Summer 2010

A model of Mandarin tone categories--a study of perception and production

Bei Yang
University of Iowa

A MODEL OF MANDARIN TONE CATEGORIES

-- A STUDY OF PERCEPTION AND PRODUCTION

by

Bei Yang

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Second Language Acquisition
in the Graduate College of
The University of Iowa

July 2010

Thesis Supervisor: Professor Richard R. Hurtig

ABSTRACT

The current study lays the groundwork for a model of Mandarin tones based on both native speakers' and non-native speakers' perception and production. It demonstrates that there is variability in non-native speakers' tone productions and that there are differences in the perceptual boundaries in native speakers and non-native speakers.

There are four experiments in this study. Experiment 1 utilizes native speakers' production data from a published speech database to explore the features of tone production by native speakers. Inter-speaker normalization is used to analyze the data. Experiment 2 synthesizes 81 tones that are carried by four sentences to measure perception by native and non-native speakers. The intra-speaker and inter-speaker normalization is used to investigate the perceptual space of T1, T2, T3, and T4. The researcher also explores the salient features distinguish native speakers' and non-native speakers' perception of the four principal tones. Experiment 3 uses both synthesized tones and natural tones that are carried by sentences to explore how pitch values of tones create overlapping areas in the perceptual map. Experiment 4 examines tone production by non-native speakers to identify the differences between native speakers' perception and non-native speakers' production; and the differences between non-native speakers' perception and their production of tones.

The results of the perception and production experiments with native speakers show the perceptual boundaries and tonal categories in the perceptual space and the production space. The difference of native speakers' perception and production shows us the perceptual cue for perception. Meanwhile, the similarities of native speakers' perception and production reveal the acoustic cues, including register and contour, for tone perception and production. The results of the perception experiments with non-native speakers indicate that there are no clear boundaries, and that tone overlap in the perceptual space. Register plays an important role in the perception of tones by non-native speakers. The results of non-native speaker production also show overlapping tones in the acoustic space. The non-native speaker production appears to be determined by the contour of the tones in contrast both the contour and register determine the tonal categories of native speaker.

Abstract Approved ______________________________
Thesis Supervisor

____
Title and Department

____
Date

A MODEL OF MANDARIN TONE CATEGORIES

-- A STUDY OF PERCEPTION AND PRODUCTION

by

Bei Yang

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Second Language Acquisition
in the Graduate College of
The University of Iowa

July 2010

Thesis Supervisor: Professor Richard R. Hurtig

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Bei Yang

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Second Language Acquisition at the July 2010 graduation.

Thesis Committee: Richard R. Hurtig, Thesis Supervisor

Jill N. Beckman

Weldon S. Coblin

Robert A. Fox

Judith E. Liskin-Gasparro

To my mother.

ACKNOWLEDGMENTS

I am grateful for all the supports I have received while researching and writing up this dissertation. My biggest thanks go to my advisor, Professor Richard Hurtig, who was always ready with a thoughtful response whenever I needed help and advice. His advice has been through every step of my dissertation research: the literature review and experimental design were developed through our weekly meetings; the tone synthesis and data analysis were under his supervision and conducted in the perception laboratory; and the results and discussion were completed through email advising after I moved to Madison. He spent considerable time revising my draft again and again with numerous and valuable comments and suggestions. Without his guidance, patience and persistent help this dissertation would not have been possible.

Special thanks go to Professor Judith Liskin-Gasparro, one of my committee members and the co-director of FLARE Program. From the class of second language acquisition theory to the dissertation project, from research design to data presentation, from the *Modern Language Journal* review section to academic conferences, she guided me into the field of second language acquisition step by step. During my five-year Ph.D. study at the University of Iowa, she helps me overcome various difficulties. Without her, it would have been impossible to complete my Ph.D. study. It is hard for me to express my appreciation verbally.

I would like to thank other committee members: Professor Jill Beckman, Professor Welden Coblin and Professor Robert Fox. I learned a lot from Professor Beckman's phonology classes and the independent study about research on tones and I benefited from her helpful comments on the literature review and research design. I also benefited from Professor Coblin's Classical Chinese classes and talking with him about traditional phonology of Mandarin tones. All of their teaching, discussion and comments enhance the quality of my dissertation. I also thank Professor Fox for kindly agreeing to serve as an external committee member of my dissertation committee. He challenged me with some interesting questions and inspired me to think some issues in depth. I really appreciate it.

Thanks especially go to Professor Mary Erbaugh who has given me many supports and encouragements since 2000. During my dissertation writing, she provided valuable comments and suggestions on the earlier version. She also helped make my wording more native-like. I also would like to thank Professor Huatong Sun, my long standing friend. She provided moral supports for me, gave me a lot of subtle suggestions and helped me adapt to the American academic culture based on her own experience. I appreciate all their helps very much. I am also indebted to all the native speakers and non-native speakers who participated in this project. Many thanks for their time and cooperation.

Moreover, I would also like to appreciate the professors, the classmates, and the friends in FLARE, Chinese program and other related programs who help me during my five-year studies at the University of Iowa.

Last but not least, I wish to thank my family: my lovely daughter Shelley, for bringing happiness and smiles into my life; my husband Bo, for his patience and thoughtfulness whenever I encountered problems; my parents, for their understanding and forbearance whenever I felt impatient with the hundreds of hours I have spent on my dissertation research.

Financial support was provided by the summer fellowship from the Graduate College and graduate research funding from the Graduate Student Senate at the University of Iowa.

ABSTRACT

The current study lays the groundwork for a model of Mandarin tones based on both native speakers' and non-native speakers' perception and production. It demonstrates that there is variability in non-native speakers' tone productions and that there are differences in the perceptual boundaries in native speakers and non-native speakers.

There are four experiments in this study. Experiment 1 utilizes native speakers' production data from a published speech database to explore the features of tone production by native speakers. Inter-speaker normalization is used to analyze the data. Experiment 2 synthesizes 81 tones that are carried by four sentences to measure perception by native and non-native speakers. The intra-speaker and inter-speaker normalization is used to investigate the perceptual space of T1, T2, T3, and T4. The researcher also explores the salient features distinguish native speakers' and non-native speakers' perception of the four principal tones. Experiment 3 uses both synthesized tones and natural tones that are carried by sentences to explore how pitch values of tones create overlapping areas in the perceptual map. Experiment 4 examines tone production by non-native speakers to identify the differences between native speakers' perception and non-native speakers' production; and the differences between non-native speakers' perception and their production of tones.

The results of the perception and production experiments with native speakers show the perceptual boundaries and tonal categories in the perceptual space and the production space. The difference of native speakers' perception and production shows us the perceptual cue for perception. Meanwhile, the similarities of native speakers' perception and production reveal the acoustic cues, including register and contour, for tone perception and production. The results of the perception experiments with non-native speakers indicate that there are no clear boundaries, and that tone overlap in

the perceptual space. Register plays an important role in the perception of tones by non-native speakers. The results of non-native speaker production also show overlapping tones in the acoustic space. The non-native speaker production appears to be determined by the contour of the tones in contrast both the contour and register determine the tonal categories of native speaker.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

English-speaking learners have considerable difficulty learning Mandarin Chinese. The greatest difficulty is the tones. Helping learners acquire the tonal system is key to an effective pedagogy of pronunciation. This chapter introduces Mandarin tones and the justification for the current study.

1.1 Mandarin Tones

Tones can be regarded as suprasegmentals and the features contrast with segmental features; that is, the features of consonants or vowels. There are two main perceptual aspects of tones: "first, the human ability to perceive the physical properties of frequency, duration, and intensity, and second, the psychological response to various acoustic stimuli" (Chun, 2002:10). The physical properties of frequency, duration and intensity play roles in tone perception. "Tones are perceived principally as differences in pitch, although intensity and duration also provide important secondary auditory cues for the identification of discrete tones" (Norman, 1987:145).

1.1.1 Tone Categories

Chinese is a tonal language. Each syllable has a lexical tone that can distinguish meaning. That is, two identical syllables with the same initials and finals[1] will mean something completely different when they differ only by tone (See Table 1). Each syllable has an initial, a final and a tone. However, an initial is not obligatory in a Mandarin syllable. The syllables without initials are called zero initial syllables, such as *ai4* (love) and *e2* (goose). Most descriptions of Mandarin Chinese posit four tones. However, a neutral tone, which has weaker and shorter pitch, can be regarded as Tone 0.

---

[1] The "initial" which is also called the onset, is the consonant at the beginning of each syllable. Final, which is also called the rhyme, is composed of the vowels and consonants at the end of each syllable. Consonants are optional in a rhyme.

Several terms are used to describe tones: pitch value, tone contour, and tone category (see Table 1 below). Pitch is a physical property of each tone with a particular sound frequency that is called fundamental frequency (F0). Phoneticians calculate a relative value for each pitch after measuring the values of F0 in a tone. This relative value is the pitch value. The pitch value conveys the actual manner of producing or pronouncing a tone, as well as the changes of a tone's pitch including level, rising, and falling. Tones that have the same pitch value belong to the same tone category. Classical Chinese phonology used eight tone categories, and modern Chinese linguists still use the names of the classical categories (e.g. Hu, 1990; Norman, 1987). However, the range and the number of modern tone categories differ from those used in ancient China. Since there are four kinds of tones in Mandarin, only four names of the classical tone categories are used to refer to Mandarin tones. They are *Yinping*, *Yangping*, *Shangsheng*, and *Qusheng*. In this research, Modern Mandarin Tones are labeled T1, T2, T3, and T4 (Table 1).

Table 1 Four tones in Mandarin

Tone Category	Example	English Meaning	Classical Tone Category Name	Pitch Value	Tone Contour
T1	tang1	'soup'	Yinping	55	Level
T2	tang2	'candy'	Yangping	35	Rising
T3	tang3	'lie down'	Shangsheng	214	Dipping
T4	tang4	'burning hot'	Qusheng	51	Falling

Tone contour is the shape of the pitch. The numerical notation of pitch value used in this study, as designed by Chao (1930), is widely used in academic research and pedagogy. The pitch values are labeled from 1 to 5. 1 is the lowest pitch value, while 5 is the highest pitch. Pitch contours are labeled by two or three numbers to designate the duration of each syllable. So a T1, with high level pitch, is labeled 55, because it starts high and ends high. A T2, which is rising, is labeled 35 because it starts at mid-range and rises to high. The pitch contours of the examples in Table 1 are included in the "Pitch Value" column, which also match the tone contours in the "Tone Contour" column. For example, T1 is described as "55" which means T1 is a high level tone, and the pitch value of T2 is "35" which indicates a rising tone.

In this study, the neutral tone, labeled T0, is a weak form of T1, T2, T3, and T4, because it lacks fixed pitch values and tone contours. Controversy remains over whether it should be regarded as a unique tonal category. This study categorizes the neutral tone as a unique tone group. The reasons for this decision are explained in Chapter 2.

The four traditional tonal categories described above are based on the concept of citation tone, while the neutral tone is a sandhi tone. Citation tones are the tones assigned to lexical items when a tone is produced without context. A tone can have different pitch values depending on the context, and can change[2] to different variations of a particular tone or other tones. The term "sandhi tone" refers to tones in context. Traditional Chinese linguists used "base tone" to refer to citation tone. Chen (2000) did not use the traditional term "base tone" since the underlying form can be either the citation tone or the sandhi tone.

[2] Tone change is a term in historical Chinese phonology. It refers to the historical change of tone, such as results from language contact. If a tone changes based on context at the synchronic level, it is termed "tone sandhi" (section 1.1.3).

1.1.2 T3 in Context

T3, whose pitch value is “214”, is the most complicated tone in the tonal system of Mandarin, because it has many contextual variations. The traditional view of T3 is that it is a dipping tone; that is, the contour first falls and then rises. This description can be found in many books and articles (e.g., Hu, 1990). Chao (1968) characterized the change of a tone, such as the most common tone sandhi — T3 changes to T2 that is a rising tone, when it comes before T3. However, T3 changes to a half dipping tone that is a low falling tone (the pitch value is 31) when it comes before T1, T2, and T4. In Chao’s interpretation, the base form of T3 is the dipping contour. Chao distinguished phonetic change from phonemic change. He regarded the low falling component of T3 as a phonetic change since it does not change the tonal category of T3, whereas the change from T3 to T2 is a phonemic change since T2 is an independent phoneme.

Chen (2000) did not distinguish the terms of phoneme and phonetic variation. He regarded all the variations of T3 as allotones, including T2. According to some perception and acoustic experiments (Chao, 1968; Shen & Lin, 1991; Shih, 1987; Xu, 1999), Chen (2000) found that T3 has three allotone variations based on different contexts. Table 2 illustrates these below based on the previous research (Chao, 1968; Shen & Lin, 1991; Shih, 1987; Xu, 1999). If T3 is at the end of a sentence, there are two allotones: the pitch value is 214, that is, a dipping tone, or 21. If T3 is in the middle of a sentence and it precedes T1, T2, or T4, its pitch value is 21. If T3 comes before another T3, then the pitch value is 35 and is similar to T2. Table 2 provides a detailed description of T3 changes systematically. Note that it is hard for us to decide the tone category for a tone whose pitch value is 35, because it can belong to either T2 or T3, according to Chen’s characterization of the allotones of T3.

Table 2 Allotones of T3

Allotone / Sandhi contexts	Pitch Values 214	21	35
Before T3	−	−	+
Utterance-final	+	+	−
Elsewhere	−	+	−

Source: adapted from Chen, M. *Tone sandhi: Patterns across the Chinese dialects*. Cambridge: Cambridge University Press. (2000:20)

Acoustically, the tone contour of T3 in context is a low falling contour since its pitch value is 21. This tone contour of T3 is similar to that of T4, which is distinct from T2 since T2 carries a rising contour. Figure 1 shows the difference between T3 and T4. The data in Figure 1 are derived from a sample collected from a native speaker, who produced two sentences which include T3 and T4 respectively. The Kay Elemetrics Multi-Speech program was used to extract F0 values from the digitized speech samples.

Figure 1 shows the pitch contour of T3 and T4 in the speech samples. The pitch values are selected at 0, 5%, 10%, ……, 100% of the time of each tone. In this figure, the vertical axis represents pitch values and horizon axis represents the percentages of the time. It shows that both T3 and T4 have falling contours. However, T3 has a lower register, or starting F0 value, than T4 does. This study will investigate the forms of tones in context. Therefore, the most common variation of T3, that is, the low falling contour, will be employed in the analysis of naturally produced tones and in the perceptual studies of synthesized tones.

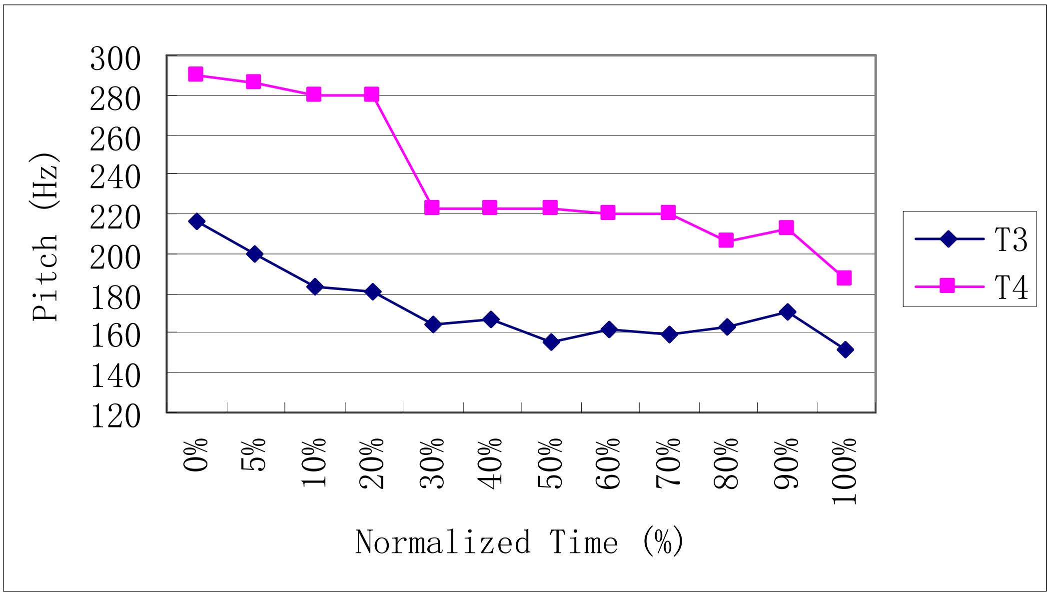


Figure 1 Different contours of T3 and T4

However, there is another problem, since T3 still exhibits the rising contour when it occurs at the end of a sentence, which is similar to T2. Shen and Lin (1991) indicated that both citation forms of T2 and T3 have a dipping contour, that is, the first part of the contour is falling and the second part of the contour is rising. The difference between them is found in the duration of the falling contour of T2, which is much shorter than that of T3. The duration of the falling contour in the first part of T2 is much shorter than the duration of the rising contour in the second part of T2; the duration of the falling contour in the first part of T3 is much longer than the duration of the rising contour in the second part of T3.

1.1.3 Tone Sandhi

Tone sandhi refers to changes of tones that are governed by a rule in some specific contexts. Gradually, the term *sandhi* has been extended to describe many phenomena, including allotonic variations and changing to other categories that are affected by intonation and other morphological or syntactical conditions (Chen, 2000; Yip, 2002).

Chen (2000) claims that Mandarin has three sandhi rules (Chen, 2000). The most well known tone sandhi in Mandarin is T3 sandhi, where a T3 switches to a rising T2 before another T3 syllable. For example, the citation tones for the word for 'fruit' are T3T3. Sandhi changes this to T2T3. But T3 sandhi is only "one of the many ways the idealized tone shape may be modified under the influence of neighboring tones and the overarching intonation pattern" (Chen, 2000:20). The second rule is T2 sandhi. T2 becomes a T1 when it is preceded by T1 and comes before another T2, as in (1) below.

(1)	*tian wen tai*	'observatory'(lit. astronomy + platform)
	T1 T2 T2	base form
	T1 T1 T2	sandhi form

T2 sandhi is not widely used. In Chen (2000:301), T2 sandhi is explained as a prosodic change; that is, a stress effect. However, if we analyze the pitch of *tianwentai* and *tianwen*, the example that Chen used to explain the difference of T2 that two *wen* carry, we find that the syllable *wen* in both words/phrases has the same tonal contours (see Figure 2). The program Praat was used to analyze this word and generate the pitch contours in Figure 2.

In Figure 2, the vertical axis represents the pitch of tones, while the horizon axis represents the time. The pitch contours between two solid lines are the tone contours of *wen* that carries T2. We can find that their pitch contours are same, that is, both tones fall first, and then rise a little bit. Therefore, the second sandhi rule that Chen proposed is worth further consideration.

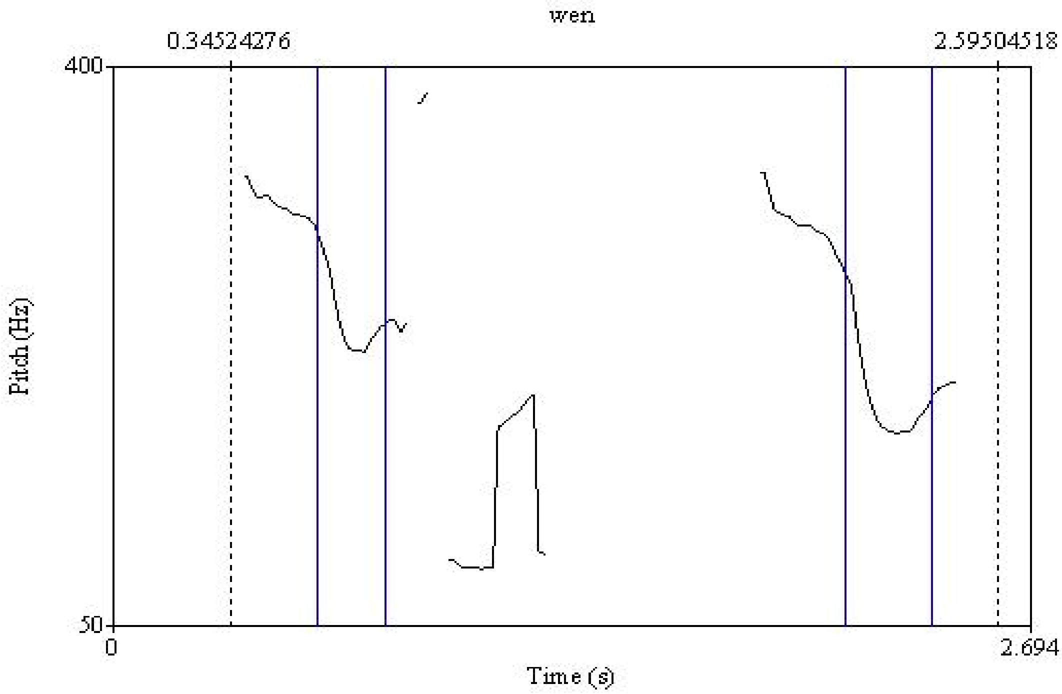


Figure 2 Contours of *wen* in different words (the pitch contours between two solid lines are the tonal contours of *wen*.)

The third rule is the neutral tone (T0) sandhi. The neutral tone can carry a falling or dipping pitch (Cao, 2002b; Chen & Xu, 2006) or assume a level pitch (Chen, 2000). Whatever the contour of the base form of the T0 is, the T0 contour is determined by the preceding full tone. The rule is illustrated in (2a) through (2d) below. In (2a), (2b), (2c) and (2d), the word *de* carries the neutral tone. If the previous tone is T1 or T2, the neutral tone that carries a middle level pitch can be perceived as a falling tone, because T1 is a high level tone and T2 is a high rising tone. If the previous tone is T3, the neutral tone that carries a middle level pitch will be perceived as a rising tone, because T3 is a low falling tone. If the previous tone is T4, the neutral tone that carries a low level pitch will be perceived as a low falling tone, because T4 is a high falling tone.

(2a) *sheng de* 'uncook + adj' 'raw'

T1 T0 (Tone category)

55. 3 (Pitch value)

(2b) *shen de* 'God + poss' 'God's'

T2 T0

35. 3

(2c) *sheng de* 'province + poss' 'province's'

T3 T0

21. 3

(2d) *sheng de* 'remain + adj' 'leftover'

T4 T0

51. 1

Although the neutral tone has numerous variations of pitch, it has a salient feature: its duration is much shorter than that of other tones. Generally, the duration of a neutral tone is around 60% of T1, T2, T3, and T4 (Cao, 2002b; Chen & Xu, 2006).

Finally, in addition to these three sandhi rules, some other sandhi patterns are morphologically conditioned. One alternation is called the "*yi-bu-qi-ba* rule", because it applies specifically to four high-frequency words: *yi1* 'one', *bu4* 'not', *qi1* 'seven', and *ba1* 'eight'. According to this rule, the tones carried by these syllables are changed to T2 before a falling T4 syllable (Chen, 2000). For example, in (3), the tone for *bu* without context is T4 whose pitch value is 53; however, when it comes before another T4, it changes to T2 whose pitch value is 35. The reason for this change may stem from a constraint on adjacent autosegments (Obligatory Contour Principle) in autosegmental theory (Goldsmith, 1976).

(3) *bu yao* 'not want'

T4 T4 (Tone category) Base tone

53. 53 (Pitch value)

T2 T4 (Tone category) *yi*-*bu*-*qi*-*ba* rule

35. 53 (Pitch value)

cf. rule does not apply before a T3, as in

bu xiang 'not think'

T4 T3 (Tone category)

53. 214 (Pitch value)

Some studies also have explored the triggers for tone sandhi (Chen, 20000; Shen, 1992; Shih, 1988; Wu, 1982; Xu, 1997). These studies suggest that tonal coarticulation, is the most important factor underlying the sandhi effects. Tonal coarticulation can be described as a natural consequence of articulatory constraints, which includes both assimilation and dissimilation.

1.1.4 Conclusion

The preceding sections introduced the Mandarin tones as they are produced in context. While the acoustic properties of tones change across different contexts, citation tones or the neutral tone, belong to one of the tone categories, that is, T1, T2, T3, T4 and T0. The current study will explore how American learners of Chinese acquire Mandarin tones in the framework of these tone categories.

1.2 The Specific Aims of the Study

The study will explore how non-native speakers (NNS) acquire tones by comparing native speaker (NS) perception with NNS production. Previous research generally focuses on perception or on production; or it analyzes NNS perception and production (Chen, 1997; Elliot, 1991; Kiriloff, 1969; Leather, 1990; Li, 1995; Sun, 1997; Miracle, 1989; Shen, 1989).

This study will develop a model of tone categories. The tonal categories for Mandarin are based on the classical Chinese tone categories developed by Chinese phonologists over the past 2,000 years. Modern researchers have studied the acoustic features of each tone (Cao, 2002b; Shen, 1990; Shih, 1987; Xu, 1997). Although acoustic features are not equivalent to perceptual cues, they have a close relationship. Categorical perception experiments have also been carried out with tones (Halle, 2004; Leather, 1987; Schwanhauber, Burnham, & Jones, 2003; Stagray & Downs, 1993). However, the results have been inconsistent, and the tone categories have not been explored systematically. This study builds on the earlier work and explores the entire tonal system. This should provide a new method for describing Mandarin tones, including the perceptual range of each tone and the nature of the neutral tone. It is hoped that such a model can also be used to explain tonal phenomena phonetically and phonologically, including some changes of tones known as tone sandhi.

The model can be used to compare, native speaker (NS) tone perception with non-native speaker (NNS) tone perception, NS tone production with NNS tone production, and NS tone perception with NNS tone production. It should reveal the gap between NNS production/perception and NS perception/production, and further identify the features of tonal acquisition by NNS.

Thus, this study would contribute to general second language acquisition (SLA) studies. The study will provide evidence to show whether NNS can acquire some categories that do not exist in their native language (L1), and how NNS acquire the tone categories.

.

CHAPTER 2 LITERATURE REVIEW

The literature review in this chapter will focus on:

(a) tonal studies in phonetics and phonology;

(b) studies relevant to general difficulties for L2 learners to acquire suprasegmental features; and

(c) studies about tonal acquisition of Mandarin as an L2.

Section one (2.1) reviews the phonetic and phonological studies of Mandarin tones that are relevant to the current study. Section two (2.2) explores why it is difficult for NNS who are speakers of a non-tone language to learn Mandarin tones. And finally, Section three (2.3) reviews the research about the tonal acquisition of Mandarin as a second language (L2).

2.1 Survey of Research about Linguistic Studies of Mandarin Tones

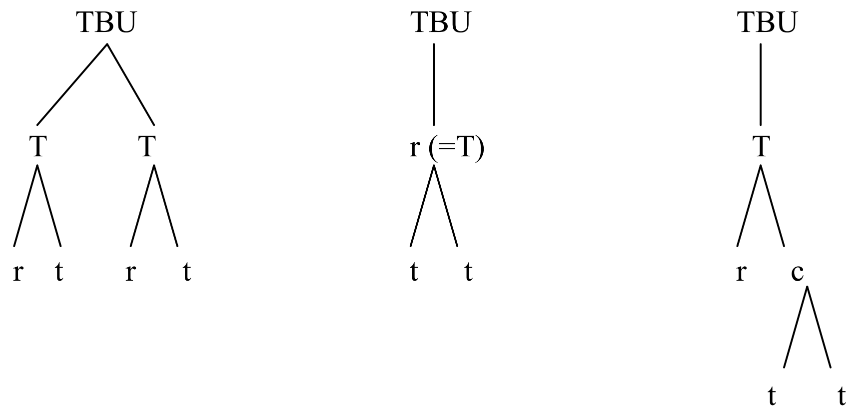
In tonal phonology there are two essential descriptive terms for tones: one denoting pitch height (H for high register and L for low register) and the other denoting pitch movement (hl for falling contour, lh for rising contour, hh for level contour, etc.)[3] (Yip, 2002). Some researchers use other features, such as [raised] to refer to tonal contour (Yip, 1989).

Generally, the phonological features are classified according to phonetic features, such as major class features, place and manner of articulation. Clements (1985) proposed "a hierarchical organization of the features into functionally related classes, grouped under nodes of a tree structure" (McCarthy, 1988:84). The theory of feature geometry is widely used to explain sound phenomena.

---

[3] The term "contour" is used in Chapter 1, which describes the shape of the tone in traditional phonology system of Mandarin. In generative phonology, it has almost the same meaning yet is used for more tone languages.

Different tone models based on feature geometry (Bao, 1990; Duanmu, 1990; Yip, 1989, 1993) make different predictions about whether features can spread independently or together (Yip, 2002:53).



A: Duanmu (1990) B: Yip (1989) C: Bao (1990)

Figure 3 Models of tonal geometry

Note: T = tone root
t = terminal tone segment
r = register
c = contour
TBU = tone bearing unit

Source: Adapted from Yip, M. *Tone*. Cambridge: Cambridge University Press. (2002)

"In the autosegmentalist view of assimilation as spreading, assimilatory processes are diagnostic of constituency" (Chen, 2000:72). For example, as Figure 3 shows, model A and model C could present register spread, because register subordinates to the tone root, whereas model B cannot because register is the tone root.

However, these geometries present two problems. First, they neglect tone duration. Identical syllables which have the same CV combination may differ lexically when the tone duration is shortened. The problem of neglecting tone duration makes it hard for researchers to deal with some special tones, such as the *Ru* tones of Wu dialects and the neutral tone that has shorter duration in Mandarin. Second, the two proposed phonological features H/L (register) and h/l (contour) share the same acoustic feature: pitch. In addition, a pitch with value 3 can be assigned to either the high register or the low register.

Although it is hard to find a fully appropriate way to deal phonologically with descriptive features like register and contour, studies from various perspectives (e.g., phonology, phonetics, and historical Chinese linguistics) have indicated that the register and the contour are the most important descriptive features of tones. Since the register and the contour share the same acoustic feature of pitch, it is worth considering using this acoustic variable—pitch—to analyze tones and to explore whether duration plays an important role during tone perception.

In Chapter 1, we introduced Chen's phonological analysis (Chen, 2000) based on some phonetic studies (such as Shih, 1988). He illustrated some allotones of T3 in context. It shows that the pitch value of T3 is 21 in most contexts, while the pitch value of T3 is 214 when it locates at the end of a sentence. . The pitch value 21 shows that T3 is a low falling tone in most contexts, and yet T3 is perceptually contrastive with T2 and does not contrast T4, which is a high falling tone. That the similar pitch contours of T3 and T4 are perceived differently indicates that other cues affect tonal perception.

. The problem of neglecting tone duration in the phonology studies shows that there may be overlapping areas in perceptual space. That is, if the duration of a tone, such as T0, is lengthened, the tone might be perceived as T1, T2, T3 or T4. If so, they lead to a question how pitch values of tones create overlapping areas (research question 1 and research question 3.1; see 2.4). The problem of T3 indicates that the acoustic cues might be different from the perceptual cues. In the case of T3, which perceptual cues cause the rising part of T3 at the end of a tone perceptually when the pitch value of a T3 is 21? This leads to research question 3.2.

2.2 Why Is It Difficult for NNS Who Are Speakers of a Non-tone Language to Learn Mandarin Tones?

Before reviewing the studies about tonal acquisition of Mandarin as an L2, we discuss two general factors which make tones difficult for NNS to acquire in this section.

All analysts agreed that learners whose native language (L1) is not a tone language, such as native speakers of English, find it extremely difficult to learn Mandarin tones. But why is it so hard? There are two reasons.

First, when we talk about the learners of Chinese, we generally refer to adults. Children have less difficulty learning tones. The principal reason given is that adults are farther away from the critical period for acquiring native-like pronunciation. There are two claims about tonal acquisition and age. Burnham (2002) claimed that there is a resurgence of tone discrimination ability in adulthood. However, this claim is inconsistent with Cutler's claim that learning prosodic features occurs early and only once (Cutler, 1994; Cutler, Mehler, Norris, & Segui, 1992). Burnham and Mattock (2008) concluded that "the resurgence is due to adults' ability to disregard the linguistic constraints that bind children, so that they can perceive tones in all (or most) of their perceptually-salient glory" (Burnham & Mattoch, 2008:267). However, both claims indicated that children's acquisition of prosodic features is different from adults'. It is harder for adults to acquire tones than children do.

The other reason lies in language typology. Generally speaking, there are three large groups of languages: (a) tone languages, such as Chinese, Thai, and Yoruba, which comprise about half of the world's languages; (b) stress languages, such as English; and (c) mora-timed languages, such as Japanese. Certain languages, such as Swedish and

Serbo-Croatian, also have pitch accent or accentual systems (Yip, 2002). Comparing with tone languages, accentual languages have only a small number of tonal contrasts. “There is no absolute division between accent languages and tone languages” (Yip, 2002:4). However, a huge difference exists between tone languages and stress languages. Tone languages use tones to distinguish the meaning of every syllable (See Table 1 above). But stress languages use stress to distinguish only a very tiny minority of lexical meanings. For example, the English word ‘record’ is a noun, if the first syllable is stressed; it is a verb if the second syllable is stressed. Learners of Chinese whose L1 is a stress language could transfer stress to tones (Chiang, 1979).

Although stress languages also have pitch, pitch contrasts do not mark semantic differences at the syllable level; they apply at paralinguistic or sentential levels. Most research indicated that the default intonation pattern is declination (Liberman, 1975; Sorensen and Cooper, 1980; Pierrehumber, 1980, 1981; Cruttenden, 1997). That is, “the pitch range narrows and drifts downwards over the course of a major phrase.” (Pierrehumber, 1981:987). English has some intonation patterns although these patterns do not cause any semantic differences. Pierrehumbert (1980) illustrated five intonation patterns for the word “Anna”. The first four patterns have a high pitch for the first syllable and a low pitch for the second syllable, which indicate an answer or a statement. The fifth one has a low pitch for the first syllable and a high pitch for the second syllable, which show typical melody for a question. However, there are some variations for the pitch patterns in the first four patterns. For example, the pitch at the end of the second syllable may rise a little bit or dramatically, which indicate that an answer is incredulous or incomplete respectively; the pitch contour could be level at the beginning of the intonation and “stop far short of the bottom of the speakers’ range”, which shows that somebody is calling out for Anna. Obviously, there are three basic phonetic pitch contours -- rising, falling and level contours, although different intonation patterns indicate different emotion, sentence type, and so on.

Furthermore, the pitch contours not only express different emotions, questions, etc, but also are related to the position of a tone in a sentence. Some researchers (Broselow, Hurtig, and Ringer, 1987) investigated such influence on learners of Chinese. Their study indicated that the boundary tone does affect the tone acquisition. The falling intonation that indicates a statement has a positive influence on perceiving T4 at the end of a sentence, yet it does not affect T4 in the middle of a sentence.

The previous studies indicate that there are three pitch contours for English intonations phonetically. However, these contours should work with other factors, such as the position of a sentence and emotion, to form intonations. Yet, the pitch contours are not related to semantic meanings, which is totally different from Chinese.

Overall, the difficulties of learning Mandarin tones lie in the nature of the languages and the distance between L1 and L2, as well as some individual learner factors.

2.3 Survey of Research on Tone Perception and Production of Mandarin as L2

A large number of studies have been carried out on the acquisition of Chinese tones. Questions about how second language learners acquire Mandarin tones has attracted a large amount of interest on the part of scholars and teachers. In this section, the major strands of that research are reviewed.

2.3.1 Introduction

Over the past several decades, a great many studies have been conducted from different perspectives on Mandarin tone acquisition by learners whose native language is not a tone language. Some research has focused on the order of tones acquired (Chen, 1997; Elliot, 1991; Kiriloff, 1969; Leather, 1990; Li, 1995; Miracle, 1989; Shen, 1989; Sun, 1997). Other studies have explored how learners transfer suprasegmental features of their native language to L2 Mandarin (Broselow, Hurtig, & Ringen, 1987; Chiang, 1979). Some research has been conducted from the perspective of phonetics and phonology

(Leather, 1990; Read, Zhang, Nie, & Ding, 1986; Shen, 1989). New technology and phonetic experiments have allowed researchers to carry out experiments from a psycholinguistic perspective (Halle et al., 2004; Leather, 1983, 1987; Schwanhauber, Burnham, & Jones, 2003; Stagray & Downs, 1993). Recently, some research has focused on neurophysiological processing (Sereno & Wang, 2008; Soares, 1982; Sussman, Franklin, & Simon, 1982; Wang, Behne, Jongman, & Sereno, 2004; Wuillemin & Richardson, 1994). Studies that are relevant to this current research have been selected for review below.

2.3.2 Order of Acquisition of Tones

The studies summarized in Table 3 (Chen, 1997; Elliot, 1991; Kiriloff, 1969; Leather, 1990; Miracle, 1989; Shen, 1989; Sun, 1997) present a proposed order of tonal acquisition although they have varying results. One underlying result is that the orders of perception and production differ.

These studies raise an important question: What aspect of tones is the most difficult to learn? Most studies agree that the difficulties for perception are different from those for production. Most of them indicate that T2 is the most difficult tone to perceive for learners whose L1 is not a tone language. But for production, the results are varied. It is necessary for us to investigate the relationship between perception and production. Then we can find the greatest difficulties for NNS to acquire Mandarin tones.

Table 3 Relative difficulty of the four tones reported in previous studies

Study	Mode	Order
Kiriloff (1969)	Perception	4<1<3<2
Elliot (1991)	Perception	4<3<1=2
	Perception: self	4<1<3<2
Sun (1997)	Perception: TIDT (stimulus)	4<1<3<2
	Perception: TIDT (response)	1<4≤3<2
Miracle (1989)	Production	1<4<3<2
Shen (1989)	Production	2<3<1<4
Leather (1990)	Production	1<4<2=3
Elliot (1991)	Production	1<4<2<3
Chen (1997)	Perception/production	1<4<2<3
Sun (1997)	Production: REPT&RDGT (stimulus)	1<4<3<2
	Production: REPT (response)	3≤4<1≤2
	Production: RDGT (response)	4<1≤3<2
	Production: TRAT (stimulus)	1<2≤3<4

Note: TIDT means "tone identification task"; REPT means "the repetition task"; RDGT means "the reading aloud task"; and TRAT means "the oral translation task".
Source: Adapted from Sun, S. H. The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese. Honolulu: University of Hawai'i Press. (1997:196)

2.3.3 The Relationship between Perception and Production

Some studies consider the relationship between perception and production (Chen, 1997; Elliot, 1991; Leather, 1990; Li, 1995). Leather (1990) reported that error patterns of production correlate with those in perceptual tests, and he claimed that the perception and production of tones are interrelated. Elliot (1991) found a moderate correlation between perception and production. The relationship is not close, however, especially for the third and fourth tones. Elliot assumed that learners have different prototype to categorize the fourth tone. Sun (1997) criticized the selection of the 33 participants in Elliot's study: "fourteen participants are native speakers of tone languages such as Vietnamese or other Chinese dialects". Elliot does not discuss "the possible effects of their native language tones on L2 acquisition". Some studies did specifically investigate learners whose native language is a tone language, such as Thai (Li, 1995). Li's study identified some effects of tonal acquisition in a second tone language, but without considering further analysis of linguistic or cognitive factors. However, the results indicated that the difficulties in production of tones differ for native speakers of tonal and non-tonal languages. Chen (1997) specifically looked into the relationship of learners' tone perception errors and their tone production errors. However, this study, like the Elliot (1991) study, has a subject selection problem. Among the six subjects, three had lived in Hong Kong for more than ten months, which suggests that the tones of Cantonese may have affected their acquisition of Mandarin tones. One subject, born in California to a Chinese American family, had visited Taiwan once for three months, but did not speak much Chinese. However, a heritage student who cannot speak Chinese may still comprehend quite a bit of Chinese. Many American-born Chinese can understand Chinese, but cannot speak it. Therefore, a subject like this one probably exhibits a high proficiency in perceiving tones.

Yang and Ankenmann (2007) studied American learners of Chinese who had studied Mandarin for one or two years, none of whom was a heritage learner. The results indicated a close relationship between tone perception and production. However, the subjects' performance in tone production was much better than their tone perception. Suprasegmental and segmental categories affect tone perception and production differently. Perception is affected by tone categories, and syllable level context (initial and final categories). But tone production errors are independent of tone categories. For perception, T2 is significantly different from the other tones, with markedly fewer correct responses than for other tones. So, T2 is the hardest tone to perceive. Yet, for production, learners did not show any obvious pattern of difficulty across the tones. The results of the study showed that tone perception has a categorical preference, whereas tone production does not. For example, it is obvious that learners make more errors in T2 when they perceive tones, while there is no clear tendency when learners produce tones. The difference in the tone perception mean between initial 1 (labials) and initial 2 (alveolars) is significant at the 0.05 level, the difference of the tonal perception mean between initial 2 (alveolars) and initial 4 (retroflexes) is significant at the 0.05 level, and the difference in the tonal perception mean between final 1 (mono-vowel finals) and final 2 (multi-vowel finals) is significant at the 0.05 level. The results indicated that it is hard for NNS to perceive tones that are carried by syllables which have alveolars and by syllables which have multi-vowel finals. However, there are no significant differences between any finals or initials when learners produce tones. The results that tone perception has a categorical preference revealed that NNS perceive tones based on the interaction between categories of segments and those of suprasegmentals. Yet, tone production does not.

The difference between perception and production may be a function of scoring methods. NNS' production is perceived and scored by NS judges, while NNS' perceptions, based on a still-emergent model, must fall into one of only four standard categories. In most situations, production is based on perception (Flege, 1985). Therefore, the perception results showing that NS and NNS have different tone categories may account for the difficulties NNS face in production. In a follow-up experiment (Yang & Ankenmann, 2007), participants were asked to listen to a target tone that had been perceived incorrectly in the first tone-identification task. Listeners compared the tone in a word read by a native speaker with the same syllable read in all four tones, and then wrote down the tone type for the target word. For example, if a participant perceived the word 'ping2' incorrectly, then s/he was asked to listen to 'ping2', followed by 'ping1', ‘ping2’, ‘ping3’, ‘ping4’, and then was to write down the tone type of the target syllable. The purpose of this task was to investigate whether external tonal referents; that is, the native tonal system, can facilitate learners’ perception. The accuracy of tone perception did improve, suggesting that learners make perceptual errors because they lack a fully developed internal representation of the tonal system that matches the native speaker’s system.

Based on the results of the previous studies, it seems important for learners to develop the correct tonal categories. The results described above indicate that it is difficult for learners to acquire the tone categories of the tonal system of the L2. If these categories are incorrect, learners will make errors during tone perception and production. The results also supported by some tonal training (Leather, 1997; Wang, et al, 1999; 2003).

“Substantial evidence demonstrates that well-designed training can improve listeners’ perceptual ability of non-native contrasts at the suprasegmental level”. “The improvement in tone perception is modest, varying from about 5% to 25% across studies” (Wang, et al., 1999). And the improvement can be retained over three months after training.

However, there are not many studies of production training in L2. Leather (1997) carried out a parallel perceptual and production training study that focused on Dutch learners of Chinese learning the four tones. Another perceptual training study indicated that perceptual training can improve production (increases of 5%) even without special production training (Wang, et al., 2003). In this study, around 60 NS evaluated NNS production and statistical method is used to calculate the production scores.

Obviously, perceptual training does have an effect on both perception and production of tones. Improvement of production could be based on the development of perception, in addition to muscular control. On the other hand, methods to improve evaluation of NNS production have also improved. It is hard to evaluate NNS tonal production from the perspective of NS perception, since different NS will evaluate the same NNS production differently. To this end, a fuller model of NS perceptual categories is needed.

Therefore, both the studies about the relationship between perception and production and the tonal training studies indicate that it is necessary for us to study the perceptual categories of NS and NNS.

2.3.4 Categorical Perception of Tones

It is controversial whether tone is perceived categorically or not. Researchers have conducted numerous perceptual experiments (Blicher, Diehl, & Cohen, 1990; Gandour, 1983; Gandour & Harshman, 1978; Shen & Lin, 1991; Stagray & Downs, 1993) to explore the tone categories in Mandarin.

An early study of Thai speakers' perception of tones (Abramson, 1979) showed that there is no categorical effect for tones. However, other studies (Chan, Chuang, & Wang, 1975; Fox & Unkefer, 1985) found categorical perception effects for NS of Mandarin, in contrast to continuous perception for American English listeners. Some studies explored tone categories in beginning Chinese learners who do not have any prior knowledge of tones (Leather, 1987; Stagray & Downs, 1993; Wang, 1967).

Some of the studies on the perception of tone categories focus on a particular tone category, such as level tone (Stagray & Downs, 1993) or a contrastive pair of tones, such as T2 versus T3 (Blicher, Diehl, & Cohen, 1990; Shen & Lin, 1991.) Others focus on cross-language tone categories (Gandour, 1983; Gandour & Harshman, 1978). These studies explore the perceptual representations across tonal languages, such as Mandarin and Thai, and non-tonal languages, such as English.

In this section, we review the studies in this field. We compare the definitions of categorical perception for tone perception and propose the term used in the current research; analyze the phonetic features that the researchers used to conduct the experiments related to this topic, and compare the two main results of the research as well.

2.3.4.1 Definition of categorical perception in tone perception

Categorical perception occurs when a physical continuum is perceived discontinuously, with sharp identification boundaries between categories, and poor or absent discrimination within the categories (Liberman, Harris, Hoffmann, & Griffith, 1957; Studdert-Kennedy, Liberman, Harris, and Cooper, 1970). Consonants are perceived categorically (Liberman, et al. 1957; Rosen & Howell, 1987), while vowels are perceived continuously (Eimas, 1963; Fry, Abramson, Eimas, & Liberman, 1962).

"A stringent definition of categorical perception requires an optimal fit between observed discrimination performance and performance predicted from identification, reflecting the strong claim that discrimination between two sounds is uniquely determined by the probability that they are labeled differently (Schouten & Van Hessen, 1992). Clearly, in many cases, such an ideal fit is not obtained (Pisoni & Lazarus, 1975; Pisoni & Tash, 1974; Wood, 1976)" (Halle et al., 2004: 415).

In the field of tonal perception, most cases did not follow the original definition of categorical perception strictly, and used more liberal view of categorical perception. Some studies focused on how to categorize the phonetic variations that base on some phonetic differences, and whether the phonetic categories match linguistic categories which related to the linguistic system of a specific language (Wang, 1976; Gandour, 1983; Stagray and Downs, 1993). Other studies that explored the category boundary were closer to the experiments of categorical perception for the segments (Abramson, 1979; Halle, 2004; Wu, 2006). However, "'Categorical boundary effects' cannot be explained entirely by psychophysical responses. In other words, the increased sensitivity to differences between the members of the pairs that straddle category boundaries at least partially reflects phonetic coding into linguistic categories that may, ..., be language-specific" (Wood, 1976; Halle et al. 2004:415). Categorical perception is related to linguistic category, phonetic category and psychophysical responses, and it connects the physical properties to the real-world segmental or suprasegmental categories.

In the current research, we use the phonetic features to explore the boundaries of the phonetic categories that match the linguistic categories, and further define the ranges of each category according to the boundaries. That is, the pitch ranges of tone categories in this study. The phonetic categories are relevant to the phonetic properties, such as the low falling tone, while the linguistic categories are related to the linguistic system of a specific language, in this case, it is Mandarin, which use tones to distinguish meanings. Meanwhile, as we mentioned in 2.3.3, the studies about the relationship between perception and production indicated that the perceptual categories of tones are important. Therefore, on the other hand, the results also showed that the perceptual categories are closely related to tone production, at least for NS. Hence, in the current study, we also explore the tone categories during production that might base on perception. Finally, the term "perceptual category" is used in the study to investigate NS and NNS tone perception and production instead of the traditional "categorical perception".

2.3.4.2 Phonetic features explored in the studies about tone perception

Early research about tone perception raised the controversial issue about whether NNS are sensitive to small phonetic difference while NS are not. Wang (1967) and Stagray & Downs (1993) found that NNS have different categories compared to NS, and that NNS are sensitive to some small acoustic differences that NS ignore. However, Leather (1987) argued that some NNS may have the similar categories as NS, and that

NNS are less sensitive to small pitch differences. However, all of the studies agree that errors come from interrelated contrastive pairs that related to a specific phonetic feature.

A complicating factor is that the studies about tone perception used different phonetic features when the researchers explored tone categories. Stagray and Downs (1993) used pitch height to explore the level tone; Blicher, Diehl, and Cohen (1990) used duration to compare T2 with T3; Shen and Lin (1991) indicated that the salient acoustic feature that can distinguish T2 and T3 is the turning point of the pitch contour; and Gandour and Harshman (1978) and Gandour (1983) used height and direction, similar to register and contour, to characterize a multi-dimensional model for the perceptual representation of tones. In the studies, subjects make similarity judgments amongst the standard tones. The dimensions that Gandour and Harshman (1978) and Gandour (1983) identified that the resultant perceptual space are close to phonetic features which, they argued, are sufficient to capture the whole system of Mandarin tones. In Halle, et al.'s (2004) experiments, "intermediate contours were obtained via interpolation between endpoints." (Halle, et al. 2004:401). Wu's (2005) study was based on Shen and Lin's (1990) results, which assumes that the turning points of total contour for T2 or T3 provide the most important cue to distinguish these two tones. Therefore, the experiments in Wu's study focused on changing the turning point to a different time. It is hard to compare categories systematically when they are based on different foci.

One problem with the acoustic studies of Mandarin tones centers on the features of T3. Most acoustic studies (e.g., Xu, 1999) and phonological studies (e.g., Chen, 2000) indicate that T3 is a low falling tone in most contexts. However, T3 is perceptually contrastive with T2 and does not contrast with T4, which is a high falling tone. Thus, there must be other cues that influence the perception of tonal categories. However, among the studies of tone perception, some researchers employed some phonetic properties that are used to contrast T2 with T3.This problem need to be further explored.

In addition to T2 and T3, we also need to use the same phonetic parameter to explore the tonal system during perception and production.

2.3.4.3 Comparing NS perception with NNS perception

Although using different phonetic parameters to study, most researchers are interested in the question whether and Why NNS have different categories of tones from NS. Recently, research has focused on this topic. Using different methodologies (see 2.3.4.2), two recent articles investigated NS and NNS categorical perception (Halle, et al., 2004; Wu, 2005). Halle et al. (2004) showed that NNS whose L1 is not a tone language do not process tones linguistically, and their perceptual judgments are based on a general psychophysical function. This research indicates that NS perceive tones in a quasi-categorical way, whereas NNS use a different strategy. The results are similar to Burnham and Jones (2002). Halle et al. (2004) used the elicitation phrase *yi ge X zi* (one + measure word + X + Chinese character). But X still works as a single syllable, independent of sandhi changes, because it is not linked with either the previous or following syllables to compose a multi-syllable word. This is not a natural context. Therefore, the results still reflect a context-free environment. Wu's (2006) study investigated two contrastive pairs: T1 and T4; and T2 and T3. The results indicate that the perceptions of T1 and T4 are categorical, whereas the perceptions of T2 and T3 are not categorical. Therefore, it is hard for learners to acquire the contrast between T2 and T3.

The two studies above indicated that there are two explanations about the topic why NNS have different categories of tones from NS. One is that NNS use different strategy to perceive tones (Halle et al. 2004), and the other lies in the tonal system in L2 (Wu, 2006). The early research by Stagray and Down (1993) also supported the first reason. However, Wu's study is also based on a context-free environment. As we mentioned above, some previous research indicated that in context, T3 is much closer to

T4, since the final rising in T3 is usually absent in context. Therefore, in context, both T3 and T4 are falling tones, and the only difference between them is the register.

Overall, these studies did not investigate all contrastive pairs, but instead focused on only some of them. Also, the experiments were conducted with stimuli presented without a context. In addition, these context-free experiments use different methods to synthesize the tones (see 2.3.4.2). It is natural that different methods may be needed to synthesize the tones if different variations in contexts are considered. Therefore, it remains necessary to explore perceptual categories in NS and NNS perception and production systematically and in context.

2.3.5 Speaker and Rate Normalization

The term "normalization" in this section refers to the concept that listeners normalize speaker pitch range and speaking rate in tonal perception. Some studies focused on the context effects of neighboring tones (e.g., Fox & Qi, 1990), while others explored whether speaker information is relevant to tone identification (Moore, 1995).

Moore's (1995) study investigated only T2 and T3. The results indicated that tone identification is influenced by changes in speaker identity. It showed that pitch range influences the recognition of T2 and T3. Moore's results (1995:130) indicated that "it is necessary for us to further clarify the relationship between acoustic variability and normalization".

Moore (1995) also examined rate normalization. These data suggested that "speaker F0 range may be more important to tone identification than rate information" (Moore, 1995:173). The results indicated that rate normalization does not influence tone identification for stimuli varying in a non-temporal dimension. Given these results, the current study focuses on pitch range and rate information will not be manipulated (Additional reasons for this decision can be found in Chapter 3).

2.3.6 Conclusion

2.3.6.1 The largest difficulties in NNS tone acquisition

Tones are the most difficult to acquire for NNS who do not have any knowledge of tones. Which aspect is the most difficult? The studies discussed above have shown that the difficulties for perception are different from those for production. Based on these studies, most researchers agree that T2 is the most difficult tone to perceive, while no specific tone stands out as the most difficult to produce.

The studies that focused on the acquisition order of tones indicate that one of the reasons that research on tonal production has yielded heterogeneous results is that some studies used acoustic features such as register to explain tone errors while others used different acoustic features, like contours. Sun (2002) stated that “the other [difficulty lies] in holding in the mind the acoustic-phonetic information of tones long enough to be able to interpret the information phonemically” (p. 201). Other early categorical perception research (Stagray & Downs, 1993) demonstrated that NNS are sensitive to small differences of acoustic features while NS are not. It is obvious that the greatest difficulty for learners lies not at the phonetic level, but at the phonological level, making it hard for learners to categorize tones. For NNS without any previous tonal experience, it is hard to acquire a whole tonal system that has clear categories. Studies about the relationship between perception and production, such as the experiment in Yang and Ankenmann (2007), also demonstrated that accessing NS’ tonal system can help NNS improve their tone perception.

The gap between the phonetic level and the phonological level indicated that acoustic cues do not simply map onto perceptual features. Not all differences in acoustic cues will be discriminated by NS. However, NNS may respond to these acoustic differences and categorize tones differently. On the other hand, if two tones have the same acoustic features, but appear in different contexts, will NS perceive them as belonging to the same tonal category? Thus what makes tone acquisition difficult for learners of Chinese whose L1 is not a tone language is not a specific tonal category, but rather that these NNS do not have a tonal system similar to the NS tonal system. Therefore, it is important for us to study the perceptual space for tonal categories.

2.3.6.2 Problems in the research on NNS tone acquisition

There are five major problems in the previous research on NNS tonal acquisition. First, the experiments in some studies (e.g., Sun, 1999; Wu, 2005) focused on context free tones. Yet, contexts are central during conversation. A tone has variations when it is in context or in isolation. For example, the citation contour of T3 first falls, then rises without a context; however, the contour is falling in context (Figure 4). If tones are investigated without a context, T3 is regarded as contrastive to T2 according to the contours. Yet in context, T3 can contrast with T4, which has the same falling contour but a different register. So the contrast here depends on register. On the other hand, each tonal category also has different tonal variations in different contexts.

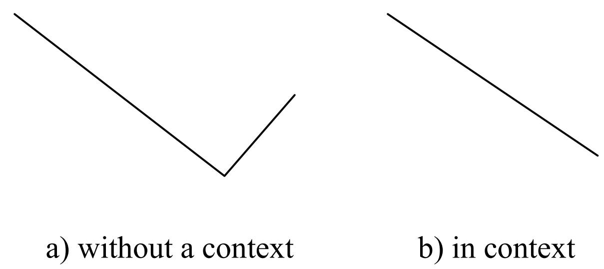


Figure 4 T3 not in a context and in a context

Take T3 as an example again. "The T3 that has the tone value 21(4)[4] is low falling that is represented by the numerical number 21 in non-final position, but acquires a rise, shown by (4) that follows 21, at the end of a phrase" (Yip, 2002:181). How should these variations of a tone category be processed during perception and production? This problem has not been resolved in NS speech, although some research has considered contexts (Broselow, Hurtig, & Ringen, 1987; Fox & Qi, 1990; Hgan & Rozsypan, 1983; Leather, 1983; Xu, 1997).

Moreover, almost all studies about categorical perception of tones (section 2.3.4) have been conducted without context. It is necessary for us to do some experiments in context.

Second, acquisition of the differential features in Mandarin is still a controversial issue. The results from the studies about acquisition order (section 2.3.2) raised an important question: what is the most difficult for learners of Chinese to acquire? The discussion in section 2.3.6.1 showed that the largest difficulty lies at the phonological level. Recall that the review of tonal representation in phonology presented in 1.1 indicated that there are two descriptive features: contour and register. White (2003:139) stated that "L2 learners can acquire feature strength which differs from the L1, as well as features which are not instantiated in the L1". As for the new functional categories, some research (Leung, 2001; Robertson, 2000) indicated that L2 learners acquire the new categories, together with the associated features. At the suprasegmental level, it is worth exploring whether learners of Chinese acquire tone categories together with the contour feature and the register feature.

Some researchers (e.g., Sun, 2000) hold the view that Mandarin does not have phonemic tonal contrasts that depend on register. This view is used to criticize Shen's (1990) results. However, this view is based on tones without a context. As mentioned

---

4 21(4) denotes the pitch value of T3. The value 4 is optional.

above, other researchers (Chen, 2000; Xu, 1997; Yip, 2002) present the falling contour of T3 when it is used in a context, thus establishing the contrast with T4, which also has a falling contour. Thus, the contrast lies in the register. Yet, the problem is that learners of Chinese do have problems in distinguishing T2 and T3, although the acoustic feature of T3 is closer to T4. This issue needs a resolution from the non-acoustic cues.

Moreover, speakers acquire tones as a system of contrasts, not as individual items. Therefore, the system of tones should be studied integrally. However, almost all of the studies of tonal acquisition by NNS focus exclusively on one of the four citation tones in isolation or focus on some contrastive pairs, which are illustrated as a problem for studies about categorical perception of tones (section 3.4). For example, in Halle et al. (2004), the researchers explore three pairs of contrastive tones: T1 and T2; T2 and T3; T3 and T4. In fact, there are five tone categories in Mandarin, although the neutral tone can be subsumed under tone sandhi. The neutral tone (T0) is widely used in connected speech. It can also distinguish meanings, as shown in 5) below.

5) *dong1 xi1* versus *dong1 xi0*
east (and) west thing

The neutral tone is always neglected in the second language acquisition area. The neutral tone is used for some specialized phrase/word structures, such as reduplications, *kan4 kan* 'look at'; *kuai4 kuai de* 'quickly'. The main phonetic properties of the neutral tone may be weaker. If it has a pitch, the pitch value depends on the phonetic environments. Namely, the neutral tone carries different tone values when it comes after a level tone, a dipping tone, a rising tone or a falling tone (Cao, 2002b). For learners of Chinese, if they cannot distinguish a neutral tone from a normal tone, they will fail to comprehend word meanings. Similarly, if they fail to produce a neutral tone, listeners may mistake their meaning, in addition to noticing a foreign accent. Research on the neutral tone indicates that it is processed as a variation of the other four tones (e.g., Li, 1981; Lin & Yan, 1980). However, other researchers claim that the neutral tone is independent (Qiao, 1956; Xu, 2004; Zhang, 1956). This problem also needs resolution. Otherwise, it is hard to place the neutral tone into the whole tonal system. This might also be the reason that few studies have been conducted on the neutral tone. Meanwhile, we also need to know what the difficulties for NNS to acquire the neutral tone is.

Another problem concerns the empirical measurement used in production experiments. Many NNS pronunciations are not clear enough for NS to judge their accuracy, as discussed in section 2.3.3 above. Early research on acquisition order generally recruited two to ten NS to judge the accuracy of NNS tones, then used inter-rater reliability checks on the consistency of the scores (e.g., Chen, 1997). Others (e.g., Shen, 1989) used acoustic features to measure the accuracy of NNS tones. Shen's study counted errors in NNS use of register. It is still a controversial issue whether Mandarin has phonemic contrasts that are conditioned by register or pitch range. As stated above, acoustic features are not equal to perceptual features. Acoustic features are at the phonetic level, while listeners perceive tones at the phonological level. Some researchers who carry out production training realize this problem, so they use statistical methods to assess NNS production. For example, Wang (2003) recruited 50 NS to evaluate learners' tones. However, whether they use inter-rater reliability or statistical calculations, researchers need many NS to participate in scoring. The more NS judges there are, the better the evaluation of tone production will be.

Even so, it is impossible for NS to judge the tones of a NNS as absolutely correct or absolutely wrong. The only thing a NS can tell is whether the tones are similar to NS tones or far from them. What the judges perceive is the NNS' production. If we developed a model of NS perception, it would be easier for us to measure to what extent NNS tone production is close to or far away from NS perceptual categories.

The last problem is that most studies have compared NS' perception with NNS' perception or NS' production with NNS' production, yet few of them have compared NS' perception with NNS' production and NNS' perception with NNS' production. This study will analyze the relationship between NNS' perception and NNS' production; compare NS' perception with NNS' perception, and NS' perception with NNS' production.

In view of the research about NNS tonal acquisition, it seems that most problems will require further study of the NS tonal system, especially NS perception of tones. Thus it is necessary for us to set up a model of NS perceptual categories. These categories can be compared with NNS' tone perception and NNS' tone production, and they can be used to evaluate production. Since this model is also used to explore NNS tone production, it should focus on two aspects: (a) the range and the boundaries of the perceptual categories of tones, and (b) the NS process of perceiving tone categories based on the tone features.

2.4 Conclusions and Research Questions

According to the review of the literature on linguistic and second acquisition studies of tones, two issues need further exploration: (a) laying out the whole tonal system of Mandarin Chinese; (b) the gap between acoustic features and perceptual results.

Given these issues, we need a study that (a) sets up a perceptual model for Mandarin tones systematically and in context; (b) explores the phonetic cues and other non-acoustic cues that could be used to contrast tones and influence tonal perception especially for T3; (c) generates a model against which we can evaluate NNS' tonal production. Chapter 3 presents the argument for the selection of the appropriate features to be addressed in this study. Therefore, we need to observe NS' tonal production, then the perceptual experiments could be designed based on the features of NS' tonal production.

The following research questions will be addressed in order to set up a model of perceptual categories for Mandarin tones.

1. What is the production space for native speakers?

2. Are tones perceived categorically by NS and NNS? If so, how can we describe them?

2-1. What are the pitch ranges of the perceptual categories of T1, T2, T3, and T4 for NS?

2-2. What are the pitch ranges of the perceptual categories of T1, T2, T3, and T4 for NNS?

2-3. What is the difference between NNS' tonal perception and NS' tonal perception?

3. How do pitch values of tones create overlapping areas?

3-1. Does pitch play an important role in perceiving the neutral tone?

3-2. Do the pitch at the end of T3 and the pitch at the beginning of the following tone play an important role in perceiving T3?

4. What are the features of NNS' tone production? What are the differences between NNS' production features and NS perception features?

CHAPTER 3 VARIABLES AND PITCH NORMALIZATION

This chapter describes the variables used in this study, introduces how to normalize pitch, and explains why five tonal categories are defined in this research.

3.1 Variables

In this section, the researcher examines the acoustic variables of tones, and selects two acoustic variables for this study. Meanwhile, two factors that could affect tone perception – finals and the position of tones, are discussed.

3.1.1 Acoustic Variables

To set up a perceptual model which can be used to categorize tones, the most important task is to select a primary feature that will distinguish the different tones. Most studies regard fundamental frequency (F0) as the primary acoustic cue for tones. However, duration is included in this study as another primary cue for the neutral tone, since short duration is its salient feature. Some research shows that T3 has a longer syllable duration than T2 does (Blicher et, al. 1990). However, most tones in connected speech are around 350 ms. The duration of individual syllables in Chinese speech has a wide dynamic range, since it is governed by multiple factors that act at different levels. These include the syllable's position in the utterance, morphological constraints, and syntactic context (Cao, 1989, 2002a). Although the tonal duration is one of the acoustic variables that distinguish T1, T2, T3 & T4, it is not sufficient distinguish the tonal categories, especially in this study. The reason is that we use the neutral model to remove the beginning part and the ending part of a tone which is relevant to the nearby pitch values and irrelevant to duration.

In tonal phonology, two essential tonal features H/L and h/l are relevant to register and contour[5] (Yip, 2002). However, using these features presents two problems. First, it neglects the tone duration. An identical syllable which has the same CV combination may differ lexically when the tone duration is shortened. Second, the two proposed phonological features, H/L (register) and h/l (contour), share the same acoustic feature: pitch. And as we have discussed earlier, the pitch value of 3 could be regarded as in either a high register or a low register.

The first issue of duration makes it hard for researchers to deal with some special tones, such as the Ru tones of Shanghai Wu dialect, and the neutral tone in Mandarin. The second problem occurs because the tonal features Yip proposed are based on acoustic features. Different tonal geometry models (Yip, 1989; Duanmu, 1990; Bao, 1990; Yip, 1993) make different predictions about which features can spread independently, and which can spread together (Yip, 2002:53). Although the most powerful model is Bao's, that model cannot account for the range of tone variations.

Almost all research uses two variables to study tones: duration and F0 (Zhu, 1995; Xu, 1997; Cao, 2002b). Some studies that focus on T2 and T3 use the turning point as an additional parameter (Shen & Lin, 1991; Moore, 1995; Wu, 2005). Most research also uses δF0 as a variable to study the F0 change of a tone (Moore, 1995; Xu, 1997). The selection of variables is based on the nature of the research. In the current study, the whole system is examined, and the pitch values at the starting point and the end point of the nucleus of a tone are selected as variables. The slope (δF0/T, T is the duration of the nucleus of a tone) is not considered in this research. There are three reasons.

[5] The term "contour," the shape of the tone, is used in some phonological research (such as Yip, 2002) to contrast with the term "level."

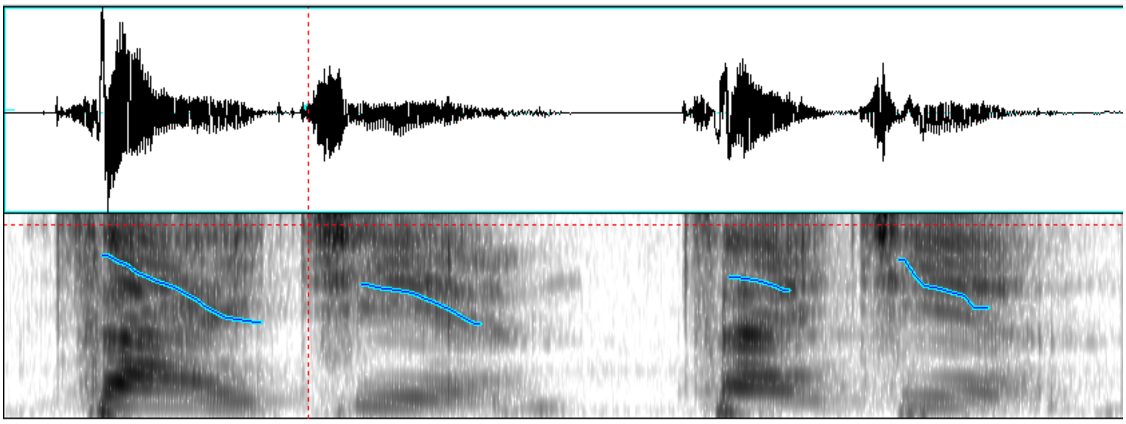


Figure 5 Pitch contours of Zai4 Jian4 (bye-bye)

First, the perception of a tone does not depend on the F0 slope. Figure 5 illustrates the pitch contours of the same word (zai4 jian4 'bye-bye') with two different rates. It is obvious that the first Zai4 in Table 4 does not have the same slope as the second Zai4. However, the slope of the first Jian4 is similar to the second Jian4. Yet, all four tones in these two pairs are T4. Therefore, even if the slopes are different (Table 4), tones in the same tonal category can still be perceived as the same. On the other hand, δF0 is also different among the four tones in figure 5. However, the distance between the starting point and the ending point of a tone can tell us the direction of each tone; that is, whether it is a level tone, a falling tone, or a rising tone. In this study, the difference in F0 between the starting point and the ending point of a tone represents the direction of a tone; and the positive or negative nature of the pitch values at the starting point or the ending point can be used to judge the pitch register of a tone.

Speaking rate is not a variable that can affect the tonal perception in this study as all the tokens are produced at the same speaking rate. Therefore, the two pitch values that define the beginning of the nucleus and the end of the nucleus of the tone are selected as our acoustic variables.

Table 4 Comparing sounds in fast speech and slow speech

	Time (ms)	δF0	Slope
Zai4 (slow)	240	97.44	0.41
Zai4 (fast)	180	20.49	0.11
Jian4 (slow)	100	56.01	0.56
Jian4 (fast)	120	69.73	0.58

This study aims to use a simplified tone to set up the pitch range of all tones in the tonal system. Not every tone has a turning point. The turning point cannot be used to describe all tones in the system. On the one hand, the turning point cannot distinguish T1 and T4 since they do not have turning points. On the other hand, T3 generally is a falling tone in context. The turning point cannot distinguish T2 and T3 in most cases. Therefore, turning point is not appropriate in this study.

In order to set up appropriate phonemic contrasts, duration, F0 at both starting and ending points are used as variables, as shown in Table 5 below. Duration is defined as the time from the beginning point of the nucleus of a tone to the ending point of the nucleus. The nucleus part is from the first turning point of a tone (if it exists) to the last turning point of a tone (if it exists).

Table 5 Variables for phonemic contrasts for tones

Variable		Phonological feature	Acoustic feature
Duration		Length	Time
F0 at the starting / ending point	F0's positive/negative feature	Register	Pitch range
	Distance between two F0s	Contour	Pitch change

Since each person has a different pitch range, F0 at both starting and ending points will need to be normalized to analyze tones produced by different speakers in this study.

3.1.2 Other Factors

Although the nucleus of a tone is used to avoid many variations of each tone that are caused by consonants, prosodic boundary, stress, and etc, two other factors still need to be considered: finals and the position of a tone in a sentence.

There are three kinds of finals in Mandarin: single vowel finals, multiple vowel finals, and nasal finals. All three kinds of finals are examined in this section. To determine if there are any differences between the tones as a function of these three kinds of finals a sample corpus is generated.

3.1.2.1 Sample corpus

The corpus contains three carriers including three tokens. Four target tones are embedded in each token.

Three syllables that have a single-vowel final, a multi-vowel final, and a nasal final are selected as stimuli. They are *da* /ta/, *tao* /t‘au/, and *wan* /uæn/. In table 3-3 there are three groups of words. The first group is the single-vowel group, which includes four words in which the first syllable of each word has a single-vowel final *a* /a/. The second group is the multi-vowel group, which includes the words in which the first syllable of each word has a multi-vowel final *ao* /au/. The third group is the nasal-final group, which includes the words in which the second syllable of each word has a nasal-final *an* /æn/. The tokens in which the target tones are embedded are delivered through three carrier sentences respectively:

1) ta1 **da**dao0 shi2me cheng2du4?

2) Ta1 yao4 **tao**qian2.

3) San1**wan** shi4 shi2me yi4si?

Table 6 Target tones and carriers for finals

Finals	Characters	Pinyin	Meaning	Carrier Sentence
Single-Vowel Final	答到	da1dao0	answer	他答到什么程度? ta1 da1dao0 shi2me cheng2du4? How well does he answer?
	达到	da2dao0	arrive at	他达到什么程度? ta1 da2dao0 shi2me cheng2du4? To what degree he reaches?
	打到	da3dao0	hit	他打到什么程度? ta1 da3dao0 shi2me cheng2du4? How well does he play (basketball)?
	大到	da4dao0	big	他大到什么程度? ta1 da4dao0 shi2me cheng2du4? How big is he?

Table 6 continued

Multi – Vowel Final	掏钱	tao1qian2	pay money	他要掏钱. Ta1 yao4 tao1qian2. He wants to pay.
	逃钱	tao2qian2	skip money	他要逃钱. Ta1 yao4 tao2qian2. He wants to avoid to paying.
	讨钱	tao3qian2	ask for money	他要讨钱. Ta1 yao4 tao3qian2. He wants to ask for money.
	套钱	tao4qian2	drag money	他要套钱. Ta1 yao4 tao4qian2. He wants to drag money.
Nasal Final	三湾	san1wan1	Sanwan (city name)	三湾是什么意思? San1wan1 shi4 shi2me yi4si? What does sanwan mean?
	三丸	san1wan2	Three (medicine)	三丸是什么意思? San1wan2 shi4 shi2me yi4si? What does sanwan mean?
	三碗	san1wan3	three bowls of …	三碗是什么意思? San1wan3 shi4 shi2me yi4si? What does sanwan mean?
	三万	san1wan4	thirty thousand	三万是什么意思? San1wan4 shi4 shi2me yi4si? What does sanwan mean?

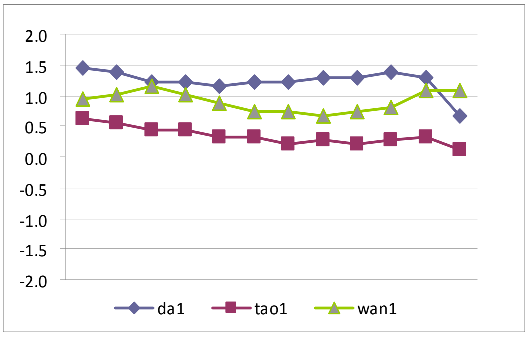
3.1.2.2 Procedure and analysis

A female speaker whose native language is Mandarin read the sentences in Table 6. Her speech rate is normal, and the sounds are recorded and analyzed via the program Multi-Speech.

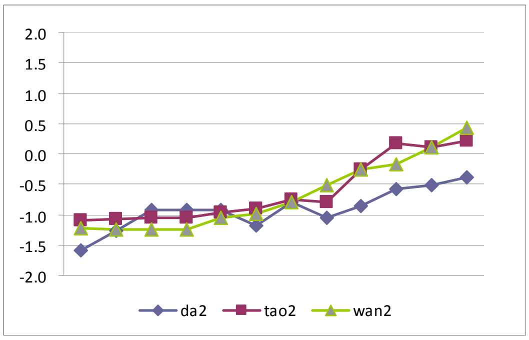
F0 values are measured at the 5% point of the tone duration and every one-tenth of the duration for the tones. Then, these F0 values are normalized via z-score (see 3.2.2).

3.1.2.3 Results

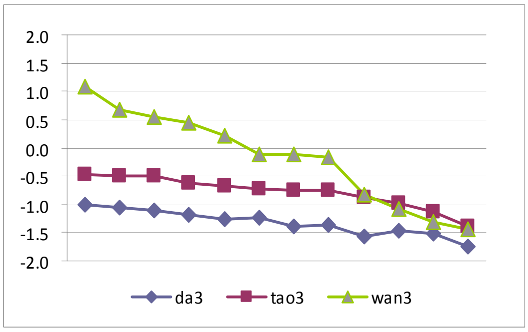
The results indicate that the types of finals do not affect the tones.



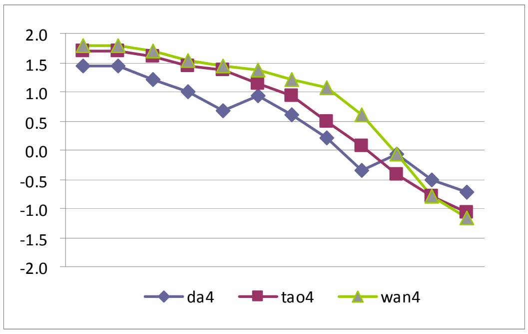
a. T1 carried by different finals



b. T2 carried by different finals



c. T3 carried by different finals



d. T4 carried by different finals

Figure 6 Comparison of tones carried by different finals

In figure 6, T1 carried by a single-vowel final, a multi-vowel final and a nasal final have the same pitch contour, and exist at the same register of the pitch space. The same thing happens to T2 and T4. As to T3, the three finals have a similar pitch contour. However, the starting point of T3 that is carried by the nasal final is higher than those of others.

The starting point of T3 that is carried by a nasal final is close to the starting point of T4 that is carried by the nasal final. Figure 7 indicates that the two tonal contours are similar.

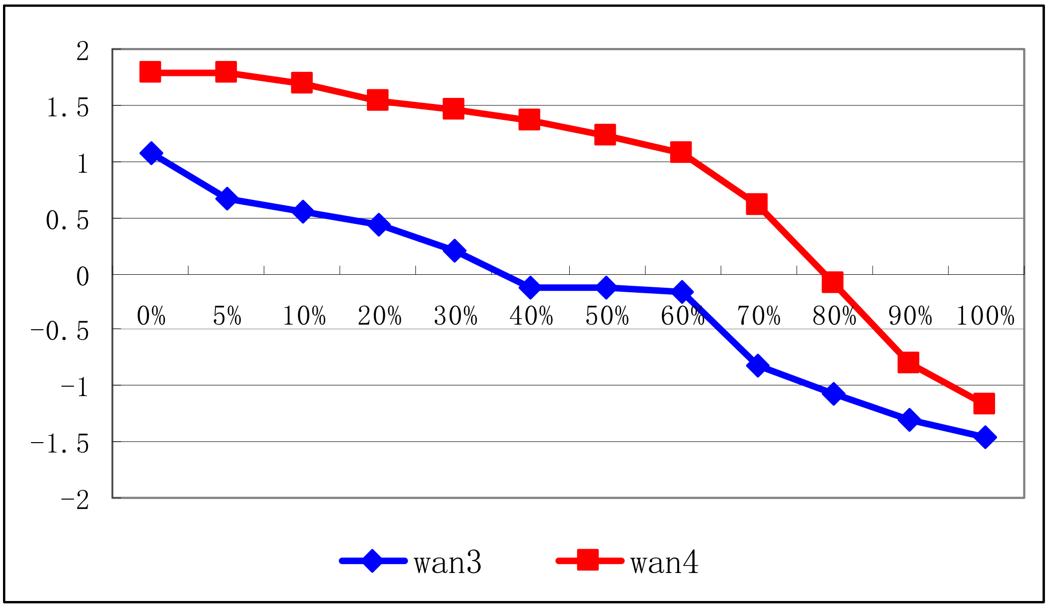


Figure 7 Comparison of T3 and T4 carried by a same nasal final

Observing three carriers sentences:

1) ta1 **<u>da</u>**dao0 shi2me cheng2du4?

2) Ta1 yao4 **<u>tao</u>**qian2.

3) San1**wan** shi4 shi2me yi4si?

We find that the syllable that comes before the token has a high-level pitch value at the end of the tone in 1), and the syllable that comes before the token has a low-level pitch value at the end of the tone in 2). However, this difference does not influence the acoustic features of tones of the token in 1) and 2). It shows that the ending point of the previous tones does not affect the target tone which is carried by the token. Although the syllable that comes before *wan* has a high-level pitch value at the end of the tone, it does not raise the pitch value at the starting point of T3. Observing the tone that comes after the target tone in the three carriers, we find that dao0 is a neutral tone and its pitch follows the previous tone in 1); qian2 has a T2 which has a low-level pitch value at the beginning of the tone in 2); and shi4 has a T4 which has a high-level pitch value at the beginning of the tone in 3). It is possible that the tone (shi4) that comes after the *wan* may cause the similarity of the acoustic features of T3 (*wan3*) and T4 (*wan4*).

3.1.2.4 Conclusion

The results show that the types of finals do not generally affect the tones. Therefore, we chose not to set the types of finals as a variable with one exception, since T3 and T4 have a similar pitch contour in the carrier 3 when they come before a tone at the high level in the pitch space.

3.2 Pitch Normalization

In order to establish a model of perceptual categories for native Mandarin speakers, certain technological and methodological problems need to be considered. One of them centers on individual speaker differences in their baseline F0. F0 is a primary acoustic parameter for this study as we showed it in 3.1.1. However, the cues that cause either intra-speaker or inter-speaker variations of F0 need to be avoided. The term *normalization* is used to deal with such variations. Two kinds of F0 normalizations are discussed in this section.

3.2.1 Intra-speaker Normalization

Generally, a person can produce different variations of a tone in which the acoustic features are influenced by contexts, such as word boundary, sentence boundary, stress, and intonation. According to Zhang and Hirose (2004:452), “through F0 observations it was found that an articulatory F0 transition may occupy a large portion of a syllable F0 contour: 30% in (Howie, 1974), 50–100 ms after the release of the initial consonant in (Rose, 1988 : 76) (accounting for about 40% proportion in the given example [1988:65]), and more than 50% when the rising tone starting from the middle of the rhyme (Shih, 1988).”

Some studies have suggested that a tone can be divided into several parts. Generally, a tone is divided into three parts: an onset[6], a central part (nucleus) and an offset. In Gottfried and Suiter’s (1997) research, “formant frequencies were measured for the speaker’s vowels at two points in the syllable: 50 ms after the initial consonant burst and 50 ms before the cessation of significant vocal energy” (Gottfried & Suiter, 1997:212). Some researchers use the first six and final eight pitch periods of the intact syllables as the onset and the offset (Lee, 2008). The problem with this method is that the durations of the onset and the offset are fixed. Yet, in real speech, the influence is different if the contexts are different. That is, the durations of onset and offset differ based on different environments, such as consonants, vowels, and prosodic cues.

Recently, researchers proposed (Zhang & Hirose, 2004; Wang, et al. 2008) a new model to handle these intra-speaker variations; this is the tone nucleus model.

---

6 Some papers use the term “segment” (Gottfried & Suiter, 1997; Lee, 2008; etc.). Since the term “segment” in phonology means vowels and consonants, I use “part” instead of “segment.”

Sub-syllabic F0 segments		
(1)	2	(3)
Tone onset		Tone offset
(Onset course)	Tone nucleus	(Offset course)
←	Syllable F0 contour	→

The F0 segments in parentheses are optional; only the tone nucleus is obligatory.

Figure 8 Proposed F0 segmental structure model of Chinese syllable F0 contours

Source: From Zhang, J. & K. Hirose. Tone nucleus modeling for Chinese lexical tone recognition. *Speech Communication*.42:447-466. (2004:450)

In the tone nucleus model, the F0 contour of a syllable may be divided into three parts: onset course, tone nucleus, and offset course (see Figure 8 above). The tone nucleus represents the main pitch contour and is obligatory; the onset and the offset courses are articulatory transitions and are optional because they are influenced by contexts easily, such as word boundary, sentence boundary, intonation, and stress. The tone nucleus of a syllable is assumed to be the target F0 contour of the associated tone; it usually conforms more closely to the standard tone pattern than the articulatory transitions do.

Zhang and Hirose (2004) indicate that the nucleus is unaffected by F0 variations that result from voice/voiceless initials, word boundary, sentence boundary, intonation and stress.

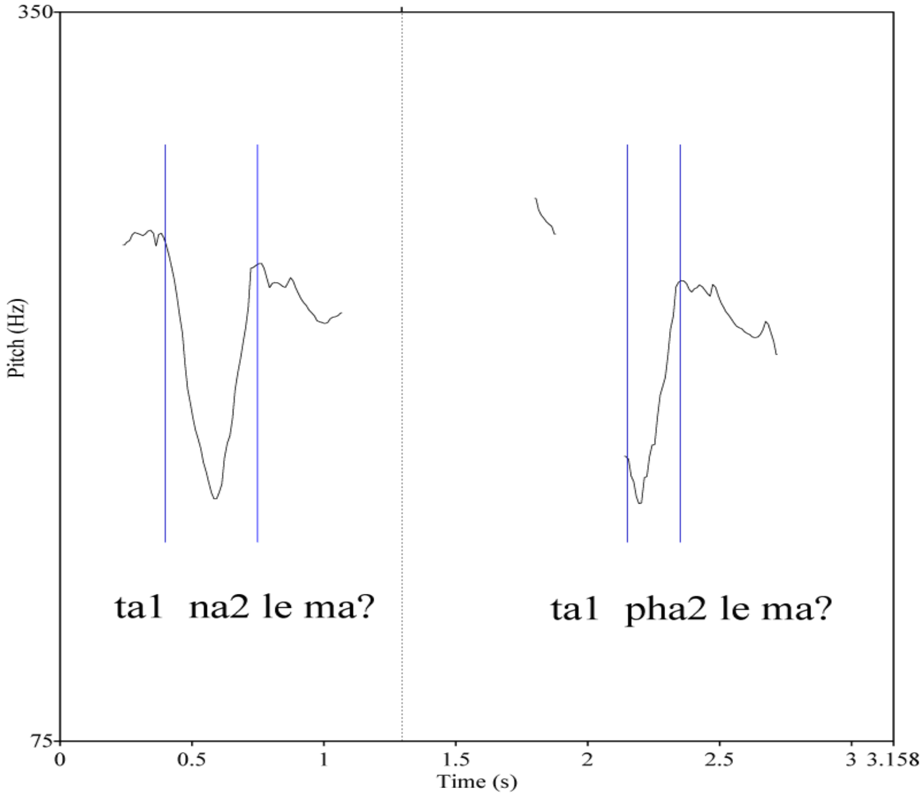


Figure 9 Tones carried by a voiced consonant and a voiceless consonant

Figure 9 illustrates F0 contours of two continua of /a^{35}/ (a2). The syllable /na^{35}/ in the sentence on the left has a voiced consonant, while the syllable /pha^{35}/ in the sentence on the right has a voiceless consonant. The thin vertical lines indicate the syllable boundaries. The pitch contour of /na^{35}/ has a dipping shape like that of Tone 3, and it tends to be mis-recognized as Tone 3. However, if the tone-nucleus model is used, the contours can be divided into three parts and the first part of the pitch contour of /na^{35}/ can be cut. Thus the pitch contour of /na^{35}/ is the same as that of /pha^{35}/. Therefore, focusing on the nucleus part of a tone avoids variations that are influenced by consonants.

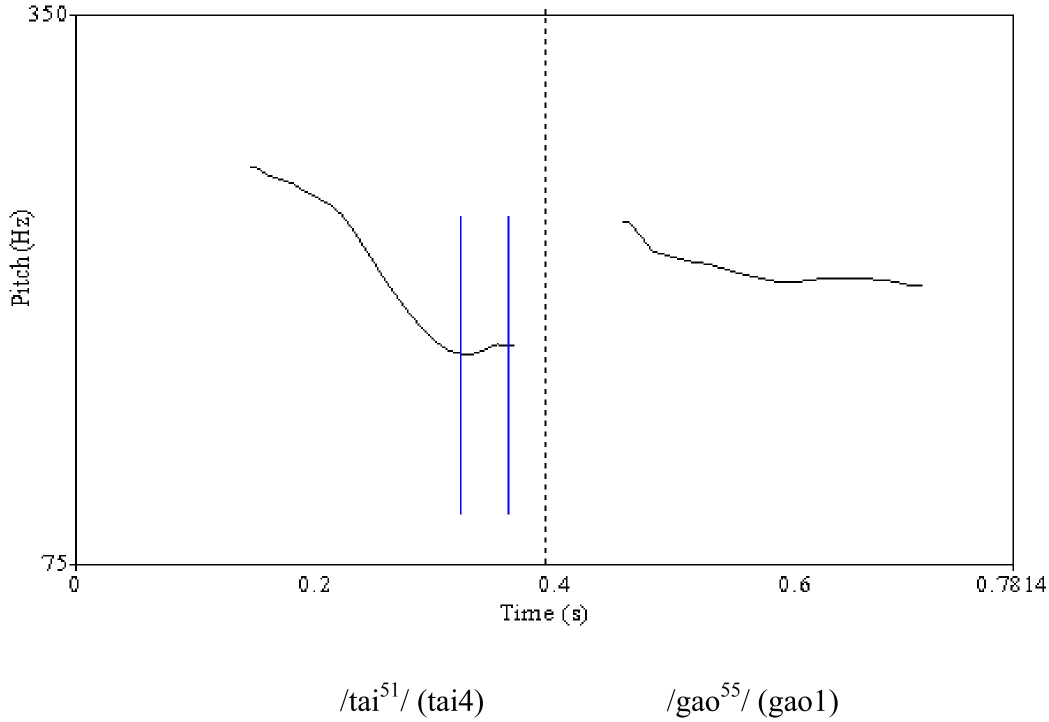


Figure 10 F0 contours of ‘‘T4 + T1”

Figure 10 illustrates F0 contour of T4 and that of T1. The contour on the left is that of T4, and the contour on the right is that of T1. The pitch contour between the thin vertical lines is a flat portion with a little rising, and it is a part of the production of T4 which is a falling tone. However, this slightly rising ending is not relevant to perception. It is obviously affected by the following T1 which has a high register beginning. The tone-nucleus model is used to remove the contour between the two vertical lines. Thus there is no effect, which comes from the second tone, on the pitch contour of T4. Therefore, using this model can avoid variations on the prosodic phrase boundary that is influenced by neighboring linguistic contexts, such as following tones.

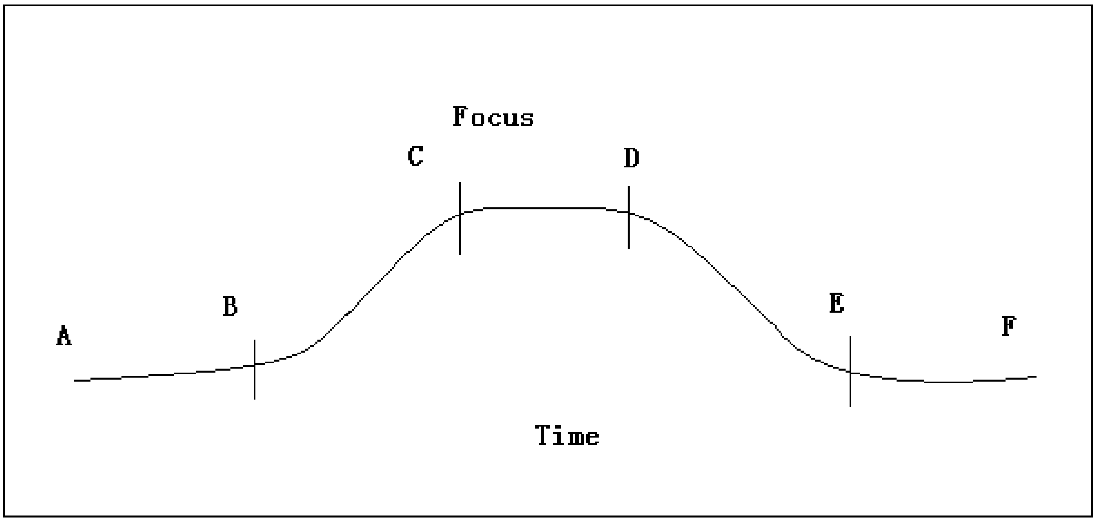


Figure 11 Contour of focusing on the second T1

Source: from Zhang, J. & K. Hirose. Tone nucleus modeling for Chinese lexical tone recognition. *Speech Communication*. 42:447-466. (2004: 453)

Figure 11 illustrates a sequence of three T1 tones where the focus of the phrase is on the second T1. This could be regarded as the contour of Xu's (1999) example "mo1 mao1 mi1" (touch the cat) which the focus is the second T1. The stress of the phrase "mo1 mao1 mi1" is the second syllable "mao1", which carry a highest tone contour in this case. We see that the focus leads to a substantial raising in the early portions of F0 contour of the second T1 and a falling F0 contour in the beginning of the third T1. The parts BC and DE in this figure are regarded as parts of the articulatory transition between two T2, according to the tone-nucleus model, while the parts AB, CD and EF represent the tone nuclei of the three T1. Compared with the whole F0 contours of the three T1 tones, the tone nuclei show a similar contour that is level. Thus we can regard the level contours represented by AB, CE and EF as the pitch contour of T1 (Zhang & Hirose, 2004:453). Therefore, using the nucleus part can also avoid the variations influenced by prosodic features, such as stress in this case.

The tone nucleus model can be used to eliminate the parts of a tone which are easily affected by the surrounding linguistic environment. In this research, we divide a tone into three parts manually. The nucleus of a tone pitch contour is identified as that part of the F0 contour between the two turning points at the beginning and the end of the tone. If there are no turning points in a tone pitch contour, then whole contour is identified as the nucleus.

If there is a falling part at the beginning of a rising contour, we regard this part as an onset course and eliminate it from the analysis; if there is a falling part at the end of a rising contour, we regard this part as an offset course and eliminate it (see Figure 12). If there is a rising part at the beginning of a falling contour, we regard this part as an onset course and eliminate it; if there is a rising part at the end of a falling contour, we regard this part as an offset course and eliminate it (see Figure 13). If a tone is a level tone, whether there are any peaks at the starting point depends on the pitch level of the ending point of the previous tonal contour. For example, if the ending point of the previous tone has a lower pitch level than the observed tone and there is a rising part at the beginning of the level contour, we regard this part as an onset course and eliminate it (see Figure 14). If the starting point of the previous tone has a higher register than the observed tone and there is a falling part at the beginning of the level contour, we regard this part as an onset course and eliminate it (see Figure 15). On the other hand, if a tone is a level tone, whether there are any peaks at the ending point depends on the pitch level of the beginning point of the following tonal contour. For example, if the starting point of the following tone locates has a higher pitch level than the observed tone and there is a rising part at the end of the level contour, we regard this part as an offset course and eliminate it (see Figure 16). If the starting point of the following tone has a lower pitch than the observed tone and there is a falling part at the end of the level contour, we regard this part as an offset course and eliminate it (see Figure 17).

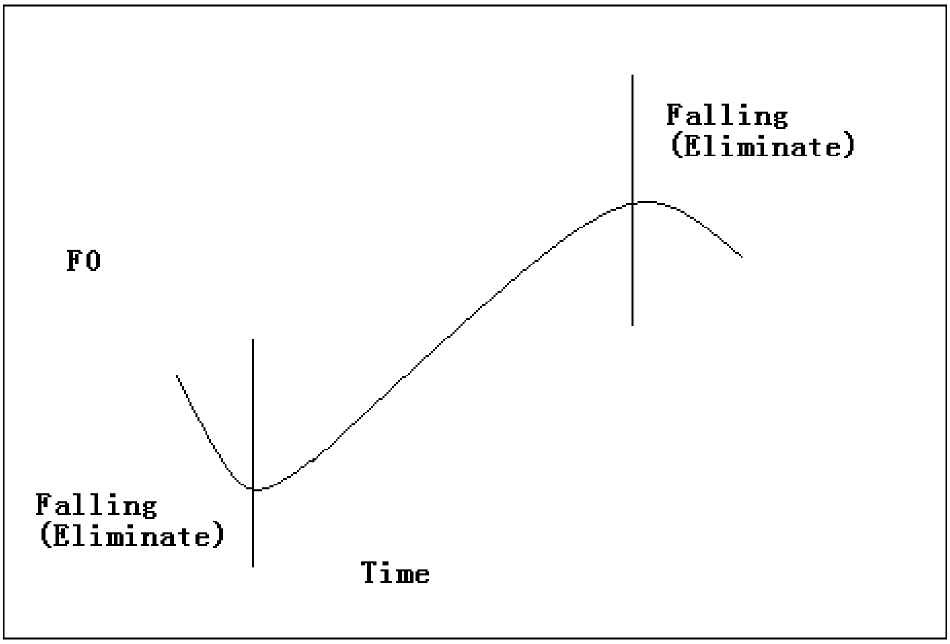


Figure 12 Eliminate the onset and offset of a rising tone

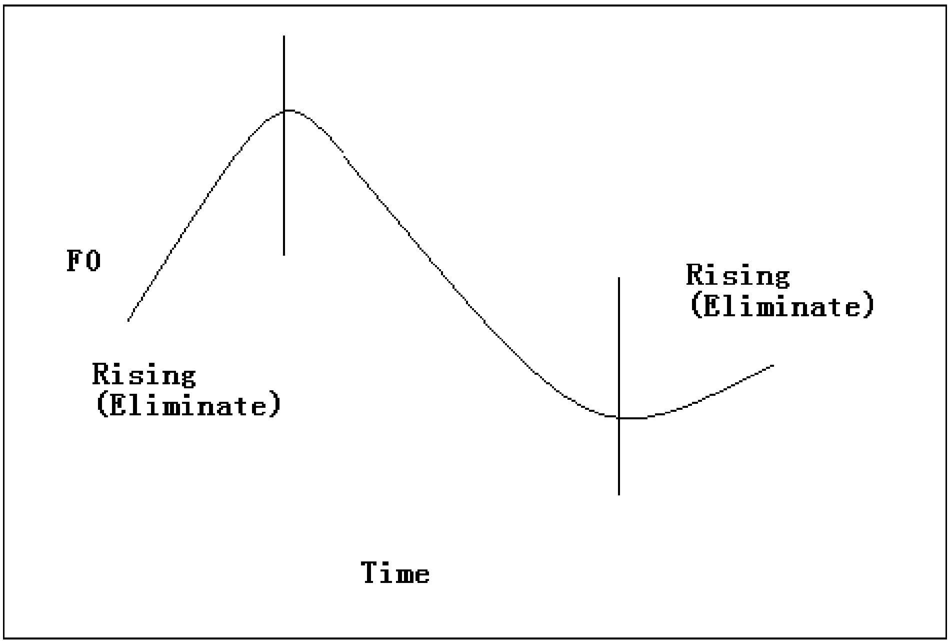


Figure 13 Eliminate the onset and offset of a falling tone

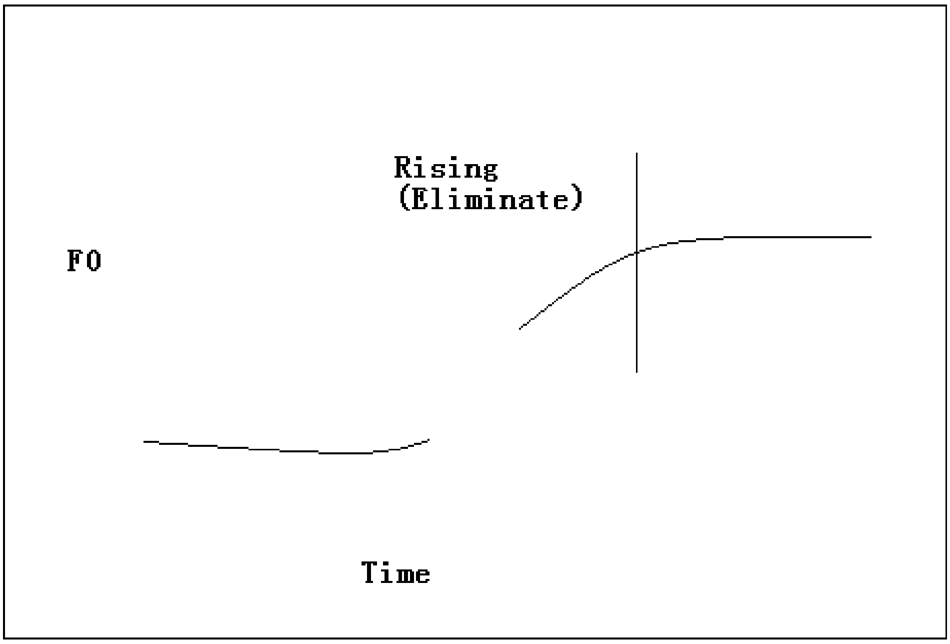


Figure 14 Eliminate a rising onset of a level tone

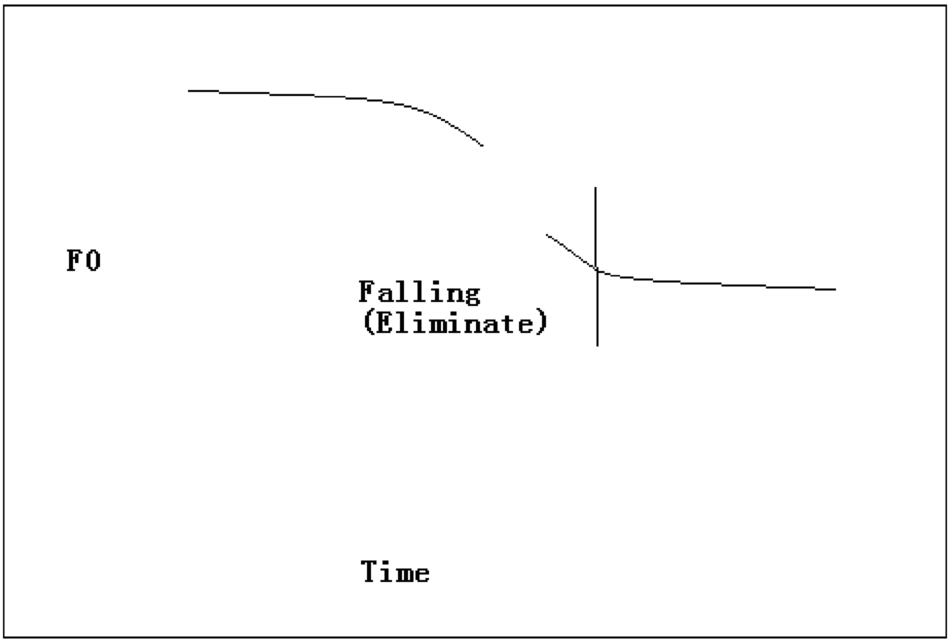


Figure 15 Eliminate falling onset of a level tone

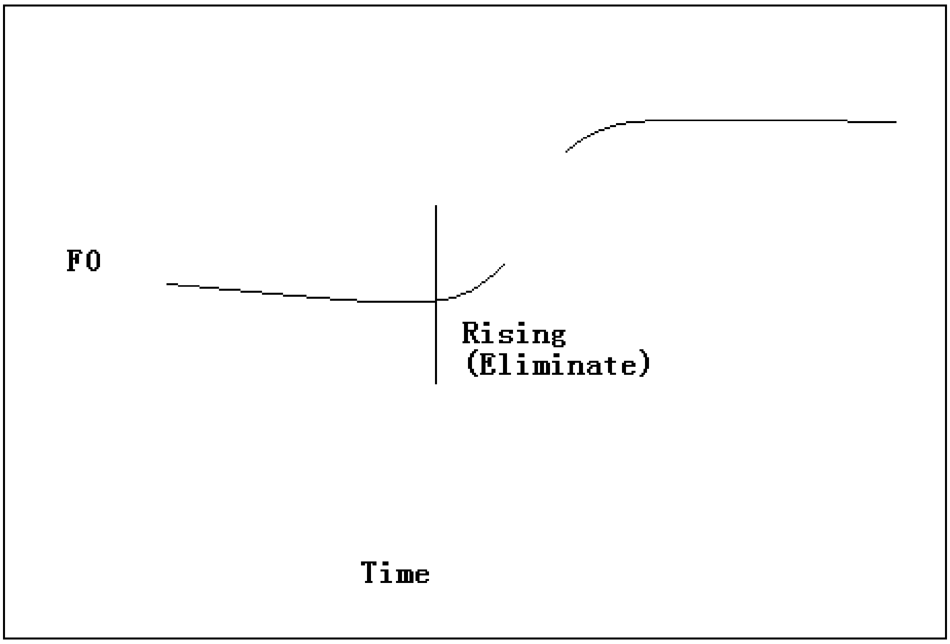


Figure 16 Eliminate a rising offset of a level tone

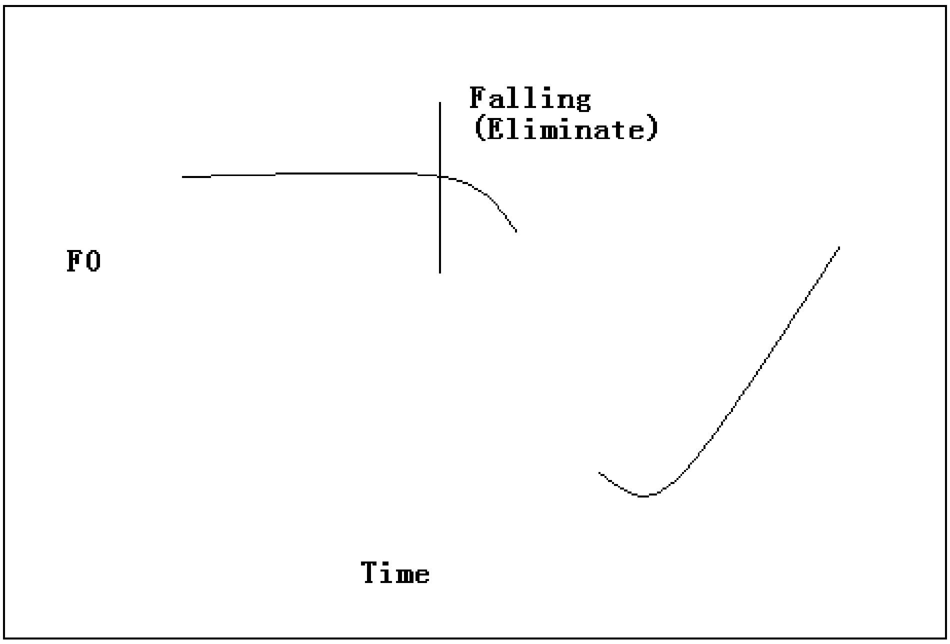


Figure 17 Eliminate a falling offset of a level tone

In conclusion, using the nucleus of a tone can avoid the intra-speaker variation of tones that are influenced by consonants, prosodic boundary, and stress. To create the model of the tones used in the current study, we avoid these effects by focusing on the tone nucleus.

3.2.2 Inter-speaker Normalization

In order to deal with inter-speaker variation, many methods of normalization have been proposed. Six main normalization strategies were surveyed by Zhu (1995)[7].

The first strategy utilizes a z-score transform (e.g. Jassem, 1971; Menn and Boyce, 1982; Rose 1987, 1989); these "express an observed F0 value as a multiple of a measure of dispersion away from a mean F0 value" (Rose, 1987:347). The F0 are calculated using the standard Formula A below:

[Formula A] $z_i = \frac{x_i - m}{s}$

X_i is an observed F0 value, m is the mean value, and s is the standard deviation (SD) . For example, if Speaker M1's F0 value is is 180 Hz, and his overall mean and SD are 160 Hz and 10 Hz respectively, the normalized value (z-score value) for 180 Hz would be (180-160)/10=2.

The second method, the fraction of range (FOR), (e.g. Earle, 1975; Ladd et al., 1985; Rose, 1982; Takefuta, 1975) expresses "an observed F0 value as a fraction of the difference between two range-defining F0 values" (Rose, 1987:347). The F0 value to be normalized (v) is calculated using Formula B below:

[7] References in these six strategies are cited from Zhu (1995).

[Formula B] $v_i = \frac{x_i - x_L}{x_H - x_L}$

x_H and x_L are respectively the up and low F0 values for determination of the range, and X_i is an observed F0 value.

The third method, the proportion of range (POR), also expresses an observed F0 value as a proportion of a range that is defined by mean and SD. The F0 value to be normalized (g) is computed by Formula C below:

[Formula C] $g_i = \frac{x_i - (m - cs)}{(m + cs) - (m - cs)}$

m is mean, s is SD, and c is a constant. Generally, c=2 or c=2.5. X_i is an observed F0 value

"Formula C looks like Formula B. However, the range-defining strategies are different. In FOR, the range is determined by two F0 values, while in POR the range is determined by the mean and SD that result from all F0 values used for normalization." (Zhu, 1995)

The fourth method is the ratio of log semitone distances (LD). It expresses an observed F0 value as a proportion of the distances of tow semitone. A semitone distance is: $D = 12 \cdot \log_2 \frac{x_2}{x_1} = \frac{12}{\log_{10} 2} \cdot \log_{10} \frac{x_2}{x_1}$ (Hart et al., 1990:24), or $x_2 = \left(\sqrt[12]{2}\right)^n \cdot x_1$ (Baken 1987:127). X_i is an observed F0 value. The F0 value to be normalized (d) is calculated by Formula D below:

Formula D

$$d_i = \left(\frac{12}{\log_{10} 2} \cdot \log_{10} \frac{x_i}{x_L}\right) \div \left(\frac{12}{\log_{10} 2} \cdot \log_{10} \frac{x_H}{x_L}\right)$$

$$= \left(\log_{10} \frac{x_i}{x_L}\right) \div \left(\log_{10} \frac{x_H}{x_L}\right)$$

$$= \frac{\log_{10} x_i - \log_{10} x_L}{\log_{10} x_H - \log_{10} x_L}$$

x_H and x_L are respectively the upper and lower delimiters of a range.

Depending on the methods that are used to define the range in terms of semi-tones, LD can consist of two sub-strategies: LD_{FOR} that uses two values and LD_{POR} that uses many values.

The fifth method is the logarithmic z-score (LZ) transforms. It expresses an F0 value as a multiple of dispersion away from a mean value. The relevant values are expressed in logarithmic terms. The F0 value is normalized by Formula E below:

Formula E

$$z_i' = \frac{y_i - m_y}{s_y}$$

$$= \frac{\log_{10} x_i - \frac{1}{n}\sum_{i=1}^{n} \log_{10} x_i}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\log_{10} x_i - \frac{1}{n}\sum_{i=1}^{n}\log_{10} x_i\right)^2}}$$

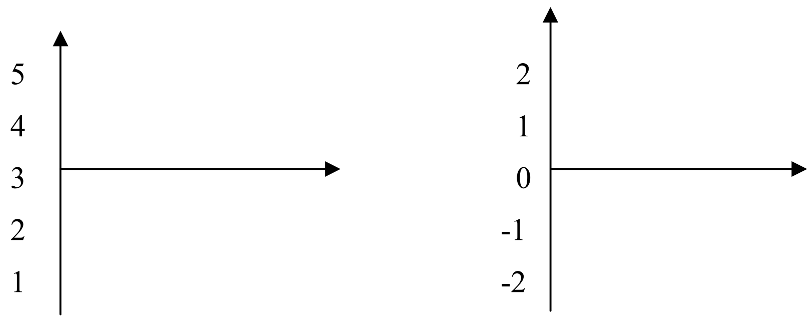
In Formular E, $y_i = \log_{10} x_i$, s_y is SD of y_i (i =1, 2, …, n); and m_y is the arithmetic mean of y_i (i =1, 2, …, n). Therefore, it is the logarithmic geometric mean of the original F0 values.

The sixth method is the logarithmic proportion of range (LPOR) normalization. It expresses an observed F0 as a proportion of a range too, yet it is defined by the mean and SD in logarithmic terms. The F0 value is normalized (g’) by Formula F below:

Formula F $g' = \dfrac{y_i - (m_y - cs_y)}{(m_y + cs_y) - (m_y - cs_y)}$

c is a constant, $y_i = \log_{10} x_i$, m_y and s_y are, respectively, the mean and SD of y_i (i=1, 2, …, n); and m_y is the logarithmic geometric mean of the observed F0 values.

Zhu used Formula E, the LZ strategy, to normalize F0 of participants' tones in his research. The normalization formula is redesigned in this research. LgZ-Score is not used in this study, because the current research focuses on the F0 at the starting point and that at the ending point. The δLZF0 at a higher level is larger than that at a lower level even if the two tones have the same δF0.



Traditional 5-scale system 5-scale system in this research

Figure 18 Comparison of traditional 5-scale system and 5-scale system in this research

When we use the 5-scale pitch notation system in the current research, the five pitch levels are from -2 to 2 associated with the z-score scale (Figure 18). The larger LZ normalized numerical value of a F0 value will be much lower than 2 (traditionally, the number 5 represents the number 2 in our 5-scale system), while the smaller LZ normalized value of a F0 value is also much lower than -2 (traditionally, the number 1 represents the number -2 in our 5-scale system) (see table 7). The results of the z-score normalization are much closer to the traditional 5-scale system. Therefore, z-score is used to normalize F0 data to avoid inter-speaker variation in this research.

Table 7 Comparison lg_z-score with z-score

F0	LZ(Lg_z-score)	z-score
303	1.78	2.01
118	-2.78	-2.22

3.2.3 Conclusion

Most of the previous research used either inter-person normalization or intra-person normalization. However, both normalization methods are employed in this study. Intra-person normalization is used to synthesize tones for the experiments of categorical perception; while both intra-person normalization and inter-person normalization are employed to normalize the NS and NNS production data for comparison.

3.3 Tone Categories in This Study

Generally, there are four tonal categories in Mandarin. However, how to deal with the neutral tone is an important issue. There are two opposing views about the nature of the neutral tone. One regards the neutral tone as an independent tonal category (Qiao, 1956; Xu, 2004; Zhang, 1956), because the neutral tone shares the acoustic features of falling tones. The other view regards the neutral tone as a dependent part of the previous tone (Cao, 2002b; Li, 1981; Lin & Yan, 1980; Shen & Lin, 1991), since the pitch contour of the neutral tone depends on the previous tonal contour.

In the current research, the neutral tone is regarded as an independent category. There are two reasons. The first reason is that the neutral tone has a salient feature, short duration, which can distinguish it from the other four tones, even though the pitch contour of the neutral tone is not stable. On the other hand, some sandhi tones could be categorized as belonging to another tonal category, for example, T3+T3 → T2+T3, the first T3 is changed to T2. Under this sandhi rule, the first T3 changes to a tone which has all of the acoustic features that T2 has. Similarly, when a normal tone (T1, T2, T3, or T4) changes to the neutral tone, the neutral tone does not have any features of the original tone. Therefore, in this study, five tone categories are studied, that is, Tone1, Tone2, Tone3, Tone4 and Tone0.

3.4 Summary

This chapter mainly introduces the method of normalization on Chinese tones. It extracts the neutral part of a tone and normalizes the F0 values. Then such pitch values can be used to analyze and compare data across individuals. This method avoids variables that are produced due to context or individual differences.

Chapters 4-7 present the four experiments that are proposed to answer the four research questions respectively. Experiment 1 is proposed to answer Research Question 1, to explore the features of tonal production by native speakers. Experiment 2 addresses Research Question 2 to determine the perceptual space of pitch values for T1, T2, T3, and T4 and explore the salient features that distinguish native speakers' and non-native speakers' perception of the four principal tones. Experiment 3 is designed to answer Research Question 3, to determine how pitch values of tones create overlapping areas in the perceptual map. Experiment 4 addresses Research Question 4 by examining tonal production by non-native speakers, to identify the differences between native speakers' perception and non-native speakers' production; and the differences between non-native speakers' perception and their production of tones.

CHAPTER 4 EXPERIMENT 1: TONAL PRODUCTION BY NATIVE SPEAKERS

4.1 Introduction

This experiment addresses Research Question 1: What is the production space for native speakers? The purpose of the experiment is to explore the features of tonal production by native speakers.

The research method, analysis and results of the experiment are presented in this chapter.

4.2 Research Method

4.2.1 Speech Sample

The data used in this research comes from the TDT4 Multilingual Broadcast News Speech Corpus (2005). It is one of the few published speech corpora that consist of natural speech from various broadcast news sources.

The researcher selected two samples from this published corpus, which were produced by two speakers. One is a female speaker from CBS; the other is a male speaker from VOA. Both of them are speakers of Mandarin Chinese. A total of 237 tones, including T1, T2, T3, T4, and T0, were selected from their broadcasts. The sampling frequency of the data is 16000Hz.

4.2.2 Pitch Normalization

There are two steps in pitch normalization. The first step is the intra-speaker normalization. That involves identifying the nucleus model to remove the part of the tones that is influenced by the surrounding linguistic contexts. The second step is inter-speaker normalization. That involves the use of the z-score to normalize the pitch values so that the pitch values of different people can be directly compared.

4.2.2.1 Intra-speaker Normalization

F0 values of each tone in every speech excerpt were extracted automatically using the F0 detection algorithm in the Multi Speech program. Then the researcher used the nucleus model (see 3. 2.1) to remove the onset or offset which is influenced by the surrounding linguistic context, so only the midpoint of the tone remains[8]. Then, the F0 values at the beginning and at the end of the nucleus of each tone are selected.

4.2.2.2 Inter-speaker Normalization

This study uses z-score transforms (see 3.2.2) to “express an observed F0 value as a multiple of a measure of dispersion away from a mean F0 value” (Rose, 1987:347). The F0 values to be normalized (z-values) are calculated using the standard Formula A below:

[Formula A] $z_i = \frac{x_i - m}{s}$

All the F0 values obtained in the previous section 4.2.2.1, that is, the F0 values at the beginning and the end of the nucleus of all tones are normalized using z-score formula. The resultant normalized values express the pitch of each tone in the 5 point-scale notation system (Chao, 1930).

4.2.3 Analysis

4.2.3.1 Procedures

The normalized pitch values of each tone are coded according to the tonal categories. Then the data of T1, T2, T3 and T4 are put into a table that represents the

8 For detailed steps please refer to section 2.1 in Chapter 3.

production space of the normal tones. A second table is generated to present the data for the neutral tone (T0).

4.2.3.2 Interval between Two Neighboring Numerical Points

In order to design each table, an appropriate interval in the normalized notation system need to be selected in order to characterize the register and pitch direction[9]. Then all of the normalized data are placed into the intervals.

Since we have chosen to use the 5-scale pitch notation system (Chao, 1930) in this study, the range of normalized pitch is from -2 to 2. The z-score values over 2 are truncated to 2 and those values below -2 are truncated to -2. To determine the interval on the 5-scale notation system an analysis of F0 distribution is conducted.

4.2.3.2.1 Decide an interval

The distribution of normalized F0 for beginning points of all the tones in the sample for both speakers is plotted using an interval of 0.1 standard units on the x-axis that is represented in Figure 19. The histograms show a peak at the modal value (0.5), which matches Moore's analysis (1995:89), and another small peak at the area of -0.5 to -0.9 at the lower end of the F0 values.

.

---

[9] Since we use the nucleus part of a tone in this research, all of the pitch contours are falling, rising, and level. Therefore, pitch direction is also used to refer to these three kinds of pitch contours.

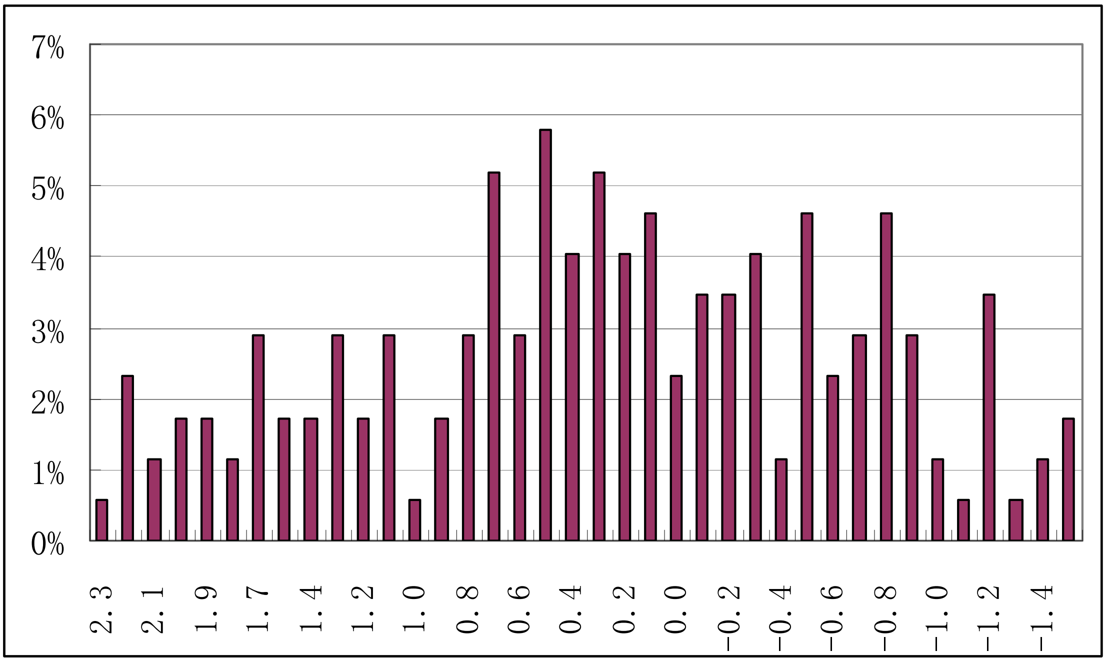


Figure 19 Distribution of normalized F0 values of the beginning point of Mandarin tones

According to Figure 19 there is no clear pattern in the distribution of beginning points. It does not show that tones are distributed evenly based on any interval, such as 0.1, 0.5 or 1. Therefore, any interval we select will have the same function. In order to present each number in a numerical unit, that is 1, in the 5-scale notation system, 0.5 is selected as the interval in this numerical notation system.

4.2.3.3 Describe production space and production map

Based on the distribution observed in figure 19, and using the interval of 0.5, we designed the table of the production space and the production map derived from the table.

4.3. Results

4.3.1 Production Space

The results are presented in two tables of production space. The first one is for T1, T2, T3 and T4. The other one is for the neutral tone. The normal tones (T1, T2, T3 and T4) have some overlapping areas with the neutral tone in the production space. Therefore, we need to present the results in two tables.

The results of the NS production experiment indicate that T1, T2, T3 and T4 are produced in different parts of the space, although there are some overlapping areas. Table 8 illustrates the tonal categories in the production space, plotting starting normalized F0 against ending normalized F0. The blue area is the production space for T1; the red area is that for T2; the green area is that for T3; and the yellow area is that for T4. The grey area represents the overlapping tone boundary area of the production space. It shows that the boundaries are located in the areas where either the beginning point or the ending point is near the value 0. On the bottom-left side of the production space, there are some overlapping areas for T1, T3 and T4. Moreover, the production space does not occupy all of the areas in the table. Some peripheral areas do not belong to any categories, such as the area where the beginning point is -2.

The grey areas and some part of production space for T1 and T4 show overlapping areas for the tones in Table 8. In the overlapping areas, a tone can be perceived as at least belonging to two categories.

Table 8 Tone categories in the production space

start / end	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2			2		1				
1.5		2	2		2			1	1/4
1		2		2	1/2	1	1	1	1
0.5			2	2	1/2	1/4	1/4	4	1/3
0		2	2	2	3	1	1/4	4	4
-0.5		2	1/2	1/2/3	2/3/4	1/4	3/4	3/4	4
-1			3	2/3/4	3/4	3/4	3	4	4
-1.5			3	3/4	3/4				
-2		3	3			4			

Table 9 illustrates distribution of T0 in the production space. The grey area is the production space for T0. Most T0s are located in the low register areas, that is, in the bottom-left corner of the production space. In the bottom-left corner, both pitch values at the beginning point and at the ending point are negative. Generally, the pitch value at the beginning point is not lower than that at the ending point. This indicates that most of the neutral tones are falling tones or level tones, which matches the results of some studies (Qiao, 1956; Xu, 2004; Zhang, 1956).

It should be noted that the production space for T0 in Table 9 overlaps with the production space of other tones presented in Table 8.

Table 9 T0 in the production space

end \ start	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2
2									
1.5									
1									
0.5						0			
0					0	0			0
−0.5			0	0	0	0			
−1		0	0	0					
−1.5					0				
−2									

4.3.2 Production Map

While the production space for the neutral tone is fairly circumscribed, there is a fair amount of overlap for the normal tones (T1, T2, T3 and T4). In Figure 20, the horizontal axis denotes normalized F0 values of the ending point of tones and the vertical axis denotes normalized F0 values of the beginning point of tones. The blue circles represent Tone 1, the red triangles represent Tone 2, the green diamonds represent Tone 3, and the yellow squares represent Tone 4. The black line in the figure denotes the function y=x.

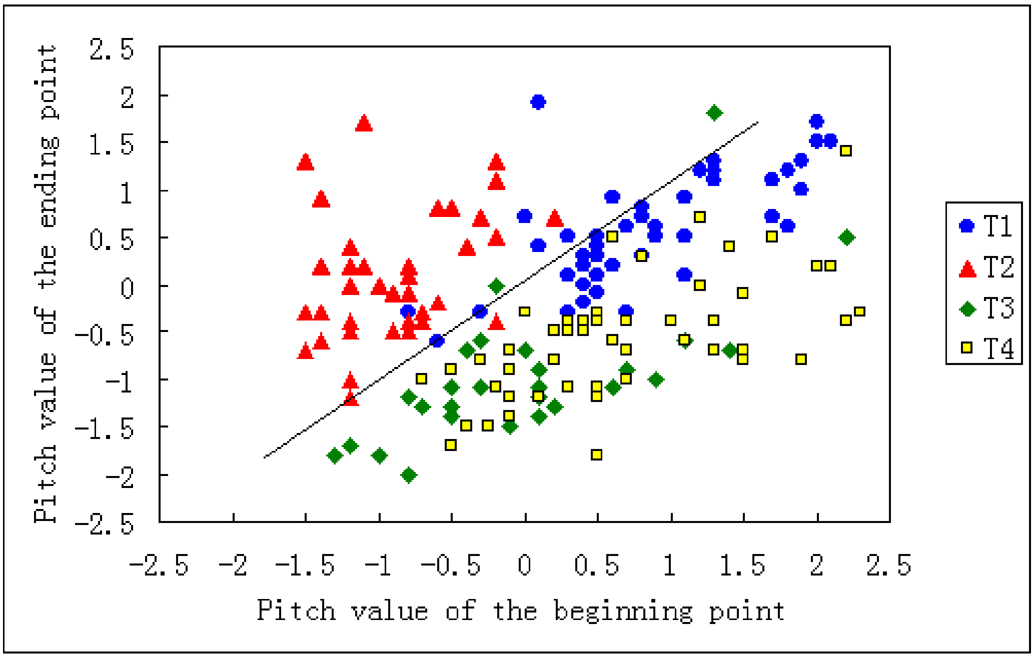


Figure 20 Production map for native speakers

It is obvious that the blue circles, green diamonds and yellow squares are under y=x. Only red triangles, that is, Tone 2 are above y=x. This indicates that only Tone 2 is a rising tone, while the other tones are level tones or falling tones. Tone 2 has few overlapping areas with other tones in the map, while Tone 1, Tone 3 and Tone 4 have some overlapping areas, especially Tone 3 and Tone 4. There are only few exemplars of each tone that appear to be marked outliers, such as the blue circles (Tone 1) in the middle of the upper area, and the green diamonds (Tone 3) in the upper-right corner of the map.

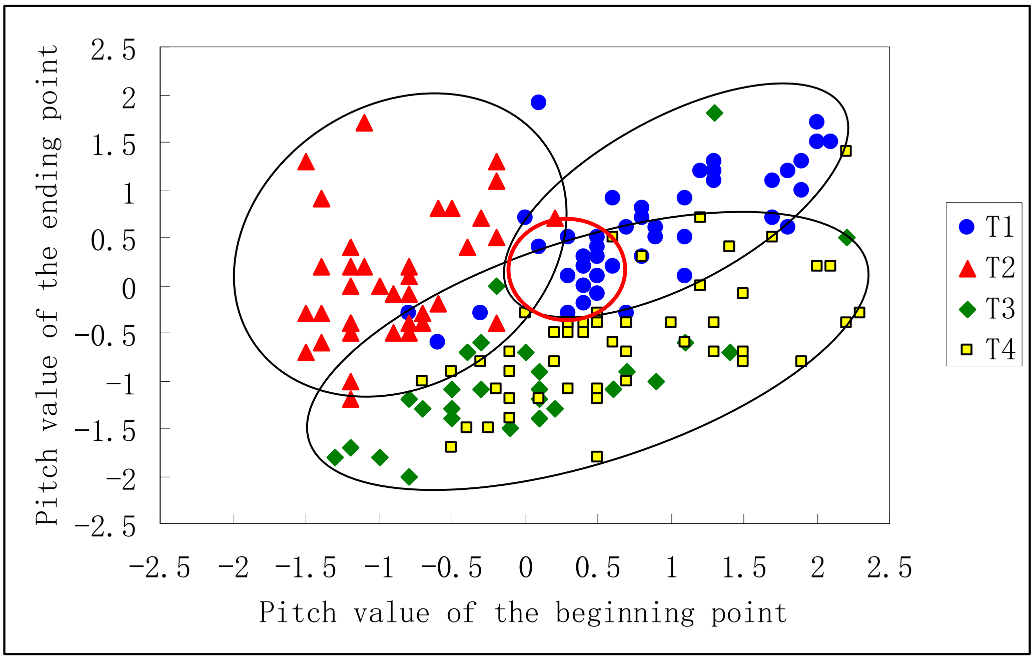


Figure 21 Three clusters in NS production map

Figure 21 indicates three main clusters of tones in the native speakers' production map. The category of Tone 2 exists in the upper-left cluster of the map. In this cluster, the pitch value of the beginning point is smaller than that of the ending point, which indicates that the tones in this cluster are rising tones. Another cluster is on the middle-right side of the map, and the tones in this cluster have positive pitch values both at the beginning point and the ending point. It also matches the nature of T1, that is, T1 is a level tone. Actually, most tones in this cluster are T1, which are represented via blue circles. The third one is an oval in the production map. It is obvious that there are both T3 and T4 in the bottom cluster in Figure 21. The range of pitch values at the beginning point of the tones in this category is [-1.5, 2] and all tones have negative pitch values at the ending point. It matches the nature of the falling tone. Actually, most tones in this cluster are T4, which are represented via yellow squares and T3 which are represented via green diamonds. Most green points have negative pitch values at the beginning point and the ending point. T3 and T4 overlap considerably. They are mixed in the middle of the circle. In the lower left, there is a smaller area for T3, and in the middle right, there is another smaller area for T4. Since the T3 and T4 mixed area is much larger than both the smaller unique area for T3 and that for T4, the three areas are circled in a same cluster. On the other hand, the cluster for T2 is above the function y=x, which indicated that the tones inside are rising tones. The cluster for T1 is beside the function y=x which shows that the tones inside are level tones. The cluster for T3 and T4 is below the function y=x, which indicates that the tones inside are falling tones.

There are three shared areas in the middle of the three clusters. They are boundaries of the tones nearby. In the middle of the boundary area, there is a unique place that only contains T1. The place is circled by red solid line in the production map.

4.4 Further Experiments

Based on this production map, the next experiments address whether NS and NNS have perceptual maps similar to the NS' production map (Research Question 2); whether there are other cues that could explain the overlapping area in production map in Figure 4-3 (Research Question 3-b); and whether there are other cues that could explain the overlapping area that represents both T0 and the other tones in Table 4-2 (Research Question 3-a).

CHAPTER 5 EXPERIMENT 2: PERCEPTUAL CATEGORIES OF MANDARIN TONES

5.1 Introduction

The purpose of this experiment is to explore the perceptual space of pitch values for T1, T2, T3, and T4. The subjects' perception of each tone will be used to define the perceptual space for Mandarin tones and their boundary areas. The researcher will explore the salient features of and the differences between native and non-native speaker perception.

Experiment 2 was developed to answer Research Question 2:

Are tones perceived categorically by NS and NNS? If so, how can we describe them?

a) What are the pitch ranges of the NS perceptual categories for T1, T2, T3, and T4?

b) What are the pitch ranges of the NNS perceptual categories for T1, T2, T3, and T4?

c). What is the difference between NS and NNS tonal perception?

5.2 Native Speaker Perception Experiment 2-a

5.2.1 Purpose of the Experiment

Experiment 2-a explores the pitch ranges for the perceptual categories of the normal tones. All of the possible tones in the perceptual space were synthesized using the Kay Elemetrics ASL program.

5.2.2 Method

5.2.2.1 Introduction

There are three main steps in this experiment. First, we calculate the F0 values for synthesis. Then, we use the Kay Elemetrics ASL Software to synthesize tones according to the F0 values from the first step. Finally, the perception experiment is conducted. Four native speakers of Mandarin listen to carrier sentences which include the synthesized tones, and are required to write down what they hear.

5.2.2.2 Calculation of the F0 value for synthesis

5.2.2.2.1 F0 value table for synthesis

The F0 values for synthesis are calculated in several steps. First, we select an appropriate interval (see 4.2.3.2) in the normalized notation system. This is used to establish a coordinate to denote the pitch register and direction.[10] Then all the synthesized tones are mapped to the normalized scale. Based on the normalized value for each tone (the beginning point and the ending point of each tone), we use normalization formula to determine the actual F0 values for each tone.

In the current research, only two pitch points are sampled to determine the tone contour: the beginning point and the ending point. Figure 22 shows the coordinates used to draw normalized pitch contours. The horizontal axis displays the beginning point and the ending point of a tone. The point on the left represents one possible F0 value at the beginning of the tone and nine points on the right represent nine possible F0 values at ending of the tone. The vertical axis denotes normalized F0 values; the interval of the values is 0.5.

---

[10] Since we use the tone nucleus in this research, all of the pitch contours are falling, rising, or level. Therefore, pitch direction is also used to refer to these three kinds of pitch contours.

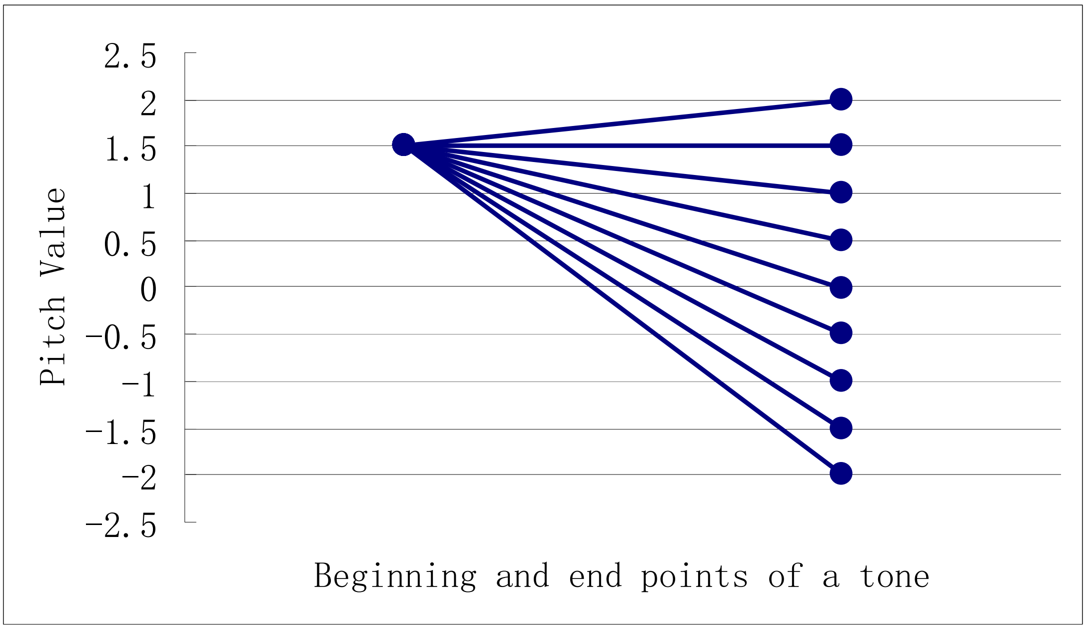


Figure 22 Coordinates for normalized pitch values

In order to generate a wide range of possible tones in the perceptual space, we synthesize 81 tones. There are 9 numerical values for a beginning point with the interval between them being 0.5. (2, 1.5, 1, 0.5, 0, -0.5, -1, -1.5, -2). Similarly, there are 9 numerical values for the ending points.

Then we record a female native speaker's speech. She is asked to say two sentences: wo3 yao4 tao1 qian2 and wo3 yao4 tao1 sui4.

Based on that sample we use the Multi-Speech program to derive 121 sample pitch values. The mean of these samples is 215.68Hz, the SD is 43.82. The mean and SD are used in the z-score formula to calculate normalized F0 values for each tone.

The following formula is used to calculate the F0 values at the beginning point and the ending point of each tone based on the 81 normalized pitch values. *m* is 215.68, and *s* is 43.82.

$$x_i = sz_i + m$$

Thus, we produce all the F0 values for the beginning and ending points to be synthesized (Table 10).

Table 10 F0 values used to synthesize tones (0% is the beginning point, 100% is the ending point)

	2		1. 5		1		0. 5		0		−0. 5		−1		−1. 5		−2	
	0%	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%	100%
2	303	303	281	303	259	303	237	303	216	303	194	303	172	303	150	303	128	303
1. 5	303	281	281	281	259	281	237	281	216	281	194	281	172	281	150	281	128	281
1	303	259	281	259	259	259	237	259	216	259	194	259	172	259	150	259	128	259
0. 5	303	237	281	237	259	237	237	237	216	237	194	237	172	237	150	237	128	237
0	303	216	281	216	259	216	237	216	216	216	194	216	172	216	150	216	128	216
−0. 5	303	194	281	194	259	194	237	194	216	194	194	194	172	194	150	194	128	194
−1	303	172	281	172	259	172	237	172	216	172	194	172	172	172	150	172	128	172
−1. 5	303	150	281	150	259	150	237	150	216	150	194	150	172	150	150	150	128	150
−2	303	128	281	128	259	128	237	128	216	128	194	128	172	128	150	128	128	128

5.2.2.2.2 Selecting appropriate F0 values for each tone

5.2.2.2.2.1 Carrier sentences

Two carrier sentences are used in this experiment to compensate for the influence of the tone following the target syllable.

Carrier 1 他 要 ___ 钱.

Ta1 yao4 TAO qian2.

He wants to ____ money.

In Carrier 1, if T1 is carried by the TAO syllable, the sentence means 'he wants to pay'. If T2 is carried by TAO, it means 'he wants to avoid paying'. For T3 TAO means 'he wants to ask for money'. T4 means 'he wants to drag money'. Only the tones on TAO are different, but they convey different meanings, much like the vowel contrasts between the English sentences 'he hid the money' versus 'he had the money'.

Carrier 2 他 要 ___ 税。

Ta1 yao4 TAO shui4.

He wants to ____ tax.

In Carrier 2, if T1 is carried by TAO, the sentence means 'he wants to pay the tax'. With T2, it means 'he wants to avoid paying the tax'. With T3, it means 'he wants to ask for a tax'. With T4, it means 'he wants to drag in the tax'. Only the tones differ, but they produce different meanings.

In chapter 4.1.2, we discussed how the tone that comes after the token is important. Therefore, two carriers are used in this experiment. In Carrier 1, the tone that comes after the token TAO is 'money' *qian* T2, which means that the beginning point of the tone is at a low level. In Carrier 2, it is 'tax' *shui* T4, which has a high-level beginning point of the tone.

In the next synthesis step, we change the original pitch value of TAO1 in Carrier 1 to a beginning point to 281 Hz with an ending point of 303 Hz, to produce a synthesized tone which is embedded in a test sentence (see Section 5.2.2.3). In the same way, we produce the other eight synthesized tones with the same pitch value – 281 Hz at the beginning. The synthesized tones are then embedded in the carrier sentences.

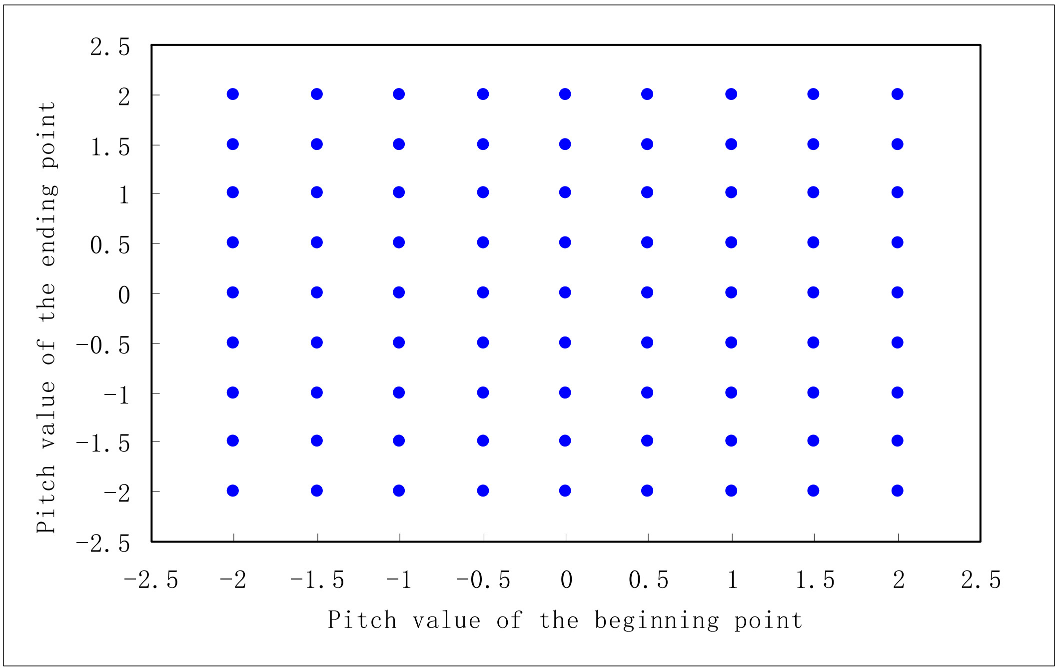


Figure 23 Stimulus map: 81 synthesized tones

Following these steps, 81 tonal variations are created (see Figure 23). Thus, we have 81 tonal variant sentences for Carrier 1. Then, another 81 tonal variant sentences are created for Carrier 2.

5.2.2.3 Synthesizing tones

5.2.2.3.1 Introduction

The Kay Elemetrics Analysis-Synthesis Laboratory (ASL) is a standard, useful program for acoustic phonetics. Specific speech parameters (e.g., fundamental frequency and formant values) can be precisely altered for speech perception experiments. In this study, ASL is used to alter pitches of tones, which are then resynthesized to produce the stimuli used to assess the perceptual boundaries of the tones. Figure 24 illustrates the steps in the process used to produce the stimuli.

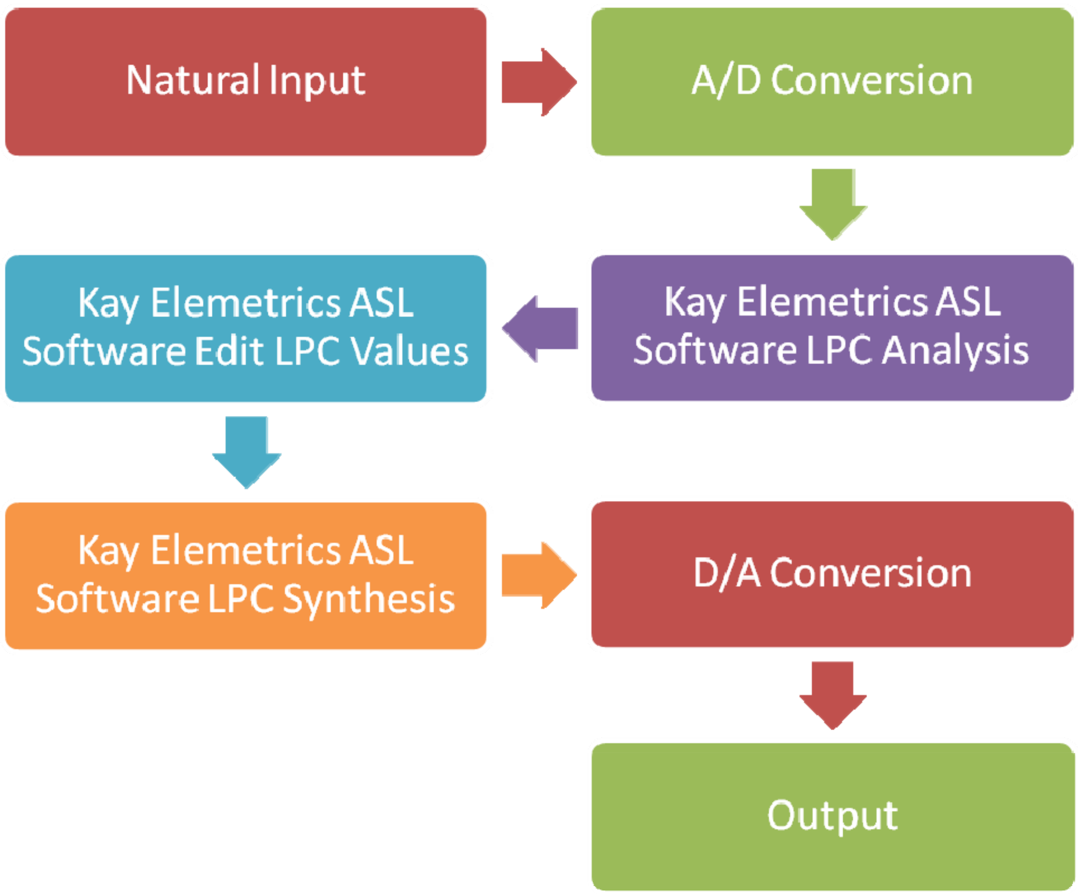


Figure 24 Steps for synthesizing tones

5.2.2.3.2 Analog signals are converted to digital signals

A female NS of Mandarin produces the Carrier 1 sentence *Ta1 yao4 TAO1 qian2*. The recording equipments consisted of a Logitech microphone and a ThinkPad X61s laptop computer. The software used for the recording is the Praat program running on the Windows XP platform. The sounds are recorded using a sampling rate of 22050 Hz and stored as single channel digital files. The recorded sentence is saved as a waveform file named 'Original_TAOQIAN'. The same recording procedure is used for the Carrier 2 sentence *Ta1 yao4 TAO1 shui4.* That sentence is saved as a waveform file named 'Original_TAOSHUI'.

5.2.2.3.3 LPC analysis and LPC values editing

The researcher uses the ASL software to do a linear predictive coding (LPC) analysis of each sample to extract the acoustic parameters (F0) and then to edit the LPC coefficients.

Figure 25 shows the LPC analysis of a sentence produced by a NS. The graph on the upper right-hand side of the screen is a waveform of the sentence. The waveform between the two blue solid lines represents the token used to carry the syllable TAO1. The graph on the upper left-hand side represents the calculated pitch values for the sentence. The F0 range on the Y-axis goes from 0 to 350Hz. The Table on the lower left-hand side shows the numerical values of generated by the LPC analysis. The first column is the number of each sample; the second column shows the time, the third column shows the F0 values of tones. Using the editor function, the numerical values in this column were modified prior to re-synthesizing the tones.

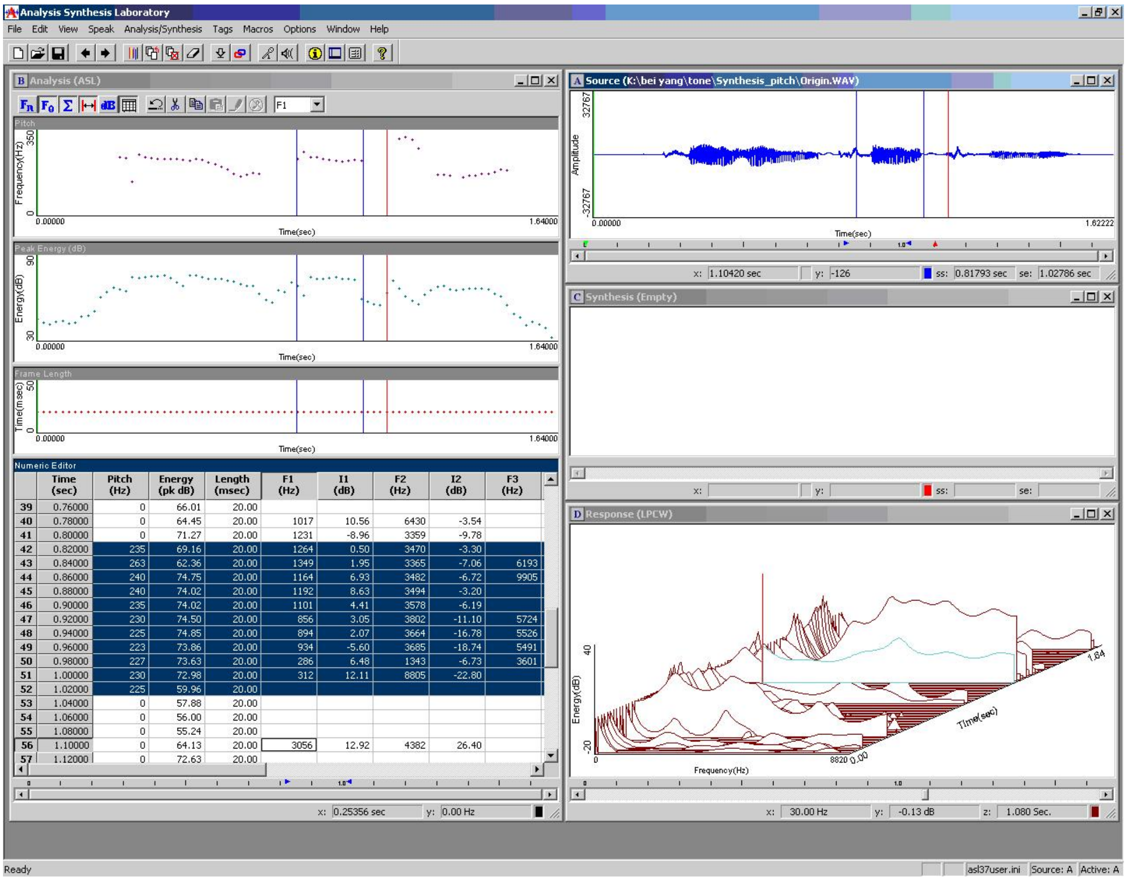


Figure 25 LPC analysis of a natural speech sample

Figure 26 shows an example of how we edit the LPC values, that is, how we change a sample F0 value of 240 hz to 302hz, then use 302 hz to synthesize a tone for the experiments.

Pitch (Hz)	Energy (pk dB)	Length (msec)
240	74.72	20.0
235	74.72	20.0
230	74.50	20.0
225	74.85	20.0
223	73.86	20.0

↓

Pitch (Hz)	Energy (pk dB)	Length (msec)
340	74.72	20.0
235	74.72	20.0
230	74.50	20.0
225	74.85	20.0
223	73.86	20.0

Figure 26 Changing the F0 values

5.2.2.3.4 LPC synthesis

In the next step, the researcher used LPC coefficients with the altered F0 values to resynthesize tones. Figure 27 illustrates the screen which contains the information subsequent to the synthesis step. The middle graph on the right side shows the waveform of the synthesized tones. The red waveform between the two blue solid lines represents the syllable whose F0 is modified. Note that the remaining parts of the sentence are unchanged. The highlighted part in the Table (lower left) shows the LPC values of the synthesized syllable.

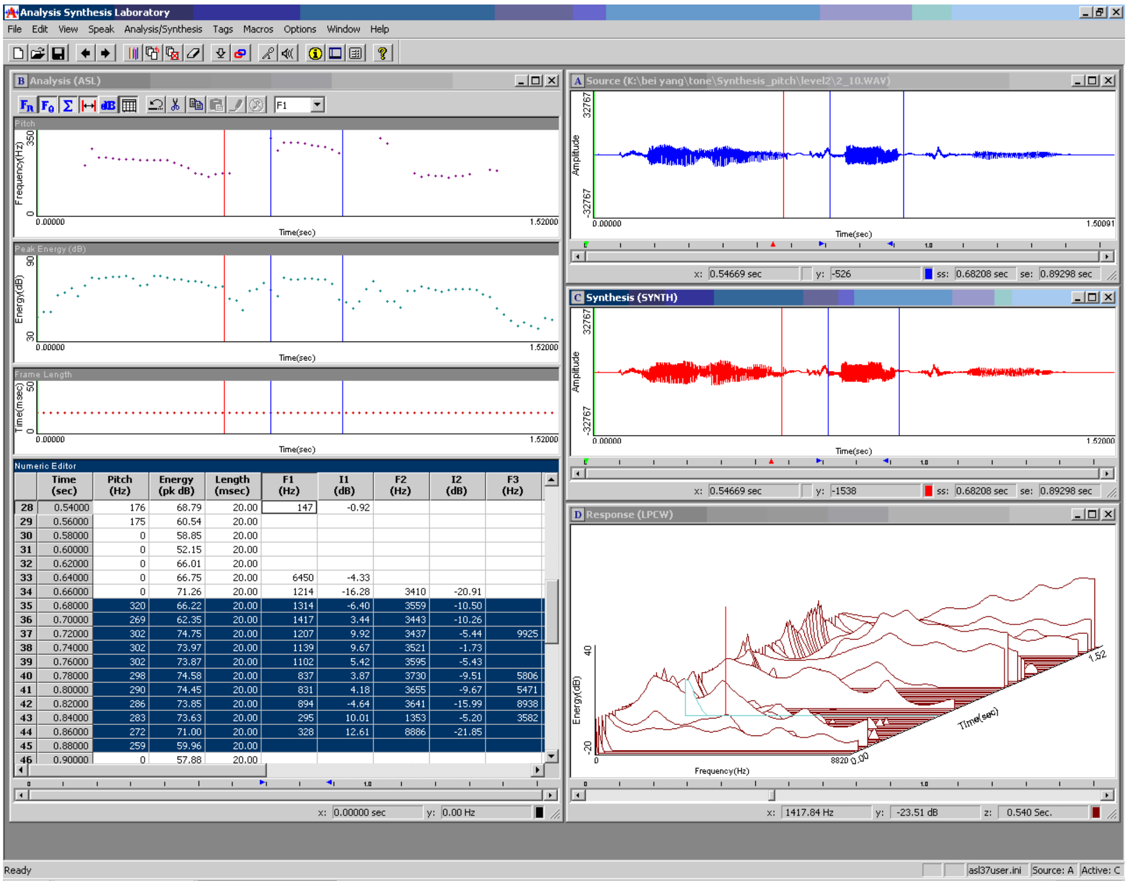


Figure 27 LPC synthesis of tones

5.2.2.4 Perception experiments

5.2.2.4.1 Native speaker subjects

Four subjects from Mainland China whose native language is Mandarin Chinese participated in this experiment. Three are female, one is male. The four subjects also speak four different Chinese dialects. Two are from northern China, and speak northern Mandarin dialects. Two are from southern China; one speaks Wu dialect and the other speaks Gan dialect. These subjects have spoken Mandarin since they were six years old. They do not have any problems when they communicate with others in Mandarin. They are bilingual. Although we do not expect that there are not any dialectal differences, it is hard to control since Mandarin is an official language in China, which has absorbed some dialectal elements, such as words, sentence patterns, and sounds from different dialects. Therefore, dialectal influence is inevitable. Their ages range from 25 to 35.

5.2.2.4.2 Procedure

The subjects are asked to complete an identification task. The 162 sentences containing all the synthesized variants of two carrier sentences are played randomly to the four subjects. Each sentence is played twice. For each sentence the subjects are provided with a response sheet that that listed the four possible interpretations (see Appendix A). The subjects are instructed to select the appropriate response item that matches the sentence they hear.

All of the answers are transcribed into four tone categories. They are T1, T2, T3, and T4. For the synthesized tones, if the subjects disagree about the tone assignment the pitch area of this tone is coded as belonging to an “obscure area”. If subjects agreed in their judgments of a synthesized tone, then the pitch area for this tone is coded as belonging to a “stable area”.

5.2.3 Results

5.2.3.1 Perceptual categories in two tonally differing carrier sentence contexts

Table 11 Tone categories perceived in NS perceptual space (For the carrier “Ta1 yao4 TAO qian2” “he wants to TAO money”)

End \ Start	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2
2	2	2	2	1/2	1/2	1	1	1	1
1.5	2	2	2	1/2	1/2	1	1	1	1
1	2	2	2	1/2	1/2	1	1	1	1/4
0.5	2	2	2	1/2	1/2	1	1	1/4	1/4
0	2	2	2	1/2	1	1/4	1/4	1/4	1/4
-0.5	1/2	2	2/3	1	1/4	4	4	4	4
-1	1/3	2/3	1/3	1/3	1/4	4	4	4	4
-1.5	1/3	1/3	1/3	3/4	3/4	4	4	4	4
-2	1/3	3	3	3/4	4	4	4	4	4

Note: The first row denotes the normalization values of the beginning point. The first column denotes the normalization values of the ending point. The other numbers in the Table denote tonal categories. 1 = T1; 2 = T2; 3 = T3; and 4 = T4. The underlined variants mean that they are different from the results of second carrier.

Table 11 reports the distribution of tone perceptual judgments when the beginning point of the following syllable is at a low register T2 (in Carrier 1 ‘money’ qian2). In this Carrier 1, a tone can be perceived as T1 if the values of both the beginning and the ending points are positive. It is perceived as T2 if the beginning point is negative, but the ending point is positive. It is perceived as T3 if both the beginning and the ending points are negative. It is perceived as T4 if the beginning point is positive and the ending point is negative. However, note that there are still some obscure areas in the NS perceptual space. The area near [-0.5, 0] at the beginning point is an obscure area. In this obscure area, tones can be perceived as either T1 or T2 if the ending point is positive. But they can be perceived as T1, T3 or T4 if the ending point is negative. Among the four tonal categories, T3 is the least stable. In the bottom-left area of the table, tones can be perceived as belonging to at least two categories.

Table 12 Tone categories perceived in NS perceptual space (For carrier2 "Ta1 yao4 TAO shui4" "he wants to TAO a tax")

End \ Start	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	2	2	2	<u>2</u>	1/2	1	1	1	1
1.5	2	2	2	1/2	1/2	1	1	1	1
1	2	2	2	1/2	1/2	1	1	1	1/4
0.5	2	2	2	1/2	1/2	1	1	1/4	1/4
0	2	2	2	1/2	1	<u>1</u>	<u>4</u>	<u>4</u>	<u>4</u>
-0.5	<u>2</u>	2	2/3	<u>1/3</u>	<u>1/2</u>	4	4	4	4
-1	<u>2</u>	2/3	<u>3</u>	<u>1/3/4</u>	<u>4</u>	4	4	4	4
-1.5	<u>2/3</u>	<u>3</u>	<u>3</u>	3/4	3/4	<u>3/4</u>	4	4	4
-2	<u>3</u>	3	3	3	<u>3/4</u>	<u>3/4</u>	4	4	4

Note: The numbers in the Table denote tonal categories. 1 = T1; 2 = T2; 3 = T3; and 4 = T4. The underlined variants mean that they are different from the results of second carrier.

If the ending point of a tone is near [-0.5, 0.5], the tone can be perceived as T1, T2, or T4. This is also an obscure area.

Table 12 reports the distribution of the normal tones when the beginning point of the following syllable is at a high register (Carrier 2; "Ta1 yao4 TAO shui4" "he wants to TAO a tax").

When the tone to be judged is in Carrier 2 sentence, a tone can be perceived as T1 if the values of both the beginning and the ending points are positive. It is perceived as T2 if the value of the beginning point is negative, but the ending point is positive. It is perceived as T3 if both the beginning and the ending points are negative. It is perceived as T4 if the beginning point is positive and the ending point is negative. However as Table 12 illustrates, there are still some obscure areas in the perceptual space. The area near [-0.5, 0] at the beginning point is an obscure area. Here, tones can be perceived as T1 or T2 if the ending point is positive. But they are perceived as T1, T3 or T4 if the ending point is negative. Among the four categories, T3 again is the least stable. In the bottom-left area of the table, we find that tones can be perceived as any category. If the ending point is near [-0.5, 0.5], the tone can be perceived as T1, T2, or T4. This is also an obscure area.

There are still some differences between the results for Carrier 1 and Carrier 2, as displayed in Table 11 and Table 12, since the tone that follows the token differs.

The underlined numbers in Table 11 and Table 12 are the categories that are perceived differently according to the contrastive following tones. A tone with the beginning point (0.5, 2) and the ending point (0) is perceived as either T1 or T4 in Carrier 1, with the rising tone at the end of Ta1 yao4 TAO qian2 'he wants to ____ money'. However, a tone with the beginning point between 1 and 2 and the ending point (0) can be perceived only as T4 if the tone is carried by Carrier 2. In the same context, a tone with the beginning point [0.5] and the ending point [0] can be perceived only as T1.

The two tables show that if the beginning point of a tone which follows the token is much higher than the ending point of the target tone, the target tone is easily perceived as a rising tone or a dipping tone. If the beginning point of a tone which follows the token is much lower, the target tone is easily perceived as a falling tone. However, if the beginning point of a tone which follows the token is similar to the ending point of the target tone, the target tone is easily perceived as a level tone.

5,2.3.2 Features of perceptual categories

5.2.3.2.1 Stable and obscure perceptual areas

Table 13 reports the ranges of perceptual categories for tones. It indicates the stable and obscure areas of each tonal category; note that T3 has a very limited stable area. In Table 13, the values in the first row denote the normalized pitch values at the beginning points; the values on the first column denote the normalized pitch values at the ending points. The numbers 1, 2, 3, and 4 in the colored areas represent T1, T2, T3, and T4.

One can see from the table that the ranges of the perceptual categories for the tones are not decided by pitch contour alone, but also by pitch register.

Table 13 Tone categories perceived in four quadrants of perceptual space (for both carriers)

End \ start	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	2	2	2	1/2	1/2	1	1	1	1
1.5	2	2	2	1/2	1/2	1	1	1	1
1	2	2	2	1/2	1/2	1	1	1	1/4
0.5	2	2	2	1/2	1/2	1	1	1/4	1/4
0	2	2	2	1/2	1	1/4	1/4	1/4	1/4
-0.5	1/2	2	2/3	1/3	1/2/4	4	4	4	4
-1	1/3	2/3	1/3	1/3/4	1/4	4	4	4	4
-1.5	1/3	1/3	1/3	3/4	3/4	3/4	4	4	4
-2	1/3	3	3	3/4	3/4	3/4	4	4	4

These results reveal a quadrant model. The perceptual space for T1 ranges from a beginning point from -0.5 to 2 to an ending point from 0 to 2. Since the pitch values at the beginning point are positive, and the pitch values at the ending point also positive, the perceptual space for T1 is described as the [+, +] area. The perceptual space for T2 ranges from a beginning point from -2 to 0 and an ending point between -0.5 to 0. Since the pitch values at the beginning point are negative, while the ending point is positive, the perceptual space for T2 is described as the [-, +] area. The perceptual space for T3 ranges from beginning points between -2 to -0.5 and ending points between -2 to -0.4. Since both the beginning and ending points are negative, the perceptual space for T3 is described as the [-, -] area. However, this area is more complicated than other areas because many tones in this area can be perceived as different tones. This T3 effect will merit further investigation. The perceptual space for T4 ranges between beginning points from -0.5 to 2 and ending points between -2 to 0.5. Since the pitch values at the beginning point are positive, and the ending point is negative, the perceptual space for T4 is described as the [+, -] area.

Boundary areas exist in the perceptual space summarized in Table 13. In the boundary areas printed in grey, the tones are perceived as “obscure”, that is, they can be perceived as at least two different tones. These boundary areas are located where either the beginning point or the ending point is near the value 0. For example, when the beginning point is [0.5, 2] and the ending point is [0.5, 2], the tones can be perceived as either T1 or T2 because this is the boundary between T1 and T2. In addition, Table 13 also indicates that if the beginning point is different, the boundary is different too. The higher the pitch value number of the beginning point, the wider the boundary area.

Figures 28-31 indicate these results graphically for all four tones. The horizontal axis denotes normalized F0 values; the vertical axis denotes the beginning and ending points; 1 means the beginning point and 2 means the ending point.

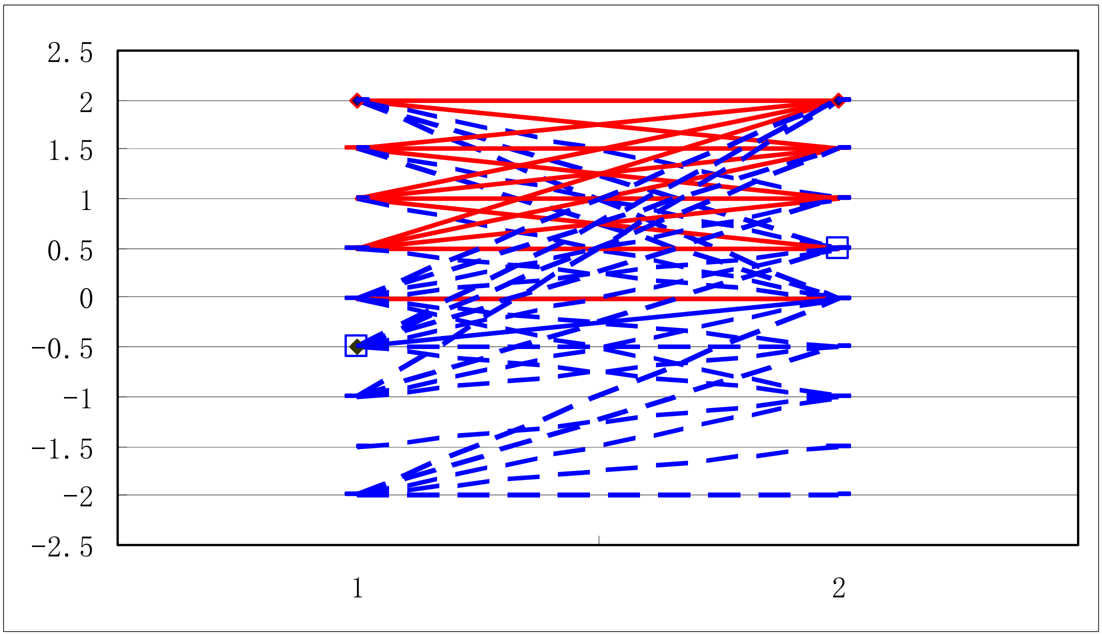


Figure 28 T1 perceptual space

In Figure 28, solid red lines indicate a stable perceptual space for T1. A tone represented by a solid line can only be perceived as T1. Dotted lines indicate obscure perceptual areas or boundary areas. A tone represented by a dotted line can be perceived as either T1 or the other tones.

Compared with Figure 29-31, Figure 28 shows that T1 has the widest perceptual space including the stable area and the boundary area.

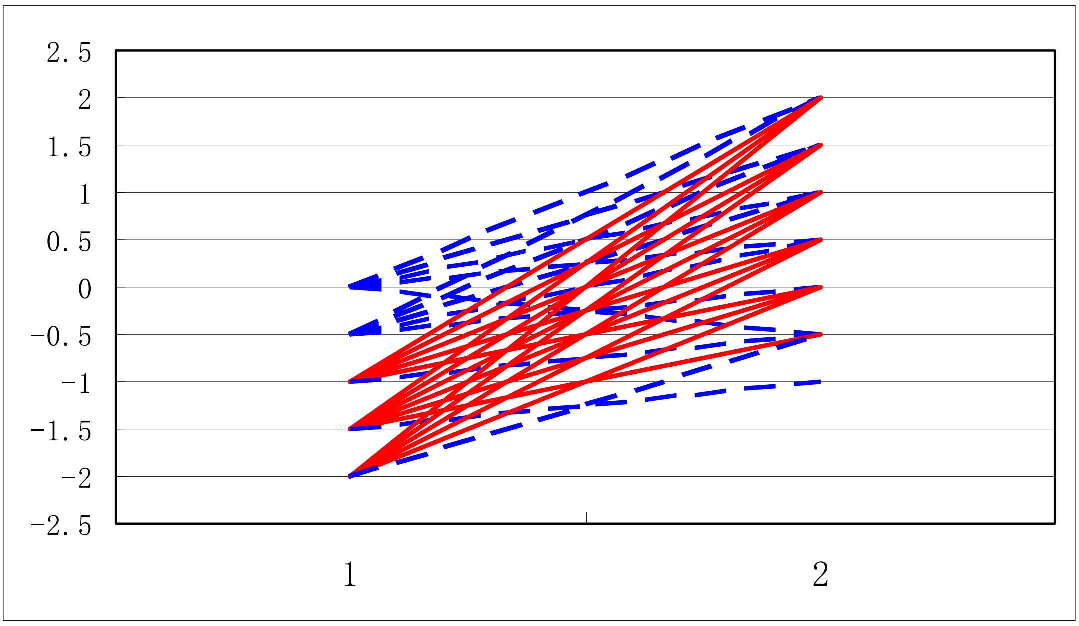


Figure 29 T2 perceptual space

For T2 the perceptual space is mapped in Figure 29 above. Solid red lines indicate stable perceptual space. A tone indicated with a solid red line can only be perceived as T2. Dotted blue lines indicate obscure perceptual areas or boundary areas. A tone represented by a dotted blue line can be perceived as either T2 or the other tones.

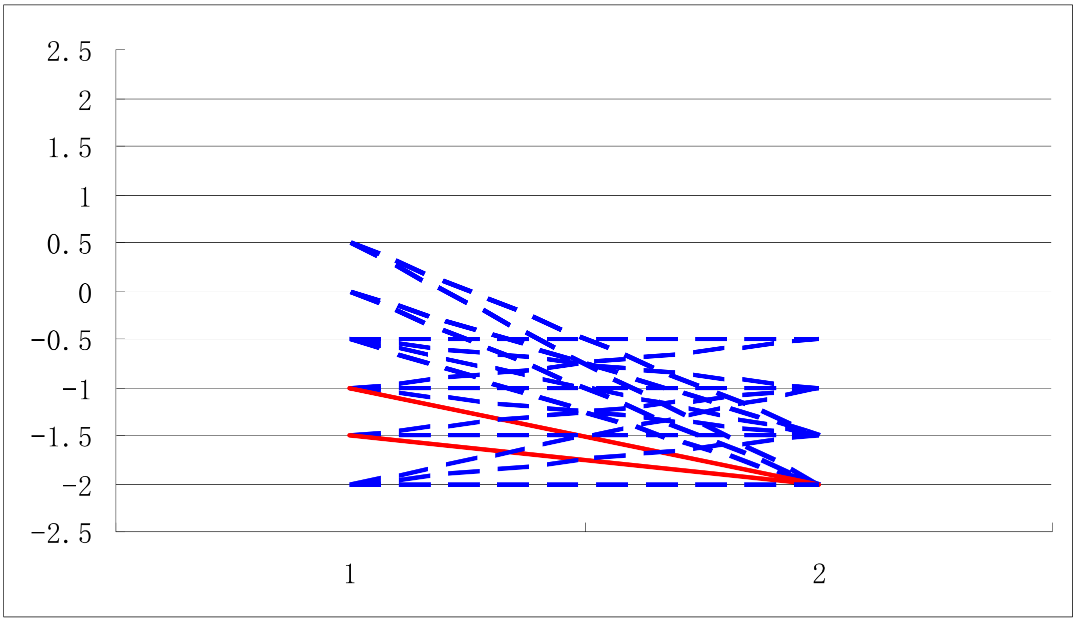


Figure 30 T3 perceptual space

For T3 solid red lines indicate a stable perceptual space in Figure 30. A tone represented by a solid red line means can only be perceived as T3. The dotted blue lines indicate obscure perceptual areas or boundary areas. A tone represented by a dotted blue line can be perceived as either T3 or other tones.

Compared with T1, T2 and T4 in Figure 28, Figure 29 and Figure 31, T3 as shown in Figure 30, has the narrowest perceptual space.

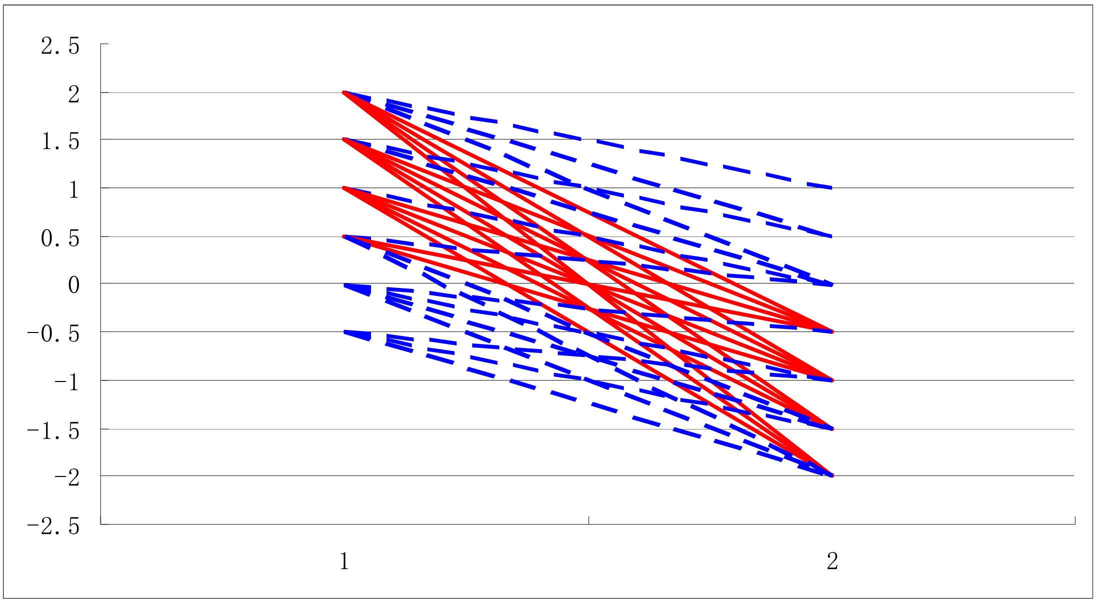


Figure 31 T4 perceptual space

For T4 in Figure 31, solid red lines indicate stable perceptual space. A tone represented by a solid red line can only be perceived as T4. The dotted blue lines indicate obscure perceptual areas or boundary areas. A tone that is represented by a dotted blue line can be perceived as either T4 or other tones.

5.2.3.2.2 T3's wide and unstable perceptual space

The wide and unstable perceptual space for T3 is a mystery area. Most points in this area share space with the other tones T1, T2 or T4. As we know, T3 has a famous tone change sandhi rule when it precedes another T3. The unstable perceptual space of T3 may cause T3 to change to other tonal categories in different contexts.

Moreover, T3 sometimes shares the same acoustic features with other tones, especially the falling contour as T4. How can these tones, which have the same acoustic features, be perceived as different tonal categories? They must have some perceptual cues that help NS normalize the tonal categories. This question will be explored further in Chapter 6.

5.2.3.3 Comparing native speaker perceptions with production

Comparing Tone Categories in the Production Space in Table 8 with the Native Speaker perceptual quadrants in Table 13, we find many similar features between the perceptual space and the production space. First, both of them are divided into separate tones which are separated by boundary areas. The positions of the four tonal categories are the same in the perceptual space as in the production space. T1 is located at the upper right of the space, T2 at the upper left side, T3 at the bottom-left side, and T4 locates on the bottom-right side. Moreover, the space for T3 is narrower than the space of the other tones. Both the perceptual and production spaces have similar boundary areas. It shares a large area with T4 in the boundary areas. The boundary area between T1 and T4 shows the same feature: the higher the pitch value of the beginning point, the wider the boundary. Figure 32 presents a summary of the NS perceptual map.

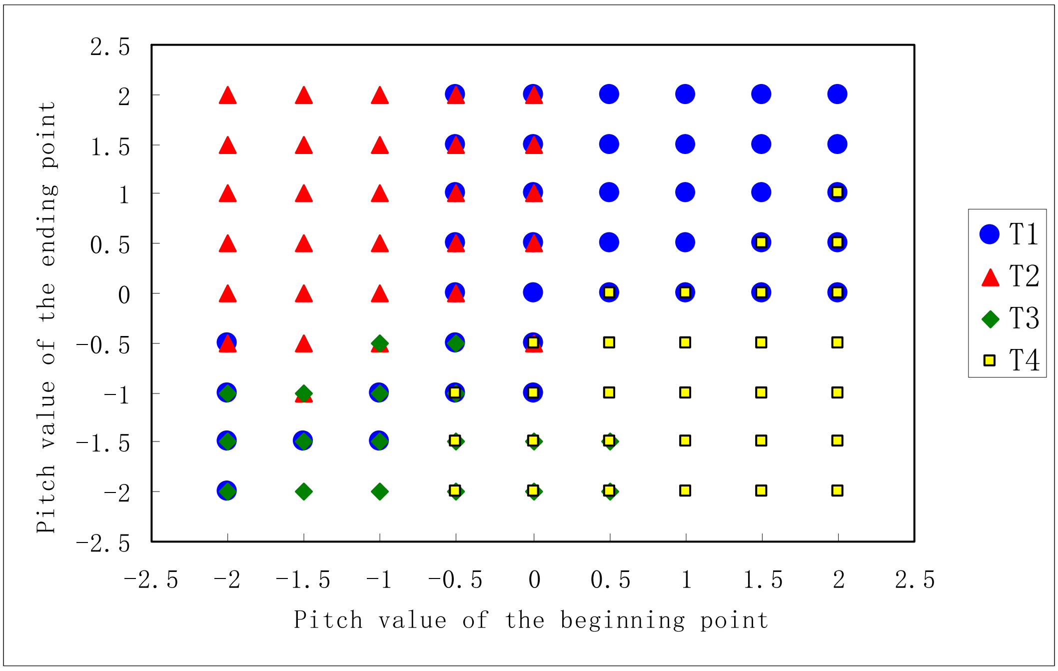


Figure 32 Perception map for native speakers

Figure 33 provides us with the comparison of the perceptual and production maps of NS. Both Figure 33 a and c indicate three main clusters in the production space. T3 either occupies a very small space in the production map, or at the bottom-left corner in the production map. However, in the perception space and perception map Figure 33 b and d, T3 is regarded as an independent category, although multiple tones exist in this category in addition to T3, such as T1, T2 and T4. Similarly, the T3 category is the smallest category among the four categories in perception space and the perception map. The boundary area between T3 and T4 is larger than other boundary areas.

Figure 33 c (production map for NS) indicates three main clusters of tones in the native speakers' production map. The Tone 2 category (red) is in the upper-left area of the map. In this cluster, the pitch value of the beginning point is lower than that of the ending point, which indicates that the tones in this cluster are rising tones. Another cluster is on the upper-right side of the production map, and the tones in this cluster have positive pitch values both at the beginning point and the ending point, and most tones in this category are T1, which are represented via blue dots. The third cluster locates on the bottom of the production map. The range of pitch values at the beginning point of the tones in this category is [-1.5, 2] and all tones have negative pitch values at the ending point. Most tones in this cluster are T4 (yellow) and T3 (green). In the perception map for native speakers (Figure 33 d), there also exist three major single-color areas. One is on the upper left side of the map, mainly containing T2; another one is on the upper right side of the map, mainly containing T1; and another is on the bottom right side of the map, mainly containing T4. The T1 area is similar to the blue cluster drawn in the production map, and the T2 area is similar to the red cluster in the production map. T3 exists in the multiple colored area which is located on the bottom left corner of the perception map, and some of them are in the boundary area between T3 and T4. It reveals that NS produces T3 and T4 similarly, however, NS perceive them differently even if T3 and T4 share the same acoustic features.

Another similarity between the production map and the perception map is that T1 occupied the central part of the both maps: (0.5, 0) in the production map and (0, 0) in the perception map.

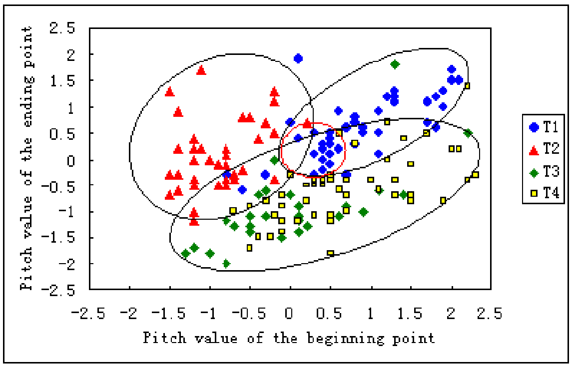
One salient difference between the perceptual and production spaces is that the perceptual space is wider than the production space. Both Table 13 and Figure 32 reveal this difference. This wider perceptual set of tone category boundaries may explain why a NS can interpret NNS' tonal production even their tones are produced differently. Thus a NNS tone production can be perceived in the NS perceptual space, even if it falls outside the NS production space.

start / end	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2			2		1				
1.5		2	2		2			1	1/4
1		2		2	1/2	1	1	1	1
0.5			2	2	1/2	1/4	1/4	4	1/3
0		2	2	2	3	1	1/4	4	4
-0.5		2	1/2	1/2/3	2/3/4	1/4	3/4	3/4	4
-1			3	2/3/4	3/4	3/4	3	4	4
-1.5			3	3/4	3/4				
-2		3	3			4			

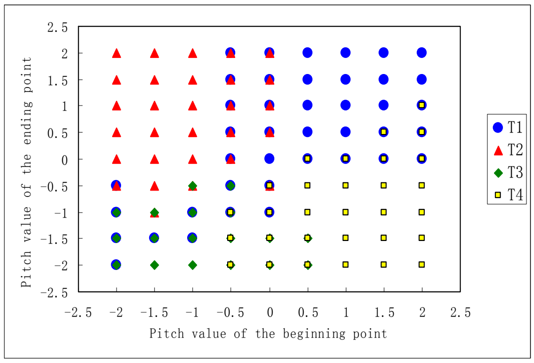
a. Tone Categories in the Production Space

start / End	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	2	2	2	1/2	1/2	1	1	1	1
1.5	2	2	2	1/2	1/2	1	1	1	1
1	2	2	2	1/2	1/2	1	1	1	1/4
0.5	2	2	2	1/2	1/2	1	1	1/4	1/4
0	2	2	2	1/2	1	1/4	1/4	1/4	1/4
-0.5	1/2	2	2/3	1/3	1/2/4	4	4	4	4
-1	1/3	2/3	1/3	1/3/4	1/4	4	4	4	4
-1.5	1/3	1/3	1/3	3/4	3/4	3/4	4	4	4
-2	1/3	3	3	3/4	3/4	3/4	4	4	4

b. Tone Categories in the Perception Space



c. Production Map of NS



d. Perception Map of NS

Figure 33 Comparing NS production with NS perception

5.3 Non-Native Speaker Perception Experiment 2-b

This section explores the features of perceptual space for NNS, and compares this with the NS perceptual space.

5.3.1 Subjects

Six American students of Chinese who are native speakers of English participated in this experiment. Three are male, three are female. All had completed first-year Chinese course, and were enrolled in second-year Chinese at the University of Iowa. The instructional goals of the first-year and second-year Chinese language courses are summarize as following:

1. After completing the first-year Chinese course, students need to reach Intermediate-High in listening, Intermediate-Mid in speaking and reading, and Intermediate-Low in writing;

2. After completing the second-year Chinese course, students need to reach Advance in listening, Intermediate-High in speaking and reading, Intermediate-Mid in writing.

Chinese OPT (Oral Proficiency Test) is used to measure the subjects' oral proficiency levels. Other proficiency levels are measured by Iowa developed tests. The results indicated that they were at the Intermediate level.

5.3.2 Method

The stimuli and task are the same as in Experiment 2-a.

5.3.3 Results

In this section, the perceptual data are individually presented for each of the six non-native speakers and then compared with the native speakers in the following sections.

5.3.3.1 Perception results for NNS-1

The perceptual results for NNS-1 are summarized in Table 14 below. They differ markedly from the NS perceptual space.

Table 14 NNS-1 tone categories in perceptual space (for both carrier 1, 2)

End \ Start	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	2/4	2	1/3	1/2	1/2	1	1/2	1/2	1/2
1.5	2/4	2	1/2	1/2	1	1/2	1/2	1/2	1
1	2/4	1/2	1	1/2	1	1	1	1	1
0.5	2/3	1	1	1/2	1/2	1	1	1	1
0	1/2	1/4	1	2	1	1	1	1	1
-0.5	2/3	2/3	4	4	1/3	3	1/4	1/4	3/4
-1	2/3	2/3	4	1/3	1/3	1/3	1/2	1/2	4
-1.5	2/3	4	4	1/3	3	3	1/3	1/3	4
-2	2/3	3/4	4	3/4	3	3	3/4	3/4	4

There are no clear boundaries in NNS-1's perceptual space. Most T3 fall in the low register, while most T1 are distributed in the high register. Most T1 are located in the upper-right space, most T2 fall in the upper-left space, most T3 fall in bottom-left space, and most T4 are in the bottom-right space. The space is overlapping. That is, most tones can be perceived as belonging to two tonal categories.

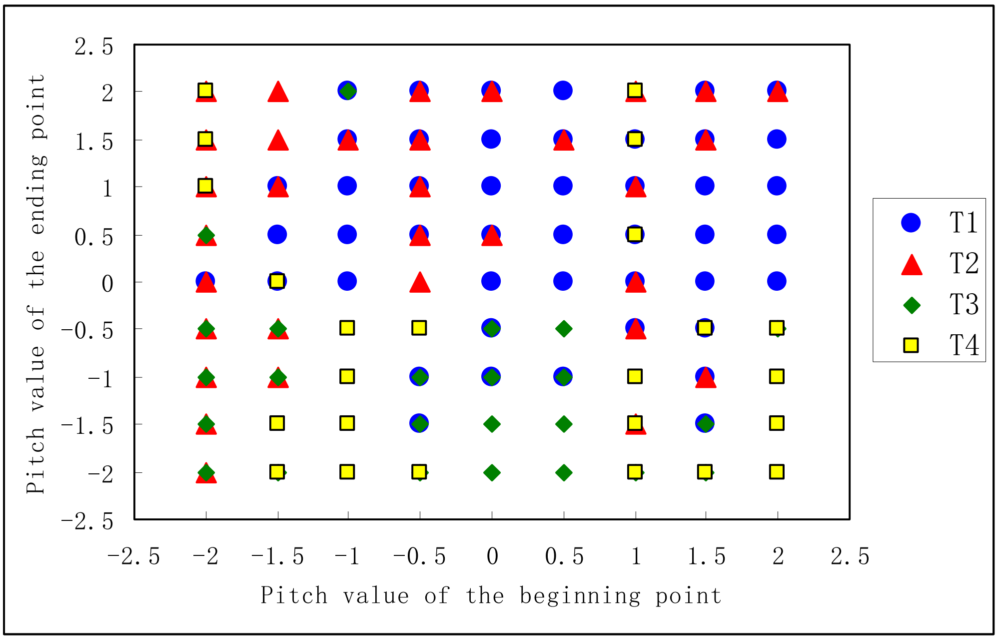


Figure 34 Perception map of NNS-1

Figure 34 further indicates the distribution of tonal categories for NNS-1. Similar to the perception space (Table 14), most T3s fall in the low register, while most T1s are distributed in the high register. Figure 34 shows that T1 and T2 occupy the upper half of the perception map, and T1 and T2 overlap in some areas. T3 and T4 occupy the lower part of the perception map. It indicates that the position of the ending point of the tone divides the tones into two large areas.

5.3.3.2 Perception results for NNS-2

The perceptual results for NNS-2 appear in Table 15 below. They show both similarities to and differences from the NS perceptual space.

Table 15 NNS-2 tone categories in perceptual space (for both carrier 1, 2)

End \ Start	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	2	2	3/4	1/3	4	4	4	1/3	1/4
1.5	1/2	2	3/4	1/4	1/2	1/2	2	1/4	1/4
1	2	2	1/4	1	1	2	2	1	2/4
0.5	2	2	1/3	2/4	1/2	2	1	1	1/4
0	2	2	2	2	2	2	1/2	2	1
-0.5	2/4	2/3	4	3/4	4	4	4	1/4	4
-1	3/4	3/4	2/4	3/4	4	4	4	2/4	4
-1.5	3/4	3/4	3	3/4	3/4	4	4	4	4
-2	3/4	3/4	3/4	1/3	3/4	3/4	3/4	3/4	4

The results for NNS-2 show something approaching a four quadrant model, although it differs from the NS perceptual space. The bottom-left area resembles the NS perceptual space for T3. However, the T1 area is obscure, that is, several tones overlap in this area. The perceptual boundary is also not clear, yet there is still a boundary between -0.5 to -1 at the beginning point of tones.

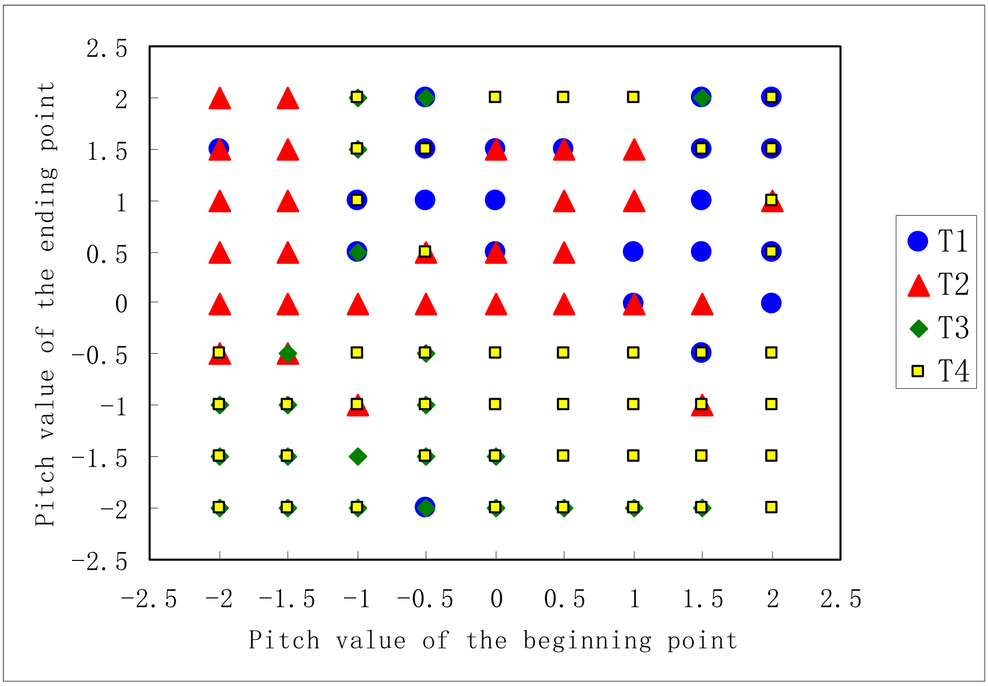


Figure 35 Perception map of NNS-2

Figure 35 further indicates the distribution of tonal categories for NNS-2. Similar to the perception space (Table 15), T1, T2 and T4 each occupy a unique space, that is, T1 is located in the upper-right area; T2 is located in the upper-left area; and T4 occupy the bottom-right area, besides the bottom-left area which is the unstable perceptual area for T3 in NS perceptual space. However, the T1 area is obscure, that is, several tones overlap in this area. Figure 35 shows that T1 and T2 occupy the upper of the perception map, and T1 and T2 overlap in some areas. T4 occupy the lower part of the perception map. It indicates that the position of the ending point of the tone divides the tones into two large areas, and that T3 does not have clear space in this perception map.

<u>5.3.3.3 Perception results for NNS-3</u>

The perceptual results for NNS-3 are displayed in Table 16 below. They show both similarities to and differences from the NS perceptual space.

Table 16 NNS-3 tone categories in the perceptual space (for both carrier1, 2)

Start / End	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	3	2/3	4	3	2	4	2/3	1/2	2/3
1.5	3/4	3	3/4	3	2/3	3	2	1	2
1	3/4	2	2/4	2/3	1/2	2/3	2	1	1
0.5	3	2	3	2/4	1	1/2	1	1	1
0	2/3	2	2/3	1/4	1/2	1/2	1/2	3	1
-0.5	4	2/4	4	4	2/4	3/4	2/4	1/4	4
-1	3/4	3/4	2/3	2/3	2/4	4	3/4	4	3
-1.5	3/4	3	3/4	3/4	3/4	3/4	3	3/4	3/4
-2	3/4	4	3/4	2/3	4	2/4	4	2/3	4

NNS-3 shows something approaching a four quadrant model, although it differs from the NS perceptual space. Most T1 are located in the upper-right space; most T2 in the upper-left space; T3 occupies a small area of the bottom-left area; while most of T4 lies in the bottom-right area. The space shows many overlaps. That is, each area can be perceived as belonging to two tonal categories. And there is no clear boundary in NNS-3’s perceptual space.

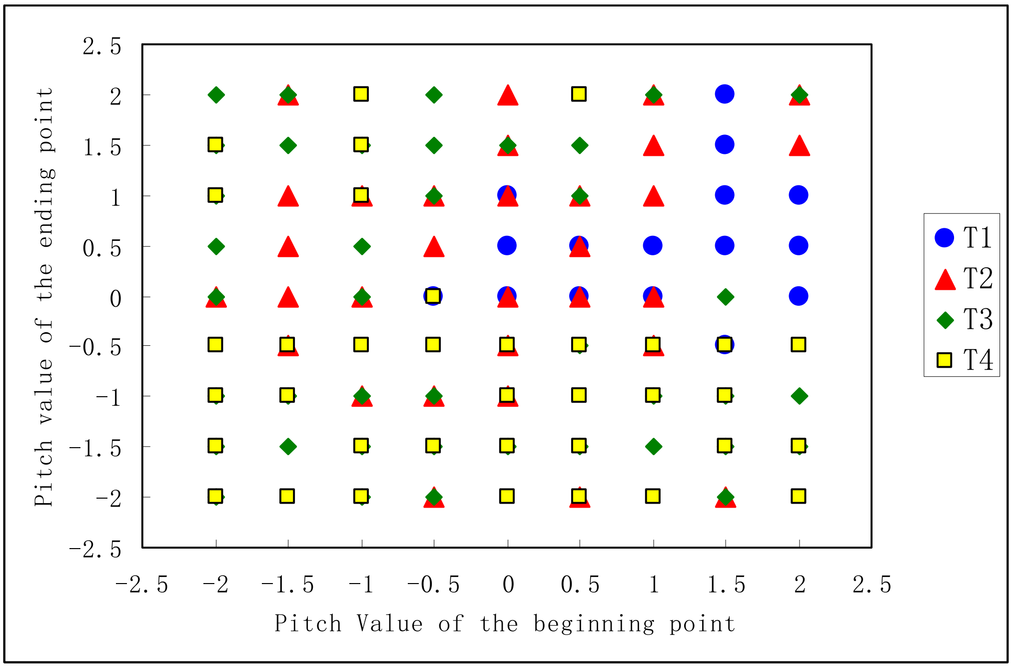


Figure 36 Perception map of NNS-3

Figure 36 further indicates the distribution of tonal categories for NNS-3. It shows that T1, T2 and T3 occupy the upper half of the perception map, and they overlap in some areas. T4 occupies almost all lower part of the perception map. It indicates that the position of the ending point of the tone divide the tones into two large areas, and T3 does not have clear space in this perception map.

5.3.3.4 Perception results for NNS-4

The perceptual results for NNS-4 are displayed in Table 17 below. These differ from NS perceptual space.

Table 17 NNS-4 tone categories in perceptual space (for both carrier 1, 2)

End \ Start	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	2/3	2/3	2	1/2	1	1/2	1/3	4	2/4
1.5	2/3	3	1/2	1/2	1/2	1/4	1/4	1/4	4
1	1/3	2	2	1	2	2/4	1/4	4	4
0.5	2/3	2	2	2	2/4	2/4	4	4	2/4
0	2/3	2	2	2	2/4	1/2	2/4	2	1/2
-0.5	1/4	1	1	1/4	1	1/2	1/2	4	4
-1	3/4	1	1	1/4	2/4	1/3	1/4	4	4
-1.5	4	3/4	1/3	1	1/3	1/4	1/3	3/4	4
-2	4	3/4	1/3	3/4	1/3	1/4	4	4	4

Most T2 are located in the upper-left area. There is no clear boundary area between the four tones.

Figure 37 further indicates the distribution of tonal categories for NNS-4. It shows that T4 occupies the right part of the perception map and T2 occupies the upper part of the perception map. It indicates that the position of the ending point of the tone decides the distribution of T2, and the beginning point of a tone decides the distribution of T4. T1 and T3 do not have clear space in this perception map.

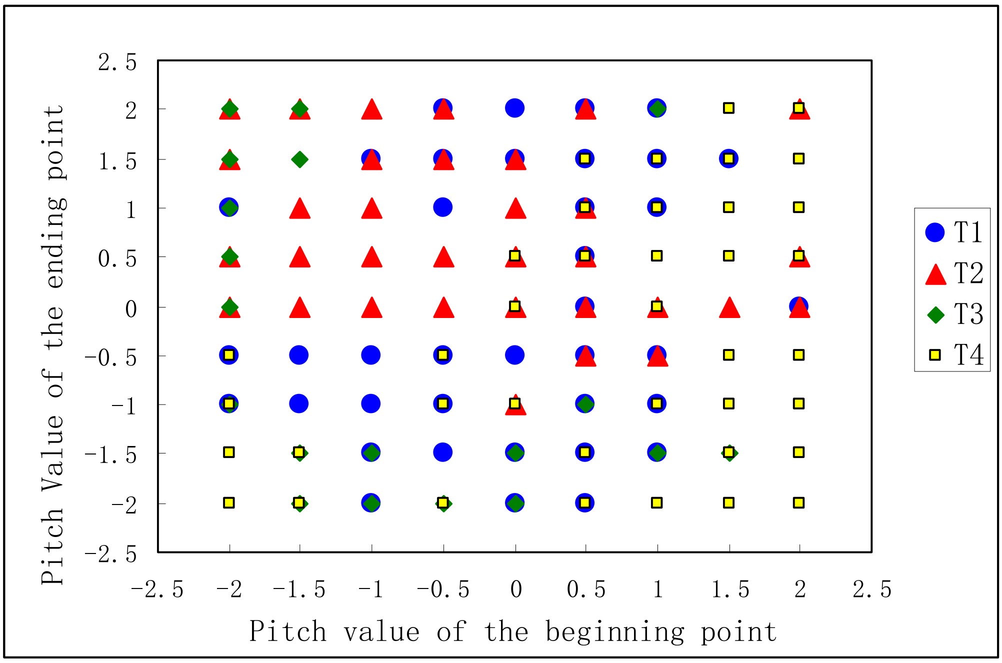


Figure 37 Perception map of NNS-4

<u>5.3.3.5 Perception results for NNS-5</u>

The perceptual results for NNS-5 are displayed in Table 18 below. These also differ from NS perceptual space.

NNS-5 showed no clear perceptual tendency. Both Table 18 and Figure 38 indicate that there is no four quadrant distribution and there are no clear boundary areas for any of the four tones.

Table 18 NNS-5 tone categories in perceptual space (for both carrier 1, 2)

Start / End	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	3/4	2/3	2/3	2/3	1	3/4	1/4	1/3	2/4
1.5	3	2/3	3/4	3/4	3	2/3	1/3	1	1/3
1	1	3/4	2/4	3/4	2/3	3/4	1/2	2/4	4
0.5	3/4	1/4	2/3	2/4	1/4	1/4	1/4	1/2	3/4
0	2/4	2	2/4	2	1/2	1/3	1/2	1/2	1/4
-0.5	2/4	1/3	2/4	3/4	4	2/4	4	3/4	4
-1	1/3	3/4	2/4	1/4	1/4	2/4	2/3	1/2	2/3
-1.5	1/3	2/3	3	4	3/4	1/3	2/4	1/4	1/2
-2	1/4	3/4	2/3	2/4	1/4	3/4	2/4	1/3	4

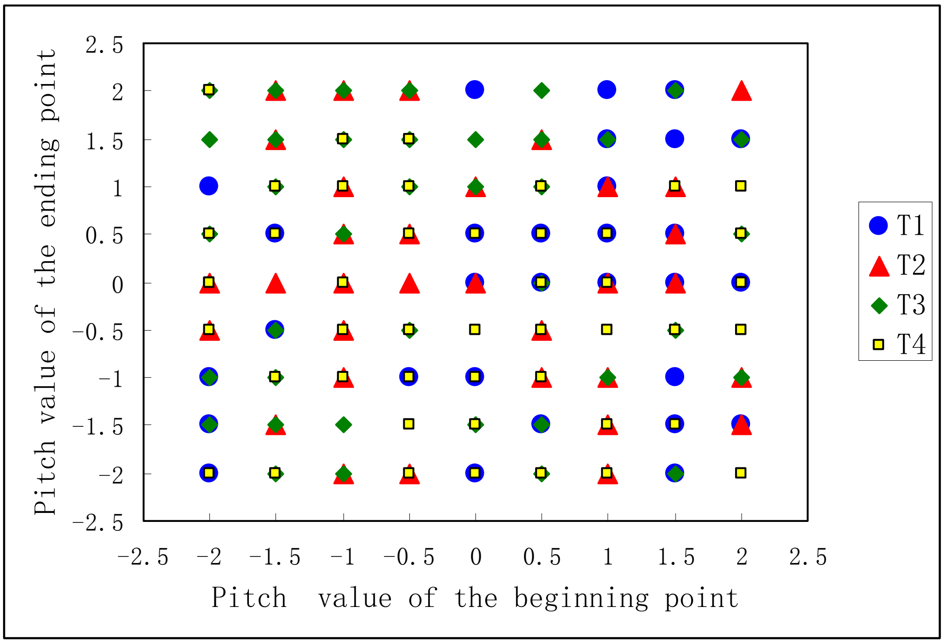


Figure 38 Perception map of NNS-5

5.3.3.6 Perception results for NNS-6

The perceptual results for NNS-6 appear in Table 19 below. They differ from NS perceptual space.

Table 19 NNS-6 tone categories in perceptual space (for both carrier 1, 2)

Start End	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
2	3/4	2/3	2/3	2/3	1	1	1/3	1	1
1.5	1/2	3	2	2/3	1	1/2	1/2	1	1/4
1	1	1/3	1	1	1	1/2	1/2	1/2	1
0.5	1/4	1/3	1/2	1/2	2	1	1/2	1	1/4
0	1/4	1/3	1	1/2	1/2	2	1/3	1	1/2
-0.5	3/4	1/2	2/3	1/2	1/3	3/4	4	4	4
-1	2/4	1/3	2/3	2/3	1/4	2	3/4	4	3/4
-1.5	1	1/2	3	2/3	2/3	4	3/4	3/4	4
-2	1/4	2/4	2	2/3	3	3/4	4	3/4	3/4

Obviously, this perceptual space lacks any clear boundary areas. Even so, most T1 is located in upper-right area, most T2 is located in upper-left area; most T3 is located in bottom-left area; while most T4 is located in bottom-right area. The spaces overlap. That is, most areas can be perceived as belonging to two tonal categories.

The register for the beginning point of each tone is clear, but the contour is not. For example, for the tone with the beginning point -1 and the ending point -2 is perceived as T2 by NNS-6. Yet the contour shows that it is a falling tone which the NS would perceive as T4.

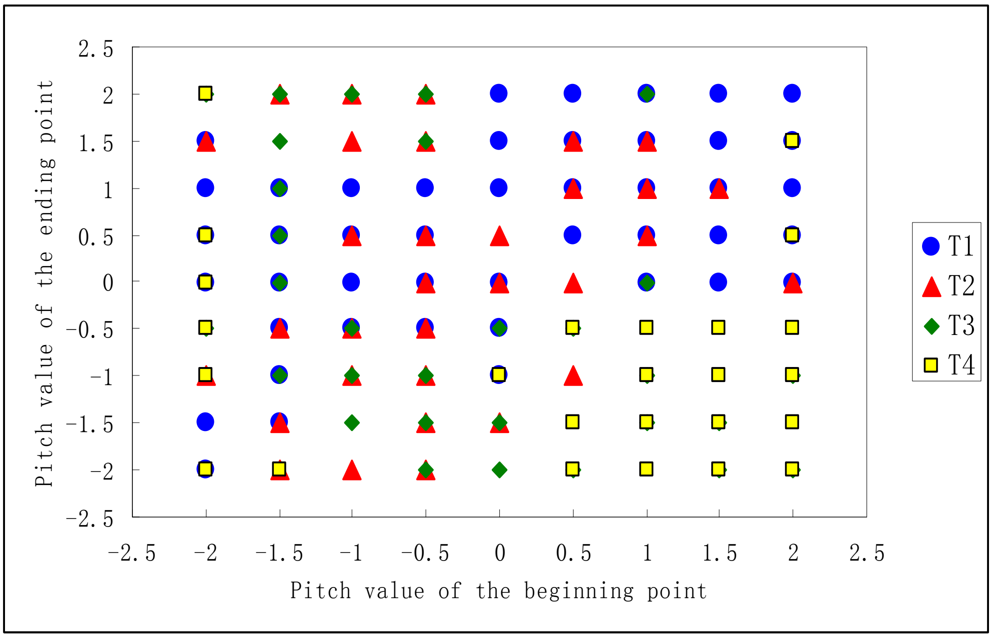


Figure 39 Perception map of NNS-6

Figure 39 further indicates the distribution of tonal categories for NNS-6. It shows that T4 occupies the bottom right area of the perception map. T1 and T2 occupy the upper part of the perception map and they overlap in some areas. The bottom left area is the mixed area, which contains all the tones.

5.3.4 Three groups of the NNS perception

Almost no boundary areas appear in most non-native speakers' perceptual space. Some students' perceptual space creates a close four-quadrant distribution for the four tones, although even this is not as clearly divided as a native speaker's perceptual space. We find only one student who does not yet have any tendency to categorize different tones. The other five students fall in between in distinctive and separate tone perceptions.

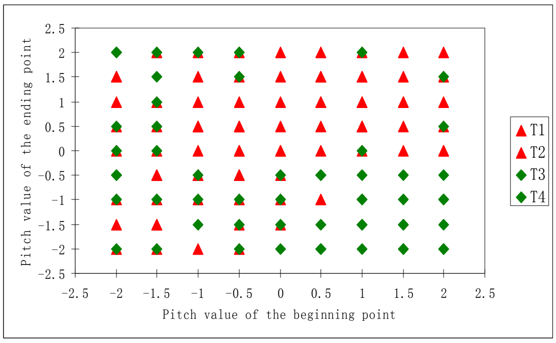
Most learners’ perceptual spaces show that, like native speakers, most T1 are located in the upper-right space; most T2 in the upper-left; most T3 in the bottom-left, and most T4 in the bottom-right. However, unlike the NS’ perceptual space, almost all areas in the NNS’ perceptual space show overlap of the tones. That is, most areas can be perceived as belonging to two tones. This is the essential difference between NS and NNS perception of tones.

The six NNS perception maps appear to reveal three patterns.,

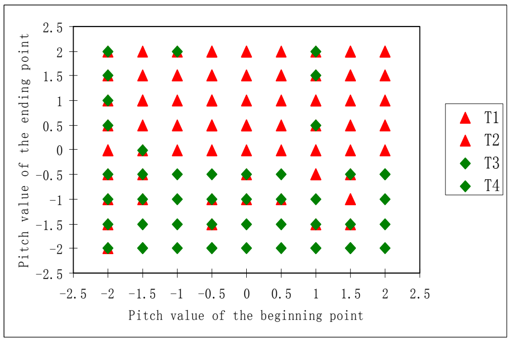
The first pattern does not show any clear distinctions in tone perception. This pattern reflects NNS5 (Figure 38).

There are four NNS perception maps which define the second pattern. This pattern can be seen in NNS1 (Figure34), NNS2 (Figure 35), NNS3 (Figure 36) and NNS6 (Figure 39), which are adapted in Figure 40.

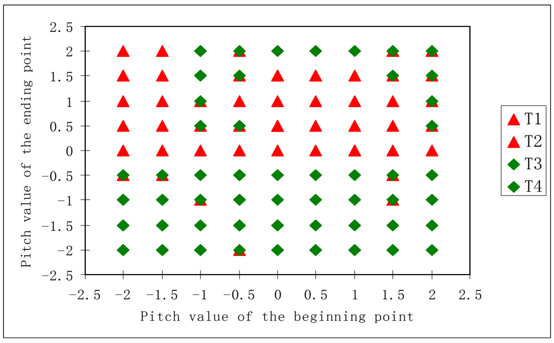
The perception maps of NNS1, NNS2, NNS3 and NNS6 are similar in that T1 and T2 are found and overlap in the upper half of the space of each map. Sometimes, T3 also appear at the upper part. It is obvious that pitch values of the ending point of a tone divide the tones into two categories: one is at the upper area of a map, and mainly contains T1 and T2; the other is on the bottom of a map, and mainly contains T3 and T4.



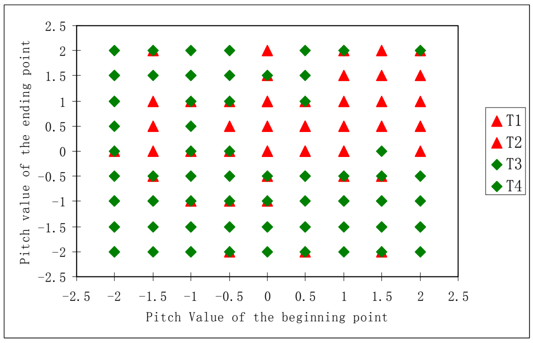
NNS1 perception map



NNS2 perception map



NNS3 perception map



NNS6 perception map

Figure 40 Perception group 2 of NNS

The third pattern of NNS perception maps is exemplified by NNS4 (Figure 37). In this perception map, pitch values of the beginning point of a tone divide the tones into two categories: one is on the right side of a map, and mainly contains T4; the other is on the left side of the map, and the upper-left part that mainly has T2 while the bottom-left part has T1, T3 and T4. The upper-right area also has some overlap of T1 and T4. This map is similar to the NS perception map. The difference is that T1 is in the upper-right area of the NS perception map while T1 occupies both the upper-right area and bottom-left area in the perception map of NNS4.

5.3.5 Summary Comparison of NS and NNS Perception

Recall that the NS perception map (Figure 32) could be characterized as having four quadrants. However, remember that the bottom-left quadrant is an overlapping area which includes all tones. Therefore, strictly speaking, there are three categories in the NS perception map. The T3 area is a mixed area in which perception of T3 overlaps with other tones.

Recall that we observed that there are three patterns in the NNS perception maps. However, the majority of the NNS, such as NNS-1, NNS-2, NNS-3 and NNS-6, had a binary map with T1 and T2 in the upper part and T3 and T4 in the bottom area of the perception maps.. One NNS (4) also had a two category perception map with T1, T2 and T3 are on the left side of the map and T4 on the right side of the map. NNS5's perception map was the only one that does not reveal any pattern.

Therefore, the salient difference between NS perception and NNS perception is that NS perceive tones based on three discrete categories (T1, T2, T4) and one more ambiguous category (T3), while most of our NNS perceived tones based on two categories: T1 and T2 in the upper half of the space and T3 and T4 in the bottom half of the space. The NNS perceptions appear to be determined by the end points of the tones; by contrast, both the pitch values of the beginning and those of the end points determine the tonal categories of the NS.

.:

CHAPTER 6 EXPERIMENT 3: OVERLAPPING PITCHES IN THE PERCEPTUAL SPACE

6.1 Introduction

The results of Chapter 4 and Chapter 5 indicate that there is an overlapping area (T3) in the production map (Figure 21) and in the perception map (Figure 32); the results of Chapter 4 also indicate that most of the neutral tones are falling tones or level tones, and it is obvious that the production space of T0 in Table 9 shares the same production space of the normal tones in Table 8. Based on these results, it is necessary to explore whether there are other cues that affect perception of the neutral tone and T3.

This chapter is to answer the Research Question 3: How do the pitch values of tones create overlapping areas?

a) Does pitch play an important role in perceiving the neutral tone?

b) Does the pitch at the end of T3 and pitch at the beginning point of the following tone play an important role in perceiving T3?

Experiment 3 was designed to explore whether and how pitch values of tones create overlapping areas. Experiment 3 includes a perceptual experiment for the neutral tone and a perceptual experiment for T3.

There are three experiments in this chapter. The first experiment (3-a) explores whether changing the pitch contour can affect the perception of the neutral tone. The second experiment (3-b) investigates whether a neutral tone with a longer duration can be perceived as a normal tone. If it can, it will indicate that the contour of some normal tone is similar to some variations of the neutral tone. The purpose of these two experiments is to investigate whether pitch contour plays an important role in tonal perception. The third experiment (3-c) proposes to explore whether there are other cues that affect the perception of T3.

6.2 Experiment 3-a: Perception Experiments for the Neutral Tone

6.2.1 Purpose of the Experiment

This experiment explores whether pitch plays an important role in perceiving the neutral tone.

Undoubtedly, short duration is the salient feature of the neutral tone (Cao, 2002b; Chen and Xu, 2006). This salient feature can be used to judge whether a tone, which is produced by NS, is a neutral tone or a normal tone. However, does this feature play an important role in tonal perception? Or does pitch contour play any role in tonal perception? Cao (2002b) indicates that the pitch contour of T0 depends on the previous tone. T0 will be a high falling tone if the previous tone is T1 or T2; it will be a low falling tone if the previous tone is T4; the contour of T0 will rise a little bit and then fall if the previous tone is T3. The research by Chen and Xu (2006) shows that all T0 share the falling nature in addition to short duration. The hypothesis in this experiment is that the perceptual space of T0 will overlap with T1, T2, T3, and T4 if pitch plays an important role in perception.

6.2.2 Subjects

Four subjects from Mainland China whose native language is Mandarin Chinese participate in this experiment. Three of them are female, and one is male. The four subjects speak four different Chinese dialects. Two of them are from northern China, and speak north Mandarin dialects. Two of them are from southern China; one speaks Wu dialect and the other speaks Gan dialect. The ages of the four subjects range from 25 to 35.

6.2.3 Stimuli

Most studies indicate that the pitch contour of the neutral tone depends on the tonal category of the previous syllable (such as Cao, 2002b). Therefore, four words in which the first syllables carry four different tones are selected (see table 20).

Table 20 Four stimuli with the neutral tone

Characters	Pinyin (Base Form)	Pinyin (Sandhi Form)	Meaning
哥哥	ge1ge1	ge1ge0	brother
格格	ge2ge2	ge2ge0	princess
葛葛	ge3ge3	ge3ge0	gege (name)
个个	ge4ge4	ge4ge0	everyone

The second syllable *ge0* in each word is cut and then combined with the first syllables of the four target words, that is, *ge1, ge2, ge3* and *ge4*. This creates four groups of words. Stimuli (see table 21) are created using Praat.

6.2.4 Procedure and Analysis

Subjects are required to listen to the above four groups of words in the carrier sentences.

First, the subjects read the English meaning of a sentence, and then they listened to a group of sentences that had the words with modified tones embedded. After that, they need to write down the number of the sentence which they think best matches the English meaning of the sentence. For example, the subjects read the English meaning "The brother went to work," and then they listen to following four sentences:

1) *ge1ge0(ge1) qu4 shang4ban1 le*

2) *ge1ge0(ge2) qu4 shang4ban1 le*

3) *ge1ge0(ge3) qu4 shang4ban1 le*

4) *ge1ge0(ge4) qu4 shang4ban1 le*

After listening, they select the number, such as 1, which they think matches the English meaning.

Table 21 Stimuli for Experiment 3-a

Group 1: Even the brother went to work.	Group 2: Even the princess went to work.
1. *ge1ge0(ge1) dou1 qu4 shang4ban1 le.*	5. *ge2ge0(ge1) dou1 qu4 shang4ban1 le.*
2. *ge1ge0(ge2) dou1 qu4 shang4ban1 le.*	6. *ge2ge0(ge2) dou1 qu4 shang4ban1 le.*
3. *ge1ge0(ge3) dou1 qu4 shang4ban1 le.*	7. *ge2ge0(ge3) dou1 qu4 shang4ban1 le.*
4. *ge1ge0(ge4) dou qu4 shang4ban1 le.*	8. *ge2ge0(ge4) dou1 qu4 shang4ban1 le.*
Group 3: Even Gege (name) went to work.	Group 4: Everybody went to work.
9. *ge3ge0(ge1) dou1 qu4 shang4ban1 le.*	13. *ge4ge0(ge1) dou1 qu4 shang4ban1 le.*
10. *ge3ge0(ge2) dou1 qu4 shang4ban1 le.*	14. *ge4ge0(ge2) dou1qu4 shang4ban1 le.*
11. *ge3ge0(ge3) dou1 qu4 shang4ban1 le.*	15. *ge4ge0(ge3) dou1 qu4 shang4ban1 le.*
12. *ge3ge0(ge4) dou1 qu4 shang4ban1 le.*	16. *ge4ge0(ge4) dou1 qu4 shang4ban1 le.*

Note: The pinyin in the parentheses is the base form of the neutral tone.

6.2.5 Results

All of the subjects selected the sentence that has the original neutral tone to match the English meaning. All of them reported that the other three sentences did not match the English meaning, and the words which include the mismatched neutral tone sounded strange to them. The results support the hypothesis. The pitch contour of the neutral tone does affect listeners' comprehension. It plays an important role in tone perception. Therefore, the perceptual space of T0 overlaps with that of other tones.

6.3 Experiment 3-b: Perceptual Experiment about the Relationship between T0 and Other Tones

6.3.1 Purpose of This Experiment

The salient features of the neutral tone are its shorter duration and numerous variations of pitch. The experiment explores whether the neutral tone can be perceived as a normal tone, such as T3, or T4, if the duration of the neutral tone is lengthened.

6.3.2 Subjects

The four subjects in this experiment are the same as those in experiment 3-a.

6.3.3 Stimuli

Cao (2002b) concludes that the neutral tone has different pitch contours depending on the tones of the previous syllable. Fig 6-1 shows four pitch contours of the neutral tone when they come after T1, T2, T3, and T4.

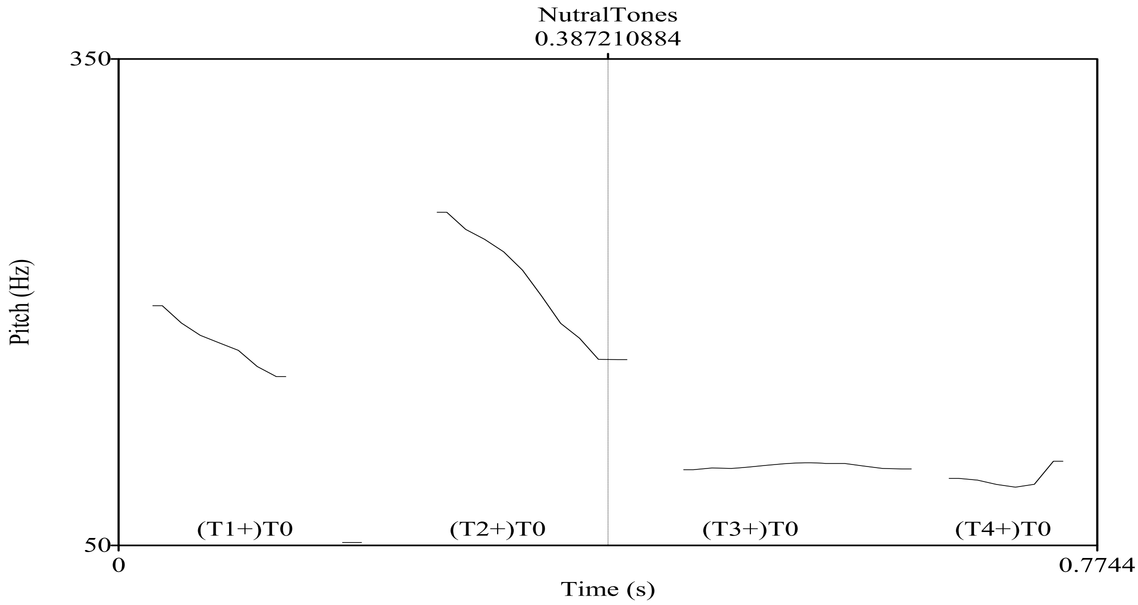


Figure 41 Pitch contours of T0 (after different tones)

It is obvious that the pitch contour and the register of T0 that comes after T1 are similar to those of T3, and the pitch contour and the register of T0 that follows T2 is similar to those of T4. The pitch contour of the other two is similar to T1, yet the register is totally different. Therefore, this experiment is proposed to compare the perception of T0 that comes after T1 and T2 with the perception of T3 and T4.

The syllables used here are the same as those in 3-a (see table 20 above). Following is the way to synthesize the duration of the tones. First, the neutral tones which come after T1 and T2 are cut from the sentences. They are labeled as T(1+)0, and T(2+)0. The two neutral tones are lengthened to double duration via the program Praat. The minimum pitch is set to 50Hz, the maximum pitch is set to 350Hz, and the factor is set to 2. Two tones with duration double that of the original neutral tones are synthesized, and are labeled as (T1+)T00 and (T2+)T00. Then (T1+)T00 is inserted into the carrier sentence in place of T3, and (T2+)T00 is inserted into the carrier sentence in place of T4.

Two sentences with the original T3 and T4 are also used for subjects to listen to. Two fillers are used in this experiment. So, in total there are 6 sentences for native speakers to listen to.

6.3.4 Procedure and Analysis

Subjects listen to the synthesized syllables (words) with the modified tones, and then select an appropriate English meaning from the four choices; that is, brother, princess, gege (name) and everyone. Then they are required to mark the degree of naturalness (natural or unnatural).

6.3.5 Results

All four subjects in this experiment selected the correct English meaning for the stimuli. The result indicates that the lengthened T0 that originally came after T1 sounds natural when they replace T3; the lengthened T0 that originally came after T2 sounds natural when they replace T4. It shows that the neutral tone and normal tones can be altered based on duration. It is obvious that the perceptual space of T0 must overlap with the normal tones, because the neutral tone and T3 and T4 share the same pitch contour and register.

6.3.6 Conclusion

The results of Experiment 3a and Experiment 3b indicate that pitch plays an important role in perception of the neutral tone. In some cases, the neutral tone and normal tones can be altered based on duration, that is, normal tones are the lengthened neutral tones. Therefore, the perceptual space of T0 overlaps with that of the normal tones.

6.4 Experiment 3-c: Perception Experiments for T3

6.4.1 Purpose of the Experiment

Do the pitch at the end of T3 and the pitch at the beginning of the following tone play an important role in perceiving T3? The hypothesis is that the perceptual space of T3 might overlap with T4 if the pitch at the beginning of the following tone plays an important role. This experiment is used to test this hypothesis.

6.4.2 Subjects

Four native speakers, who are the same as those in Experiment 3b, participate in this experiment.

6.4.3 Stimuli

Table 22 shows two pairs of words used in the experiment.

Table 22 Stimuli: two pairs of words relevant to T3

	Original word	Word in which the syllable positions are changed	Possible meanings
Pair 1	Hen3 ai4	Ai4 hen	a) very much love b) love and hate c) love very much
Pair 2	Kan3 zai4	Zai4 kan	a) cut on b) be looking (at) c) be cutting

Select two words that are combinations of T3 and T4: *hen3ai4* (very much love) and *kan3zai4* (cut on). A native speaker reads these two words, and the sounds are recorded. Then, use program Praat to change the position of the syllables in each word. Thus, two other words are created: *ai4hen* and *zai4kan*. We have two pair of words, that is, *hen3ai4* and *ai4hen*, *kan3zai4* and *zai4kan* (Table 22). Another pair of words is used as filler in this experiment. So there are three pairs of words.

6.4.4 Method

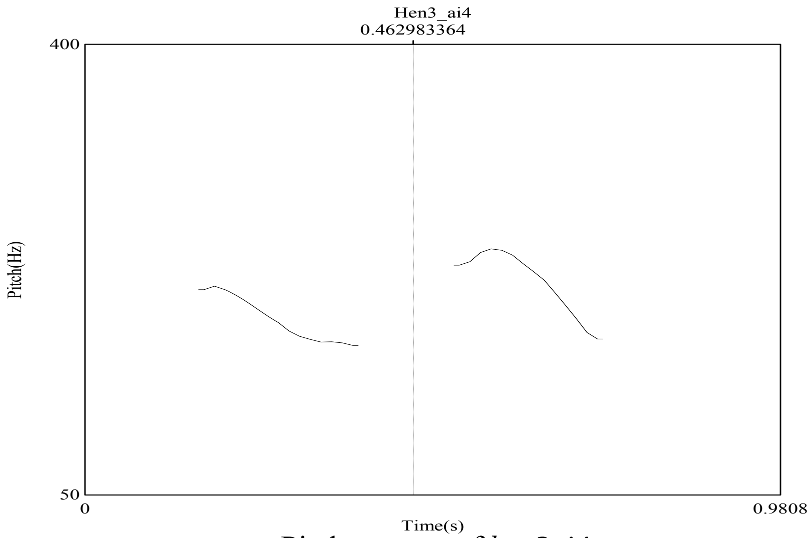
Four native speakers listen to these three pairs of words. Three English meanings for each pair are provided to the native speakers (see Table 22), and then they are required to select the one which is most close to the word perceived.

6.4.5 Results

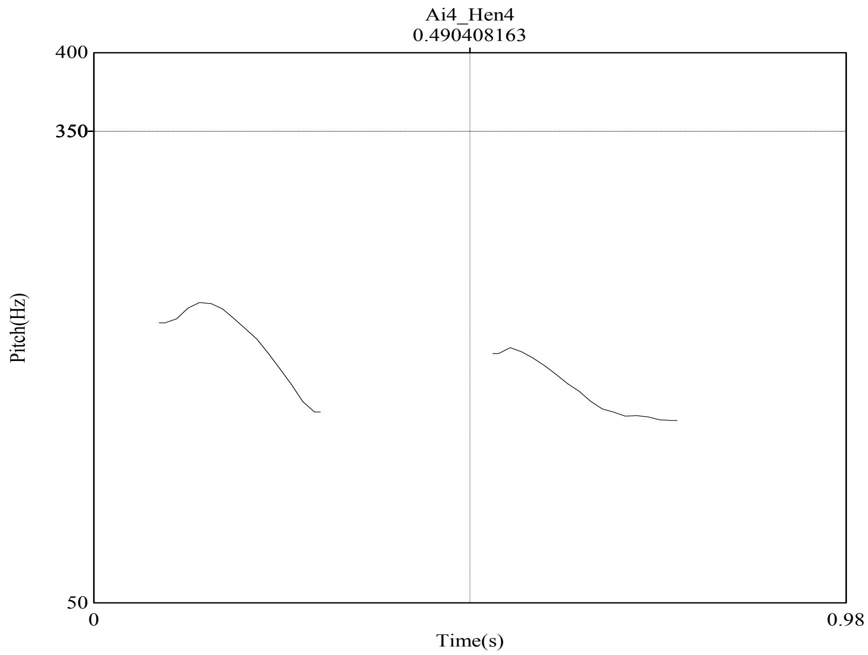
All native speakers select the second meaning for the word that is modified. That is, *Ai4 hen* means 'love and hat'; and *Zai4 kan* means 'be looking (at)'. The results indicate that all subjects perceive *hen* and *kan* as T4 of which the acoustic features are the features of T3.

The reason that NS perceive *hen* and *kan* as T4 is that no syllables come after the words, so there is no high register at the beginning point of the following tone which could be used to construct the rising part of *hen* and *kan*.

Figure 42 (a-d) indicates the pitch of each words. The pitch contour and register of the first syllable *hen3* are the same as those of the second syllable *hen4* in the second word. The pitch contour and register of the first syllable *kan3* are the same as those of the second syllable *kan4* in *zai4kan4*.



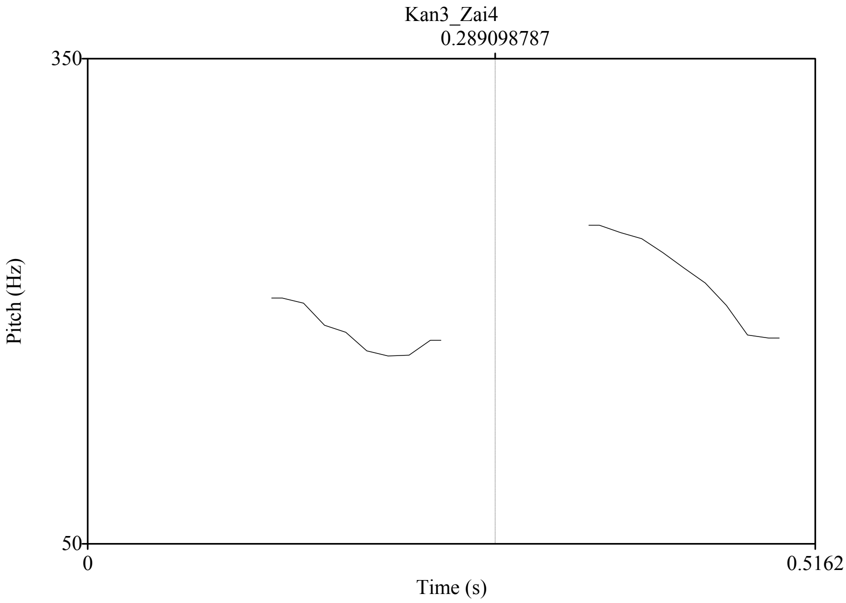
a. Pitch contour of *hen3ai4*



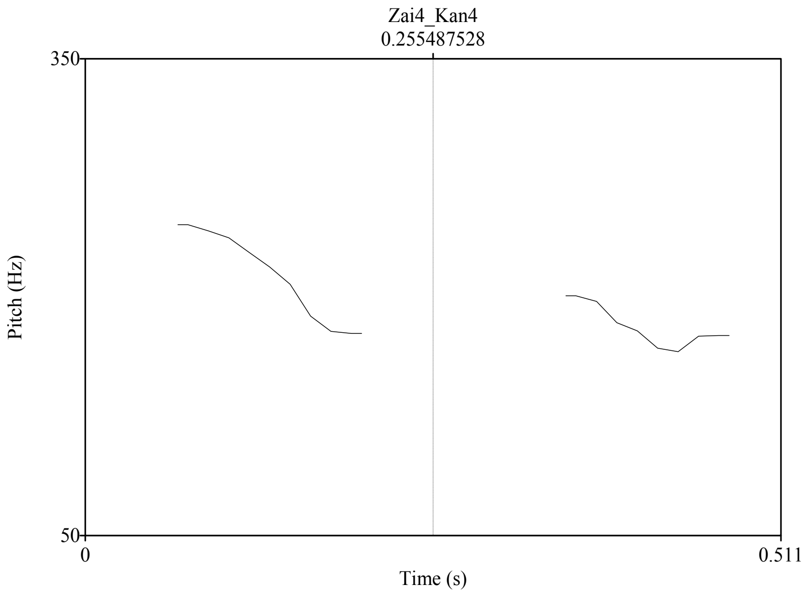
b. Pitch contour of *ai4 hen4*

Figure 42 Pitch contour of two pairs of words

Figure 42 continued



c. Pitch contour of *kan3 zai4*



d. Pitch contour of *zai4 kan4*

It is obvious that T3 can be perceived as T3 or T4, according to the linguistic environment nearby, especially the neighboring tones, even though the acoustic features or F0 values are not changed. As we know, T3 is regarded as a dipping tone in the Mandarin system. However, the acoustic features of T3 in the connected speech do not show this perceptual feature. The acoustic feature of the following tone decides whether a low-register tone can be perceived as T3. If the beginning point of the following tone is not in the high register, or T3 is in the word or phrase boundary, a low-register tone could be perceived as a falling tone, that is T4. Thus, T3 are overlapping with T4 in the low-register. This is the reason that T3 and T4 are overlapping in the NS production map, and they are circled in the same group. Also, in the perception map, T3 has it own perceptual space although it occupies the least area. Meanwhile, other tones are overlapping with T3 in the low-register perceptual space. If the beginning point of the following tone is in the high register, the ending point of T3 and beginning point of the following tone create a rising perceptual cue. Therefore, we hear a dipping tone: falling first, then rising.

In conclusion, the pitch at the end of T3 and the pitch at the beginning of the following tone play an important role in perceiving T3. These two pitch points let listeners perceive the rising part of T3, which makes T3 as the dipping tone. The results support the hypothesis that the perceptual space of T3 overlaps with T4, and the pitch at the beginning of the following tone plays an important role in perceiving T3.

CHAPTER 7 EXPERIMENT 4: NON-NATIVE SPEAKER TONAL PRODUCTION

7.1 Introduction

This experiment investigates tonal production by non-native speakers, and explores the differences between native speakers' perceptions and non-native speakers' production. Some studies (Yang, etc, 2007) have shown that non-native speakers' production is not consistent with their perception. However, this study does not purely focus on the differences between non-native speakers' perceptions and production.

This chapter mainly focuses on Research Question 4:

What are the features of NNS' tonal production? What are the differences between NNS' production features and NS perception features?

Experiment 4 is designed to answer this question.

7.2 Subjects

Eleven American students of Chinese who are native speakers of English participated in this experiment. All were enrolled in second year Chinese at the University of Iowa. Six of them also participated in Experiment 2-b. Therefore, the researcher only uses the data from these six learners in Experiment 4, in order to compare their perception with their production.

7.3 Method

This section describes how the production data are elicited, how F0 is normalized, and how the results are analyzed.

7.3.1 Elicitation

Eleven non-native speakers are asked to speak for one minute on the topic of gender equality (See appendix B: Production Data Collection Instruments). The subjects complete the task in the language laboratory which is equipped with a Sanako Lab 100 digital audio system. Only one subject at a time is in the lab. The researcher controls the recording system and recorded each speech. Since the production data needed to be compared with the perception data, only data from the six non-native speakers who participated in Experiment 2-b are used.

7.3.2 Pitch Normalization

The method is the same as that described in Chapter 3. Both intra-speaker normalization and inter-speaker normalization are conducted. Here is a brief description of the two kinds of normalization.

The first step is intra-speaker normalization. This uses the nucleus model to remove the sensitive part of a tone which is influenced by the surrounding linguistic contexts. Pitch values for each tone in every student speech are extracted automatically by the Multi Speech Program. The researcher uses the nucleus model (see 3.2.1) to remove tiny amounts of sound at the beginning and end of each tone with an onset or offset, so only the midpoint of the tone remains[11]. Then, the F0 value at the beginning and end of the nucleus of each tone are selected.

The second step is inter-speaker normalization. That is, the individual pitch values are normalized by a z-score method so that different speakers can be compared at a similar level. The normalization controls for inter-speaker variation, since different people can produce different variations of a tone in the same context.

These z-score transforms (see 3.2.2) "express an observed F0 value as a multiple of a measure of dispersion away from a mean F0 value" (Rose, 1987:347). The F0 values to be normalized (z-values) are calculated using the standard Formula A below:

[Formula A] $$z_i = \frac{x_i - m}{s}$$

[11] The detailed steps please refer to Chapter 3, section 3.2.1.

X_i is an observed F0 value, m is the mean value, and s is the standard deviation (SD) of a given sample. For example, if a speaker's F0 value at the 5% point of a tone is 180 Hz, and his overall mean and SD are 140 Hz and 16 Hz, respectively, the normalized value for 180 Hz would be calculated as (180-140)/16=2.5.

All the F0 values at the beginning of the nucleus of all tones and those at the end of the nucleus of all tones, which are produced by six learners, are normalized using z-score formula. Thus we get the normalized values of a relevant pitch in the 5-scale notation system (Chao, 1930)

7.3.3 Analysis

Two native speakers listened to these six student speeches, and coded every tone. If a tone belongs to T0, it is not further analyzed by the perceptual model, since the pitch space overlaps with that of the normal tones. If a tone could not be perceived as belonging to any categories of normal tones or could be perceived as belonging to two categories of normal tones, the listener coded it as "ambiguous" on the evaluation sheet (see Appendix C: Native Speaker's Evaluation Sheet). The two native speakers disagreed on the assignment of 11% of the tones. Additional three native speakers listened to these divergent tones, so these divergent tones were coded by five native speakers in total. A tonal category of "ambiguous" is marked for a tone which at least three NS found ambiguous or disagreed about.

Then the normalized pitch values of these tones are compared with the results from Experiment 2-a. The researcher categorizes each tone, according to Table 13 Tone Categories in Perceptual Space. This produces the results from the perceptual model.

The next step compares human native speakers' perceptions to the results from the perceptual model. Some of the results are the same, yet others differ. So, we define four groups to code the data (see Table 23). The group "same" means that the results from human and from the model are the same. The group "boundary" means that the pitch values of the tones are in the boundary area in the perceptual model regardless of whether the results from a speaker and from the model are the same or not. The group “T3” means that the original tone is T3, or the result from a speaker or the model is T3, regardless of whether the results from a speaker and from the model are the same or not. The group “other” means that the results from human and from the model differ, and these tones do not belong to either “boundary” or “T3”.

Table 23 Coding results from human and machine judgments

	Same Results from Human and Machine Judgments	Different Results from Human and Machine Judgment
Same	Include in sample	Exclude
Boundary	Include	Include
T3	Include	Include
Other	Exclude	Include

Then, the individual differences are analyzed and the production maps are further drawn.

7.4 Results

The results in this experiment are presented from three perspectives. The first one is the error analysis for all subjects, which mainly focuses on error distribution. The second one is individual error analysis, which mainly focuses on register and tonal contour. The third one is NNS production maps. Through these results, the features of NNS production of tones are revealed.

7.4.1 Production Errors by Six Learners

This section compares the human perception results with the perceptual model. The results show the error distribution of NNS production. The results also indicate whether the perceptual model works for assessing non-native speakers' production.

7.4.1.1 Error distribution of NNS' production

Six speeches produced by six non-native speakers are analyzed in Experiment 4. Since every participate gives a speech, each produces a different number of tones. We obtained a total of 447 tones, which include 396 normal tones and 51 T0. We obtained 66 tokens from NNS-1, including 8 T0; 73 tokens came from NNS-2, including 2 T0; 78 tokens from NNS-3, including 13 T0; 62 tokens came from NNS-4, including 6 T0; 102 tokens came from NNS-5, including 13 T0; 66 came from NNS-6, including 9 T0, as summarized in Table 24 below.

Table 24 Number of tones produced by NNS

	Normal Tones T1-4	T0	Total
NNS-1	58	8	66
NNS-2	71	2	73
NNS-3	65	13	78
NNS-4	56	6	62
NNS-5	89	13	102
NNS-6	57	9	66
Total	396	51	447

Only the results of the normal tones T1-T4 are analyzed here.

According to the categories defined in Table 23, we calculate the number of the errors in which the human perception assessment differs from the perceptual model. We also calculate the number of errors which are relevant to either boundary or T3. The error distribution is summarized in Figure 23 below. In 50% the human assessment is the same as the perception model. In 45% the results are different, and fall either in the boundary areas or are related to Tone 3. The remaining 5% of results are 'other', and involve T1, T2, or T4.

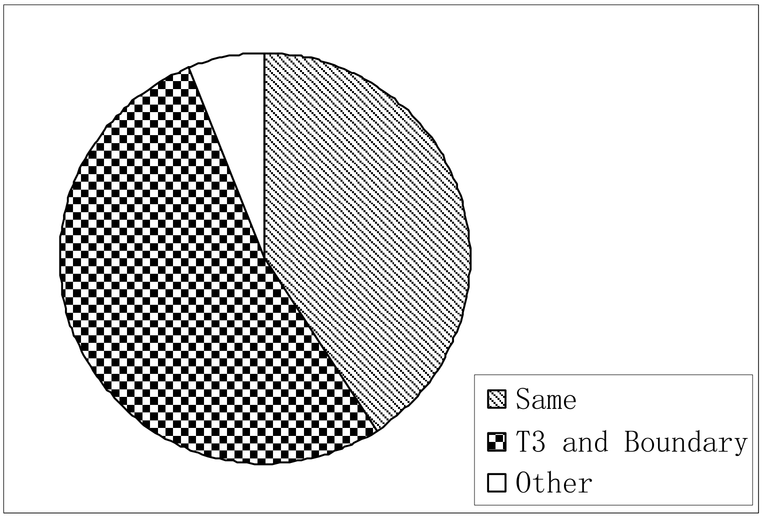


Figure 43 Error distribution of NNS production

Table 25 shows the errors by individual speakers according to the categories defined in Table 23.

Table 25 Individual error distribution

	Same	Boundary	T3	Other	Total	Error (%)
NNS-1	6	9	11	0	19	33%
NNS-2	24	11	15	0	35	49%
NNS-3	15	10	12	2	27	42%
NNS-4	9	11	11	2	21	38%
NNS-5	15	18	21	3	44	50%
NNS-6	18	11	8	1	28	49%
Total	87	70	78	8	174	

The total error numbers and error percentages in the last two columns show results that are assessed by both human and the perception model. If both the human and the model's judgments are the same, only 1 error is counted; if human and the model judgments differ, 2 errors are counted. Obviously, 'other' errors caused by other factors are fewer for the NNS. For most non-native speakers' errors, the human and the perception model assessments agree, coded as 'same'. Most divergent results are in the boundary areas or related to T3, because in the boundary areas, a tone can be perceived as belonging to at least two tonal categories, even if they have the same pitch values. However, some individual differences emerge in Table 25. For example, NNS-1 and NNS-4 have fewer "same" results than the other speakers.

In conclusion, the perception model is reliable for assessing NNS production. Only 5% of the difference between human and the model judgments cannot be accounted for, except boundary area and T3. However, some individual differences remain.

7.4.1.2 Errors in boundary areas and/or relevant to T3

The model is employed in this section to analyze how many errors are made in the boundary areas and/or relevant to T3.

Table 26 Errors in boundary areas and/or relevant to T3

	Boundary+T3	Total Number of Errors	Boundary+T3 (%)
NNS-1	13	16	81.25%
NNS-2	21	34	61.76%
NNS-3	12	23	52.17%
NNS-4	15	20	75.00%
NNS-5	34	44	77.27%
NNS-6	18	28	64.29%

Most errors (ranging from 52.17% to 81.25%, and average is 71.43%) are made in the boundary areas and/or relevant to T3.

7.4.1.3 Conclusion

The production results for the six learners indicate that the assessments made by humans and those based on the perceptual model are close (5% difference), except for the tones in the boundary areas and those relevant to T3, that is, the overlapping areas. 71.43% of errors, which are assessed by the perceptual model, are in the boundary areas or relevant to T3. Some individual differences remain. Therefore, the results for each individual NNS are presented in Section 7.4.2 below.

7.4.2 Individual Error Analysis

In this section, all the errors of normal tones made by each NNS are coded into tonal categories, that is, T1, T2, T3 and T4. The results for each tone are analyzed at two levels: contour and register. The results indicate that if the contour is wrong, one of the registers at the beginning/ending point are also wrong. However, if the register of a tone is incorrect, the contour may still be correct. The results also reveal the similarities between NNS productions and NNS perceptions.

7.4.2.1 Tone errors for NNS-1

The results for the production analysis of NNS-1 are summarized in Table 27 below. Only 4 errors relate to contour. But all of the mis-produced tones have register errors.

Table 27 Tone errors for NNS-1

Tonal category	Contour Errors	Register Errors	Total No. of Errors
Tone 1	Falling	3 End point	3
Tone 2	3 falling	6 Beginning point	6
Tone 3	No	6 Beginning point	6
Tone 4	No	1 End point	1

The perceptual results for NNS-1 indicate that most T2 and T3 appear in the low register, while most T1 and T4 are distributed in the high register. The spaces overlap. That is, each area can be perceived as belonging to two tones. There are no clear boundaries in NNS-1's perceptual space (see Table 14 in Chapter 5).

Since NNS-1 has not-bad perceptual space distinctions, not many errors appear in her tonal production. The main problem is the register errors. The reason is that NNS-1 lacks a clear perceptual space, and there are not any stable areas. Generally, without a good foundation in perception, it is easy for a NNS to mis-produce tones in wrong register. Yet, NNS-1 does not make many contour mistakes.

7.4.2.2 Tone errors for NNS-2

The results for the production analysis of NNS-2 are summarized in Table 28 below. Many fewer errors occur for T1 and T4 than for T2 and T3. All the mis-produced tones have register errors. However, almost all the T2 errors relate to contour. That is, 17 out of the 29 contour errors are on T2, more than on the other three tones combined.

Table 28 Tone errors for NNS-2

Tonal category	Contour Errors	Register Errors	Total No. of Errors
Tone 1	2	Both points	2
Tone 2	17	Both points	19
Tone 3	1	Beginning point	10
Tone 4	2	Both points	3

The perceptual results for NNS-2 showed a four-quadrant model of tonal perception, which is similar to NS perception. The bottom-left area resembles the NS perceptual space for T3. However, the T1 area is obscure, that is, several tones overlap in this area. The perceptual boundary is not clear too, yet there is still a boundary between -0.5 to -1 at the beginning point of tones (see Table 15 in Chapter 5).

Since NNS-2 has good perceptual space, few errors appear in her tonal production, except on T2. The majority of errors are relevant to T2 and T3. However, the main problem for T2 is contour, while the main problem for T3 is register. Two explanations exist for T2 contour errors. The first is that NNS-2 does not know that some syllables should be produced as T2, so they are mispronounced as T4. The second possibility is that NNS-2 does know that these syllables carry T2, but mispronounced them with the made contour errors. As for T3, NNS-2 made 10 errors, but only one relates to contour, while all relate to register. NNS-2 made mistakes on T2 and T3 for different reasons. Therefore, NNS-2 does not confuse T2 with T3, even though most of the errors are on T2 and T3.

7.4.2.3 Tone errors for NNS-3

The production analysis of NNS-3 is summarized in Table 29 below. Many fewer errors appear for T1 and T4 than those for T2 and T3. All the mis-produced tones have register errors. But, all errors for T2 relate to contour, much more than for the other three tones.

Table 29 Tone errors for NNS-3

Tonal category	Contour Errors	Register Errors	Total No. of Errors
Tone 1	1	both points	1
Tone 2	11	both points	11
Tone 3	1	both points	9
Tone 4	1	both points	2

The perceptual results for NNS-3 show something approaching a four quadrant model, although it is differs from NS perceptual space. Most T1 are located in the upper-right space; most T2 in the bottom-right space; T3 occupies small area of the bottom-left area; while most T4 are in the upper-left area. The space overlaps. That is, each area can be perceived as belonging to two tonal categories. And there is no clear boundary in NNS-3's perceptual space (see Table 17 in Chapter 5).

Since NNS-3 has a not-bad perceptual space, not many tone errors appear. Most errors are relevant to T2 and T3. However, the main problem with T2 is contour, while the main problem with T3 is register. There are two possible explanations for T2 contour errors. The first is that NNS-2 does not know some syllables should be T2, and so mispronounces them as T4. The second possibility is that NNS-2 does know that these syllables carry T2, but mis-pronounced them with contour errors. As for T3, NNS-3 made 9 errors. Only one relates to contour, but all relate to register. NNS-3 made mistakes on T2 and T3 for different reasons. Therefore, NNS-3 is not confused between T2 and T3, even though of the errors are on T2 and T3.

7.4.2.4 Tone errors for NNS-4

The production analysis of NNS-4 is summarized in Table 30 below. No errors appear on T1. However, NNS-4 makes more errors on T3 and T4 than on T2. All the mis-produced tones have register errors. Almost half of the errors relate to contour. NNS-4 mis-pronouced four T4 falling syllables as rising tones.

Table 30 Tone errors for NNS-4

Tonal category	Contour Errors	Register Errors	Total No. of Errors
Tone 1	n/a	n/a	0
Tone 2	2	beginning point	4
Tone 3	1	beginning point	9
Tone 4	4	both points	7

The perceptual results for NNS-4 show a clear distinction between T2 and T4. Most T2 are located in the bottom-right area, and most T4 in the upper-right area. However, there is no clear boundary area separating the four tones (see Table 18 in Chapter 5).

Since NNS-4 has good perceptual map, few tonal errors appear. Most errors relate to T3 and T4. Half of the T4 errors relate to contour. There are two possible explanations for T4 contour errors. The first is that NNS-4 does not know some syllables should be T4, so mispronounces them as T2. The second possibility is that NNS-4 does know that these syllables carry T4, but s/he made contour errors in production. T3 and T4 do have similar contours: the only difference between them is register. Therefore, it is reasonable that NNS made register errors for T3 and T4.

7.4.2.5 Tone errors for NNS-5

The production analysis for NNS-5 is summarized in Table 31 below. Many fewer errors appear for T1 than those for T2, T3 and T4. Most T3 errors concern contour. And all the errors relate to register. NNS-5 mis-produced six rising tones for T4 falling tones.

Table 31 Tone errors for NNS-5

Tonal category	Contour Errors	Register Errors	Total No. of Errors
Tone 1	2	beginning point	3
Tone 2	2	beginning point	11
Tone 3	12	both points	17
Tone 4	6	both points	13

The perceptual results for NNS-5 showed no clear perceptual tendency. There is no four quadrant distribution and no boundary areas to separate the four tones (see Table 19 in Chapter 5).

Since NNS-5 lacks a similar perceptual space to a NS, s/he lacks the good perceptual foundation for tonal production. NNS-5 made more production errors than any other subjects. Most of the T3 and T4 errors relate to contour. There are two possible explanations. The first is that NNS-4 does not know some syllables should be produced as T3 or T4, so mispronounces them as rising tones. The second possibility is that NNS-5 does know that these syllables carry T3 or T4, but mis-pronounces them with contour errors. To conclude, NNS-5 made the most tonal production errors which are relevant to both the contour and register. The production results match the perception results.

7.4.2.6 Tone errors for NNS-6

The production analysis of NNS-6 is summarized below in Table 32. Many fewer errors occur for T1 and T4 than for T2 and T3. All the mis-produced tones have register errors. However, most errors for T2 relate to contour. For T3, half of the errors relate to contour, as do all the errors for T4.

Table 32 Tone errors for NNS-6

Tonal category	Contour Errors	Register Errors	Total No. of Errors
Tone 1	1	both points	1
Tone 2	8	both points	10
Tone 3	7	both points	13
Tone 4	4	both points	4

The perceptual results for NNS-6 indicate that this learner's perceptual space lacks any boundary areas. Even so, most T1 are located in upper-right area; most T2 in bottom-right area; most T3 locate in bottom-left area; while most T4 are located in upper-left area. The spaces overlap. That is, each area can be perceived as belonging to two tonal categories. Meanwhile, the register for the beginning point of each tone is clear, but the contour is not. For example, the tone with the beginning point -1 and the ending point -2 is perceived as T2 by NNS-6. Yet the contour shows that it is a falling tone which a NS would perceive as T4 (see Table 20 in Chapter 5).

Since NNS-6 has not-bad perceptual space, not many errors appear in his/her tonal production. Most errors relate to T2 and T3. Most production errors relate to contour. The production results match the perception results. NNS-6 has an obvious contour problem.

7.4.2.7 Conclusion

The analysis of the individual tonal production yields several results. First, few errors appear for T1. Generally, both T1 and T4 have fewer errors than T2 and T3 do. This indicates low register tones are harder to produce than high register tones.

Second, all the errors relate to the register of either the beginning point or the ending point, yet most have the correct contours. However, T2 has more contour errors than the other tones. NNS often produce T2 as falling tone. As mentioned above, this may be for two reasons. First, NNS-2 may not know that some syllables should be T2, so they mispronounce them as T4. The second reason is that NNS-2 does know that these syllables carry T2, but they made contour errors when they producing these tones. Because falling intonation is a dominant stress marker for sentence intonation in English, it is reasonable that English-speaking NNS mis-pronounce T2 as a falling tone.

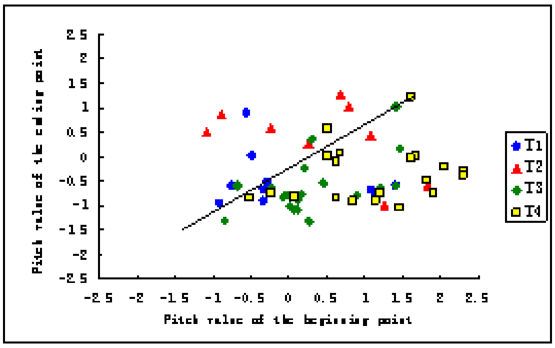
Moreover, although T2 and T3 produce more errors than T1 and T4, this does not mean that NNS are confused by a contrastive feature between T2 and T3. T2 and T3 mistakes occur for different reasons. Most T2 errors relate to contour, while most T3 errors have the correct contour but the wrong register. Therefore, T2 and T3 are perceived inaccurately for different reasons.

Finally, the results from individual NNS production are consistent with their individual perceptual spaces. For example, NNS-6 has an obvious contour problem, both in production and perception. Another example is NNS-5, whose perceptual space lacks both a four quadrant distribution as well as boundary areas between the four tones. Consistent with his/her perceptual space, NNS-5 made most tonal production errors relative to both contour and register.

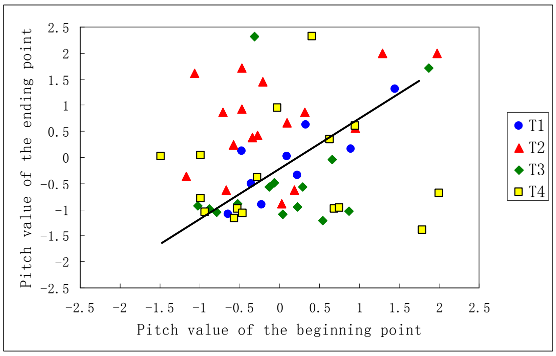
In conclusion, all errors relate to register. More errors appear for T2 and T3, although T2 and T3 mistakes occur for different reasons. On the other hand, the results from individual non-native speakers' production are consistent with their individual perceptual spaces. In the following sections, we will further observe the features of NNS production of tones through production maps, and explore the relationship between NS perception and NNS production in the following section (section 7.5).

7.4.3 NNS Production Maps

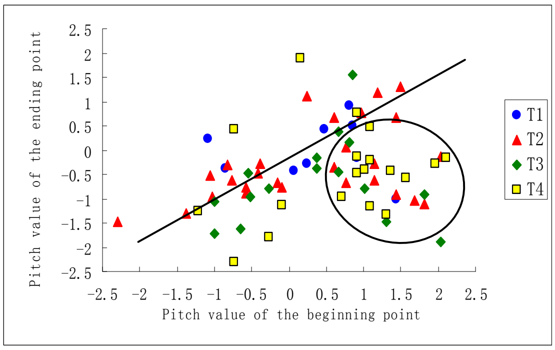
The results of this experiment also show that the features of NNS production of tones are different from those of NS production of tones. Figure 44 shows the production maps of six NNS.



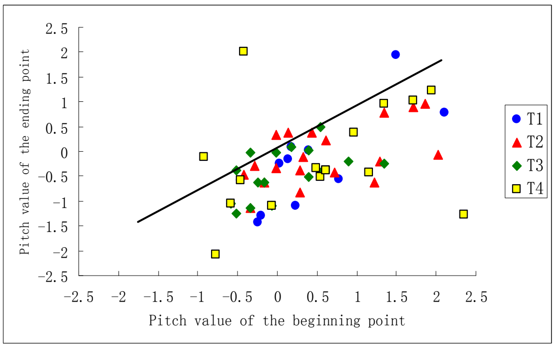
NNS1



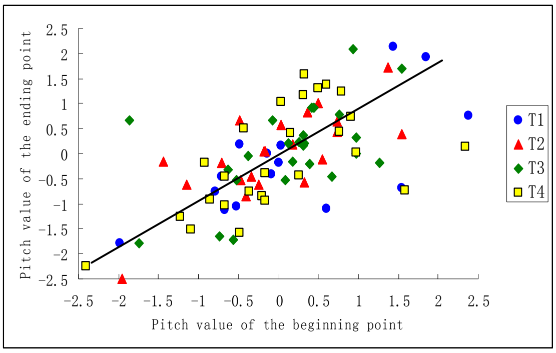
NNS4



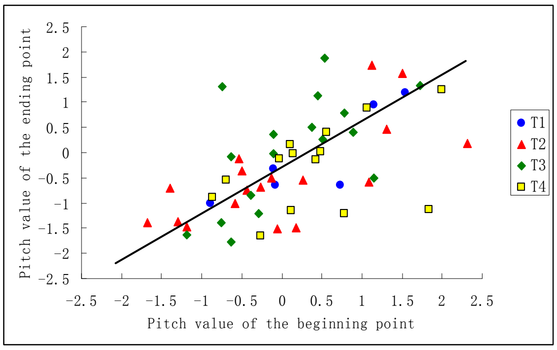
NNS2



NNS3



NNS5



NNS 6

Figure 44 Production maps of NNS

The production maps of NNS could be divided into three groups. The first group includes the production map of NNS1 and that of NNS4, the second group includes the production maps of NNS2 and NNS3, and the third group includes the production maps of NNS5 and NNS6.

<u>7.4.3.1 First group of production maps</u>

The production maps in the first group divide the production space into two categories. One is a T2 space where most points representing T2 are above the function y=x; the other is T1, T3 and T4 space where most points representing T1, T3 and T4 are under the function y=x. This means that most T2 are rising tones, while most T1, T3 and T4 are falling tones. However, there is still a little difference between the production of NNS1 and that of NNS 4. Figure 45 and Figure 46 enlarge the production maps of NNS1 and NNS4, which show the detailed difference between the two NNS.

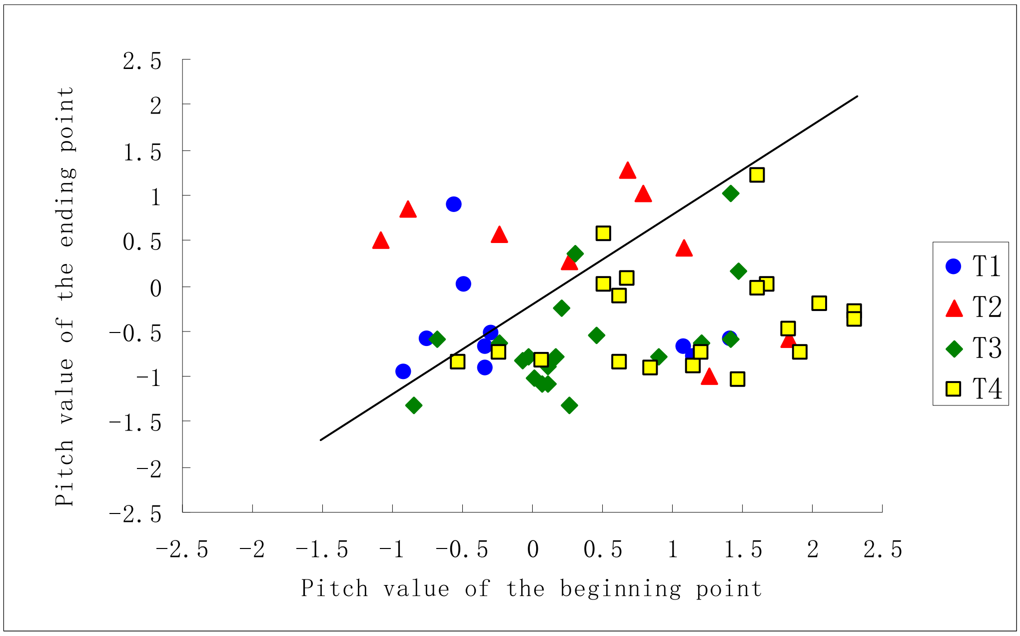


Figure 45 Production map of NNS1

Figure 45 shows the production map of NNS1. Most T2 are above y=x, and T3 and T4 are mixed under y=x, which is similar to the production map of a NS. T1 is near y=x (x≤0, y≤0), which is different from the NS production map. It means that T1 is produced in the low register area.

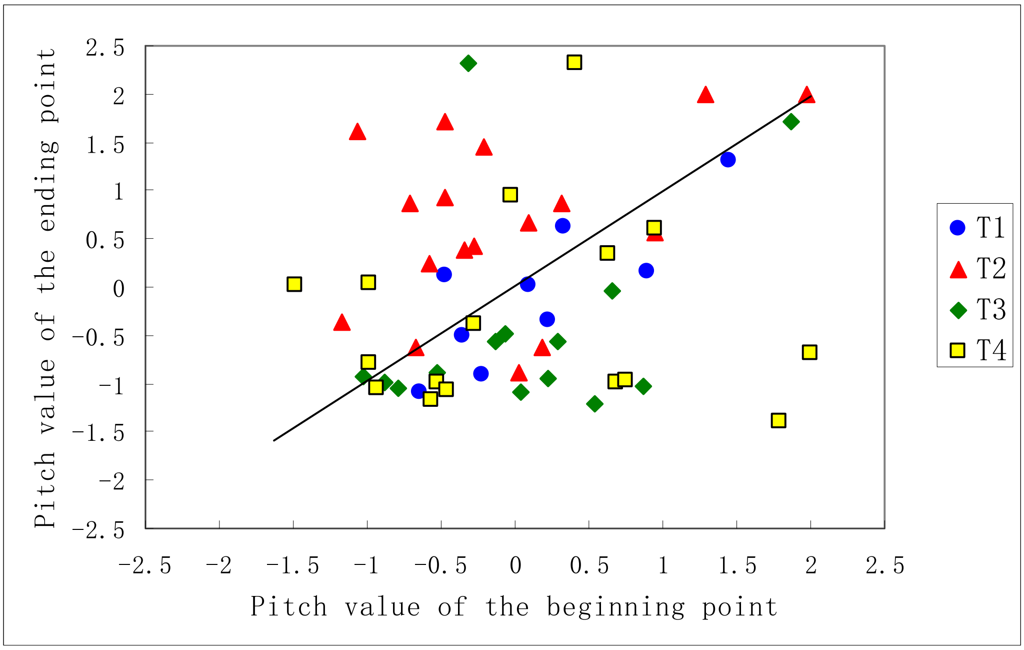


Figure 46 Production map of NNS4

Figure 46 shows the production map of NNS4. Most T2 are above y=x. T1 is mixed with T3 and T4 in the map. Most T3 and T4 are under y=x, which is similar to the NS production map, yet most T3 and T4 are produced in the lower register area. T1 are near the function y=x. which is different from the production map of NS.

7.4.3.2 Second group of production maps

The production maps in the second group have features in common, such as, most tones are under the function y=x. This means that the NNS intend to produce falling tones.

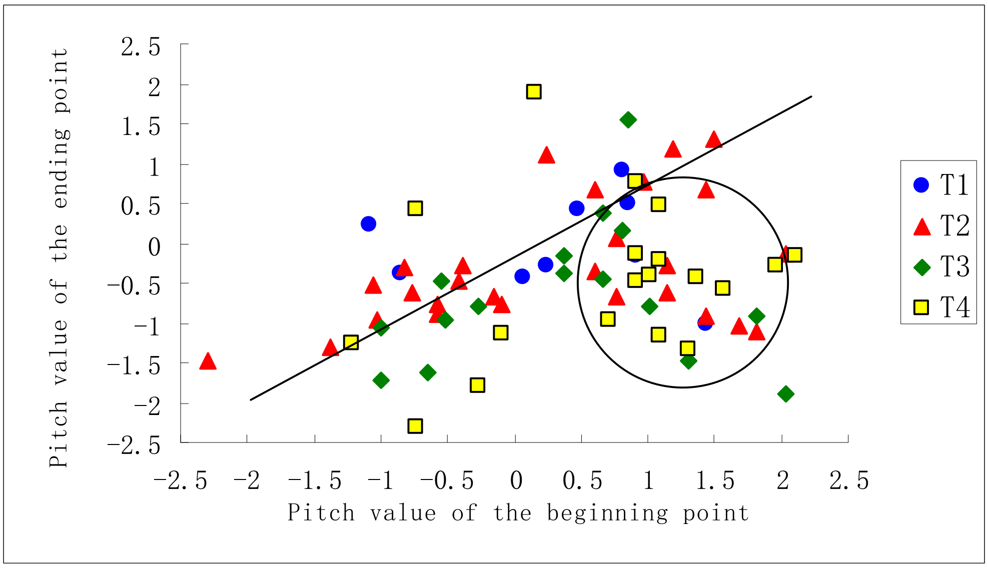


Figure 47 Production map of NNS2

The production map of NNS2 is different from that of NNS3. Almost all the tones produced by NNS3 are under the function y=x. However, there is a special area in the map of NNS2. Figure 47 enlarges the production map of NNS2. In this map, most tones are near or under the function y=x. However, the points which are circled are most T2 and T4, and the points in this circle represent falling tone. As we know, T2 is a rising tone. Recall that NNS-2 made 19 errors relevant to T2, while 17 of them are contour errors. The results match the production map. On the other hand, T4s are produced as falling tone in this circle, that is, the beginning points are positive, and the ending points are negative. Again, because falling intonation is a dominant stress marker and sentence intonation in English, it is possible that English-speaking NNS mis-pronounce T2 as a falling tone and it is easy for them to produce T4 well.

In this group, NNS intend to produce falling tones instead of other tones because falling intonation is a dominate stress marker in English – the first language of non-native speakers.

7.4.3.3 Third group of production maps

The production maps in the third group have a feature, that is, most tones are near the function y=x. This means that NNS intend to produce level tones instead of rising tones and falling tones. The results indicate that NNS in this group produce most errors of tones. It seems that NNS in this group have less tonal categories.

7.4.3.4 Conclusion: comparison between NNS production and NS production

The results of NNS production show overlapped tones in the acoustic space, while NS production data indicate three main categories (one of them is the mixed categories of T3 and T4) in the acoustic space, with overlapped tones in the boundary areas.

The results of the production maps indicate that some NNS can categorize the tones into two groups: a rising tone which includes T2, and a falling tone which includes T1, T3 and T4. Some NNS cannot categorize the tones during production. Therefore, they mainly produce a single category: which is either falling tones that is the default intonation in their L1 or level tones.

The NNS production appears to be determined by the contour of the tones by contrast both the contour and register which are used to determine the tonal categories of NS.

7.5 Comparison between NNS Production and NS Perception

Based on the results, we further discuss the differences and similarities between NNS production and NS perception.

Recall the results of perception experiment for NS. The results of the NS perception of tones reveal a quadrant model, and the T3 is a mixed category since mixed tones exist in T3 space. Boundary areas exist in the perceptual space. In the boundary areas, the tones are perceived as "obscure", that is, they can be perceived as at least two different tones. These boundary areas are located where the beginning point or the ending point is near the value 0.

First, the production results for the six learners indicate that the assessments made by humans and that based on the perceptual model are close (5% difference), except for the tones in the boundary areas and those relevant to T3, that is, the overlapping areas. Average 71.43% errors, which are assessed by the perceptual model, are in the boundary areas or relevant to T3. In the boundary areas and T3 space of NS perceptual space, the tones are perceived as "obscure", that is, it can be perceived as at least two different tones. Therefore, it is hard for NS to perceive the tones that are produced in the boundary areas or relevant to T3 by NNS.

Second, the results of NNS production maps indicate that NNS do not have three or four categories that match NS perception. The first group of NNS production has two categories: T2 vs. T1, T3 and T4. T3 and T4 locate in the right space that is similar to the production space of NS. T1 is different form that in NS space. However, recall that T1 has the widest perceptual space in NS perception space. This is the reason that T1 produced by NNS in the first group can still be perceived by NS. The results also indicate that NNS in this group produced least tonal errors. The second group of NNS production mainly has one category. That is, most tones are falling tones. There is few similarity of NNS production in this group to NS perception. The third group of NNS production mainly has one category too. That is, most tones are level tones. There is few similarity of NNS production in this group to NS perception. Therefore, NNS in the second and third groups produced much more errors than those in the first group.

It is obvious that NNS production of tones is based on contour, while NS produce tones based on both contour and register. The features of tonal production by NNS in the first two groups are relevant to the perception features of L1. As mentioned before, because falling intonation is a dominate stress marker and sentence intonation in English, it is reasonable that English-speaking NNS produce every tone as a falling tone. On the other hand, it is common and frequent to use rising intonation to ask questions in English. Therefore the English-speaking learners of Chinese can distinct rising tone and falling tone and further categorize them into two groups. Thus, it is also reasonable that NNS in the first group produce two categories of tones. However, as for the third group, most tones are produced as level tones by NNS. It seems that the default intonation in English has the acoustic feature of a falling tone, yet the perceptual feature of a level tone, so these tones are perceived as level tones by NNS.

Therefore, in the next chapter, we will further comparing L1 perception and production with L2, how NNS generate tones, and some pedagogical implications of the results.

CHAPTER 8 DISCUSSION

8.1 From the Classical Model to the Current Model

As we mentioned in Chapter 1, classical Chinese phonology used eight tonal categories, and modern Chinese linguists still use the names of the classical categories. However, the range and the number of modern tonal categories differ from those used in ancient China. There are four tonal categories based on the classical Chinese phonology. They are *Yinping*, *Yangping*, *Shangsheng*, and *Qusheng*, which are represented by T1, T2, T3 and T4 in this research. T1 is a level tone, T2 is a rising tone, T3 is a dipping tone, and T4 is a falling tone.

The current study lays the groundwork for a model of Mandarin tone based on both perception and production for NS. It shows the pitch range of productions in the acoustic space and the perceptual boundaries and tonal categories in NS' perceptual space. Both spaces are divided into separate tones, with separate boundary areas. The positions of the four tonal categories are the same in the perceptual space as in the production space. T1 is located at the upper right of the space, with the other tones arrayed counter-clockwise: T2 at the upper left, T3 at the bottom-left, and T4 locates on the bottom-right. Moreover, the space for T3 is narrower than for the other tones. Both the perceptual and production spaces have similar boundary areas. The boundary area between T1 and T4 shows the same feature: the higher the pitch value of the beginning point, the wider the boundary. The results of NS perception and production experiments also reveal that NS produces T3 and T4 similarly. Yet NS perceive them differently even if T3 and T4 share the same acoustic features. The reason is that the pitch value at the end of T3 and the pitch value at the beginning of the following tone play an important role in perceiving T3. These two pitch points allow listeners to perceive T3 as a dipping tone. The results of Experiment 3 support the hypothesis that the perceptual space of T3 overlaps with T4; and the pitch at the beginning of the following tone plays the crucial role in perceiving T3. That is, the lowest F0 is the significant invariant cue to T3. However, it is worth further studying the varying amplitude as well as F0 to see their impact on categories.

The current perceptual model describes the tones at both acoustic and perceptual levels. The similarities of NS perception and production reveal the acoustic cues for tone perception and production. These acoustic cues are phonetic features including register and contour. The difference between NS perception and production features, that is, three tone categories in the acoustic space and four categories in the perceptual space, shows the perceptual cue for perception. The perceptual cue is that the pitch at the end of T3 and the pitch at the beginning of the following tone let listeners perceive the rising part of T3, even though no real contour exists between these two points. Nonetheless, listeners perceive T3 as a dipping tone. Therefore, NS not only process the acoustic cues for perception and production, but also use the perceptual cue to perceive tonal categories, or produce tonal categories for others to perceive.

This chapter compares three perception models (SLM, NLM and PAM) to discuss how NNS come to use the acoustic cues and perceptual cues that NS do, and what causes difficulty for NNS in generating Mandarin tones. Finally, some pedagogical implications are discussed.

8.2 How NNS Generate Tones

Some studies of L2 sound acquisition indicate that adults' perception and production are strongly influenced by their native language (L1), while others do not. The current study supports the first view, that L1 does influence L2 acquisition. In the case of Mandarin tones, the phonological categories of tones do not exist in the native language, English, for American learners of Chinese. However, some of the phonetic features and phonetic categories of intonations do exit in English. These help NNS to organize new phonetic categories of tones in L2, even though their interlanguage models may differ from L1 phonological categories and NS phonological categories. In Chapter 2, We generalized three intonation contours for English intonation at the phonetic level. In this section, the Speech Learning Model (SLM) (Flege 1986, 1990, 1995), Native Language Magnet model (NLM) (Grieser and Kuhk, 1989, Kuhl 1991, 1992), and Perceptual Assimilation Model (PAM) (Best, 1994, 1995; Best et al. 1988) are used to discuss the difference between the tone perception and production of NNS and the influence of their L1.

8.2.1 Two Phonetic Features of Tones: Register and Contour

As we discussed in the first chapter, two phonetic features that are sometimes used as phonological features: register and contour. Register describes the pitch level of a tone, such as a low falling tone like T3 or a high level tone like T1. Contour means the shape of a tone, such as a level tone like T1, a rising tone like T2, and a falling tone like T4.

In section 8.1, we discussed the similarities between NS perception and production, which reveal crucial acoustic cues for both tone perception and production. These cues are the phonetic features including pitch register and pitch contour. NS rely on both register and contour.

For NNS perception, in contrast, no clear boundaries have developed, as the tones overlap in perceptual space. Most of them perceive the four tones as only two categories: T1 and T2 in the upper half of the space, and T3 and T4 in the bottom half of the space. The NNS perception appears to be determined by the height of the end points of the tones by contrast both the pitch values of the beginning point and the end point that determine the tonal categories of NS. Register plays the central role in perception, while contour is perceived very little. That is, high register contrasts to low register. When NNS perceive tones, there are no contrasts by reference to the contour feature. Yet, NS perceive tones depending on not only register but also contour which are decided by the pitch values of the beginning point and the end point.

For NNS production, tones also overlap in acoustic space. Some NNS can categorize the tones into two groups: a rising tone which includes T2 as contrasted with a falling tone which includes T1 T3 and T4. Some NNS cannot categorize tones during production. They mainly produce one category: either a falling tone derived from the default intonation in their L1, or a level tone. The NNS production appears to be determined by the contour of the tones by contrast both the contour and register that determine the tonal categories of NS. NNS tone productions depend more on tone contours. That is, the rising contour of T2 contrasts to the falling contour of T1, T3 and T4 in some NNS production, or there are no contrasts which results in a single tone category in other NNS production.

8.2.2 Three Previous Models: Accounting for the Contour Feature

The three models are used in this section, SLM, NLM and PAM. All are designed to apply to segments, that is, consonants and vowels. However, it is not clear whether these models fit suprasegmentals, for suprasegmentals have different phonological functions. In English, suprasegmentals are intonations which distinguish sentence types, emotions, etc. such as statements and questions. In Chinese, they are tones which create contrasting lexical meaning for syllables. Both intonation and tone share the same acoustic parameter – pitch, in addition to some phonetic features, such as pitch contour.

In section 8.2.1, we mentioned two phonetic features of tones – pitch contour and pitch register. Yet, English has no lexical contrast at the pitch register level, and the intonations differ at the contour level. Therefore, in this section, NNS perception and production are compared with the intonation patterns in their L1 according to the pitch contour.

8.2.2.1 Speech Learning Model

The Speech Learning Model (SLM) is to explain or analyze why NNS can or cannot acquire some contrastive sounds of L2. “Learners of an L2 may fail to discern the phonetic differences between pairs of sounds in the L2, or between L2 and L1 sounds, either because phonetically distinct sounds in the L2 are assimilated to a single category, because the L1 phonology filters out features (or properties) of L2 sounds that are important phonetically but not phonologically, or both. The model claims that without accurate perceptual ‘targets’ to guide the sensorimotor learning of L2 sounds, production of the L2 sounds will be inaccurate. …… The model does not claim, however, that all L2 production errors are perceptually motivated. …… A basic tenet of the model is that many L2 production errors have a perceptual basis”. (Flege, 1995:238)

Flege emphasized the relationship between perception and production. SLM has four postulates and seven hypotheses. Most of the hypotheses relate to L1 transfer. SLM claims that “Mechanisms and processes used in learning L1 sounds remain intact over the life span. …… Phonetic categories established for L1 evolve over the life span to reflect the properties of all L1/L2 sounds identified as a realization of each category” (Flege 1995: 236).

In the current study, most NNS produce two categories of tones: a rising tone which includes T2 and a falling tone which includes T1, T3 and T4. In NNS L1 (English), three primary intonations exist: a rising intonation for questions, a falling intonations which is used in most contexts of narrating or declaring a statement, and a level intonation that signals non-finality. Therefore, L1 intonation categories can help NNS to create tone categories during production. On the other hand, some NNS cannot categorize tones during production. Therefore, they mainly produce one category: either falling tones that reflect the default intonation for statements in their L1 or a level tone which signals non-finality in English. Recall that we introduced three intonation contours in Chapter 2. In this case, the phonetic categories, such as rising or falling intonations established for L1 do not evolve. They remain in unchanged in the L2 phonetic system. The only difference is that the phonological functions are different, such as rising and falling intonations in L1 and rising and falling tones in L2. However, the number of the tone categories for NNS production is less than that for NS production. It is obvious that NNS simplify the tone categories. They simply categorize T1, T3 and T4 into one group and T2 into the other group, or categorize four tones into a group. Tones are processed at the phonetic level, including phonetic features and phonetic categories here.

Overall, NNS production of tones is influenced by their L1's intonation contours. The results are consistent with other research (Chiang, 1979). Three phonetic categories: rising tone, falling tone and level tone appear in NNS L2 production. However, it is hard to use this model to explain the contrast of two phonetic categories which both NS and NNS have, just as Best pointed out (Best, 2001). The reason is that SLM focuses on individual phonetic categories. It compares individual phonetic categories of NNS with the equivalents in the sound system of NS. It does not compare two contrastive individual phonetic categories in the sound system of NNS. However, Perceptual Assimilation Model (PAM) compares two individual phonetic categories in the sound system of NNS, and further compares the pair with the phonetic categories in the sound system of NS. This will be illustrated in 8.2.2.3.

8.2.2.2 Native Language Magnet Model

The Native Language Magnet model (NLM) proposes that people develop acoustic prototypes for L1 phonetic categories. L1 prototypes have magnet like effects, which "shrink" the nearby perceptual space. L2 learners found that it is difficult to discriminate phonetic variation near prototypes than those near non-prototypes, or poor exemplars, of the same category. In the current study, three English prototypes are adapted to Chinese tones: rising intonation, falling intonation and level intonation. Therefore, in L2 production, the tone as we illustrated in 8.2.2.1 are closely derived from these three prototypes. However, again, it is hard to use this model to explain the contrast of two phonetic categories, because it focuses on individual tonal categories, and compares the individual tonal categories with the intonation prototypes in learners' L1.

On the other hand, the NLM does assume that speech perception involves general auditory mechanisms rather than specifically phonetic information. We will discuss this in the section about non-linguistic information below.

8.2.2.3 Perceptual Assimilation Model

According to the PAM (Best, 1995, 2001), "a given non-native phone may be perceptually assimilated to the native system of phonemes in one of three ways: (1) as a categorized exemplar of some native phoneme, for which its goodness of fit may range from excellent to poor; (2) as an uncategorized consonant or vowel that falls somewhere in between native phonemes (i.e. roughly similar to two or more phonemes); or (3) as a non-assimilable non-speech sound that bears no detectable similarity to any native phonemes. Adults' discrimination of a non-native contrast is predicted to depend on how each of the contrasting phones is assimilated. Several pair wise assimilation types are possible. The non-native phones may be phonetically similar to two different native phonemes and assimilate separately to them. This was termed Two Category assimilation (TC). Both may, instead, assimilate equally well or poorly to a single native phoneme, termed Single Category assimilation (SC). Or both might assimilate to a single native phoneme, but one may fit better than the other. This was termed a Category Goodness difference (CG). Alternatively, one non-native phone may be Uncategorized, as defined above, while the other is Categorized, forming an Uncategorized-Categorized pair (UG). Or both non-native phones might be Uncategorized speech segments (UU). Finally, the two phones' articulatory properties may both be quite discrepant from any native phonemes, and be perceived as Non-Assimilable (NA) non-speech sounds." (Best, 2001:777)

It is hard for us to explain the results by reference to the PAM because tones have multiple way contrasts even for a single tonal feature, such as the contour of T1 contrasts to not only the contour of T2 but also the contour of T4. The PAM mainly compares the segments in a pair.

In the current study, most NNS produce two tone categories: a rising tone which includes T2 and a falling tone which includes T1, T3 and T4. The two tone categories that these NNS produce in Mandarin are similar to English rising and falling intonation at the phonetic level.

In addition, some NNS cannot categorize tones during production. Therefore, they simply produce one category: a level tone or a falling tone. Neither one contrasts to other contours. In this case, the contrasts between two phonetic categories, such as rising or falling intonations established for L1 do not involve in L2 production. The intonation patters remain unchanged in the L2 phonetic system. The only difference is that only one category appears in L2 tonal system. This corresponds to a Single Category assimilation at the phonetic level.

However, at the phonological level, intonations and tones have different functions. Actually, NNS cannot use the phonetic category of falling tones to distinguish the meanings of the syllables that have the same segments but carry T1, T3 and T4 respectively, even though their production maps show two contrastive phonetic categories: rising tones and falling tones. In this sense, "it is temping to label tone contrasts as UU (uncategorized-uncategorized)" (Halle, etc. 2004:417).

8.2.2.4 Conclusion

Discussion of our results in the context of the three previous models illustrates how NNS produce tones are based on the intonation categories in L1. It shows that the three previous models can not only apply for segments, but also apply for suprasegmentals partially, with some limitations, such as tones. The reason that we can use these models to discuss the results is that the tone categories we get from the production data are based on the contours. The intonation categories in L1 help NNS to create similar tone categories in L2, as well as the contrastive tone categories at the phonetic level.

8.2.3 Non-linguistic Information: Accounting for the Register Feature

Recall that the NS perception map (Figure 32) has four quadrants. Remember that the bottom-left quadrant for T3 is an overlapping area which includes all tones. Therefore, strictly speaking, there are three discrete categories in the NS perception map. The T3 area is a mixed area in which perception of T3 overlaps with the other tones.

For NNS, recall that we found three patterns in the perception maps. The majority of the NNS (1,2,3 & 6) had a binary map with T1 and T2 in the upper part and T3 and T4 in the bottom area of the perception maps. One NNS also had a two-category perception map, but with T1, T2 and T3 on the left side and T4 on the right side. NNS5 perception map was the only one that did not reveal any pattern.

Therefore, the salient difference between NS perception and NNS perception is that NS perceive tones based on three discrete categories (T1, T2, T4) and one more ambiguous category (T3), but most of our NNS perceive only two categories of tones: T1 and T2 on the upper half of the space and T3 and T4 on the bottom half of the space. In this case the NNS perceptions appear to be determined by the end points of the tones, in contrast with the pitch values of the beginning and of the end points by NS. In this case, register plays an important role for NNS perception of tones. However, in their L1, the contrast at the register level does not exist in English intonation. How can NNS perceive tones according to the register feature? Since register does not cause phonetic/phonological contrasts in the intonation system, it is natural that we will look for additional non-linguistic information.

The NLM assumes that speech perception involves general auditory mechanisms rather than specifically phonetic features. Listeners fail to develop prototypes for non-native categories, because they lack some relevant acoustic experience. Best (2001) suggested that infants progress developmentally from detection of only nonlinguistic information in speech, to recognition of how phonetic variants fit into (or fail to) language-specific phonetic classes. Therefore, it is possible for NNS to use non-linguistic information to acquire new categories in L2. In this study, the focus will be register.

Register contrasts could help us distinguish gender. Generally, women's voices have a higher fundamental frequency, that is, are at higher register, while men's voices are at lower register. This indicates that listeners utilize register to distinguish non-linguistic contrasts. People also perceive and produce register contrasts in music. Register is not a contrastive feature in the English intonation system, but NNS can perceive two main categories in L2 according to the register. One of the possibilities is that NNS access register features through non-linguistic and acoustic experience, such as gender identity and music.

8.2.4 How NNS Generate Tones

In the previous sections, we discussed how NNS access phonetic features. They can either transfer intonation contours from their native language – English, or adapt non-linguistic information, such as NNS acoustic experience in distinguishing gender or musical tones.

NNS use different ways to create register contrasts and contour contrasts at the phonetic level. The phonetic categories they perceive or produce are based on either equivalents in their native language or on other non-linguistic categories. Therefore, the tonal categories that NNS perceive or produce are different from NS categories. On the other hand, for NNS, the phonological categories of intonation in L1 are different from the phonological categories of tones in L2. The phonological function for intonation is to distinguish mode at the sentence level; yet the phonological function for tones in L2 is to distinguish meaning at the syllable level. In the current study, the results of NNS production indicate that some NNS produce T4 which is a falling tone instead of T2 which is a rising tone, or produce T2 which is a rising tone instead of T4 which is a falling tone. This phenomenon reveals that NNS can produce the right phonetic categories of tones. But they cannot match the phonetic categories to phonological function. In this case, NNS cannot use the appropriate phonetic category to distinguish meanings.

One can posit three levels for generating tones, so that tones can be used for appropriate phonological functions.

The first level is to transfer the phonetic features to tones, that is, contour and register from NNS native language (L1) or other non-linguistic categories based on NNS acoustic experience. However, there is only one contrast in English intonation which is relevant to tonal contrast in Mandarin, that is, falling contour versus rising contour. In Mandarin, there is a multi-way contrast in tone, that is, falling contour, rising contour and level contour. All three intonation contours exist in English at the phonetic level, but only the falling contour and the rising contour are contrastive in distinguish intonation of questions and that of statements. Level contour signals non-finality in a sentence in English. Therefore, most NNS production reveals two-way contrast (falling versus rising), though some NNS produce a single tone category with only level or falling tones.

At the first level, the perceptual cue also participates in tone perception. NS utilize both acoustic features and perceptual cues to perceive tones. However, NNS mainly perceive tones by register. For NS, the perceptual cue comes from the pitch at the ending point of T3 and the pitch at the beginning point of the following tone, this transition creates a rising contour of T3. The virtual rising contour can be perceived. It could be interpreted as a contour feature. For NNS, the real contour feature does not work well in perception. It is extremely hard for NNS to use the perceptual contour to perceive tones. The results of NNS production data show that most errors relate to T3. This phenomenon documents difficulty which NNS have in using perceptual cues to perceive tones at the first level.

The second level coordinates the two features to organize different tone categories. This is hard for NNS. The results in this study indicate that NNS may perceive and produce different categories. The categories which are perceived by NNS are based on default acoustic experience, but their production categories are based on L1's linguistic experience. NNS can access the phonetic features of tones. However, they can only use one of the features to decide tone categories. Meanwhile the features, such as the contour feature, that have multi-way contrasts, such as rising contour versus falling contour, level contour versus rising contour and level contour versus falling contour, is different from the segmental features that generally have two-way contrasts, such as voice versus voiceless. Therefore, NNS intend to simplify the tone categories in Mandarin into one or two category according to the two-way contrast. That is, falling contour versus rising contour, in their L1 English. Or NNS simply categorize these tones into a big category, either falling contour tones or level contour tones.

The third level is to match phonetic categories to appropriate phonological categories. That is, different tone categories could distinguish the meanings for contrastive syllables. Since NNS can only categorize simplified tone category according to a single feature and the two-way contrast or non-contrast, it is hard for NNS to use these limited number of the tone categories to distinguish lexical meanings, that is, it is hard for NNS to match the phonetic categories in the NNS speech to the appropriate phonological categories that is relevant to the lexical meanings, with the same CV combinations. In addition, some NNS still make mistakes at this level, for example, T2 is mis-produced as T4. The main reason is that phonetic categories of pitch are intonation categories in learners' L1, yet the phonetic categories of pitch are tone categories in L2. Therefore, their phonological functions are totally different.

8.2.5 Can NNS acquire native-like tone categories?

The sections 8.2.2 and 8.2.3 analyzed how NNS perceive and produce tones depending on the phonetic features. Perception is mainly influenced by their acoustic experience and the L2 system, that is, Mandarin, since NNS perceive tones depending on the register feature that does not exist in English. NNS production is influenced by their L1 system, that is, English, since they perceive tones depending on the contour feature. The section 8.2.4 further concluded three levels that NNS generate tones based on the feature analysis. It seems that the most difficulty for learners to categorize tones is that they cannot attain the phonetic contrasts. Although they can access the phonetic features, they cannot set up the multi-way contrasts that are similar to NS. For example, T1, T3 and T4 are in the same category when NNS produce tones. However, it does not mean that they cannot access the contour or the register feature. The level contour and the falling contour exist in their L1, yet they are not contrastive. On the other hand, they experienced how to use the register feature to distinguish female sounds and male sounds. Yet, this contrast, such as the boundary area, is different from the contrast in the Mandarin tonal system. Recall that we concluded in the chapter of literature review (Chapter 2), and found that the greatest difficulty for learners of Chinese to acquire tones is that they cannot attain the correct tone categories. According to our results and analysis, we can further extend the discussion on this topic. Because learners do not have the contrastive elements based on the phonetic features, such as the level contour versus the rising contour, they cannot acquire the phonetic categories that match the linguistic functions completely. Thus the incomplete tone categories affect the perception and the production of the learners. Stagray and Down (1993) found that NNS are sensitive to the phonetic details, while Leather (1987) argued that NNS are less sensitive to small pitch differences. The results from the current study indicate that NNS do not focus on the phonetic details, yet they intend to categorize tones into the simplified categories which means the number of the categories is less than that in the L2 or the categories that are similar to their L1 because they have not set up all contrasts at the phonetic level.

The subjects in this study are second year Chinese students. If they learn more Chinese, and reach the advanced level, will they attain the same patterns as that of NS?

NS can identify words even if the tones are in the boundary area, because they do not need bottom up as much and can rely on top-down syntactic and semantic processing. NNS cannot control the top-down processing, so mostly depend on bottom-up processing, and given their perception maps they have significant trouble. We predict that they would need to get to certain proficiency with syntax and semantics before they would have enough left over resources to attend to processing tones.

However, top-down processing can only help some advanced learners to acquire tone categories similar to that of NS. In the earlier studies of categorical perception using voice onset time (VOT) for voiceless and voiced consonants, researchers (Eimas, Miller, and Jusczyk, 1987) found that infants are sensitive to the boundaries of different languages. However, such ability decreases after they acquire their native language. This experiment indicated that categorical perception is not innate, it can be learned. However, after a certain period, it might be hard for people to set up a second contrast after they have some contrasts in their early language acquisition. Flege (1995) showed that bilinguals have different boundaries of the contrastive sounds from NS, which is in the mid-point of the boundaries of L1 and L2. However, the sounds that they produce are still regarded as native-like sounds. The judgment of native-like sounds depends on NS perception. From the perspective of the acoustic features, there is a point (points) or a range (ranges) of an associated value (associated values) to divide the contrastive sounds into two or more segments (suprasegmentals), such as pitch values. In the case of Mandarin tones, learners could use top-down processing or other factors to acquire similar categories or perceive the same categories with the development of their language proficiency. However, they have to acquire the contrasts among the tone categories and attain same or very close values at the phonetic level or the phonological level, and then they can produce native-like tones. However, it is really hard for NNS to achieve this. Therefore, it is difficult for most learners to attain the same pattern as that of NS.

8.3 Pedagogical Implications

We have discussed three levels of acquiring tones for NNS. As we discussed above, at the phonetic feature level, some phonetic features do not have contrastive elements in NNS L1, therefore, it is hard for them to set up the multi-way contrast in the L2 tonal system. At the phonetic category level, it is hard for NNS to assign two phonetic features to different phonetic categories. They can only perceive or produce one phonetic feature to distinguish some tone categories. At the phonological function level, it is also hard for NNS to match the phonetic categories of tones to the phonological functions. That is, NNS sometimes cannot use tonal categories to distinguish syllable meanings.

8.3.1 Utilize Existing Pedagogy to Improve Tone Acquisition

In this section, we mainly discuss how one might improve NNS tone perception and production by using existing pedagogical techniques.

At the phonetic feature level, NNS can access the contour and register features from L1 and non-linguistic information. However, it is hard for NNS to acquire the multi-way contrast of contours. Therefore, it is important for Chinese teachers to help students acquire multi-way contrast of contours. We can develop production practice exercises that use minimal pairs to contrast the falling contour and the level contour, the falling contour versus the rising contour, and the rising contour versus the level contour. This may help learners of Chinese get sense of the multi-way contrast of tonal contours. To permit generalization, teachers could also provide different variations of syllables, including syllables with nasal consonants, multiple vowel finals, etc.

Moreover, teachers need to treat perception and production differently. According to our results, the tones which are perceived by NNS are based on default acoustic experience, and the tones which are produced by NNS are based on L1's linguistic experience. This means that NNS cannot perceive contour correctly, and cannot produce register correctly. Therefore, teachers should design different teaching methods to train/teach tone perception and production respectively. As for tone production, teachers can use musical register training, which could help students produce tone categories within every specific register range. As for perception of tones, the focus should be on contours. In order to teach learners to acquire the contrasts of contours, teachers could exaggerate their pronunciations when they demonstrate tones. At the same time, teachers could also use minimal pairs to help learners of Chinese practice tones, and further acquire the multi-way contrasts.

At the phonological function level, teachers need to strive to help students match the phonetic categories of tones to phonological functions. Teachers could provide some characters which have different meanings yet have the same consonants and vowels with different tones. This could help learners distinguish meanings based on the different tonal categories.

Overall, educators could provide different methods to focus on specific problems of tones which students have.

8.3.2 How to Develop New Techniques to be Adapted for Tonal Training

The instruments and method (5.2.2 and 5.3.2) used for the perception experiments in this study could be used to assess the tonal perception of learners. The results of the assessment may reveal which kind of contrasts that a learner cannot distinguish at the phonetic feature level, such as T3 versus T4 at the register level.

The instruments and method used for the NNS production experiment in this study could also be used to assess the tone production of learners. The results of the assessment may also reveal which kind of contrasts that a learner cannot distinguish at the phonetic feature level, such as T2 versus T4, T1 versus T2, or T2 versus T3 at the contour level.

According to the results and discussion in this study, further training of tones based on the above assessment should focus on three levels: phonetic feature, phonetic category and phonological function. Future tonal training software should provide feedback from these three aspects to learners, and report the progress that learners make from these three perspectives as well.

8.4 Conclusion

The current study lays the groundwork for a model of Mandarin tone based on both NS perception and production. It demonstrates that there is variability in NNS tone productions in the acoustic space and that there are differences in the perceptual boundaries and tonal categories in NS and NNS.

This study was conducted to answer four research questions.

1. What is the production space for native speakers?

The results of the NS production experiment indicate that T1, T2, T3 and T4 are produced in different parts of the space, although there are some overlapping areas. On the bottom-left side of the production space, there are some overlapping areas for T1, T3 and T4. The overlapping tone boundary area of the production space shows that the boundaries are located in the areas where either the beginning point or the ending point is near the value 0 (Table 8 in 4.3.1). Moreover, the range of tone productions does not occupy the full range of the acoustic space, and some peripheral areas do not contain any members of the tonal categories

The results are presented in two kinds of production space. The first one is for T1, T2, T3 and T4. The other one is for the neutral tone. The normal tones (T1, T2, T3 and T4) have some overlapping areas with the neutral tone in the production space.

Overall, there are three main clusters of tones in the native speakers' production map (Figure 21 in 4.3.2). The category of Tone 2 which is the rising tone exists in the upper-left cluster of the map, T1 that is the level tone is on the middle-right side of the map; both T3 and T4 which overlap considerably are in the bottom of the map, and the pitch value at the starting point of T3 or T4 is larger than that at the ending point which matches the nature of the falling tone. The clusters have a shared overlap that falls in the middle of the acoustic space. They are boundaries of the tones nearby.

2. Are tones perceived categorically by NS and NNS? If so, how can we describe them?

The data from experiments X and Y demonstrate that tones are perceived categorically by NS and non-categorically by NNS.

2-1. What are the ranges of the perceptual categories of T1, T2, T3, and T4 for NS?

The results of the NS perception of tones reveal a quadrant model. The tone perceptual space is divided into four separate tone regions which are separated by boundary areas. In the perception map for native speakers, there also exists three major single-tone areas. One is on the upper left side of the map, mainly containing T2; another one is on the upper right side of the map, and mainly contains T1; and another is on the bottom right side of the map, and mainly contains T4. T3 exists in the multiple tone area which is located on the bottom left corner of the perception map, and in the boundary area between T3 and T4. Boundary areas exist in the perceptual space. In the boundary areas, the tones are perceived as "obscure", that is, they can be perceived as at least two different tones. These boundary areas are located where the beginning point or the ending point is near the value 0 (Table 13 in 5.2.3.2.1).

2-2. What are the ranges of the perceptual categories of T1, T2, T3, and T4 for NNS?

The results of NNS perception data indicate that there are no clear boundaries and tones overlap to a great extent over the entire perceptual space. NNS appear to perceive tones based on only two categories: T1 and T2 in the upper half of the space and T3 and T4 in the bottom half of the space. It appears that register plays an important role during the perception of tones by NNS. The NNS perceptions appear to be determined by the end points of the tones by contrast with the tonal categories of NS which appear based on the pitch values of the beginning point and the end point.

2-3. What is the difference between NNS' tonal perception and NS' tonal perception?

The salient difference between NS and NNS perception is that NS perceive tones based on three discrete categories (T1, T2, T4) and one more ambiguous category (T3) while most of our NNS perceived tones based on two categories.

3. How do pitch values of tones create overlapping areas?

The results in the NS perception experiments indicate that T0 and T3 overlap with other tones, except in boundary areas.

3-1. Does pitch plays an important role in perceiving the neutral tone?

The results indicate that pitch plays an important role in perception of the neutral tone, and the neutral tone and normal tones can be altered based on duration, that is, neutral tones are shortened versions of T3 or T4. Therefore, the perceptual space of T0 appears to overlap with that of the normal tones.

3-2. Do the pitch at the end of T3 and pitch at the beginning of the following tone play an important role in perceiving T3?

The pitch at the end of T3 and the pitch at the beginning of the following tone play an important role in perceiving T3. These two pitch points let listeners perceive the rising part of T3, which makes T3 as a dipping tone.

4. What are the features of NNS tone production? What are the differences between NNS production features and NS perception features?

The analysis of the individual tonal productions yielded several results. All errors appear to be related to register. Few errors appear for T1 and low register tones are harder to produce than high register tones. More errors appear for T2 and T3, although T2 and T3 mistakes occur for different reasons. On the other hand, the results from individual non-native speakers' production are consistent with their individual perceptual spaces. The results of NNS production also show that their tone productions overlap in the acoustic space. The NNS production appears to be determined by the contour of the tones by contrast both the contour and register determine the tonal categories of NS.

It is hard for NS to perceive the tones that are produced by NNS in the boundary areas. . The results seen in NNS production maps indicate that NNS do not produce three or four categories that match NS perception. Most NNS production spaces have two categories: T2 and. T1, T3 and T4, or only one category: the falling tone or level tone.

REFERENCES

Abramson, A. S. (1979). The co-articulation of tones: An acoustic study of Thai. In T. L. Thongkum, V. Panupong, P. Kullavanijaya, & M. R. K. Tingsabadh (Eds.), *Studies in Thai and Mon-Khmer phonetics and phonology: In honour of Eugenie Henderson*. Bangkok: Chulalongkorn University Press. Pp. 1-16.

Baken, R. J. (1987). *Clinical Measurement of Speech and Voice*. Boston: College-Hill Press.

Bao, Z. (1990). *On the nature of tone.* Unpublished Ph.D. dissertation, MIT.

Best, C. T. (1994). Learning to perceive the sound pattern of English. In Advances in C. Rovee-Collier and L. P. Lipsitt edited *Infancy Research*. Ablex, Norwood, NJ.

Best, C. T. (1995). A direct realist perspective on cross-language phonological influences in infants: a perceptual assimilation model. In H. C. Nusbaum edited *The development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. MIT, Cambridge, MA.

Best, C. T., G. W. Mcroberts & E. Goodel. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of Acoustic Association of America*. 109 (2): 775-794.

Best, C. T., G.W. McRoberts & N. M. Sithole. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology. Human Perception and Performance.* 4, 45-60.

Blicher, D. L., R. L. Diehl, & L. B. Cohen. (1990). Effects of syllable duration on the perception of the Mandarin tone2/tone3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*. 18, 37-49.

Broselow, E., R. Hurtig, & C. Ringen. (1987). The perception of second language prosody. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system.* Cambridge, MA: Newbury House. Pp. 350-361.

Burnham, D. (2000). Excavations in language development: Cross-linguistic studies of consonants and tone perception. In D. Burnham, S. Luksanneeyanawin, C. Davis

& M. Lafourcade (Eds.), *Interdisciplinary approaches to language processing. The international conference on human and machine processing of language and Speech.* Bangkok: Chulalongkorn University Press. Pp. 44-69.

Burnham, D., & C. Jones. (2002). Categorical perception of lexical tone by tonal and non-tonal language speakers. Paper presented at the 9th International Conference on Speech Science and Technology. Melbourne, 2-5 December.

Burnham, D., & K. Mattock. (2008). The perception of tones and phones. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech searning. In honor of James Emil Flege.* Amsterdam: John Benjamins (Series: Language Learning and Language Teaching). Pp. 259-280.

Cao, J. (1989). Temporal distribution of the bisyllabic words in Standard Chinese: An evidence for relational invariance and variablility from natural speech, *RPR-IL(CASS),* pp.38-56.

Cao, J. (2002a). The duration patterns of syllables in standard Chinese. *现代语音研究与探索 (Research on Phonetics and Phonology of Chinese).* Beijing: 商务印书馆 (Shangwu Publishing). Pp. 289-298.

Cao, J. (2002b). 普通话轻声音节特性分析. (Features of weak syllables in Mandarin). *现代语音研究与探索 (Research on Phonetics and Phonology of Chinese).* Beijing: 商务印书馆 (Shangwu Publishing). Pp. 169-180.

Chan, S. W., C. K. Chuang, & S.-Y. W. Wang. (1975). Cross-linguistic study of categorical perception for lexical tone. *Journal of the Acoustical Society of America*. 58, S119.

Chiang, T. (1989). Some interferences of English intonations with Chinese tones. *International Review of Applied Linguistics in Language Teaching*. 17, 245-250.

Chao, Y. R. (1930). A system of 'tone letters'. *Le Maitre Phonetique*. 45, 24-27.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Berkeley: University of California Press.

Chen, M. (2000). *Tone sandhi: Patterns across the Chinese dialects*. Cambridge: Cambridge University Press.

Chen, Q.-H. (1997). Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers' Association*. 32(1), 21-39.

Chen, Y., & Y. Xu. (2006). Production of weak elements in speech-evidence from F0 patterns of neutral tone in standard Chinese. *Phonetica.* 63, 47-75.

Chiang, T. (1979). Some interferences of English intonations with Chinese tones. *International Review of Applied Linguistics*. 17(3):245-250.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*. 2, 225-252.

Connell, B., J. Hogan, & A. Rozsypal. (1983). Experimental evidence of the interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*. 11, 337-351.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice.* Amsterdam: John Benjamins.

Cruttenden, A. 1997. *Intonation.* England: Cambridge University Press.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*. 92, 81-104.

Cutler, A., J. Mehler, D. Norris, & J. Segui. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*. 24, 381-410.

Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages*. Unpublished Ph.D. dissertation, MIT.

Earle, M.A. (1975). *An Acoustic Phonetic Study of Northern Vietnamese Tones*. Speech Communications research Laboratory monograph No.11. Santa Barbara, California.

Eimas, P. D. (1963). The relationship between identification and discrimination along speech and nonspeech continua. *Language and Speech*. 6, 206-217.

Eimas, P.D., J. L. Miller, and P. W. Jusczyk.(1987). On infant speech perception and the acquisition of language. In Harnad, S. edited *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press. pp. 161-195.

Elliot, C. E. (1991). The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawai'i Working Papers in ESL.* 10(2), 177-204.

Flege, J. E. (1986). The production and perception of foreign language speech sounds. In H. Winitz edited *Human Communication and Its Disorders*. Ablex, Norwood, NJ. Vol. 2, pp. 224-401.

Flege, J. E. (1990). Perception and production: the relevance of phonetic input ot L2 language learning. In Ferguson and T. Heubner edited *Crosscurrents in Second language Acquisition and Linguistic theories*. John Benjamin, Philadelphia.

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience issues in cross-language research*. Baltimore: York Press. Pp. 233-272.

Fox, R. A., & J. Unkefer. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics.* 13, 69–90.

Fox, R., & Y.-Y. Qi. (1990). Context effects in the perception of lexical tone. *Journal of Chinese Linguistics*. 18, 261-283.

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

Gandour, J. (1983). Tone perception in far eastern languages. *Journal of Phonetics*. 11, 149-175.

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

Goldsmith, J. (1976). *Autosegmental phonology*. Unpublished Ph.D. dissertation, MIT.

Gottfried, T. & T. Suiter. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*. 25, 207-237.

Grieser, D. L., and P. K. Kuhl (1989). Categorization of speech by infants: support for speech-sound prototypes. *Developmental Psychology*. 25, 577-588.

Halle, P., Y. Chang, & C. Best. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*. 32, 395-421.

Hart, J. R. Collier & A. Cohen. (1990). *A Perceptual Study of Intonation: An Experimental Phonetic Approach to Speech Melody*. Cambridge University Press.

Howie, J.M. (1976). *Acoustic studies of Mandarin vowels and tones.* Cambridge: Cambridge University Press.

Hu, Y. (1990). *Xiandai Hanyu (Modern Chinese).* Shanghai: Shanghai Gaodeng Jiaoyu Chubanshe.

Jassem, W. (1971). Pitch and compass of speaking voice. *Journal of International Phonetic Association.* 12, 59-68.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica.* 20, 63-64.

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*. 50(2), 93-107.

Kuhl, P. K. (1992). Speech prototypes: studies on the nature, function, ontogeny and phylogeny of the centers of speech categories. In Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka *edited Speech Perception, Production and Linguistic Structure*. Ohmsha, Tokyo. Pp.239-264.

Ladd, D. R.; K. Silverman, F. Tolkmitt, G. Bergmann & K. Scherer. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America.* 78, 435-444.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics*. 11, 373-382.

Leather, J. (1987). The interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process*. Berlin: Mouton de Gruyter. Pp. 75-101.

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: Proceedings of the 1990 Amsterdam symposium on the acquisition of second language speech*. University of Amsterdam. Pp.72-97.

Lee, C., L. Tao & Z. Bond. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics*. 36(4): 537-563.

Leung, Y.-K. I. (2001). The initial state of L3A: full transfer and failed features? In X. Bonch-Bruevich, W. Crawford, J. Hellerman, C. Higgins and H. Nguyen (eds.), *The past, present and future of second language research: selected proceedings of the 2000 Second Language Rearch Forum* (pp.55-75). Somerville, MA: Cascadilla Press.

Li, H. (1995). Phonetic errors in Thai students learning Mandarin. *International Chinese Pedagogy*. 32, 66-71.

Liberman, A. M. (1975). *The Intonation System of English*. Ph.D. dissertation. Massachusetts Institute of Technology.

Liberman, A. M., K. S. Harris, H. S. Hoffman, & B. C. Griffith. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*. 54, 358-368.

Lin, M., & J. Yan. (1980). 北京话轻声的声学性质. (The acoustic features of the neutral tone in Beijing dialect.) *Fangyan (Chinese dialects)*. 3, 166-178.

Linguistic Data Consortium (2005). *TDT4 multilingual broadcast news speech corpus (electronic resource)*. Philadelphia: Linguistic Data Consortium.

Menn, L. & S. Boyce. (1982). Fundamental frequency and discourse structure. Language and Speech. 25, 341-383.

Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers' Association*. 24(3), 49-65.

Moore, C. B. (1995). *Speaker and rate normalization in the perception of lexical tone by Mandarin and English listens*. Unpublished Ph.D. dissertation, Cornell University.

Norman, J. (1987). *Chinese*. Cambridge: Cambridge University Press.

Qiao, S. 1956. 汉语的字调、停顿与语调的交互作用. (The interaction between tone, pause and intonation in Chinese). *Zhongguo Yuwen*. 10-13.

Read, C., Y. Zhang, H. Nie, & B. Ding. (1986). The ability of manipulate speech sounds depends on knowing alphabetic writing. *Cognition*. 24,31-44.

Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*. 16:135-72.

Rose, P. (1982). Acoustic characteristics of the Shanghai-Zhenhai syllable types. In David Bradley edited, *Papers in South-East Asian Linguistics. No 8: Tonation, Pacific Linguistics.* A-26, 1-53.

Rose, P. (1987). Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication*. 6, 343-351.

Rose, P. (1988). On the non-equivalence of fundamental frequency and pitch in tonal description. In D. Bradley, E. J.A. Henderson & M. Mazaudon edited, *Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg*. Pp. 55-82. Pacific Linguistics, C-104.

Rose, P. (1989). Phonetics and phonology of Yang tone phonation types in Zhen-hai. *CLAO.* 18(2), 229-245.

Rosen, S. & P. Howell. (1987). Auditory, articulatory and learning explanations of categorical perception of speech. In S. Harnad, editor, *Categorical Perception: the Groundwork of Cognition*. Pp. 113–160. Cambridge University Press, Cambridge, UK.

Schwanhauber, B., D. Burnham, & C. Jones. (2003). Categorical perception of tone by tonal and non-tonal language speakers. Paper presented at the 14th Australian Language and Speech Conference, Brisbane, Australia.

Sereno, J. A., & Y. Wang. (2008). Behavioral and cortical effects of learning a second language: The acquisition of tone. In O. Bhon & M. Munro (Eds.), *Language experience in second language speech learning: In honor of Manes Emil Flege*. Amsterdam: John Benjamins . Pp. 239-258.

Shen, X-N. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers' Association.* 24(3), 27-47.

Shen, X.-N. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics.* 8, 281-295.

Shen, X.-N. S., & M. Lin. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*. 34, 145-156.

Shen, X. –N.S. (1992). On tone sandhi and tonal coarticulation. *Acta Linguistica Hafniensia*. 24, 131-152.

Shih, C. (1987). *The phonetics of the Chinese tonal system*. Technical memorandum. Florham Park: AT&T Laboratories.

Shih, C. (1988). Tone and intonation in Mandarin. Ms., Conell University and AT&T Laboratories.

Shih, C. (1997).Mandarin third tone sandhi and prosodic structure. In J. Wang & N. Smith (Eds.), *Studies in Chinese phonology*. Berlin: Mouton de Gruyter. Pp. 81-123.

Soares, C. (1982). Converging evidence for left hemisphere language lateralization in bilinguals. *Neuropsychologia*. 20, 653-659.

Sorensen, J. and W. Cooper. (1980). Syntactic coding of fundamental frequency in speech production. *Perception and Production of Fluent Speech*. Edited by R. A. Cole. Lawrence Erlbaum, Hillsdale. P.399-440

Sproat, R. & C. Shih. (1992). Mandarin morphology still is not stratum-ordered. Unpublished manuscript. AT&T Bell Labs.

Stagray, J. R., & D. Downs. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*. 21(1), 143-163.

Studdert-Kennedy, M.. A. M. Liberman, K. S. Harris, and F. S. Cooper. (1970). Motor theory of speech perception: a reply to Lane's critical review. *Psychological review*, 77, pp.234-249.

Sun, S. H. (1997). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese.* Honolulu: University of Hawai'i Press.

Sussman, H. M., P. Franklin, & T. Simon (1982). Bilingual speech: Bilingual control? *Brain and Language*. 15, 125-142.

Takefuta, Y. (1975). Method of acoustic analysis of intonation. In S. Singh edited *Measurement Procedures in Speech Hearing and Language*. 363-378. Baltimore: University Park Press.

Wang, S.-Y. W. (1967). Phonological features of tone. *International Journal of American Linguistics.* 33, 93-105.

Wang, X., K. Hirose, J. Zhang & N. Minematsu. (2008). Tone recognition of continuous Mandarin speech based on tone nucleus model and Neural Network. *IEICE TRANS. INF. & SYST.* Vol.E91–D(6), 1748-1755.

Wang, Y., D. Behne, A. Jongman, & J. A. Sereno. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*. 25, 449-466.

Wang, Y., A. Jongman, & J. Sereno. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America.* 113 (2), 1033-1043.

Wang, Y., M. Spence, A. Jongman, & J. Sereno. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America.* 106 (6), 3649-3658.

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Wu, X. (2005). *(Non-)categorical perception of Mandarin tones: A comparison between speakers of tone and non-tone languages*. Unpublished M.A. thesis, Victoria: University of Victoria.

Wu, Z. (1982). Pitch changes in the connected speech of standard Mandarin. *Zhongguo Yuwen*. 439-450.

Wuillemin, D., & B. Richardson (1994). Right hemisphere involvement in processing later-learned languages in multilinguals. *Brain and Language*. 46, 620-636.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics.* 25, 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*. 27, 55–105.

Xu,Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics. Language and Linguistics*. 5, 757-797.

Yang, B., & R. Ankenmann. (2007). The gap between the perception and production of tones by American learners of Mandarin. Presented at American Association for Applied Linguistics annual conference.

Yip. M. (1989). Contour tones. *Phonology*. 6(1), 149-74.

Yip, M. (1993). Tonal register in East Asian languages. In V. Hulst& K. Snider (Eds.), *The phonology of tone: The representation of tonal register*. Berlin: Mouton de Gruyter. Pp. 245-268.

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

Zhang, J. & K. Hirose (2004). Tone nucleus modeling for Chinese lexical tone recognition. *Speech Communication*.42:447-466.

Zhang, X. (1956). 北京话里轻声的功用. (The function of the neutral tone in Beijing dialect). *Zhongguo Yuwen*. 30.

Zhu, X. (1995). *Shanghai Tonetics*. Dissertation. Australian National University.

APPENDIX A ANSWER SHEET FOR EXPERIMENT 2

Section 1

You will hear nine groups of sentences. In each group, there are nine sentences. Four optional meanings are provided for you to select, which are listed in the following. Select an appropriate meaning for each sentence you hear. Then fill the selected letter which represents the meaning of the sentence you hear in the following table.

a) 他要掏钱。

b) 他要逃钱。

c) 他要讨钱。

d) 他要套钱。

	1	2	3	4	5	6	7	8	9
Group 1									
Group 2									
Group 3									
Group 4									
Group 5									
Group 6									
Group 7									

Group 8									
Group 9									

Section 2

You will hear nine groups of sentences. In each group, there are nine sentences. Four optional meanings are provided for you to select, which are listed in the following. Select an appropriate meaning for each sentence you hear. Then fill the selected letter which represents the meaning of the sentence you hear in the following table.

a) 他要掏税。

b) 他要逃税。

c) 他要讨税。

d) 他要套税。

	1	2	3	4	5	6	7	8	9
Group 1									
Group 2									
Group 3									
Group 4									

Group 5
Group 6
Group 7
Group 8
Group 9

APPENDIX B PRODUCTION DATA COLLECTION INSTRUMENTS FOR EXPERIMENT 4

Topic: 男女平等

You are discussing the issue 男女平等 with your friend. S/he holds the idea that 女的比男的做的家务多; while you think that 男女平等. Provide evidence to support your opinions.

* Please speak on the topic of gender equality.

*****You has **only one minute** for your oral performance.

Sentence Patterns you should use in your dialogue:

1. 反过来
2. 同样 and 一样
3. 实际上
4. 就拿…来说
5. 甚至…

APPENDIX C NATIVE SPEAKER'S EVALUATION SHEET FOR EXPERIMENT 4

Please write down each character and the tone for each character. Then, transcribe the learner's tone for this character. Please use the symbol T1, T2, T3, T4 and T0. T1 represents Tone 1; T2 represents Tone 2; T3 represents Tone 3; T4 represents Tone 4; and T0 represents the neutral tone. If a tone could not be perceived as belonging to any categories of normal tones or could be perceived as belonging to two categories of normal tones, please code it as "ambiguous".

Character	Original Tone for the character	NNS Tone