling-rising contour, respectively. Therefore, a turning point may occur to the rising tone (see Table 3-5 below). Second, the selected tones needed to exhibit the contrastive tone duration patterns typical of the four tones, i.e., T3 has the longest duration and T4 has the shortest, while T2 and T1 have intermediate duration, with T2 being longer than T1 (Xu, 1997). Table 3-5 shows the F_0 pattern and duration of the selected tones (a discussion of the turning points for Tone 1 see footnote 3 in this chapter).

² For the purpose of choosing tone 3, tokens were screened for voice quality to avoid choosing tokens of tone 3 with creaky voice.

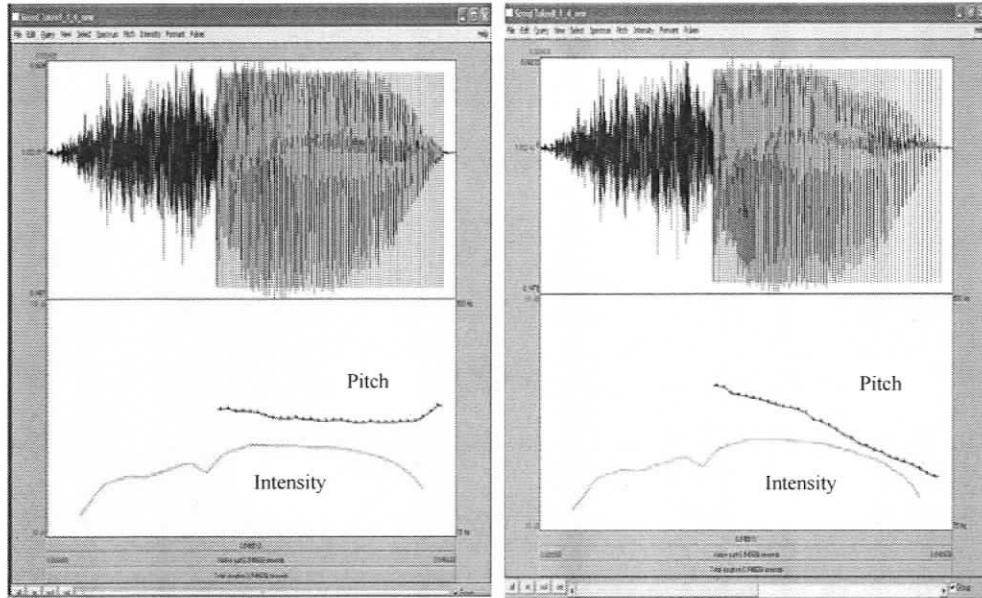
Table 3-5. F₀ pattern and duration of selected “best” tones.

Best Tone	Duration	F ₀ pattern		
		Beginning point	Turning point	Ending point
Tone 1 (High level)	382	316.5	292.6	318.4
Tone 2 (Rising)	388	241.7	213.5	334.5
Tone 3 (Falling-rising)	423	211.6	166.2	260
Tone 4 (Falling)	192	342.6	-	169

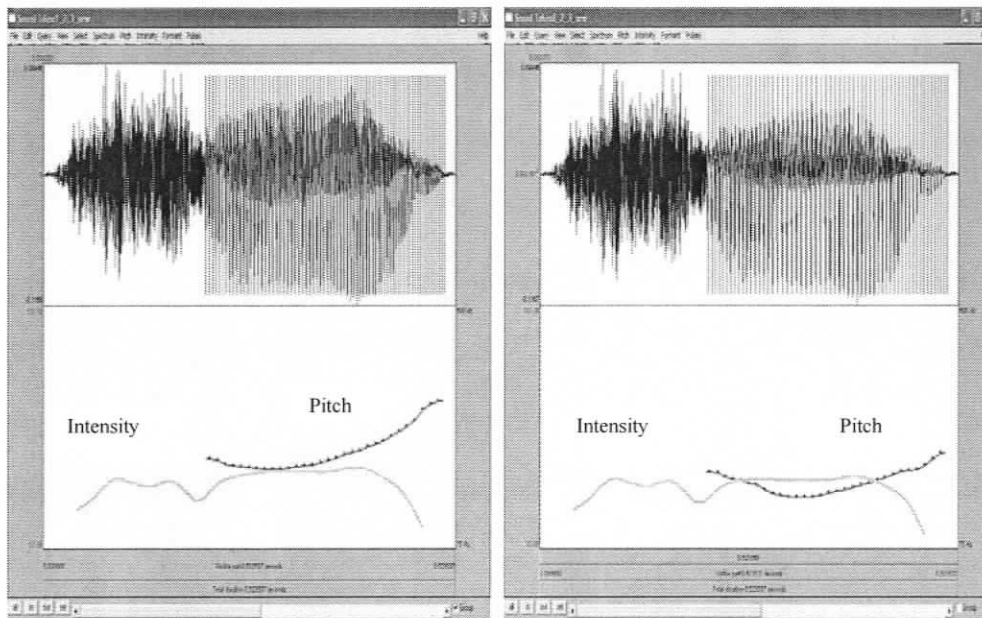
Using the four best tones, two tone continua, one ranging from high level to high falling (T1-T4), and the other ranging from rising to falling-rising (T2-T3), were chosen and created for the syllable *shi*. Since Study I was carried out as a pre-study for Study II (a categorical tone perception study), the choice to create the T1-T4 and T2-T3 continua was governed primarily by the focus of Study II (see Chapter four). Nevertheless, the two continua contained all four Mandarin tones; thus, Study I provides sufficient information to determine the correspondence between all the Mandarin and Thai tones. Each tonal continuum was composed of eight tokens (represented as t1, t2, t3, and so on) interpolated by equal steps between the end points, which were the natural tones. All the tokens on each continuum have the same amplitude envelope and duration. Praat (Boersma & Weenink, 2006) was used to equalize the duration and amplitude of the syllables (see Fig. 3-1). The duration of tone contour portion was set to 300ms, which is 50ms longer than the tones in Hallé et al. (2004). This duration was chosen because: (1) 300ms is a mean value between the length of the longest tone (T3) and the shortest tone (T4) (see Table 3-4); (2) the stimuli in Hallé et al. (2004) appeared in a carrier sentence, whereas the stimuli in this study were in citation form. A Mandarin tone in citation form or final position is known to be longer than one embedded in a carrier sentence or in non-final position (Lin, 1996). Therefore, a tone contour with 300ms duration was more appropriate for this study than 250ms.

Fig. 3-1. (A-B) Intensity and pitch envelopes of t1 and t8: (A) in T1 (left) and T4 (right), (B) in T2 (left) and T3 (right).

(A)



(B)



Along the 300ms range of the first and last tokens, thirty-one points were set at 10ms intervals (30 intervals \times 10ms). The corresponding points between the first and last tokens were interpolated

by equal steps, i.e. the first point on the first token corresponds to the first one on the second token, and so on. In this way, the duration of the tone contour for all the syllables was equalized to 300ms. Subsequently, the amplitude envelope of the first token was applied to that of tokens 2-8. Table 3-6 shows the F_0 range and duration of the manipulated four endpoint tones (cf. Table 3-5).

Table 3-6. F_0 range and duration of the manipulated tones.

Manipulated Tone	Duration	F_0 pattern		
		Beginning point (Hz)	Turning point (Hz)	Ending point (Hz)
Tone 1 (High level)	300	313.1	279.82	316.1
Tone 2 (Rising)	300	244.6	214.83	336.9
Tone 3 (Falling-rising)	300	212.7	165.5	254.5
Tone 4 (Falling)	300	334.5	-	170.5

Table 3-7 (A-B) provides F_0 values for the first and last pitch points, as well as those appearing at the turning points on each token, along with their time of occurrence (in bold, complete tables see Appendix III). The numbers in the rows refer to 31 equal steps at 10ms intervals along each token (tone) contour. Praat (Boersma & Weenink, 2006) was used to mark these points. The values in the columns represent eight equal intervals between t_1 and t_8 . As the initial and final values were obtained from the real tones, those values for the six tokens (t_2 - t_7) in between were calculated through an equation. At each point in time, $t_n = t_{n-1} + (t_8 - t_1) \div 7$ ($n = 2, 3, 4 \dots 8$). For instance, the F_0 value for t_2 on the T1-T4 continuum at point 1 was calculated as $t_2 = t_1 + (t_8 - t_1) \div 7 = 313.13 + (334.46 - 313.13) \div 7 = 313.13 + 21.33 \div 7 = 313.13 + 3.05 = 316.18$. Turning points usually occur in complicated contour tones, such as T3 (falling-rising tone). However, in real speech, simple tones may also contain turning points. The typical example is T2 (rising tone) which has the same tone contour (a falling-rising contour) as T3 in natural speech. This is one of the main reasons why T2 is often misperceived as T3. The T1 recorded for this study

has a falling contour, but with a very shallow slope. In this study, only T4 is a straight falling tone without any turning point. Therefore, as can be seen in Table 3-7 (A), on the T1-T4 continuum, F_0 increases at certain points for tokens 2-6, but not for tokens 7-8. As shown in Table 3-7 (B), on the T2-T3 continuum, the turning point occurs at the same pitch point for tokens 1-4; therefore, the falling portions of tokens 1-4 have the same length. The contours of all the eight tokens in continua T1-T4 and T2-T3 are provided in Fig. 3-2.

Table 3-7 (A-B). The first and last points and turning points (bold) along tone contours as well as occurrence time in the two continua: (A) in T1-T4, (B) in T2-T3.

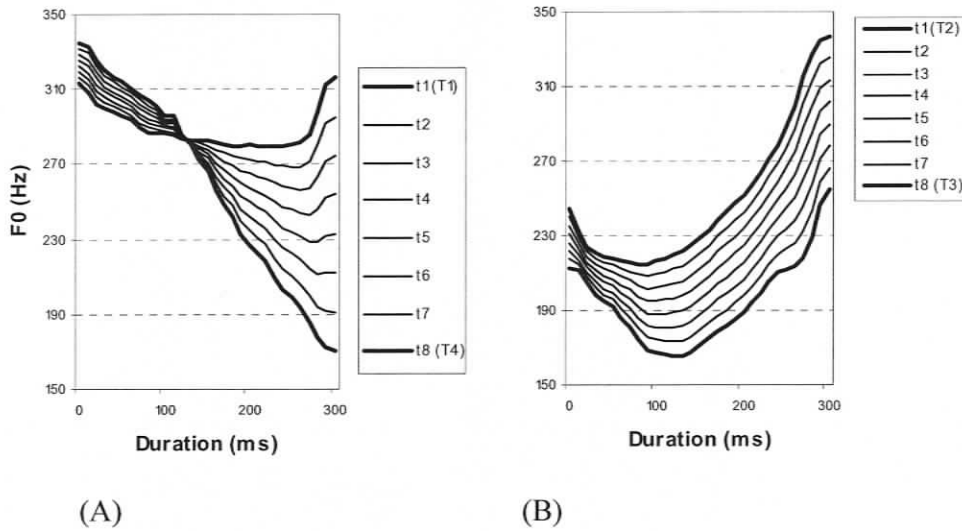
(A)

Point	Time (ms)	t1(T1) (Hz)	t2 (Hz)	t3 (Hz)	t4 (Hz)	t5 (Hz)	t6 (Hz)	t7 (Hz)	t8(T4) (Hz)
1	230	313.13	316.18	319.22	322.27	325.32	328.37	331.41	334.46
23	450	279.82	271.07	262.32	253.57	244.82	236.07	227.32	218.57
27	490	281.15	268.71	256.28	243.85	231.42	218.98	206.55	194.12
28	500	285.61	271.30	256.98	242.66	228.34	214.02	199.71	185.39
29	510	298.24	280.92	263.61	246.30	228.99	211.68	194.36	177.05
31	530	316.15	295.34	274.53	253.71	232.90	212.09	191.28	170.47

(B)

Point	Time (ms)	t1(T2) (Hz)	t2 (Hz)	t3 (Hz)	t4 (Hz)	t5 (Hz)	t6 (Hz)	t7 (Hz)	t8(T3) (Hz)
1	208	244.60	240.05	235.50	230.95	226.40	221.85	217.30	212.75
9	288	214.83	209.09	203.36	197.62	191.88	186.15	180.41	174.67
11	308	216.27	209.26	202.24	195.23	188.22	181.20	174.19	167.18
12	318	217.80	210.44	203.08	195.73	188.37	181.01	173.65	166.29
13	328	219.97	212.19	204.40	196.62	188.84	181.06	173.28	165.50
31	508	336.86	325.09	313.33	301.56	289.80	278.04	266.27	254.51

Fig. 3-2. (A-B) Eight tokens (t1, t2, t3, etc.) for the syllable *shi* in the two continua: (A) in the T1-T4 continuum, (B) in the T2-T3 continuum.



The 16 tokens along the T1-T4 and T2-T3 continua were presented in a random order to the subjects in this study. The inter-stimulus interval between tokens was set to four seconds. While this study employs synthesized tone continua, the synthesized tones preserve more than 90% of the information from the natural tones (Lin, 1988; Abramson, 1974). Thus, the correspondence between the synthesized Mandarin tones and the Thai natural tones should approximate that between natural Mandarin tones and natural Thai tones, which is the correspondence relationship that this study aims to investigate.

3.3.3 Procedure

The stimuli were presented to the subjects on a personal computer in a quiet university language lab. Each subject answered questions on a prepared response sheet (see Appendix I). The response sheet included English instructions at the top and a table with three columns. Ordinal numbers of the presented Mandarin tones (tokens) were listed in the first column; the corresponding Thai tones were to be written down in the second column; and the similarity

between the presented tokens and their Thai counterparts was to be ranked in the third column.

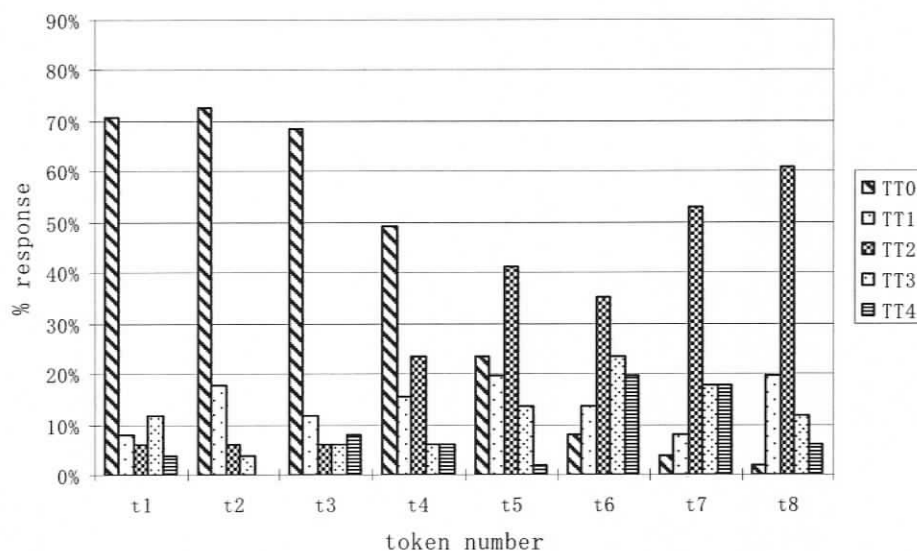
Participants were asked to choose a Thai tone as the counterpart of the tone they heard and to write down the corresponding number for the Thai tone, based on the traditional description of Thai tones, which is widely known by educated Thai speakers (0 = mid tone; 1 = low tone; 2 = falling tone; 3 = high tone; 4 = rising tone). This description was used in this test to help the subjects differentiate their native tones and to help the experimenter to interpret these tones. Subsequently, the participants were asked to rank the similarity between the presented tones and the Thai tones (1 = least similar; 2 = intermediately similar; 3 = very similar; 4 = same). The experimenter first translated the English instructions into Thai and verified that each participant understood the procedure.

3.4 Results

3.4.1 Correspondence of tokens on the T1-T4 continuum to Thai tones

Fig. 3-3 shows the percent response for the correspondence between the eight tokens on the Mandarin T1-T4 continuum and the five Thai tones. Tokens 1 to 4 were predominantly labelled as counterparts to the Thai mid tone (TT0), while tokens 6 to 8 were perceived as the Thai falling tone (TT2). The cross-over between token 4 and token 5 indicates a discrimination boundary around that region.

Fig. 3-3. Correspondence of Mandarin tonal variants on the T1-T4 continuum to Thai tones.



T1, one of the two endpoint tones along the T1-T4 continuum, was most often classified as the Thai mid tone (TT0). This result was statistically significant: a one-way ANOVA indicated that Thai listeners chose Mandarin T1 as the counterpart to the Thai mid tone significantly more often than to any other Thai tone ($p < 0.001$). As shown in Table 3-8, 35 of a total of 51 participants matched T1 with TT0, compared to 4, 3, 6 and 2 participants who matched T1 with responding the Mandarin as TT1, TT2, TT3, and TT4, respectively. As predicted, The other endpoint tone, T4, was matched with the Thai falling tone (TT2), which has an identical F_0 value, more often than it was matched with any other Thai tone; this result was statistically significant ($p < 0.001$). Nonetheless, as shown in Table 3-8, the data on the average similarity rankings does not perfectly replicate the results of the correspondence judgments. For example, the highest average similarity ranking, 3.33, is for TT4 as MT4, as opposed to 2.35 for TT2 as MT4. However,

the small number of responses for TT4 as MT4 most likely accounts for this result: only three participants labeled MT4 as TT4.

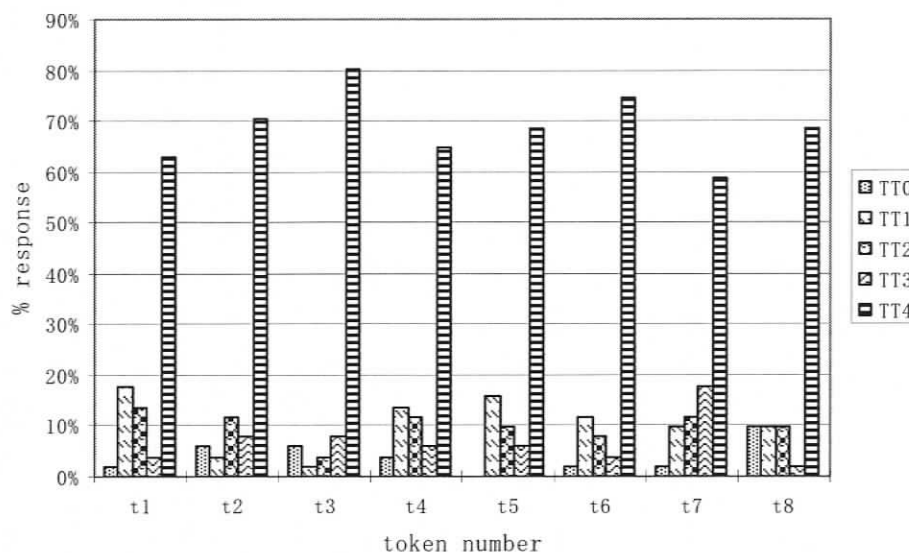
Table. 3-8. Number of responses and average similarity ranking between T1 and T4 and the Thai tones.

Thai tone	T1		T4	
	Number of responses	Average similarity ranking (SD)	Number of responses	Average similarity ranking (SD)
0	36	2.47 (0.97)	1	1
1	4	2.75 (0.96)	10	2.7 (1.06)
2	3	2.67 (0.82)	31	2.35 (0.75)
3	6	2.33 (0.81)	6	2.83 (0.41)
4	2	2 (0)	3	3.33 (1.15)

3.4.2 Correspondence of tokens on the T2-T3 continuum to Thai tones

While the results for the T1-T4 continuum (see Fig. 3-3 above) showed a clear discrimination between the two endpoint tones, the results for the T2-T3 continuum revealed a different assimilation pattern (see Fig 3-4 below). No discrimination boundary emerged between the two endpoint tones along the T2-T3 continuum. Rather, the results indicate that for most listeners, all eight tokens sounded closest to the Thai rising tone (TT4).

Fig. 3-4. Correspondence of Mandarin tonal variants on the T2-T3 continuum to Thai tones.



The endpoint tone on the rightmost end of the T2-T3 continuum was labeled by most participants (see Table. 3-9) as the Thai rising tone (TT4), which has the same F_0 value as the Mandarin T3 (214 in Ladefoged's description). While it would be reasonable to suppose that the Mandarin rising T2 (35 in Chao's description) might be judged as closer to the Thai high tone (TT3, 45 in Ladefoged's description) because of the similarities in F_0 direction and height, this result was not found; this finding is discussed in the next section. The ANOVA and post hoc test (Tukey HSD) showed that the correspondence between MT2 and MT3 to the Thai rising tone (TT4) was significantly higher than that of MT2 and MT3 to any of the other Thai tones ($p < 0.001$). The data on average similarity rankings did not provide additional evidence for the correspondence between TT4 and MT2 and MT3. As noted in the results of similarity rankings for the T1-T4 continuum, however, the highest values in Table 3-9 are not necessarily reflective of the

correspondence between the tones, given the extremely low number of responses obtained for the highest average similarity rankings.

Table. 3-9. Number of responses and average similarity ranking between T2 and T3 and the Thai tones.

Thai tone	T2		T3	
	Number of responses	Average similarity ranking (SD)	Number of Responses	Average similarity ranking (SD)
0	1	3	5	2 (1.22)
1	9	2.11 (0.6)	5	2.4 (0.89)
2	7	2.29 (1.38)	5	2.8 (1.64)
3	2	2 (0)	1	3
4	32	2.63 (0.94)	35	2.48 (1.04)

3.5 Discussion

The results of this study indicate that Thai listeners assimilate non-native tones with similar F_0 contours and pitch heights to native categories. For example, Mandarin T4 and T3 were identified as Thai falling tone and Thai rising tone, respectively. According to Gandour (1983), for Thai listeners, a non-native tone with a different F_0 value tends to be assimilated to a native tone with a similar F_0 contour, but not to a native tone with a similar F_0 height. For instance, the Mandarin high level tone (55 in Chao's description) was assimilated to the Thai mid tone category (32 in Ladefoged's description), but not to the Thai high level tone category (45 in Chao's description), which has a higher pitch value. As shown in Fig. 3-1, the actual F_0 contour of T1 starts with a slightly falling

contour³ (see also the graphs in Hallé et al., 2004; Wang et al., 1999; Xu, 1997), which resembles the Thai mid tone in terms of F_0 direction, but contrasts with the Thai high level tone in terms of F_0 contour (the latter has a rising contour). As shown in Figure 3-1, the Mandarin rising tone (35 in Chao's description) exhibits a falling-rising contour (see also the graphs in Shen & Lin, 1990; Blicher et al., 1990; Hallé et al., 2004), which is similar to the F_0 contour of the Thai rising tone (214 in Ladefoged's description). By contrast, the Thai high level tone, which contains a straight rising contour, is less similar to Mandarin T2 in terms of F_0 direction.

Based on the results presented in this study, predictions can be made about the categorization of Mandarin tones by inexperienced Thai listeners in Study II. The Mandarin T1-T4 continuum and T2-T3 continuum should be perceived in different ways. On the one hand, since the listeners labeled Mandarin T4 and T1 as two distinct Thai tone phonemes, it is expected that Thai listeners' perception of the T1/T4 contrast will be categorical and that the category boundary will fall between tokens 4 and 5 (see Fig. 3-3). Given that Mandarin T2 and T3 were identified as corresponding to a single tone phoneme in Thai, it is anticipated that the Thai listeners' perception of the T2/T3 contrast will be non-categorical. From this point of view, it is likely that the English listeners will perceive both Mandarin tone contrasts non-categorically, given that English does not

³ In this study, although the naturally produced T1 ended with a short rising contour [70ms on a 300ms tone, see Fig. 3-2 (A) and Table 3-7 (A)], compared to the rising contour in T2 (220ms) and T3 (180ms), the end rising contour should not be included when discussing the F_0 of T1.

employ lexical tone contrasts.

Chapter Four

STUDY II: IDENTIFICATION AND DISCRIMINATION OF TONE

CONTINUA

4.1 Introduction

This chapter describes the focal study of this thesis, which includes three experiments set up to determine the effects of linguistic experience and training on categorical perception of tone. The first two experiments are discrimination and identification experiments designed for a categorical perception task. The stimuli are eight synthetic tonal variants on each of two tone continua, as described in the previous chapter. The participants are distinguished according to L1 background, particularly tone experience. The third experiment is a tone perceptual training in which natural stimuli were presented to some subjects but not to others. Performance of the trained and untrained subjects was compared. The research questions raised in Chapter 1 are addressed in this study. With respect to these questions, I predict that Mandarin speakers will perceive both tone contrasts categorically; that Thai speakers will perform better than English speakers in the categorization of Mandarin tones; and that in each group, trained

subjects will perform better than untrained subjects in the categorization of Mandarin tones.

4.2 Method

4.2.1 Subjects

A total of 36 subjects in three language groups participated in this study. The Mandarin group included 12 native speakers, 10 of whom were born and raised in Beijing or other areas of Northern China; the remaining two participants lived in Beijing for more than five of their teenage years. The Mandarin group included 3 male and 9 female¹ participants, aged 22-38 years, who were either undergraduate or graduate students at the University of Victoria. None of the participants had been exposed to Thai or other tone languages or had received formal musical training. All participants spoke English as their second language.

The Thai group consisted of 12 native speakers originally from Bangkok. This group included 6 male and 6 female participants, aged 17-41 years old. None of the participants had been exposed to Mandarin or other tone languages; nor had they received any formal musical training. Like the Mandarin group, these participants all reported having English as their second language. All but the 17-year-old subject, who had come to Canada to

¹ Compared to the Thai and English groups, the Chinese group is unbalanced with respect to gender. In this study, the gender ratio was a secondary factor in the selection of participants (Francis et al. 2003) compared to the more primary issue of geographic origin. For example, a female native Mandarin speaker who was born and grew up in Beijing was selected over a male native Mandarin speaker who had lived in northern China for only a couple of years.

attend high school, were either graduate or undergraduate students at the University of Victoria. None of the 12 participants in this study participated in Study I.

The English group was composed of 12 native speakers of Canadian English; two of them were born in Ontario and the other 10 were born and grew up in the province of British Columbia. The group included 8 male and 4 female participants aged 22-46 years old, none of whom had any previous experience with tone languages. Eleven of the participants were students at the University of Victoria and the twelfth worked in Victoria. Six of the 12 participants reported having received formal musical training.

All 36 participants from the three groups reported normal speech and hearing. All of the subjects voluntarily participated in this experiment and received no financial compensation for their participation.

4.2.2 Stimuli

4.2.2.1 Stimuli in the categorical tasks

In Study I, eight tokens on each of two Mandarin tone continua, T1-T4 and T2-T3, were created for a labeling and ranking test to determine the correspondence between Mandarin and Thai tones. The reasons that T1/T4 and T2/T3 contrasts were chosen for the sake of Study II were that: (1) these two tone pairs are known to be the most confusing tone pairs in L1 and L2 acquisition (Chao, 1980; Kriloff, 1973; Wang et al., 2003; Sun, 1997; Li & Thompson, 1977; Blicher et al., 1990; Huang, 2001) and the similar starting F_0 of T1 and T4, and the rising F_0 contour shared by T2 and T3, are

considered to be the main reasons for the confusion between them (Hwang, 2001); (2) to my knowledge, no previous study within the framework of categorical perception has examined these two tonal contrasts. I chose these two tonal contrasts with two main goals in mind. The first goal was to determine whether, in the identification and discrimination of both these tonal contrasts, native and non-native listeners are affected in the same way by the potentially confusable cues. The second goal is to investigate differences between the perception of level-contour (T1-T4) and contour-contour (T2-T3) tone continua, as discussed by Wang (1976), Hallé et al. (2004), and Francis et al. (2003). According to these authors, the transition from one tone contour to another affects tone perception because of the coincidence of category boundaries with regions of natural auditory sensitivity. It is proposed that the category boundary is more auditorily salient for the transitions from rising to level and from falling to rising pitch contours, but not for those from level to level and from low-rising to high-rising contours.

These tokens were reorganized in Study II for two perception tests: an AX (same/different) discrimination test and an AXB identification test. Accordingly, these tokens were arranged in pairs for the discrimination test and in sets of three for the identification test. The stimuli used in the discrimination test consisted of all possible pair-wise combinations of individual syllable tokens, separated either by a zero step, such as token 1-token 1 (t1-t1) or by two steps, such as token 1-token 3 (t1-t3), along the continuum. The inter-stimulus interval (ISI) was set to 500ms. The reason for choosing this ISI is that listeners' discrimination reaches a maximum between 500ms and 1000ms

but falls as the ISI increases further (Cowan & Morse, 1986; Pisoni, 1973; Van Hesse & Schouten 1992; cited from Gerrits & Schouten, 2004). The choice of an interval of 500ms rather than 1000ms, as adopted in Hallé et al. (2004), was primarily due to time limitations. This procedure yielded a total of 20 pairs, including 8 identical pairs (t1-t1, t2-t2, t3-t3 etc.) and 12 two-step pairs (t1-t3, t3-t1, t2-t4, t4-t2, etc.). In the AXB identification test, A and B were two endpoint tones in the two possible orders. X varied from one endpoint to the other along the eight steps of the continuum. In the identification test, the inter-stimulus interval was also set to 500ms. This test included a total of 16 triplets (t1-t1-t8, t8-t1-t1, t1-t2-t8, t8-t2-t1, etc.)

Instructions for the discrimination and identification tests were recorded in Mandarin, Thai, and English by a native speaker of each language in a sound-insulated booth in the Phonetics Laboratory of the Department of Linguistics at the University of Victoria. The three speakers read instructions (the English version see Appendix II) in their native language into an AKG C-1000S microphone.

4.2.2.2 Tone training stimuli

The four Mandarin tones were grouped and presented to all the subjects in a natural tone training session. By contrast with the target syllable used in the categorical perception task (see 3.2.2), the syllables used in the training did not correspond to real word forms so as to avoid lexical interference in the training. Three artificial syllables, *cei*, *gian*, and *ra*, were recorded by a Beijing-born female native speaker of Mandarin, aged

22 years, in the same sound-insulated recording room noted above.

The syllables carrying each of the four Mandarin tones were read five times in a random order and the best tones were selected in the manner described for Study I. The three syllables were then arranged in pairs with tonal contrasts, for instance, *cei1 cei2* (1 and 2 refer to T1 and T2, respectively). In total, there were 36 combinations of tone pairs. The inter-stimulus interval was 500ms.

Instructions for the training were recorded by the same three native speakers who recorded instructions for the categorical perception tests. The instructions (the English version see Appendix II) included a description of the pitch movement of each tone and a demonstration of the four tones imposed on two example syllables: *ma* and *fan*.

4.2.3 Procedure

4.2.3.1 Identification test

All the subjects participated in an AXB identification task in which they were forced to answer whether the second token (token “X”) in a triplet sounded more like the first token (token “A”) or the third token (token “B”). An AXB identification task was chosen instead of the labelling task more commonly adopted in categorical perception experiments because: (1) a labelling task would be inappropriate for non-native subjects who were not acquainted with Mandarin tones (Hallé et al. 2004); and (2) for purposes of comparison, the Mandarin group, which could have completed a labelling task, needed to perform the same identification tests as the Thai and English groups.

Subjects who were assigned to the subgroups to discriminate stimuli along the T1-T4 or T2-T3 continuum identified stimuli along the same continuum; for instance, a subject assigned to discriminate stimuli along the T1-T4 continuum would identify stimuli along the same continuum. Each subject received 112 trials (16 triplets \times 7 repetitions) for each continuum in a random order, with 16 trials in one block and seven blocks in total. As with the discrimination test, the inter-trial interval was set to 3s and the inter-block interval to 8s. The subjects were not informed that the first block would not be scored.

The instructions recorded for the identification tests were also presented to the subjects audio-visually. Following the instructions, for each trial, the subjects responded by clicking one of the buttons labeled “1” or “3,” depending on whether they thought the second stimulus (X) sounded more like A than B. Responses in the identification test were collected automatically by the same computer-based application used in the discrimination test and scored by calculating the times that each stimulus was identified as the first or the third stimulus.

4.2.3.2 Discrimination test

The discrimination test was a two-step AX discrimination task, in which subjects were forced to choose between “same” or “different” for each presented stimulus pair. Six subjects from each group were randomly assigned to a subgroup to discriminate stimuli along either the T1-T4 or the T2-T3 continuum. Each subject heard 140 trials (20 pairs \times

7 repetitions) presented in a random order, with 20 trials in one block, seven blocks in total. The inter-trial interval was set to 3s and the inter-block interval to 8s. The subjects were not told that the first block would not be scored.

Discrimination tests were administered individually in the aforementioned Phonetics Laboratory by the author. Each participant took approximately 12 minutes to complete the test. The participants were presented with the instructions in their native language, followed by the stimuli, through SONY MDR-V600 headphones. The script of the instructions was also shown to the subjects in their native language on a computer. Subjects were asked to click one of the buttons labeled “same” or “different” in their native writing system. Before the participants began clicking the buttons, the author reiterated that the label “same” was intended to mean “identical,” while the label “different” could mean “similar”². Responses were collected automatically by a pre-designed computer application and the responses of “same” and “different” were summed separately.

4.2.3.3 Perceptual tone training

As noted above, one principal goal of this component of the study was to identify the influence of natural tone training on categorical tone perception. Accordingly, half of the subjects who were assigned to continuum T1-T4 or T2-T3 (three from each subgroup)

² The results of the pilot study indicated that this repetition was necessary. Without this reminder, 99% of responses were “same” for most participants in the pilot study.

were quasi-randomly selected to take the training before completing the categorical perception tests. As a result, there were very few subjects in each condition, a factor which may limit the significance of the experimental results.

Tone training was conducted on an individual basis in the aforementioned Phonetics Laboratory. The stimuli and instructions were presented to participants at a comfortable level over SONY MDR-V600 headphones. In order to help inexperienced subjects gain a better impression of each learned tone and its contour shape, the author provided each selected participant with an extra introduction to Mandarin tones, which included drawings of the tone contours and their movements³. The “extra” introduction was given in English.

The format of the tone training duplicated the high-variability procedure developed by Logan et al. (1991) and adopted by Wang (2002). With the high-variability procedure, listeners were trained to identify stimuli appearing in a variety of phonetic contexts in natural words, as produced by a variety of talkers (Wang, 2002, p.89). Since all the stimuli in this study were produced by one talker, this experiment only partially duplicated the training pattern of the high-variability procedure. The training task adopted in this study was a forced choice identification task, as described in Wang (2002, p.94).

The stimuli were arranged first by randomly assigning the 36 tone pair combinations (see 4.1.2.2) to one block. A trial was formatted as a tone pair followed by a 3s interval,

³ Previous studies, such as Leather (1990), Hazan et al. (2005), and So (2005) also employed visual aids in tone perception training.

which was followed by feedback. During the 3s interval, subjects were forced to choose a tone for the first word they heard by clicking one of the four buttons horizontally labeled 1-4 from left to right. Each number represented a corresponding tone, i.e., number 1 was T1, and so on. The feedback was recorded in subjects' native language. For example, a syllable *cei* carrying T1 and T4, together with the feedback, was presented to the English subjects as follows:

Mandarin speaker: *cei* 1 *cei* 4
 (Subject's response for the first word)
 English speaker: The first word is tone 1.
 Mandarin speaker: *cei* 1
 English speaker: The second word is tone 4.
 Mandarin speaker: *cei* 4

The above process was considered as a training trial. The trainees received 72 trials (36 pair combinations \times 2 repetitions), presented in a random order and blocked by 2. As in the identification and discrimination tests, the inter-trial interval was set to 3s and the inter-block interval to 8s.

4.3 Analyses and results

In this section, results of the identification and discrimination tests are reported and discussed on the basis of data showing: (1) the intercept and slope of category boundaries in the identification tests; and (2) the discrimination accuracy and location of peak points in the discrimination tests.

4.3.1 Identification test

4.3.1.1 PROBIT

The results of the identification tests were examined through narrow-range PROBIT analysis (Best & Strange, 1992), which fits the (negative) cumulative normal distribution to the curve defined by the probabilities (percent response) versus the tokens. A weighted regression is used so that probabilities close to 0 and 1 have less effect on the fitted line (MacKain et al., 1981; Best & Strange, 1992; Hallé et al., 1999, 2004; Dr. Lesperance, personal communication). According to this analysis, the category boundary refers to the 50% intercept of the best-fitted three-point ogives, the three points closest to the 50% crossover on identification curves (Fig.4-1). The slopes of these ogives ($1/SD$) indicate “the peak rate of change in category labeling at the crossover, and were used as a reflection of the steepness of the category boundaries, i.e., larger values indicate steeper functions” (Best & Strange, 1992, p313). To choose the three best points that are closest to the 50% crossover, a general principle of “two above and one below” was used in this study (Dr. Hallé, personal communication). Fig. 4-1 (A-B) shows group identification curves for each continuum indicating percent response of each token as t_1 , i.e., T1 and T2, respectively. Table 4-1 presents the data on the intercepts and slopes of category boundaries pooled across participants in each group.

Fig. 4-1. Identification curves of the Mandarin, Thai, and English groups: (A) on the T1-T4 continuum; and (B) on the T2-T3 continuum.

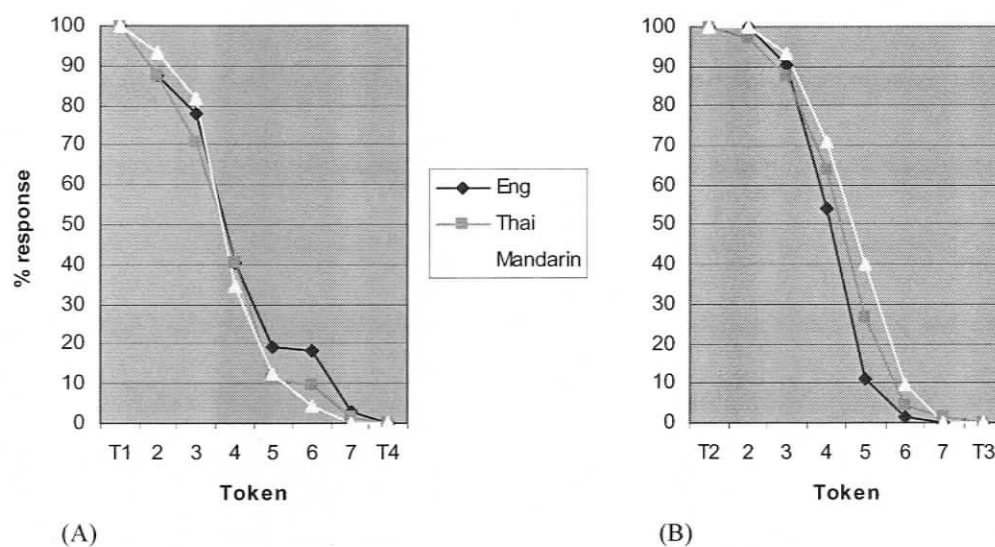


Table 4-1. Intercept and slope for Mandarin, Thai, and English participants' data.

Continuum	Intercept (token number)			Slope (1/SD value)		
	Mandarin	Thai	English	Mandarin	Thai	English
T1-T4	3.72	3.76	3.79	1.64	1.24	1.24
T2-T3	4.74	4.38	4.11	1.32	1.21	1.51
Average	4.23	4.07	3.95	1.48	1.23	1.37

4.3.1.2 Pooled results

As seen in Figure 4-1 (A-B), the identification curves exhibit sigmoid shapes and steep slopes on both continua across all the three groups, which was likely the result of using an AXB identification task (Hallé et al., 2004). By applying the AXB procedure, the identification curves may well reveal the usual sigmoid shape and cover the full range of identification from 0-100% for each endpoint stimulus by non-native listeners, while

showing the non-native intercept location and steepness.

Although the data for all three groups yielded sigmoid shape identification curves, compared to the Thai and English groups, the Mandarin group shows a smoother descending curve from left to right for the T1-T4 continuum (Fig. 4-1 (A)). The Thai and English groups both identified token 6 as token 1 more often than did the Mandarin group. The figures were 9.72% for the Thai group, 18.06% for the English group, but only 4.17% for the Mandarin group. In other words, token 6 sounded more like T4 to the Mandarin group than to the Thai and English groups.

The identification data presented in Table 4-1 show category boundaries falling on the left of the centre point (4.5 for an eight-token continuum) for the T1-T4 continuum, meaning that most of the tokens across the boundary were identified as the endpoint tone (T4). The data also indicate that the intercept falls farthest away from the centre point for the Mandarin group (3.72) and closer to the centre point (3.79) for the English group than for the Thai group (3.76). The difference among the groups, however, is not significant ($F(2, 15) = 0.13, p = 0.987$). The slope data for the T1-T4 continuum reveals to some extent the difference between the native and non-native groups. Although an ANOVA did not show a significant difference among the three language groups at the 0.05 significance level ($F(1, 16) = 2.691, p = 0.12$), it is readily apparent that the Mandarin group has a steeper slope than the Thai and English groups (1.64 vs. 1.24), suggesting that the identification function is highest for the native Mandarin speakers.

The results for the identification of the T2-T3 continuum are shown in Fig. 4-1 (B)

and Table 4-1. Fig. 4-1 (B) does not exhibit any difference in the shape of the identification curve among the three groups. The locations of the intercept, however, are not as convergent as those shown in Fig. 4-1 (A). The data are given in Table 4-1: Mandarin 4.74, Thai 4.38 and English 4.11. That is, the intercept of the category boundary falls to the right of the center point (4.5 for an eight-token continuum) for the Mandarin group, while it falls to the left of the centre point for the Thai and English groups, and furthest away from the centre point for the English group. The locations of the intercept indicate that the English listeners were more likely to identify most of the tokens as T3, whereas the intercepts tended to cluster around the centre point for the Mandarin and Thai groups. However, this result was not statistically significant.

The slopes of the boundaries by token number are also given in Table 4-1: Mandarin 1.32, English 1.51 and Thai 1.23. The English group exhibits a steeper slope than the Mandarin group does, while the data for the Thai group indicates a relatively weak identification function compared to the Mandarin and English groups. These data suggest that the Thai group had difficulties in identifying Mandarin T2 and Mandarin T3, and that the native Mandarin speakers did not have much of an advantage over the non-native speakers in identifying these two tones. Nevertheless, the Mandarin speakers still showed an advantage over the Thai speakers.

Further ANOVAs performed on the steepness of the slopes failed to show any significance on the effect of the continuum among the three groups. This result may be attributable to the small number of participants, as mentioned above. However, relative to

the slope for the T1-T4 continuum, there was a clear tendency for the identification function to decline for the Mandarin and Thai groups (from 1.64 to 1.32, and 1.24 to 1.21, respectively) along the T2- T3 continuum, and to increase for the English group (from 1.24 to 1.51). This pattern suggests that for the Mandarin and Thai groups, T1 and T4 are more readily distinguishable than T2 and T3, but the reverse is the case for the English group.

To sum up, the results of the identification tests show differences among the three groups in terms of the shape of the identification curve and the location and steepness of the category boundaries. For the Thai and English groups, token 6 on the T1-T4 continuum was identified as T1 more often (9.72% and 18.06%, respectively) than for the Mandarin group (4.17%). Accordingly, the identification curve for the Mandarin group descends more smoothly compared to the Thai and English groups.

Despite the lack of strong statistical results for the differences in the location and slope of category boundaries among the three groups, tendency analyses revealed some differences and consistencies in the intercepts and slopes. In the T1-T4 identification test, the Mandarin group showed a larger identification function than the non-native groups in the identification of T1 and T4 (1.64 vs. 1.24), but the English group performed better in identifying T2 and T3 compared to the Thai and Mandarin groups; the Thai listeners showed very weak performance in identifying T2 and T3. In the T2-T3 identification test, the English group classified most tokens near the boundary as T3 (4.11). The Mandarin and Thai groups, however, identified most tokens near the center point (4.74 and 4.38).

On the identification curves of the T1-T4 continuum, the category boundaries emerged on the left of the centre point across all three language groups. Thus, in this continuum, all three groups identified most tokens around the crossover as T4.

4.3.1.3 Comparison of trained and untrained subjects

As mentioned above, in order to determine the effect of perceptual training on tone perception, half of the subjects in each continuum subgroup received intensive training in natural tone perception before participating in the categorical perception tasks. The pooled results discussed in the previous section are divided into trained versus untrained groups here. The intercept and slope data are shown in Table 4-2.

Table 4-2 Intercepts and slopes for trained and untrained Mandarin, Thai, and English groups.

Continuum	Intercept (token number)				Slope (1/SD value)			
	Mandarin	Thai	English	Average	Mandarin	Thai	English	Average
Untrained								
T1-T4	3.57	4.22	4.19	4.00	1.40	1.13	0.99	1.17
T2-T3	4.90	4.41	4.00	4.44	1.24	1.30	1.47	1.34
Trained								
T1-T4	3.86	3.30	3.39	3.52	1.89	1.36	1.49	1.58
T2-T3	4.57	4.34	4.22	4.38	1.39	1.13	1.55	1.35

The intercept data indicate that for the T1-T4 continuum, the location of the category boundary moved closer to right endpoint after the training session for the Mandarin group, but in the opposite direction for the Thai and English groups. The raw data suggest that the trained Mandarin participants did not change much with training on tone

identification, and that with training, the Thai and English speakers became more like the Mandarin speakers; that is, their boundaries shifted towards the Mandarin boundary (in the untrained group, 4.22, 4.19, and 3.57; in the trained group, 3.30, 3.39, and 3.86 for the Thai, English, and Mandarin groups, respectively).

For the T2-T3 continuum, the intercept moved toward the left endpoint for the trained Mandarin and Thai groups, but toward the right endpoint (although still on the left side) for the trained English group. Again, the data indicate that the change occurred in the trained Thai and English groups. The category boundaries moved closer to that of the Mandarin group (in the untrained group, 4.41, 4.00, and 4.90; in the trained group, 4.34, 4.22, and 4.57 for the Thai, English and Mandarin groups, respectively).

The slope data show a general increase for the three groups after the training. The identification function increased for the trained Mandarin and English groups for both continua (for the Mandarin group, 1.40 vs. 1.89 and 1.24 vs. 1.39 for the T1-T4 continuum and T2-T3 continuum, respectively; for the English group, 0.99 vs. 1.49 for the T1-T4 continuum and 1.47 vs. 1.55 for the T2-T3 continuum, respectively). For the trained Thai group, identification increased for the T1-T4 continuum (1.13 vs. 1.36), but declined for the T2-T3 continuum (1.30 vs. 1.13). Therefore, the identification of both tone contrasts improved after training not only for the non-native groups but also for the native group. Nevertheless, the T2-T3 contrast was equally confusing for the Thai participants with and without training.

Although the tendency revealed by the data in Table 4-2 is obvious, with the

non-native intercepts shifting closer to the native intercepts and a general increase in the slopes after the training, none of the differences between the trained and untrained groups were found to be significant in ANOVAs, with the exception of a marginally significant difference between the pooled slopes for the T1-T4 continuum across all the subjects ($F(1, 16) = 3.257, p=0.09$) before and after training.

To sum up, the results of this study indicate that training in tone perception plays a positive role in the identification of lexical tone contrasts. The improvement in the identification function is nearly consistent across all three groups of listeners. This improvement is more statistically obvious for the T1-T4 contrast than for the T2-T3 contrast. However, tone perception training revealed a complicating result for the Thai group in the perception of the T2-T3 contrast: although the location of the intercept approached that of the native group, the slope did not increase, but dropped after training.

4.3.2 Discrimination test

4.3.2.1 Pooled results

The discrimination results were examined from three perspectives: (1) the average discrimination level across all the tokens; (2) the existence of one or two adjacent peak points; and (3) the correspondence of the location of peak point(s) and the identification intercept. According to previous studies, native speakers' performance in discrimination will be better than that of non-native speakers; therefore, the average discrimination level is higher for native speakers. Peak points are one of the most important indices in the

judgment of whether or not perception is categorical. That is, only when the discrimination level of one or two continuous points is significantly higher than the points adjacent to them can perception be considered to be categorical. The other important criterion for categorical perception is that the location of the peak point(s) must fall in the vicinity of the identification intercept, i.e. the location of the peak point(s) and the identification intercept should nearly overlap. Table 4-3 and Figure 4-3 (A-B) show the mean performance levels and discrimination curves across all the three language groups.

Table 4-3. Mean correct discrimination levels (%) for Mandarin, Thai, and English groups on the AX discrimination task.

Continuum	Mean (SD)		
	Mandarin	Thai	English
T1-T4	70.92 (0.14)	70.14 (0.16)	73.28 (0.16)
T2-T3	61.89 (0.1)	59.67 (0.1)	71.5 (0.13)

The data on the discrimination levels do not show any advantage for the Mandarin group over the other two groups, especially the English group (70.92, 70.14, and 73.28; and 61.89, 59.67, and 71.5 for the Mandarin, Thai and English groups for the T1-T4 and T2-T3 continua, respectively). For both continua, the pooled results for the English group are higher than both the Mandarin and the Thai groups. Further study was carried out for the English subjects with musical training in order to determine whether the high discrimination level exhibited by the English group resulted from the participation of the musically trained subjects.

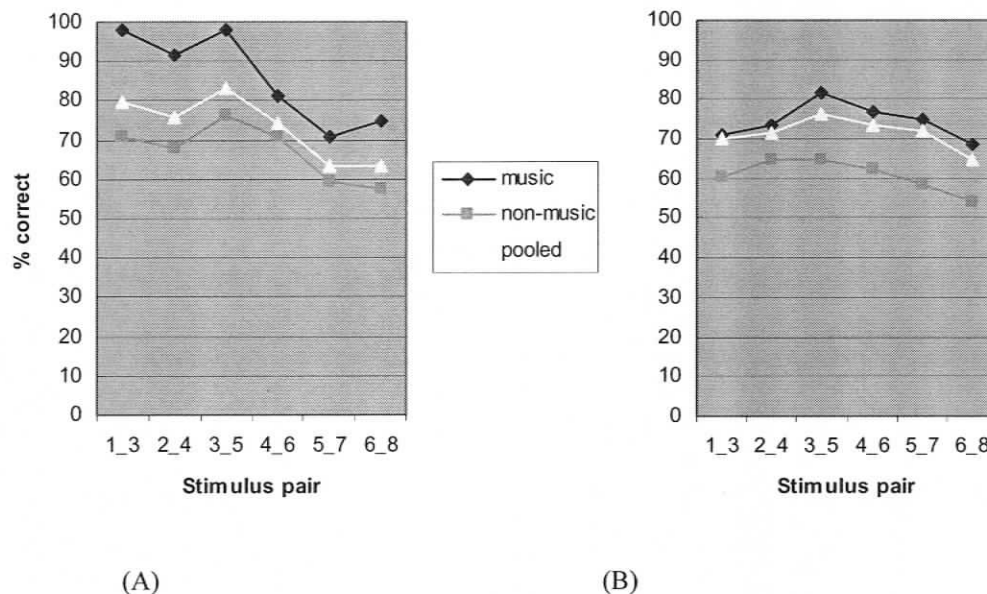
Table 4-4. Mean correct discrimination levels (%) for the English group and the musically trained English subgroups on the AX discrimination task.

	T1-T4		T2-T3	
	Mean (SD)	Number of subjects	Mean (SD)	Number of subjects
Pooled	73.28 (0.16)	6	71.5 (0.13)	6
Music: Pooled	85.83 (0.17)	2	74.4 (0.12)	4
High level	95.17 (0.07)	1	83.5 (0.13)	1
Non-high level ⁴	76.5 (0.9)	1	72.12 (0.1)	3
Non-music	67 (0.12)	4	57 (0.08)	2

The performance of the musically trained versus musically untrained English subjects in the discrimination of tones along the T1-T4 and T2-T3 continua were compared. As shown in Table 4-4 and Fig. 4-2, the musicians outperformed the non-musicians significantly in discrimination ($F(1, 34) = 14.85, p < 0.001$ and $F(1, 34) = 12.14, p = 0.001$ for T1-T4 and T2-T3 continuum, respectively).

⁴ Two of the English musician subjects who reported having more than 10 years of musical training were grouped into the “high level” category; the remaining four were categorized as “non-high level”.

Fig. 4-2 Two-step discrimination curves for the pooled English group and subgroup of participants with and without musical training: (A) on the T1-T4 continuum and (B) on the T2-T3 continuum.

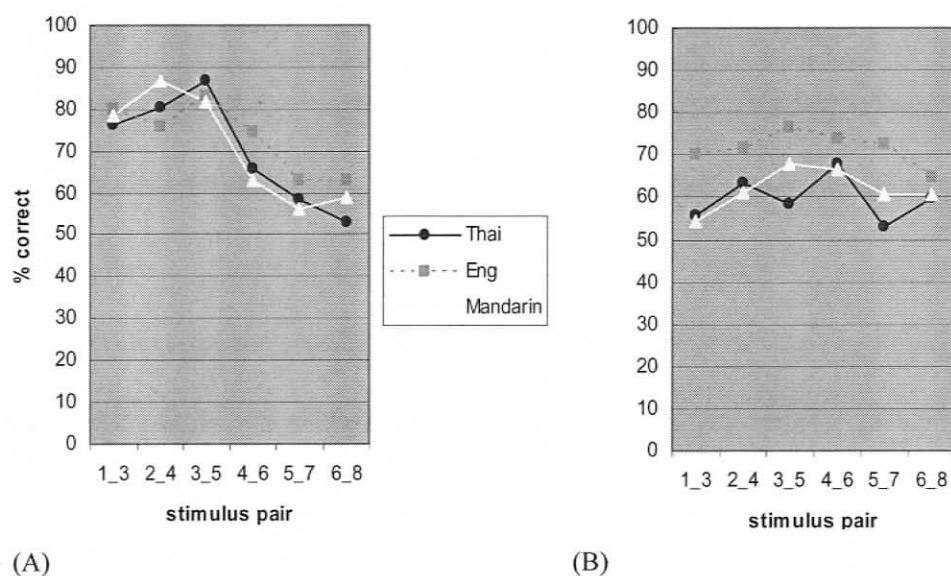


Based on the data presented in Table 4-4 and the better performance of the musicians shown in Fig. 4-2, the higher discrimination levels within the English group are most likely attributable to the effects of musical training, particularly the effect of advanced musical training reported by some subjects. An ANOVA indicated that performance did not differ significantly according to L1 background for the T1-T4 continuum; and that the performance of the Mandarin and Thai groups is significantly lower than that of the English group in the discrimination of the T2-T3 continuum ($F(1, 70) = 13.04, p = 0.001$; and $F(1, 70) = 18.45, p < 0.0001$ are ANOVA statistics for the Mandarin versus English and the Thai versus English groups, respectively). These results suggest that tone language speakers do not outperform non-tone language speakers in tone discrimination when the latter speakers are musically trained. However, according to the data presented

in Tables 4-3 and 4-4, tone language speakers outperform non-tone language speakers who have no musical training (70.92 and 70.14 vs. 67, 61.89 and 59.67 vs. 57 for the T1-T4 and T2-T3 continua, respectively). These data suggest that L1 tone experience may be the crucial factor to affect tone discrimination for listeners who have had no musical training.

The effect of the tone continuum was also statistically examined with ANOVA. As shown in Table 4-3, the Mandarin and Thai groups performed better in discriminating the T1-T4 continuum than the T2-T3 continuum; this difference was significant for both groups ($F(1, 70) = 9.945$, $p = 0.002$ for the Mandarin group; and $F(1, 70) = 10.807$, $p = 0.002$ for the Thai group). However, despite their high performance levels, for the English group, the effect of the continuum was not significant ($F(1, 70) = 0.264$, $p = 0.609$). This pattern suggests that for the Mandarin and Thai groups, the T2-T3 continuum was more confusable than the T1-T4 continuum, whereas for the English group, the two continua were equally confusable or distinctive. In the following section, I will present the results regarding the emergence of discrimination peaks and the correspondence of the peaks and the boundary locations found in the identification tests.

Fig. 4-3 (A-B). Two-step discrimination curves for the Mandarin, Thai, and English groups in the two continua: (a) in T1-T4, (B) in T2-T3.



As shown in Figure 4-3 (A-B), discrimination peaks can be found on some curves, but not on others. A discrimination curve for a two-category continuum usually has one peak point or two continuous peak points (Hallé et al., 2004). In other words, a bell-shaped curve is the prerequisite for the presence of discrimination peak(s). For this reason, the discrimination of T2-T3 by the Thai group does not meet the criterion for categorical perception, since the Thai discrimination curve shows three climaxes.

However, the points on the other discrimination curves with the highest correct responses are not necessarily the “peak points” defined for categorical perception. To qualify as a categorical peak point, the discrimination level associated with a given peak should be significantly higher than that associated with all the stimulus pairs, including the two adjacent pairs. An ANOVA showed that neither of the peak points on the

discrimination curves for the English group represented a significantly higher level of discrimination than any of the other stimulus pairs ($F(5, 30) = 1.76, p = 0.153$; and $F(5, 30) = 0.52, p = 0.76$ for the T1-T4 and T2-T3 continua, respectively). For the Thai group, the location of the peak point on the discrimination curve of the T1-T4 continuum is somewhat difficult to interpret. As shown in Figure 4-3 (A), the peak point falls on the stimulus pair 3-5, which is associated with an 87% accuracy level; the adjacent stimulus pairs are 2-4 and 4-6. ANOVA first showed that the discrimination level varied significantly according to the stimulus level ($F(5, 30) = 9.119, p < 0.0001$). Furthermore, an LSD post hoc test revealed a significant difference between the percent values of pairs 3-5 and 4-6 ($p = 0.023$), but not between the values for pairs 3-5 and 2-4 ($p = 0.912$). The results for the English group suggested that for these listeners, the perception of both these Mandarin tone contrasts was non-categorical. By contrast, it seems that the Thai listeners had no difficulty in categorizing Mandarin T4 (the endpoint tone on the right), but they had problems in categorizing Mandarin T1 (the endpoint tone on the left).

The Mandarin group shows a discrepancy in the discrimination of the native tone contrasts T1-T4 and T2-T3. Although both discrimination curves show peak points (pairs 2-4 (87%) and 3-5 (82%) for the T1-T4 continuum; and pairs 3-5 (70%) and 4-6 (74%) for the T2-T3 continuum), these discrimination peaks are only significant for the T1-T4 continuum, not for the T2-T3 continuum. An ANOVA showed that the stimulus effect is significant for the T1-T4 continuum ($F(5, 30) = 14.974, p < 0.0001$). The post hoc LSD indicated that the discrimination accuracy rate of pair 2-4 is marginally significantly

higher than that of pair 1-3 ($p = 0.091$); for pair 3-5, which also corresponds to a peak point, the LSD post hoc showed a significantly higher discrimination accuracy than pair 4-6 ($p = 0.003$). For the T2-T3 continuum, no significant difference was found between the discrimination levels of the peak points and any of the stimulus pairs in the continuum ($F(5, 30) = 1.923, p = 0.12$). These results are counter-intuitive, given that lexical tone phonemically differentiates word meanings for native Mandarin speakers. In other words, it would have been reasonable to expect that significant peak points would have occurred in the discrimination curves of both tone continua for the Mandarin group, rather than for merely one of the continua.

The last question concerns the correspondence between the locations of the peak points and the identification intercepts. Since the perception of only the T1-T4 contrast was potentially categorical for the Mandarin and Thai groups, other findings are not discussed here. As shown in Table 4-1, the locations of the intercepts on the identification curves for the T1-T4 continuum are around 3.72 and 3.76 (in stimulus number), for the Mandarin and Thai groups, respectively. The peak points fall between stimulus pairs 2-4 and 3-5 for the Mandarin group, and on the stimulus pair 3-5 for the Thai group. Therefore, the identification intercepts and discrimination peaks tend to coincide. From this perspective, it is likely that for the Mandarin and Thai groups, the perception of the T1-T4 contrast tends to be categorical. However, the situation is different between the Mandarin and the Thai groups. As revealed in Study I, Mandarin T1 and T4 were predominantly identified as Thai mid tone and falling tone, respectively, by the Thai

listeners. Thus the Thai listeners were able to distinguish Mandarin T1 from T4 and to identify the boundary between the two tones. However, due to the absence of a tone identical to Mandarin T1 in Thai, the Thai listeners were not able to recognize Mandarin T1 as a native category (see the above analyses). In this sense, despite the coincidence of the peak point with the identification intercept, the perception of the Mandarin T1/T4 contrast is non-categorical for the Thai listeners.

To sum up, the results of the discrimination tests showed that the discrimination of T1 and T4 was categorical only for the Mandarin group, but not for either the Thai or the English group. Nevertheless, the performance of the Thai group differed from that of the English group to some extent. Due to the existence of the falling tone in Thai, which corresponds to Mandarin T4, the Thai listeners assimilated this Mandarin tone category to a Thai tone category. However, the Thai group failed to assimilate the other endpoint tone, Mandarin T1, to a Thai category. For English speakers, neither of the two tones was categorized.

The results of the discrimination test for T2 and T3 indicate that none of the groups succeeded in perceiving the two tones categorically, not even the native Mandarin listeners. An explanation for the non-categorical perception of T2 and T3 contrast by the Mandarin listeners will be proposed in Chapter 5. As discussed in Chapter three with respect to Study I, due to the correspondence between Mandarin T2 and T3 and the Thai tone system, the Thai listeners identified T2 and T3 as a single category and therefore found it difficult to perceive the difference between these two tones. Discrimination

between T2 and T3 was significantly better for the English group than for either the Mandarin or Thai group, an effect which is likely attributable to the fact that many English participants (four out of six) had a background in music. Nevertheless, it is unlikely that they perceived T2 and T3 as distinct phonetic categories.

4.3.2.2 Comparison between trained and untrained subjects

As discussed above, tone perception training has a tendency to improve performance in identification tasks. The effect of tone training is even more apparent in discrimination tasks. Table 4-5 shows a consistent pattern of improved discrimination performance across the language groups and the tone continua for those participants who received training in tone perception.

Table 4-5. Mean correct discrimination levels (%) for trained and untrained Mandarin, Thai, and English listeners.

Untrained		Mandarin	Thai	English	Average
	T1-T4	68	67	72	69
	T2-T3	61	58	70	63
Average					66
Trained		Mandarin	Thai	English	
	T1-T4	74	74	74	74
	T2-T3	63	62	73	66
Average					70

As can be seen, the trained groups, whether native or non-native, performed better in discriminating both of the tone contrasts than the untrained groups. The training effect is statistically significant for the pooled results across all subjects and continua (66% vs.

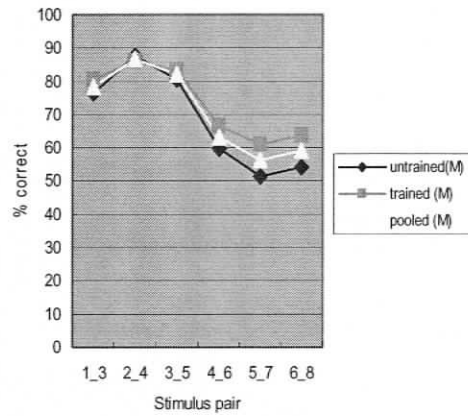
70% for the untrained and trained groups, respectively ($F(1, 214) = 4.051$ $p = 0.045$).

Although the effect of training was not statistically significant for any individual group, subjects who received training in each group consistently improved in discriminating both tone continua. For the trained Mandarin group, discrimination increased from 68% to 74% for the T1/T4 contrast, and from 61% to 63% for the T2/T3 contrast. For the trained Thai group, discrimination improved from 67% to 74% for the T1/T4 contrast, and from 58% to 62% for the T2/T3 contrast. For the trained English group, discrimination levels increased from 72% to 74% and from 70% to 73% for the T1/T4 and T2/T3 contrasts, respectively.

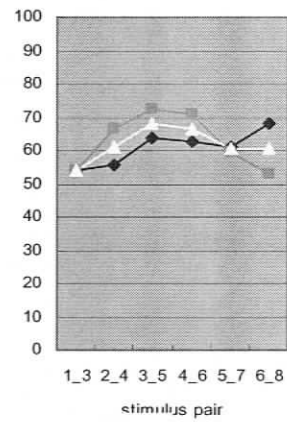
However, these increased discrimination levels do not necessarily represent increases in categorization. Therefore, it is important to determine whether or not the location and the relative level of the peak points (see above discussion) differs between the trained and untrained subjects. The panels below (Fig.4-4 (A-C)) show the discrimination curves for the trained and untrained subjects in each of the three language groups for both the T1-T4 and the T2-T3 continua.

Fig. 4-4 (A-C) Discrimination curves for trained and untrained subjects in Mandarin (Panel A, M refers to Mandarin), English (Panel B, E refers to English), and Thai (Panel C, T refers to Thai) groups: (a) on the T1-T4 continuum; (b) on the T2-T3 continuum.

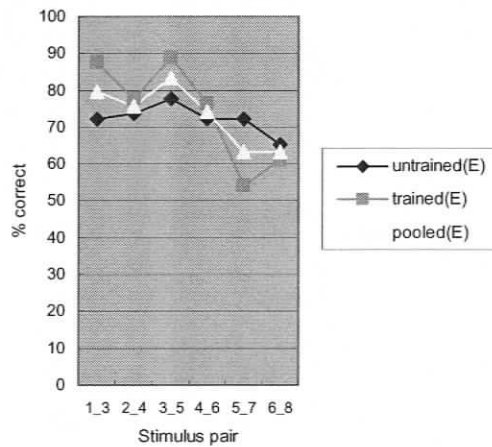
(A) (a)



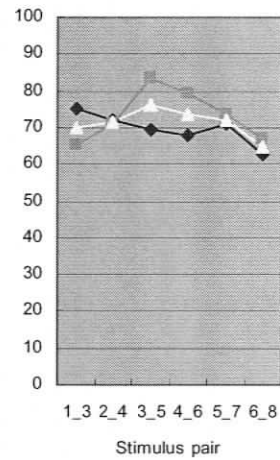
(b)

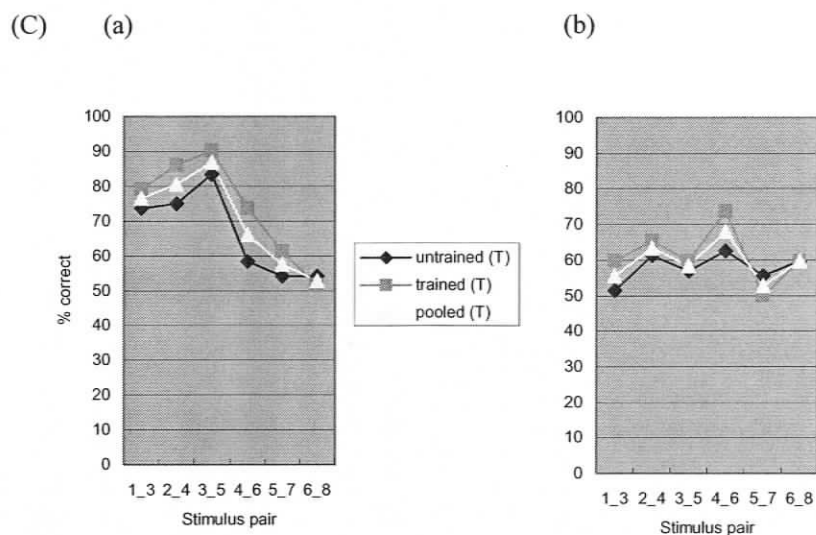


(B) (a)



(b)





The discrimination curves in each panel revealed that: (1) the trained groups performed with a higher accuracy level compared to the untrained groups; and (2) the peak point(s) fell on the same positions (stimulus pairs) for the trained and untrained groups. Thus, training did not significantly affect the locations of the peak points, i.e., the training did not produce any new peak points. The statistical results also showed that the difference between the discrimination level of the peak points and their adjacent points for the trained groups was consistent with the pooled results noted above. Nevertheless, for the Mandarin group, the relationship between the discrimination level and the stimulus pair was found to be marginally significant for the trained group, but not for the untrained group or the pooled group, in the discrimination of the T2-T3 continuum ($F(5, 12) = 2.53, p = 0.087$ vs. pooled result: $F(5, 30) = 1.923, p = 0.12$). Overall, these results suggest that the training improved discrimination performance to some degree. However, training did not significantly increase categorical perception across all the language

groups and the tone continua despite the fact that some improvement occurred in the Mandarin group on the categorization of the T2-T3 continuum.

4.4 Summary and discussion

In this chapter, the perception of two tone continua was discussed in order to determine whether L1 background, perceptual tone training, and musical background affect Mandarin tone perception. The results of the discrimination and identification tests for the T1-T4 and T2-T3 continua are summarized below.

For the T1-T4 continuum, the Mandarin listeners succeeded in categorizing the two endpoint tones. In the identification test, the Mandarin group performed better than the other two groups with a steeper slope (1.64 vs. 1.24). The location of the identification intercept also coincided with the discrimination peaks. The discrimination level for the T1-T4 continuum was also fairly high for the Mandarin group. These phenomena are representative of the phenomenon of categorical perception. The performance of the Thai group reflected their L1 tone language background: the discrimination level at the boundary was significantly higher than the adjacent pair on the right side; and the right endpoint tone was T4 (falling tone), which has an equivalent in Thai. This pattern suggests that the Thai listeners perceived T4 as a native or near-native category. The fact that their discrimination of the stimulus pair at the left side of the boundary did not show any significant difference suggests that they had difficulties in categorizing Mandarin T1 as a native phoneme. The English listeners' perception of the T1-T4 contrast yielded the

same result as found in Hallé et al. (2004) and Wang (1976): listeners who are native speakers of non-tone languages are unable to categorize lexical tones.

Furthermore, the shapes of the identification curves for the T1-T4 continuum indicate there is a discrepancy between native and non-native listeners in the identification of token 6 along the T1-T4 continuum. The native listeners tended to classify token 6 as a falling tone (token 8) more often than the non-native listeners. An account of this phenomenon will be presented in Chapter five.

The results for the T2-T3 continuum appear to be consistent across the three groups. Neither the native nor the non-native listeners succeeded in categorizing T2 and T3. The Mandarin and Thai listeners (native speakers of tone languages) performed much worse than the English listeners (native speakers of a non-tone language), both in terms of identification (1.32 and 1.21 vs. 1.51) and discrimination (61.89 and 57.67 vs. 71.5). The latter was very likely due to the participation of subjects with music training in the English group, since the results of the discrimination tests showed that the Mandarin and Thai listeners outperformed the English listeners without music training. This finding to some extent supported the influence of previous tone experience on tone discrimination.

The discrimination and identification data for the trained and untrained subjects reflected the influence of perceptual training on tone perception. The trained subjects were better than the untrained subjects in the identification and discrimination tests. Nevertheless, the data did not significantly affect the status of the discrimination peak points between the trained and untrained groups, especially for the non-native groups.

This finding suggests that brief perceptual training plays a positive role in increasing general discrimination and identification levels, but does not obviously produce categorical perception of tone.

At the end of Chapter three, I made predictions for Study II based on the findings of Study I, in which the native Thai listeners classified Mandarin T1 and T4 as Thai mid tone and falling tone, respectively, but classified Mandarin T2 and T3 as Thai rising tone. Based on these correspondences, I assumed that Thai listeners would perceive Mandarin T1 and T4 categorically, but perceive Mandarin T2 and T3 non-categorically. While the latter hypothesis was confirmed in Study II, there is a discrepancy between the results of Study I and Study II with respect to the former prediction. Nevertheless, as noted in the discussion of Study I, although Mandarin T1 was assimilated to the Thai mid tone, the phonetic distinctions between the two tones are also obvious, especially F_0 height. In Chao's description, Mandarin T1 has the highest pitch level (55), but Thai mid tone starts from a mid-level pitch and falls to a low-level pitch (32). Therefore, the response of the Thai listeners in Study I may be interpreted as meaning that for the Thai listeners, the mid tone sounds closest to the Mandarin T1 compared to other Thai tones. This perception does not necessarily mean, though, that the similarity between the Thai mid tone and the Mandarin high level tone is comparable to the similarity between the Thai falling tone and the Mandarin falling tone. Overall, the results of Study I are consistent with the Thai listeners' categorization of Mandarin T1 and T4 in Study II: Thai listeners perceived Mandarin T4 as a native or native-like category, but perceived Mandarin T1 as a

non-native category, which helps to explain why their perception on the T1 and of the T1-T4 continuum was not categorical.

Chapter Five

GENERAL DISCUSSION

The results presented in Studies I and II combine to reveal the effects of L1 background, perceptual tone training, and musical experience on the perception of lexical tones. It was found that the Mandarin listeners perceived one native tone contrast (T1/T4) categorically, but perceived another native tone contrast (T2/T3) non-categorically. The Thai listeners' perceptual patterns were based primarily on the correspondence between the Mandarin and Thai tone systems. That is, they perceived the familiar Mandarin T4 categorically and the unfamiliar Mandarin T1 non-categorically. In addition, since Mandarin T2 and T3 both correspond to a single category in Thai (see Chapter 2); the Thai listeners perceived these two tones non-categorically. Although the English listeners exhibited fairly strong sensitivity to the differences between tone contrasts, their perception was non-categorical for both tonal contrasts because of the lack of lexical tonal distinctions in English. However, a number of issues regarding the findings remain to be discussed. This chapter groups the discussion into two main issues: the question of the categorical perception of tone and that of the tone assimilation patterns.

5.1 Categorical perception of tone

Francis et al. (2003, p1042) propose three main factors to account for the perception patterns that have emerged in studies of the categorical perception of tone: (1) the presence of regions of natural auditory sensitivity; (2) learned associations between particular acoustic (F_0) patterns and linguistic categories; and (3) the estimation of a speaker's tone space on the basis of intrinsic and extrinsic acoustic information. These three factors are discussed below.

According to Holt et al. (2004) and Kuhl & Miller (1975, 1977), regions of natural auditory sensitivity are coincident with the location of phonetic categories. However, not all category boundaries coincide with regions of equal sensitivity. For example, for lexical tones, the transitions from level to level and from low- rising to falling- rising pitch contours are less auditorily salient than those from falling to rising and from rising to level pitch contours (Klatt 1973; Schouten 1985). This factor may help explain why the perception of level tone contrasts is non-categorical, as found by Abramson (1979) and Francis et al. (2003).

The second proposed factor that influences tone perception is learned association between particular acoustic (F_0) patterns and linguistic categories. In a review of the findings presented in Wang (1976, see Chapter 2, section 2.3.2.1), Francis et al. (2003) pointed out that speakers with acquired lexical tone experience are likely to perceive tones categorically, while speakers without such experience are unlikely to do so. They also proposed that Mandarin speakers with advanced psychophysical training experience

perceived native tone contrasts primarily on the basis of auditory sensitivity. In other words, their categorization of the native tones was negatively affected by their training experience.

Finally, extrinsic factors, such as the context of tone co-occurrence, are also proposed to account for tone perception patterns. For example, in distinguishing tone contrasts that lack intrinsically contrasting acoustic properties, such as level tone contrasts, listeners may rely on extrinsic information, such as a speaker's pitch range. Such information may help listeners to locate category boundaries.

In discussing the results of the present study, I will draw upon the three factors¹ proposed by Francis et al. (2003). However, given the complexity of the participant groups in this study (Mandarin, Thai, and English) and the different tone combinations examined herein (level/falling and rising/falling rising tones), it is unlikely that any one of these three factors can fully account for my findings. Therefore, I suggest that the presence of auditory sensitivity, acquired F_0 patterns (L1 tone experience), and learned acoustic experience interact to create various tone perception patterns.

5.1.1 Presence of auditory sensitivity

According to previous studies (Holt et al., 2004; Williams, 1977; Kuhl & Miller, 1975, 1977), regions of auditory sensitivity are associated with category boundaries across human and non-human listeners: regions of perceptual discontinuity tend to

¹ This study does not examine the third (context) factor since it is not used as a variable.

coincide with the locations of category boundaries and thus the perception of those contrasts with salient category boundaries tends to be categorical. This sensitivity, however, varies with the acoustic realization of the phones. With respect to tonal contrasts, the factor of auditory sensitivity is more salient if the transition is from rising to level or from falling to rising pitch contours (Klatt, 1973; Schouten, 1985; Francis et al., 2003). The order of the presented tone contours may not be crucial. As noted by Klatt, "greatest sensitivity occurs when one ramp (tone contour) increases and the other decreases" (1973, p. 8). By contrast, tones are more difficult to distinguish when presented with two level contours or two differently rising contours (Wang, 1976; Francis et al., 2003). However, the results of Hallé et al. (2004) did not support this hypothesis in that the native Mandarin speakers perceived the Mandarin T1-T2 (level to rising) and T2-T4 (rising to falling) tone contrasts categorically, while the non-native French speakers perceived both tone contrasts non-categorically. This result suggests that even though the non-native speakers had detected the category boundary, presumably because of the heightened natural sensitivity at this location, they were still not able to categorize the non-native tones. Wang (1976) reported a similar finding: the Mandarin rising and level tone contrast (presented from rising tone to level tone) was not categorized by the English speakers. Neither of the tone contrasts examined in the present study (T1-T4 and T2-T3) contain the aforementioned auditorily salient pitch contour transitions: level to rising contour and falling to rising contour. The alternation from T1 to T4 is from a level to a falling contour and the transition from T2 to T3 is from a rising to a falling-rising contour. The results of

Study II, however, were consistent with those reported in Hallé et al. (2004) and Wang (1976), in that the perception of lexical tone contrasts by non-tone language speakers tended to be non-categorical. Therefore, it seems that the presence of regions of auditory sensitivity cannot clearly account for the failure of the English listeners to categorize the lexical tones. Nevertheless, in line with the proposals of Abramson (1977) and Francis et al. (2003), the presence of auditory sensitivity does help to still provide an account for the perceptual patterns exhibited by the native speakers of tone languages when presented with tone transitions with less intrinsic information, such as level tone contrasts and differently rising tone contrasts.

5.1.2 Acquired F_0 pattern

Although native and non-native speakers and even human and non-human listeners share the same auditory system, the results of previous studies and the present study indicate that a psychophysical explanation is not sufficient to account for all tone perception patterns. As noted by Hallé et al. (2004), tone perception is influenced by linguistic experience, especially L1 tone experience. Native speakers of tone languages tend to perceive tones categorically, despite the confounding influence of other variables (Wang, 1976). By contrast, native speakers of non-tone languages, such as French and English, do not usually perceive lexical tones categorically (Hallé et al., 2004; Wang, 1976). The results of the present study confirm this tendency. Although the native Mandarin listeners succeeded only in categorizing T1 and T4, but not T2 and T3, the

English listeners perceived neither tone contrast categorically. Zue's (1976) findings suggested that English listeners perceived tones no differently than native listeners. As cited by Gandour (1978) and Blicher et al. (1990), Zue's study showed both Mandarin and American English listeners perceived the Mandarin T2-T3 contrast categorically. The perceptual patterns of the Mandarin and English listeners were also comparable in terms of the location of the category boundary and its correspondence to the discrimination peak. However, as pointed out by Shen and Lin (1990), "the suitability of the physical continuum that he [Zue] used is questionable, because the portion before the rise actually is a falling instead of a sustained F_0 for both Tone 2 and Tone 3 (as used by Zue)" (p. 146). Therefore, it is arguable that the perception of "T2" and "T3" in Zue's study was purely psychophysical and that none of the stimuli were in fact categorized as Mandarin tones [cf. the base tones on T2-T3 continuum in this study and Shen and Lin (1990)]. The similar perceptual patterns of the Mandarin and English listeners, then, reflected the fact that they all divided the psychophysical continuum into a "predominantly rising" group of stimuli (tokens 1-4) and a "predominantly level" group of stimuli (token 6-9) (Gandour, 1978, p61; see also the literature review in Chapter two for a discussion of how the tone continuum was generated in Zue, 1976). Nevertheless, Zue's (1976) study is valuable because it suggests the influence of auditory sensitivity: if the surveyed tones are not native to any group of subjects, perception patterns may be consistent across native and non-native listener groups.

The factor of acquired F_0 patterns can help account for the perceptual patterns seen

among the Thai listeners in this study. As indicated in Study I, the Mandarin high level and falling tones have counterparts in Thai. However, only the falling tone has an identical equivalent (with regard to F_0 contour and height), whereas the high level tone has a counterpart with only a similar F_0 contour (see Chapter three). Accordingly, as seen in Study II, the Thai listeners succeeded in categorizing the Mandarin falling tone, but not the Mandarin high level tone. Their perception of the Mandarin T2-T3 contrast provides further evidence for the influence of acquired F_0 pattern on tone perception. Mandarin T2 and T3 correspond to a single tone in Thai. Thus, Thai listeners would tend not to classify these two tones as distinctive categories.

The influence of acquired F_0 patterns can also explain why, in Study I, the Mandarin listeners failed to categorize the native T2-T3 tone contrast. This failure may be attributable to the partial absence of the perceptual cues that are necessary to distinguish T2 and T3. Although it is widely accepted that fundamental frequency plays a critical role in tone perception (Abramson 1979; Lin 1988), other perceptual cues, such as duration and the location of the turning point, have been proposed as required cues in the perception of the Mandarin falling-rising tone and the Mandarin rising tone, which shows a falling-rising contour in natural speech (Blicher et al., 1990; Shen & Lin 1990). Shen and Lin (1990) proposed a threshold for the timing of the turning point, based on the results of their study, in which they presented native Mandarin speakers with two T2-T3 tone continua which differed in the location of the turning point (see Chapter two). The results of the acoustical analysis showed that “for Tone 2, the average occurrence of the

turning point is at 14% of the total duration and the mean ΔF_0 between tonal onset and the turning point is 17.5 Hz; for Tone 3 the average occurrence of the turning point is 48% of the total duration and the mean ΔF_0 between tone onset and the turning point is 38.6Hz” (Shen & Lin, 1990, p152). In the present study, the tone duration was equalized to 300ms for all four tones; T2 starts from 208ms; and the F_0 is 244.6 Hz. After a period of falling, the pitch rises at 288ms and the F_0 at the turning point is 214.83Hz. T3 also starts at 208ms and the F_0 is 212.75Hz. The turning point occurs at 328ms and the F_0 is 165.5Hz (see Table. 3-6 (B)). In this study, then, for T2, the occurrence of the turning point is at 27% of the total duration and the F_0 between the tonal onset and the turning point is 29.8Hz. For T3, the occurrence of the turning point is at 40% of the total duration and the F_0 between the tonal onset and the turning point is 47.25. Therefore, based on Shen and Lin’s (1990) criteria, the T2 and T3 stimuli presented to the subjects in this study are out of the “strictly defined” locations specified for T2 and T3, as shown in Table 5-1.

Table 5-1. Comparison of the occurrence of the turning point (Occurrence) and F_0 between the tonal onset and the turning point (F_0) of T2 and T3 in Shen and Lin (1990) and the present study.

	Tone 2		Tone 3	
	Occurrence	F_0 (Hz)	Occurrence	F_0 (Hz)
Shen and Lin (1990)	14%	17.5	48%	38.6
This study	27%	29.8	40%	47.25

In sum, because of the partial loss of perceptual cues (duration differences and the timing of the turning points), the tokens on the T2-T3 continuum in this study were

difficult for the native Mandarin speakers to identify and discriminate accurately.

However, this finding makes an important contribution to the body of knowledge on tone perception in helping to identify the most confusable aspects of the T2-T3 tone contrast.

Nevertheless, the partial lack of perceptual cues cannot account for the similar pattern of non-categorical perception of the T2-T3 tone contrast for the Thai and English listeners. For the Thai listeners, the Mandarin rising tone and falling rising tone would likely be categorized as the Thai rising tone regardless of the locations of the turning points, since it is primarily the similar F_0 values that the Thai listeners would attend to in assimilating the non-native to the native tones. For the English listeners, there was no difference in the perception of the T1-T4 and T2-T3 continua in terms of the identification function, discrimination accuracy level, or occurrence of peak points. Therefore, as mentioned above, the lack of tone distinctions in their L1 is the primary reason for their failure to categorize the Mandarin tones.

The results of the identification tests in Study II revealed a discrepancy between the native and non-native groups' ability to identify tokens along the T1-T4 continuum (Fig. 4-1. (A)). The Thai and English groups identified token 6 as token 1 9.72% and 18.06% of the time, respectively; only 4.17% of Mandarin listeners gave this response. In other words, token 6 sounded more like T4 (token 8) to the Mandarin listeners compared to the Thai and English listeners. As shown in Table 3-6 (A), tokens 1-6 start with a slight falling contour and end with a rising contour. By contrast, tokens 7 and 8 consist of a straight falling contour. The difference between tokens 6 and 7 was observed by both the

Thai and English listeners, as reflected in the fact that the percent response of token 6 as token 1 was relatively high for both groups. Fig. 4-1 (A) and Table 3-7(A) are repeated here as Fig. 5-1 and Table. 5-2.

Fig. 5-1. Identification curves of the Mandarin, Thai, and English groups on the T1-T4 continuum.

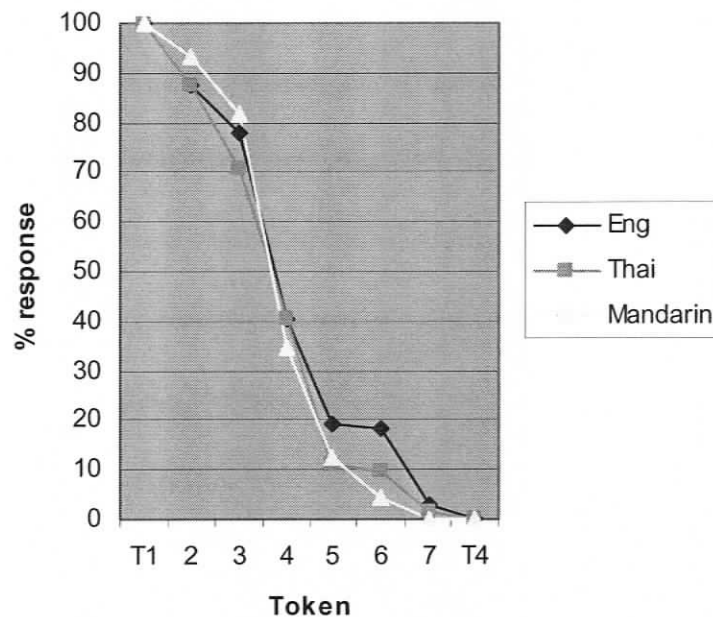


Table 5-2. The first and last points and turning points (bold) along tone contours as well as the occurrence time in the T1-T4 continuum.

Point	Time (ms)	t1(T1) (Hz)	t2 (Hz)	t3 (Hz)	t4 (Hz)	t5 (Hz)	t6 (Hz)	t7 (Hz)	t8(T4) (Hz)
1	230	313.13	316.18	319.22	322.27	325.32	328.37	331.41	334.46
23	450	279.82	271.07	262.32	253.57	244.82	236.07	227.32	218.57
27	490	281.15	268.71	256.28	243.85	231.42	218.98	206.55	194.12
28	500	285.61	271.30	256.98	242.66	228.34	214.02	199.71	185.39
29	510	298.24	280.92	263.61	246.30	228.99	211.68	194.36	177.05
31	530	316.15	295.34	274.53	253.71	232.90	212.09	191.28	170.47

Xu (2002) provides one potential explanation for the Mandarin listeners' relative

insensitivity to the end rise of token 6. Xu (2002) argued that “if something is not observed in the acoustic signal, either it is linguistically unimportant, or it is deliberately avoided due to perceptual constraints” (p. 91). As shown in Table 3-6 (A), the end rise at token 6 occurs at 280ms (from 230ms to 510ms) on a 300ms- length tone. Therefore, native speakers might have ignored the end rise portion in token 6 because it was not linguistically important. On an eight-token continuum, token 6 was thus identified as token 8. This insensitivity towards the end of the continuum was characteristic only of the native Mandarin listeners: the non-native listeners identified t6 as t1 more often than the native listeners. Therefore, it is likely that the discrepancy between the identification curves of the native versus non-native listeners reflects the effect of L1 experience.

5.1.3 Learned acoustic experience

Learned acoustic experience is the third factor that can account for categorical and non-categorical perception of tone contrasts. Learned acoustic experience includes musical or psychophysical training and any other experience that serves to enhance listeners' sensitivity to acoustic features. Wang (1976) found that native Mandarin speakers differed in the degree to which they were influenced by their native language in categorization tasks. Speakers who had never participated in psychophysical experiments (untrained subjects) succeeded in categorizing native tone contrasts, as reflected in the presence of discrimination peaks at category boundaries. By contrast, highly trained speakers no longer perceived native tone categories phonemically. Rather, they exhibited

high discrimination accuracy along the tone continuum, but no peak at the category boundary.

Musical training is known to influence tone perception. Studies by Gottfried et al. (2000, 2004) and Alexandra et al. (2005) both found that musicians have an advantage in discriminating and identifying natural Mandarin tones compared to non-musicians. In the present study, the effects of musical training among some of the English participants were clearly evident in the discrimination task (see Table. 3-7). The performance of musicians has not previously been discussed within the framework of categorical perception. The question of whether musicians' performance is closer to that of native speakers in terms of categorical perception is worth considering. In analyzing the results of this study, this question was discussed (see Fig. 4-2 in chapter four and Table 5-2 here). I found that while English musicians outperformed English, Mandarin, and Thai non-musicians in the discrimination of synthesized tones in a continuum, they had no advantage in identifying these tones. Thus, my tentative conclusion is that the lack of tone distinctions in their native language prevented the musicians from perceiving tones categorically. In this respect, their performance was no different from that of non-musicians.

Table 5-3. Intercepts and slopes for English listeners with and without musical training.

	T1-T4		T2-T3	
	Intercept	Slope	Intercept	Slope
Musicians	4.34	0.85	4.1	1.46
Non-musicians	3.51	1.44	4.13	1.61

Relative to the influence of musical training on the categorical perception of tone, the perceptual tone training conducted in this study showed an explicit positive effect on both tone discrimination and identification. The results pooled across all listeners show significantly higher discrimination accuracy for the trained participants. The tone training was also associated with improved performance in the identification of tones, i.e. the slopes at the boundary are steeper. The intercept of category boundaries for the trained Thai and English participants was closer to that of the trained Mandarin subjects compared to the untrained participants (see Table 4-2). These results are consistent with the findings of Best and Strange (1992), who found that the intercept and slope of category boundaries and discrimination levels for experienced Japanese subjects were intermediate between those of American listeners and inexperienced Japanese listeners in the perception of English approximants (for a review of this study, see Chapter two, section 2.2.1). These results suggest that perceptual training with natural speech stimuli differs from psychophysical training (Wang 1976) or musical training. Perceptual training with natural stimuli is closer to a developmental learning experience, in which learners are exposed to natural speech materials. With the benefit of such training, participants appear to perform more like native listeners relative to non-native participants who do not receive the training.

5.1.4 Summary

Drawing on the proposals of Francis et al (2003), in this section I have described

three factors that influence categorical tone perception: psychophysical factors, L1 tone experience, and training (tone learning). With respect to psychophysical factors, regions of auditory sensitivity tend to coincide with the location of tonal category boundaries, although this effect is more salient for some tone transitions than others. The transitions from level to rising and from falling to rising are auditorily salient, while the transitions from level to level and from low-rising to high-rising are less salient. The latter contrasts contain less intrinsic acoustic information at the category boundary; thus, it is difficult for listeners to perceive them as distinctive categories. This study also presented some interesting findings in relation to the second factor, L1 tone experience. The native Mandarin listeners perceived tone contrasts phonemically, but their perception was affected by the presence of perceptual cues. The non-native listeners with a different L1 tone language (Thai) perceived categorically only those tones with an equivalent F_0 in their own native phonological system. When presented with tones that did not correspond to tones in their native language, the Thai listeners performed no better than the English listeners who were not able to categorize tones because of the absence of lexical tones in their phonological system. With respect to the third proposed factor, training (learning), this study suggests that there is a difference between natural tone training and purely psychophysical training (e.g., training in detecting general pitch differences) or musical training. While musical and psychophysical training improve tone discrimination, they have no effect on tone categorization. By contrast, natural tone training was found to potentially assist all listeners in tone categorization, bringing their performance closer to

that of native speakers. These three factors interact to influence listeners' perception and categorization of tones.

5.2 Perceptual Assimilation Model (PAM)

5.2.1 Perceptual assimilation patterns

The Perceptual Assimilation Model (PAM) developed by Best and her colleagues (Best et al., 1988; Best, 1994; Best, 1995) attempts to account for assimilation patterns in terms of the phonetic correspondence between native and non-native phonetic categories. The original model incorporates four possible assimilation patterns: the "two categories" (TC) type, the "single category" (SC) type, the "category goodness" (CG) type, and the "nonassimilable" (NA) type. In the TC type, there is a direct correspondence between two categories in the L1 and the L2, while in the NA type, there is no correspondence between a particular L1 and L2 category since the sounds presented are not heard as speech. In the SC pattern, two L2 phones are perceived as equally good tokens of a signal L1 category, and are therefore assimilated to it. In the CG pattern, two L2 phones are assimilated to a single L1 category, but differ in the degree of phonetic similarity to the L1 phone. PAM predicts that in the perception of phones, those in which the correspondence between L1 and L2 are of the TC type pose no difficulty for adult non-native listeners, while those exhibiting the SC pattern (rather than the NA type) are the most difficult to discriminate, with CG-type correspondences posing moderate difficulties. In the revised PAM model (Best, 1995), two more patterns are added to the

original four-pattern model: Both Uncategorizable (UU) and Uncategorized versus Categorized (UC). In the UU pattern, both non-native sounds fall within phonetic space but outside of any particular native category and can vary in their discriminability as uncategorizable speech sounds. In the UC pattern, one non-native sound is assimilated to a native category, while the other falls outside the phonetic space captured by the native categories. Discrimination for the UU pattern is expected to range from poor to very good, and to be very good for the UC pattern. In the following sections, I draw on PAM to account for the perception of the Mandarin tones by the Thai and English listeners.

5.2.2 Assimilation patterns of Thai listeners

The results of Study I indicate that the Thai listeners perceived the four distinctive Mandarin tones as three corresponding tones in Thai: Mandarin T1 and T4 correspond to the Thai mid and falling tones, respectively, while Mandarin T2 and T3 both correspond to the Thai rising tone. However, the status of T1 and T4 differed in that T4 has an equivalent in Thai, while T1 does not. Therefore, despite the fact that most of the Thai listeners labeled T1 as the Thai mid-level tone, they were not able to categorize this Mandarin tone. Hence, the Thai listeners' perception of Mandarin T1 and T4 fits the UC (Uncategorized versus Categorized) pattern proposed by PAM: Mandarin T4 was categorized, while T1 was not. According to PAM, discrimination is expected to be very good for UC assimilation. This prediction was confirmed in Study II: although the Thai listeners had difficulties in categorizing Mandarin T1, their successful categorization of

Mandarin T4 suggests that they found it easy to distinguish between these two tones. This result can be compared with the notion of two-category assimilation which is also associated with good discrimination between non-native phones. This current study found that even though the Thai listeners had no problem distinguishing Mandarin T1 and T4 when presented with individual tokens (as in Study I), their discrimination of the similar but not identical category (in terms of F_0 values) was challenged within a categorical perception framework. This finding conflicts with the pattern found in Best and Strange (1992), in which Japanese listeners were presented with English approximant contrast: /w-j/. Despite the dissimilarities in the incidence of /w/ and /j/ in English and Japanese, Japanese listeners not only perceived the /w-j/ contrast categorically, but they also showed peaks when discriminating tokens along the /w-j/ continuum, although the Japanese discrimination peaks were smaller than those of the American listeners (p.317). In my account, the discrepancy between the performance of the Japanese listeners in Best and Strange (1992) and the Thai listeners in the present study stems difference between the two assimilation patterns: TC (Two Categories) and UC (Uncategorized versus Categorized). Although both patterns lead to good discrimination, the two endpoint stimuli along a continuum can be categorized only if they fit the TC pattern. In other words, if they fit the UC pattern, the perception will be non-categorical.

The Thai listeners' perception of the Mandarin T2-T3 contrast is perfectly accounted for within PAM. In Study II, Thai listeners categorized both T2 and T3 as the Thai rising tone, fitting the single-category assimilation pattern (SC). As predicted by PAM, this

assimilation pattern is associated with poor performance in discrimination. Study II showed that the Thai listeners perceived Mandarin T2 and T3 non-categorically, echoing the results of Study I.

5.2.3 Assimilation pattern of English listeners

The existing literature contains very few discussions of PAM in relation to tone category assimilation by speakers of non-tone languages (Hallé et al., 2004; Ko, 2005a, 2005b). Hallé et al. (2004) point out PAM's limitations in characterizing French listeners' perceptions of lexical tones. Of the six proposed patterns, non-assimilable (NA) and uncategorizable (UU) come closest to capturing the perception of lexical tones by native speakers of non-tone languages. English listeners' perception of Zulu click consonants has been described as fitting the NA pattern (Best et al., 1988), in that English does not employ Zulu clicks and more importantly, in that English speakers do not hear them as speech. However, as pointed out by Hallé et al. (2004), pitch changes are employed in almost all languages of the world. Tone languages employ pitch not only at the sentential or phrasal level, but also at the lexical level. Non-tone languages, however, also use pitch changes, although they use them primarily to generate intonation patterns in phrases and sentences. In English, for example, different pitch contours are used to form five different intonation patterns: level, falling, rising, falling-rising and rising-falling (Halliday, 1967). Therefore, it is unlikely that tones are perceived purely as non-speech sounds, that is, unassimilable for English listeners. In the revised PAM, Best (1995) added the

uncategorizable (UU) and uncategorized versus categorized (UC) patterns to her original four-pattern model. Aoyama (2003) adopted these two “new” conceptions in her study. She labeled the English nasal contrasts /m-n/ and /m-ŋ/ in syllable-final position as the UC type and the syllable-final /n-ŋ/ as the UU type for native Japanese listeners. Her study confirmed PAM’s prediction that the UC pattern is associated with good discrimination, and the UU with poor discrimination².

Hallé et al. (2004) and So (2005a) classify tone assimilation as fitting the UU pattern for native speakers of non-tone languages (French and English, respectively). In their studies of Mandarin tone perception, untrained speakers of French or English performed worse than native Mandarin speakers and native speakers of other tone languages. Nevertheless, the French and English speakers were not completely insensitive to tone contrasts. While the French listeners failed to categorize any of the Mandarin tones presented in the identification and discrimination tests, they did exhibit psychophysical sensitivity to the Mandarin tones (Hallé et al., 2004). This phenomenon was also observed in the performance of some of the English listeners in the present study. The English subjects (with the participation of musicians) outperformed even the native Mandarin listeners in tone discrimination, despite the absence of tonal distinctions in their L1. Therefore, the non-tone language speakers’ performance on tone perception ranged from fair to good, as predicted by PAM for the UU assimilation pattern, depending on the

² The UU pattern is not necessarily associated with poor discrimination. Discrimination for the UU pattern is expected to range from poor to very good.

perceived salience of the phonetic differences involved.

5.2.4 Summary

In conclusion, the findings of the present study provide evidence for the applicability of the Perceptual Assimilation Model (PAM) to the study of tone assimilation. This study advances the investigation on the tone assimilation patterns of speakers of a different tone language (Thai) as well as a non-tone language (English) within the framework of PAM. Although segmental studies have widely employed PAM to explain assimilation of non-native phonetic features (Aoyama et al., 2004; Strange et al., 2004; Best et al., 2003; Best et al., 2001; Hallé et al., 1999; Best et al., 1995 etc.), very few studies have used PAM to account for the assimilation of suprasegmental features in a given tone language by speakers of different tone languages and/or by speakers of non-tone languages. In three different experiments (the correspondence experiment, discrimination experiment, and identification experiment), this study has provided data on the assimilation patterns of Thai and English listeners in the perception of Mandarin tones and has provided further support for PAM within the context of tone perception (categorization).

Chapter Six

CONCLUSION

In this section, I summarize the findings of Study I and Study II and highlight directions for future research on the categorical perception of tone. The final section of this chapter presents the overall conclusions of this thesis.

6.1 Summary of the findings

This thesis examined the perception of two Mandarin tone contrasts by speakers who differ in their experience with tone. Speakers of Mandarin and Thai are familiar with pitch changes at the lexical level in their native languages, while speakers of English are not familiar with this use of pitch. Other participant variables included perceptual tone training and musical training.

Native language background was found to be one of the most influential factors in tone categorization. The Mandarin listeners outperformed the Thai and English listeners in the categorization of the Mandarin tones, especially the T1/T4 contrast. Their perception of T1 and T4 showed discrimination peaks at the category boundary and deep

identification slopes.

Thai speakers performed well in discriminating the Mandarin T1/T4 contrast. According to the results of Study I, Mandarin T1 and T4 corresponded to the Thai mid level tone and the falling tone, respectively. In the categorical perception task, Mandarin T4 was perceived by the Thai listeners as an individual category, but T1 was not; the difference in the Thai listeners' perception of these tones was statistically significant. The English listeners failed to categorize the Mandarin tone contrasts despite the presence of high discrimination levels.

Mandarin, Thai, and English listeners all perceived the T2/T3 contrast non-categorically. The inability of the Mandarin speakers to categorize native tone contrasts may be attributable to the partial loss of perceptual cues. According to Shen and Lin (1990) and Blicker et al. (1990), the period from the tone onset to the turning point and the location of the turning point are crucial cues that distinguish T2 and T3. While the synthesized T2 and T3 used in this study retained the majority of the F_0 information available in natural speech, the critical cues noted above were lost (see Chapter 5, section 5.1.2).

While it is possible that the Thai listeners were also affected by the loss of these cues in their perception of the T2/T3 contrast, the results of Study I suggest a different explanation. According to Study I, both Mandarin T2 and T3 correspond to the Thai rising tone, which suggests that Mandarin T2 and T3 may not sound different to the Thai listeners. It is unlikely that the Thai listeners would be able to differentiate two

non-native tones which sound the same as a single native tone.

None of the above explanations accounts for the non-categorical perception of T2 and T3 by the English group. For the pooled English listeners, there was no statistically significant difference between the results of the discrimination and identification tests for the T1/T4 and T2/T3 contrasts. The English speakers' lack of L1 tone experience is the most likely explanation for their failure to make a categorical distinction between T2 and T3 as well as T1 and T4.

Musical experience was found to affect the performance of the English participants in the discrimination task. Participants with musical training significantly outperformed those without musical training. However, there was no significant difference between the location of the category boundary and the steepness of the boundary, which suggests that despite their good performance in discrimination, the English listeners with a musical background still perceived lexical tones non-categorically.

The effect of perceptual tone training was not as obvious as that of L1 background and musical training. Nevertheless, despite the fact that tone training did not produce statistically significant differences in the discrimination and identification tasks for individual language group, the pooled result and trend analyses showed an enhancement in performance: participants from all three language groups who received the training performed better in both discrimination and identification tests.

Based on the findings on the influence of native language background, musical experience, and natural tone training, three factors were proposed to account for the

gradient of categoricalness: regions of auditory sensitivity, acquired F_0 experience, and learned acoustic experience. The presence of auditory sensitivity at the location of the category boundaries did not help the non-tone language speakers in the categorization of Mandarin tones. This factor may be more meaningful in helping to account for the failure of native Mandarin speakers to categorization of level tones or differently rising tone contrasts. Acquired F_0 experience essentially reflects the listeners' L1 tone experience. This factor explains why listeners without L1 tone experience perceived tone contrasts non-categorically, whereas listeners with L1 tone experience perceived native tones or non-native tones that were identical or near identical in terms of F_0 values to native tones, categorically. This account provides evidence for the importance of sufficient perceptual cues for native listeners. The partial loss of perceptual cues is responsible for the non-categorical perception of the T2-T3 continuum by Mandarin speakers. This factor also suggests the importance of the information provided in the initial phase of presented tones to the native Mandarin listeners (Xu, 2002).

Learned acoustic experience affects tone categorization in two main ways. On the one hand, if the learning experience focuses on increasing listeners' sensitivity to pure pitch changes, as is the case with musical training and laboratory acoustic training, tone categorization is negatively or least affected. Participants with musical training (as in this study), as well as those with experience in phonetics experiments (as in Wang, 1976), perceived all the stimulus pairs at a generally high level (above chance level) but failed in categorizing tone contrasts. On the other hand, if listeners receive training in natural tone

perception, their performance in tone categorization is enhanced. The performance of trained non-native listeners, even those without L1 tone experience, begins to resemble that of native speakers.

The results of this study support the perceptual assimilation model (PAM) developed by Best and her colleagues (Best et al., 1988; Best, 1994; Best, 1995). Mandarin T4 was perceived as a native category by the Thai listeners, while T1 was not. Nevertheless, T1 was assimilated into a similar Thai tonal category. Therefore, the discrimination of the two tones was excellent, but it was difficult for the Thai listeners to categorize Mandarin T1. Mandarin T2 and T3, which assimilated into a single category by the Thai listeners, were poorly discriminated by the Thai listeners, as predicted for the single-category assimilation pattern (SC). For the English listeners, lexical tone contrasts may be perceived according to the “both uncategorized” (UU) pattern of assimilation. The results of Study II suggested this pattern, in that neither the T1/T4 nor the T2/T3 tone contrast was perceived categorically by the English listeners. Nevertheless, the English listeners were not insensitive to the Mandarin tone contrasts; their perception of the lexical tones was most likely psychophysically based.

6.2 Limitations and further studies

Although a large body of research has investigated the influence of linguistic experience on L1 and L2 speech perception, this study is the first to examine the categorization of the Mandarin tones by speakers from three different language

backgrounds: Mandarin, Thai, and English. Furthermore, this study is also the first to investigate tone perception by these speakers within the framework of categorical perception. Finally, this study is the first to discuss tone assimilation in relation to PAM. As such, this thesis makes an important contribution to our understanding of the categorical perception of tone.

Nevertheless, it is important to note the limitations of this study. The first limitation is the small number of participants. Twelve subjects were recruited for each language group. Within each of the language groups, six participants were assigned to the discrimination and identification of one of the two continua. Within these subgroups of six, three participants were selected to receive training in natural tone perception. Given the small number of subjects in each (sub-) group, it was difficult to obtain statistically significant results, which makes the findings less convincing. Therefore, future studies of categorical tone perception should recruit a larger number of participants.

The second limitation of this study is the minimal weight of the training in natural tone perception. While this training tended to enhance tone perception across all three language groups, the effect was not comparable to that reported in Best and Strange (1992), which revealed more obvious native-like categorization results for the experienced non-native speakers than for the inexperienced non-native speakers. In future studies, the training component, including the period of learning and/or variety of training, should be strengthened.

The third limitation relates to the control of perceptual cues. This study

unintentionally highlighted the importance of durational and other cues in the perception of the Mandarin T2-T3 contrast. However, future studies on tone categorization should take care to ensure that listeners are presented with tone stimuli that are in fact categorizable by native speakers or the examiner should explicitly test and control for this factor. In other words, the primary perceptual cues should be provided in the stimuli and researchers should consider that for some special phonetic features, other important cues may be involved.

Finally, although hemispheric lateralization of Mandarin tone by speakers of another tone language was not the focus of this study, the Thai listeners' perceptions of the Mandarin tone contrasts partially support and extend the dichotic findings reported in Wang et al. (2004, see section 2.3.3). In the latter study, Norwegian and English listeners processed Mandarin tones in the same way, despite the fact that Norwegian is a tone language and English is not. Future studies on Mandarin tone processing by Thai listeners could incorporate a dichotic experiment and compare the results to those reported in Wang et al. (2004). Based on the findings of the present study, I would expect that the results of the suggested experiment would be consistent with those reported in Wang et al. (2004).

6.3 Concluding remarks

This study examined factors that affect the perception of tones as phonemic categories. The findings of Study I and Study II confirm the importance of L1

background, a factor that has been widely discussed in previous research. However, L1 experience is not the only factor that influences tone perception. Innate psychophysical sensitivity and learned acoustic experience also affect tone categorization from different perspectives.

This study has also provided strong evidence for the perceptual assimilation model (PAM). Since it was first proposed by Best and her colleagues, this model has provided a framework for the design and discussion of many studies on the assimilation of segmental contrasts by non-native listeners (Aoyama et al., 2004; Strange et al., 2004; Best et al., 2003; Best et al., 2001; Hallé et al., 1999; Best et al., 1995 etc.). As demonstrated in this thesis, PAM makes many successful predictions on tone assimilation, providing further evidence for the analogous relationship between the perceptions of segmental and suprasegmental contrasts (Hsien, 2001).

REFERENCES

- Abramson, A.S. (1961). Identification and discrimination of phonemic tones. *Journal of the Acoustical Society of America*, 33, 842 (abstract).
- Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21(4), 319-325.
- Abramson, A. S. (1977). The noncategorical perception of tone categories in Thai. *Haskins Laboratories Status Report on Speech Research*, 51-52, 91-100.
- Abramson, A. S. (1974). The tones of central Thai: Some perceptual experiments. *Haskins Laboratories Status Report on Speech Research*, 39-40, 91-104.
- Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: cross-language tests. *6th ICPHS Prague*, 569-573.
- Abramson, A. S., & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Haskins Laboratories Status Report on Speech Research*, 29 /30, 15-25.
- Alexander, J. A., Wong, P. C. M. & Bradlow, A. R. (2005). Lexical tone perception by musicians and non-musicians. *Proceedings of Interspeech*. Lisbon, Portugal.
- Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research*, 19, 3, 251-265.
- Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *Journal of the Acoustical Society of America*, 116(3), 1739-1748.
- Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 207-232. Timonium, MD: York Press, Inc.
- 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*, 171-204. Timonium, MD: York Press, Inc.

- Best, C. T. (1994). The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. In Goodman, J. C. and H. C. Nusbaum (Eds.) *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, 167-224. MIT Press, Cambridge, MA.
- Best, C. T., & Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science*, 10(1), 65-70.
- Best, C.T., McRoberts, & Gerald W. (2003). Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways. *Language and Speech*. 46, 2-3, 183-216.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 45-60.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775-794.
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20(3), 305-330.
- Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the mandarin tone 2/tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18(1), 37-49.
- Burnham, D. & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed.) *South East Asian Linguistic Studies in Honour of Vichin Panupong (Science of Language Vol. 8.)*, 29-47. Bangkok: Chulalongkorn University Press.
- Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.
- Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.

- Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, 65(7), 1029-1044.
- Fromkin, V. A. (1978). *Tone: A linguistic survey*. New York: Academic Press.
- Gandour, J. (1978). The perception of tone. In Fromkin, V. A. (Ed.), *Tone: A Linguistic Survey*, 41-76. New York: Academic Press.
- Gandour, J. (1983). Tone perception in far eastern languages. *Journal of Phonetics*, 11(2), 149-175.
- Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, 12(2), 235-261.
- Gandour, J., Dziedzic, M., Wong, D., Lowe, M., Tong, Y., & Hsieh, L. et al. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language*, 84(3), 318-336.
- Gandour, J., Potisuk, S., Ponglorpisit, S., & Dechongkit, S. (1991). Inter- and intraspeaker variability in fundamental frequency of Thai tones. *Speech Communication*, 10(4), 355-372.
- Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, 21(1), 1-33.
- Gerrits, E. & Schouten, M.E.H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics*, 6, 363-378.
- Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25, 207-231.
- Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *Journal of the Acoustical Society of America*, 115, 2545.
- Gottfried, T. L., & Riester, D. (2000). Relation of pitch glide perception and Mandarin tone identification. *Journal of the Acoustical Society of America*, 108, 2604.
- Greenberg, S. & Zee, E. (1979). On the perception of contour tones. *University of California Working Papers in Phonetics*, 45, 150-164.

- Guion, S. G., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors affecting stress placement for English non-words include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, 46(4), 403-427.
- Hallé, P. A., Best, C. T. & Levitt, A. (1999). Phonetic versus phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, 27, 281-306.
- Hallé, P. A., Chang, Y., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32(3), 395-421.
- Halliday, M.A.K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.
- Hazan, V., Sennema, A., Iba, M. & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants in Japanese learners of English, *Speech Communication*, 47, 360-378.
- Holt L.L., Lotto A.J., & Diehl R.L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *The Journal of the Acoustical Society of America*, 116 (3), 1763-1773.
- Hsieh, L. (2001). A crosslinguistic PET investigation of lexical tone perception. Ph.D Dissertation. Purdue University.
- Huang, T. (2001). The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In E. Hume and K. Johnson (eds.) *Studies on the Interplay of Speech Perception and Phonology, Ohio State University working papers in Linguistics*, 55, 23-42.
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*, 118, 3267-3278.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., & Kettermann, A. et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-B57.
- Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, 20, 63-67.

- Klatt, D. (1973). Discrimination of fundamental frequency contours in synthetic speech duplications for models of pitch perception. *Journal of Acoustical Society of America*, 53, 8-16.
- Krauss, M. E. (1979). Athabaskan tone. In Hargus, Sharon & Keren Rice (Eds.), *Athabaskan Prosody*. Amsterdam/Philadelphia: John Benjamins.
- Kuhl, P.K. & Miller J.D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72.
- Ladefoged, P. (2001). *A course in phonetics (3rd Edition)*. Harcourt Brace College Publishers.
- Lasky, R.E., Syrdal-Lasky, A. & Klein, R.E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 20, 213-225.
- Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (Eds.), *New Sounds*, 90, 72-97, University of Amsterdam.
- Lee, Y., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5), 527-542.
- Levi, S.V. & Winters, S.J. (2005). External factors affecting the perception of foreign accents in L2 speech. 1st ASA Workshop on Second Language Speech Learning, Vancouver, Canada.
- Liberman, A.S., Harris, K.S., Hoffman, H., & B. Griffith. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.
- Liberman A. M., Cooper F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- Lin, H. (1996). *Mandarin tonology*. Taiwan: Pyramid Press.
- Lin, H. (1998). Tonal elements and tone sandhi in Putonghua. *Zhongguo Yuwen*, 1, 31-39.

- Lin, H. (2001). *A grammar of Mandarin Chinese*. Languages of the world\Materials 344, Lincom, Europa.
- Lin, M. (1988). Acoustic properties and perceptual cues of tone in Putonghua (Putonghua shengdiao de shengxue tezheng he zhijue zhengzhao). *Zhongguo Yuwen*, 3, 182-193.
- Lisker, A.L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Words*, 20, 384-422.
- Lisker, A.L., & Abramson, A.S. (1970). The voicing dimension: some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967*, 563-567.
- Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89, 874-886.
- Mackain, K.S, Best, C.T. & Strange, W. (1981). 'Distinguishing universal and language-dependant levels of speech perception: evidence from Japanese listener's perception of English', *Applied Psycholinguistics*, 2, 369-90.
- Major, R. & E. Kim. (1996). The similarity differential rate hypothesis. *Language Learning*, 46, 465 – 496
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18(5), 331-340.
- Moen, I. (1993). Functional lateralization of the perception of Norwegian word tones--evidence from a dichotic listening experiment. *Brain and Language*, 44(4), 400-413.
- Pike, K. L. (1948). *Tone languages*. Ann Arbor: University of Michigan Press.
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, 52(1), 37-52.
- Pollock, L. & Pisoni, D. (1971). On the comparison between identification and discussion tests in speech perception. *Psychonomic Science*, 24, 299-300.

- Schouten, M. E. H. (1985). Identification and discrimination of sweep tones. *Perception & Psychophysics*, 37, 369-376.
- Shen, X. S. & Lin, M. A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, 34, 2, 145-156.
- Silverman, D. (1997). Laryngeal complexity in Otomanguean vowels. *Phonology*, 14(2), 235-261.
- So, C. (2005a). How do L1 prosodic backgrounds influence Cantonese and English speakers to perceive Mandarin tones? 1st ASA Workshop on Second Language Speech Learning, Vancouver, Canada.
- So, C. (2005b). The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. 149th Meeting of the Acoustical Society of America, Vancouver, Canada.
- Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a non-tone language. *Journal of Chinese Linguistics*, 21(1), 143-163.
- Strange, W. (1995). Cross-language studies of speech perception: A historical review. In Winifred (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 3-45. Timonium, MD: York Press, Inc.
- Sun, S. H. (1998). The development of a lexical tone phonology in American adult learners of standard mandarin Chinese. Ph.D Dissertation. Hawaii University.
- Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press.
- Van Lancker, D. & Fromkin, V.A. (1973). Hemispheric specialization for pitch and "tone": Evidence from Thai. *Journal of Phonetics*, 1, 101-109.
- Wang, W.S.-Y. (1976). Language Change. In S.R. Harnad, H.D. Steklis, & J. Lancaster (Eds.), *Origins and evolution of language and speech*, *Annals of the New York Academy of Sciences*, 280, 61-72.
- Wang, Y. (2002). Perception, production, and neurophysiological processing of lexical tone by native and non-native speakers. Ph.D Dissertation. Cornell University.
- Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The Role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, 25 (3), 449-466.


- Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2), 1033-1043.
- Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of mandarin tones by Chinese and American listeners. *Brain and Language*, 78(3), 332-348.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6), 3649-3658.
- Wang, Y., & Kuhl P. K. (2003). Evaluation the "Critical Period" hypothesis: Perceptual Learning of Mandarin Tones in American Adults and American Children at 6, 10, and 14 Years of Age. *15th ICPhS Barcelona*, 1537-1540.
- Wayland, R.P. & Guion, S.G. (2003). Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. *Applied Psycholinguistics*, 24, 113-129.
- Wayland, R.P. & Guion, S.G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54, 681-712.
- Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception and Psychophysics*, 21, 289-297.
- Wu, Z.J. (1986). *Hanyu Putonghua Danyinjie Yutuce* [Spectrographic Album of Monosyllables in Standard Chinese]. Beijing: Zhongguo Shehui Kexue Chubanshe. [Social Science Press].
- Xu Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61-83.
- Xu Y. (2002). Articulatory constraints and tonal alignment. *Proceedings of the First International conference on Speech Prosody*, 91-100. Aix-en-Provence, France.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- Zue, V. (1976). Some perceptual experiments on the Mandarin tones. *Paper presented at the 92nd Meeting of Acoustical Society of America*, San Diego, California.

Appendix I: Question sheet for Thai listeners to judge and rank the similarity between tokens along two Mandarin tone continua (T1-T4 and T2-T3) and the Thai five tones

You will hear 16 words¹ with different tones. Please select a Thai tone (0 = Mid tone, 1 = Low tone, 2 = Falling tone, 3 = High tone; 4= Rising tone) which sounds closest to the tone you hear. You must select one. Right after you make your choice, please rank the similarity between the Thai tone and the tone of the word you hear (1 = Least similar, 2 = Intermediately similar, 3 = Very similar, 4 = Same).

Word	Corresponding Thai tone	Similarity
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Appendix II: Instructions for discrimination, identification tests and the training session (from top to bottom) (in English):

▶▶ Welcome! In this session, you will hear two words presented in the same group. Both words have the same syllable but their tones are either identical or not quite identical. If you think the two tones are identical, please click the first button, labeled ; if you think the two tones are similar but not identical; please click the

¹ From a linguistic perspective, “syllable” is a more accurate interpretation than “word” in this context. However, considering that none of the listeners were linguistics majors and that monosyllabic words are also common in Thai, “word” was used here instead of “syllable”.

second button, labeled . Each group of words will be played at even time intervals. There are no warnings for the appearance of each word. Please answer as quickly as possible. There will be six 8-second breaks in the whole recording.

▶≡ In this session, you will hear three words presented in the same group. They have the same syllable but differ in tone. The second word may sound more like the first word and less like the third word, or vice versa. If the second word sounds more like the first word, please click the first button, labeled ; if the second word sounds more like the third word, please click the second button, labeled . Each group of words will be played at even time intervals. There are no warnings for the appearance of each word. Please answer as quickly as possible. There will be six 8-second breaks in the whole recording.

▶≡ Welcome! This time you will hear Mandarin words that differ in tone. Mandarin has four different tones. Tone 1 is high but has no changes in its movement (ˉ), as in *mā* (mother) and *fān* (sail). Tone 2 starts low but moves very high (ˊ), as in *má* (hemp) and *fán* (common). Tone 3 starts low but moves to very low and ends fairly high (ˇ), as in *mǎ* (horse) and *fǎn* (opposite). Tone 4 is very high first but falls right away and ends very low (ˋ), as in *mà* (curse) and *fàn* (rice). Please hear the four tones again: *mā má mǎ mà; fān fán fǎn fàn*.

Next you will hear two words with different tones presented in the same group. There will be three different syllables: *Cei*, *Gian*, *Ra*. Each time, you will hear one syllable with two different tones. Please choose the tone for the first word you hear. If it's Tone 1, please click the first button, labeled ; if it's Tone 2, please click the second button, labeled ; and so on. After you make your choice, I will tell you the tone of the first word and the tone of the second word. Each group of words will be played at even time intervals. There are no warnings for the appearance of each word. But there will be one 8-second break in the whole recording.

Appendix III: Thirty-one pitch points along tone contours as well as occurrence time in the two continua: (A) in T1-T4, (B) in T2-T3.

(A)

Point	Time (ms)	t1(T1) (Hz)	t2 (Hz)	t3 (Hz)	t4 (Hz)	t5 (Hz)	t6 (Hz)	t7 (Hz)	t8(T4) (Hz)
1	230	313.13	316.18	319.22	322.27	325.32	328.37	331.41	334.46
2	240	308.21	311.74	315.27	318.80	322.33	325.87	329.40	332.93
3	250	302.53	305.85	309.18	312.51	315.84	319.16	322.49	325.82
4	260	300.05	302.93	305.82	308.70	311.58	314.47	317.35	320.24
5	270	298.22	300.83	303.44	306.05	308.67	311.28	313.89	316.50
6	280	296.06	298.65	301.24	303.84	306.43	309.02	311.61	314.21
7	290	293.88	296.27	298.66	301.06	303.45	305.84	308.23	310.62
8	300	289.73	292.24	294.75	297.26	299.78	302.29	304.80	307.31
9	310	287.13	289.49	291.85	294.21	296.57	298.93	301.29	303.65
10	320	287.12	289.08	291.04	293.00	294.96	296.91	298.87	300.83
11	330	286.99	288.21	289.43	290.65	291.87	293.09	294.31	295.53
12	340	285.75	287.15	288.55	289.94	291.34	292.74	294.13	295.53
13	350	283.67	284.02	284.37	284.72	285.08	285.43	285.78	286.13
14	360	282.59	282.27	281.94	281.61	281.28	280.96	280.63	280.30
15	370	282.70	281.26	279.83	278.39	276.96	275.52	274.09	272.65
16	380	282.25	279.98	277.71	275.44	273.17	270.89	268.62	266.35
17	390	281.33	277.80	274.27	270.74	267.21	263.68	260.15	256.62
18	400	280.69	276.11	271.52	266.93	262.35	257.76	253.18	248.59
19	410	279.94	274.54	269.14	263.74	258.34	252.94	247.54	242.14
20	420	279.94	273.21	266.48	259.76	253.03	246.30	239.58	232.85
21	430	280.14	272.54	264.94	257.34	249.74	242.14	234.54	226.94
22	440	279.98	271.78	263.59	255.39	247.19	238.99	230.79	222.59
23	450	279.82	271.07	262.32	253.57	244.82	236.07	227.32	218.57
24	460	279.90	269.80	259.69	249.59	239.48	229.38	219.27	209.17
25	470	279.86	268.94	258.01	247.08	236.16	225.23	214.30	203.38
26	480	280.41	268.77	257.12	245.47	233.83	222.18	210.53	198.88
27	490	281.15	268.71	256.28	243.85	231.42	218.98	206.55	194.12
28	500	285.61	271.30	256.98	242.66	228.34	214.02	199.71	185.39
29	510	298.24	280.92	263.61	246.30	228.99	211.68	194.36	177.05
30	520	311.76	291.84	271.91	251.99	232.06	212.14	192.21	172.29
31	530	316.15	295.34	274.53	253.71	232.90	212.09	191.28	170.47

(B)

Point	Time (ms)	t1(T2) (Hz)	t2 (Hz)	t3 (Hz)	t4 (Hz)	t5 (Hz)	t6 (Hz)	t7 (Hz)	t8(T3) (Hz)
1	208	244.60	240.05	235.50	230.95	226.40	221.85	217.30	212.75
2	218	232.15	229.15	226.16	223.16	220.16	217.16	214.16	211.17
3	228	223.99	221.37	218.75	216.13	213.52	210.90	208.28	205.66
4	238	221.04	217.85	214.66	211.47	208.28	205.09	201.89	198.70
5	248	219.07	215.60	212.13	208.66	205.19	201.72	198.25	194.78
6	258	218.00	214.28	210.56	206.83	203.11	199.39	195.66	191.94
7	268	216.89	212.51	208.14	203.76	199.39	195.01	190.64	186.26
8	278	215.78	210.85	205.91	200.98	196.04	191.11	186.17	181.24
9	288	214.83	209.09	203.36	197.62	191.88	186.15	180.41	174.67
10	298	214.84	208.24	201.65	195.05	188.46	181.86	175.27	168.67
11	308	216.27	209.26	202.24	195.23	188.22	181.20	174.19	167.18
12	318	217.80	210.44	203.08	195.73	188.37	181.01	173.65	166.29
13	328	219.97	212.19	204.40	196.62	188.84	181.06	173.28	165.50
14	338	221.89	213.85	205.81	197.76	189.72	181.68	173.63	165.59
15	348	224.72	216.59	208.45	200.32	192.19	184.05	175.92	167.79
16	358	228.58	220.37	212.16	203.96	195.75	187.54	179.34	171.13
17	368	232.84	224.62	216.39	208.17	199.95	191.73	183.50	175.28
18	378	238.47	229.93	221.40	212.86	204.32	195.78	187.25	178.71
19	388	243.37	234.51	225.64	216.78	207.92	199.05	190.19	181.33
20	398	247.23	238.31	229.39	220.47	211.55	202.62	193.70	184.78
21	408	251.70	242.71	233.72	224.72	215.73	206.74	197.75	188.76
22	418	256.95	247.92	238.88	229.84	220.81	211.77	202.74	193.70
23	428	264.34	255.09	245.83	236.58	227.33	218.07	208.82	199.57
24	438	270.56	261.30	252.04	242.78	233.52	224.25	214.99	205.73
25	448	278.67	268.87	259.06	249.25	239.45	229.64	219.83	210.03
26	458	287.45	276.63	265.81	255.00	244.18	233.36	222.54	211.72
27	468	299.26	286.99	274.72	262.45	250.18	237.91	225.64	213.37
28	478	314.70	300.90	287.09	273.29	259.49	245.69	231.88	218.08
29	488	327.26	313.26	299.25	285.25	271.24	257.24	243.23	229.23
30	498	334.84	322.14	309.44	296.74	284.04	271.34	258.64	245.94
31	508	336.86	325.09	313.33	301.56	289.80	278.04	266.27	254.51