PERCEPTION OF MANDARIN CHINESE TONE 2 / TONE 3 AND THE ROLE OF CREAKY VOICE

By

RUI CAO

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

UNIVERSITY OF FLORIDA

2012

To my Mother and late Father

ACKNOWLEDGMENTS

It has been a very long journey and there are many people I need to give my thanks to. First of all, I thank my advisor and the chair of my committee, Dr. Ratree Wayland, without whom this dissertation would not have been possible. I am thankful for her wisdom and guidance throughout my studies at the University of Florida. She has given me so many valuable insights and feedback on my research. I have grown professionally as well as personally following her as a mentor and role model. I would also like to thank my co-chair, Dr. Edith Kaan, and my other committee members, Dr. Wind Cowles, and Dr. Danling Fu. They have given me very helpful suggestions on my study and provided valuable feedbacks.

I thank Dr. Caroline Wiltshire, Dr. Diana Boxer, Dr. Eric Potsdam, Dr. Gary Miller, Dr. Virginia LoCastro, Dr. Theresa Antes, and Dr. Ann WehMeyer for teaching me all the wonderful subjects in linguistics. I am also thankful to the English Language Institute, the Department of Languages, Literatures, and Cultures at the University of Florida, and the Defense Language Institute in Monterey, California for giving me opportunities to teach English as a Second Language and Mandarin Chinese.

I will always be grateful to the professor who offered me an assistantship ten years ago when I graduated from college and made it possible for me to come to the United States for graduate studies—Dr. Douglas Coleman, from the University of Toledo (Ohio). He is also a gator himself and it was him who introduced me to the Linguistics at UF. I thank him for making my dream come true.

I am also obliged to many of my friends, especially Ming Ren, Weilin Zhang, Yan Yang, Jing Li, Louisa Chang, Donruethai Laphasradakul, Priyankoo Sarmah, Mutsuo Nakamura, Jimmy Huang, and Shengming Yang, for their friendship and moral support

all these years. Special thanks should go to Tom Ratican, who has treated me like a daughter of his own. I am thankful for his genuine care for me, and for providing valuable feedback in editing my writing.

I would like to express my greatest gratitude to my mother, Keru Pang, my husband, Taran Champagne, and other family members. Without their love, understanding, and encouragement, this work would not have been possible.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

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

PERCEPTION OF MANDARIN CHINESE TONE 2 / TONE 3 AND THE ROLE OF CREAKY VOICE

By

Rui Cao

August 2012

Chair: Ratree Wayland
Cochair: Edith Kaan
Major: Linguistics

Research has shown that lexical tones, a suprasegmental feature, are processed by native speakers as linguistic elements just like other segmental information. Among the four tones of Mandarin Chinese, in particular, Tone 2 and Tone 3 are very similar in their pitch contour shapes and thus can be difficult to distinguish in native and nonnative perception. A phonation type, creaky voice, has been reported to be associated with production of Tone 3. This study investigated the perception of Mandarin Tone 2 and Tone 3 and the role of creaky voice in the perception of Tone 3 among native and nonnative listeners. The results showed that creaky voice does not serve as the major cue in perception of Mandarin Tone 3. In fact, the presence of creaky voice seems to pyschoacoustically obscure the actual duration of the falling portion in the tonal contour, making it shorter in perception among both native and non-native listeners.

CHAPTER 1
INTRODUCTION

Research Questions

This study aims to investigate the perception of Mandarin Tone 2 and Tone 3 and the role of creaky voice in the perception of these tones among native and nonnative listeners.

In tone languages, lexical tones serve as “tonemic” distinctions at a suprasegmental level, an integral part attached to a syllable as an indispensable cue to determining a word’s meaning. Mandarin Chinese has four contrastive tones in its phonological inventory: Tone 1 (high level), Tone 2 (high rising), Tone 3 (low-dipping) and Tone 4 (high-falling) (Chao, 1948). For instance, the syllable *yi* has four different meanings depending on the tone; yi (high-level) ‘one’, yi (high rising) ‘aunt’, yi (low-dipping) ‘chair’, and yi (high-falling) ‘to recall; to remember’.

According to Abramson (1978), phonemic tones can be classified as static (or level) and dynamic (or contour) tones. In the three contour tones (Tone 2, Tone 3, and Tone 4) of Mandarin Chinese, Tone 2 and Tone 3 share a certain similarity in their physical fundamental frequency contours. Although Tone 2 is a high rising tone, there is a short initial falling in the tone contour before rising in its phonetic realization. This initial falling makes Tone 2, in reality, similar to a Tone 3 contour to a certain extent, a contour that also consists of a falling and rising pattern. The timing of the turning point, then, seems to be critical in perceiving these two tones. What are the perceptual category boundaries? What cues do native listeners use to mark the boundaries? Do non-native listeners perceive these two tones in the same way as natives?

Native speakers of tone languages usually use one or more acoustic and perceptual cues in hearing tones, such as the fundamental frequency (F_0) (Howie 1972; Gandour 1984), amplitude contour (Chuang et al., 1972), and sometimes duration (Chuang et al., 1972). The timing of the turning point plays a significant role in determining the shape of the tone contour (Jongman and Moore, 1997). Therefore, one of the focuses of the present study is the effect of timing of turning point on Tone 2 and Tone 3 perception.

The current study not only investigates the range of the turning point that alters people's perception from one tone to the other, but also explores a particular perceptual cue in Tone 2 and Tone 3 recognition—creaky voice. Creaky voice is characterized by reduced intensity in waveform, lower fundamental frequency, and less frequent pitch periods with irregular duration (Gordon & Ladefoged, 2001). This phonation type has been reported to be associated with the production of Mandarin Tone 3 (Belotel-Grenie, A. & Grenie, M., 1994). However, the role of creaky voice in Mandarin tone perception has not been studied extensively and it is worthwhile to examine it in finer detail, especially its role in distinguishing Tone 2 and Tone 3 if any.

The present study, therefore, explores the perceptual boundary between the two tones among native and non-native listeners, as well as the role of creaky voice in perception of these two tones. The research questions that are addressed in the study are:

Research Question 1: Is creaky voice associated with a certain tone(s) produced by native Mandarin Chinese speakers (NC)?

Research Question 2: How does the presence of creaky voice affect the perception of tones by native Chinese speakers (NC) and native English speaking L2 learners of Chinese (NE), respectively?

Research Question 3: What is the perceptual boundary of the Tone 2-Tone 3 continuum perceived by native Chinese speakers (NC)? How does it differ from native English speaking L2 learners of Chinese (NE) with different proficiencies? What is the role of creaky voice in their perception?

Research Design

The overall design of this study involved testing two groups of listeners, native Mandarin Chinese (NC) speakers and native (American) English (NE) speakers who were second language learners of Mandarin Chinese. The NE speakers were further divided into two subgroups based on their ability to identify Mandarin tones in a baseline tone identification task. The main task for both NC and NE listeners was a tone categorization task, designed to examine the perception of Tone 2 and Tone 3, as well as the role of creaky voice in the identification of Mandarin Tone 3. The frequency of occurrence of creaky voice in the production of Mandarin tones among NC speakers was also examined. Since Belotel-Grenie & Grenie (1994)'s study did not provide a conclusive picture of creaky voice with production of Tone 3, it is necessary to provide more evidence for this matter.

Since one of the purposes of the perception experiment was to examine the role of creaky voice in the categorical perception of Tone 2 and Tone 3, the two sets of stimuli used in the experiment were consistent in all parameters except for the creaky part. Two sets of 34 stimuli were generated. The contour of the creaky tokens was the same as for the non-creaky stimuli, but with creakiness created at the dip of each stimulus. The prominent characteristics of creaky voice include irregular periodicity and a sudden decrease in fundamental frequency and intensity. The creakiness was generated by randomly assigning low pitch points at the dip to create a "jitter". Each of the participants did a baseline task and a perceptual task, while native Chinese speakers also

performed a production task. In the perception experiment, the tone tokens were presented by computer and the participants needed to make judgments regarding which tones they heard. The purpose of the production task was to decide whether there is an association between creaky voice and a certain tone in Mandarin Chinese.

Outline

This dissertation is organized in the following manner. In Chapter 2, several cross language speech models are reviewed, as well as literature on Tone languages, on Mandarin Chinese tones, and phonation types. The detailed methods of the study are described in Chapter 3. Results from production data and perception experiments are presented in Chapter 4, followed by a discussion of findings, suggestions for future research, and concluding remarks in Chapter 5.

CHAPTER 2
BACKGROUND

In this chapter we review the literature relevant to the present study. First, several cross-language speech models are discussed, since the research reported herein includes comparisons between native and non-native Chinese speakers. Perception in tone languages is then reviewed, particularly in Mandarin Chinese, (including native as well as non-native perspectives). Following this we review phonation type including how creaky voice is described in previous research and how it is associated with Mandarin Chinese. Finally, a summary of the rationale behind the present study is given.

Cross-Language Speech Perception

In the middle of the 19th century, when the treaty of Nanjing was signed, ending the first Opium War between the British and Qing Empires, the five ports of China were opened for trade. Shortly afterwards, the first English textbook in China, *Devil's talk*, appeared in one of the port cities, Guangdong (Canton). The book collected some common English words and phrases with pronunciation written in Chinese characters. For example, the entry "How do you do?" was marked by the Chinese words "好肚油肚", which reads similarly to "haw du yo du", and literally means "good belly oil belly".

As amusing as it is, this anecdote has indeed, to some extent, depicted how people have made an effort to hear and speak foreign words. The difficulties in mastering a nonnative language have always been of interest to people and in recent decades researchers have begun to gain a better understanding of the basic perceptual mechanism of speech sounds. In recent years particularly, with advanced technology that has helped research, exploration of the neurological mechanism of speech perception has also emerged as a new area to complement previous behavioral studies.

Though our understanding of the neural mechanism is still limited, there are some influential models proposed in phonetics research to predict or account for patterns of cross-language speech perception, mostly focusing on segmental features. The three dominant models are reviewed in this section: 1) the Speech Learning Model (**SLM**) by James Flege, 2) the Perceptual Assimilation Model (**PAM**) by Catherine Best, and 3) the Native Language Magnet model (**NLM**) by Patricia Kuhl. In addition, these models will be extended to account for the acquisition of suprasegmental features—lexical tones by non-native speakers, on the basis of which, predictions on phonological acquisition of Mandarin Chinese tones by native English speakers will be made.

Speech Learning Model (SLM). Flege (1995, 2002) proposed the speech learning model in order to account for how individuals learn or fail to learn how to perceive and produce L2 consonants and vowels in a native-like fashion. The major questions that SLM focuses on are whether or not certain L2 sounds are learnable, whether they are learnable only by children, and some issues in the perception of speech sounds. This model was developed with the background of the Critical Period Hypothesis (CPH) in the early 1980s, which claims “the earlier, the better”, and the Contrastive Analysis (CA) hypothesis, which claims that L2 phonemes that are similar to L1 are easier than those that are different.

While Best’s (1995) PAM is focused on naïve learners of a foreign language, the target subjects described in SLM are bilinguals who have had some experience in speaking their second languages. Derived from four postulates, Flege’s (1995) SLM had seven hypotheses on second language sound acquisition.

The model starts with a controversial postulate that the mechanism establishing the L1 sound system remains intact across the life span and can be accessible for L2 learning. Later Flege (2002) pointed out that this notion does not mean adult learners will ultimately achieve the same proficiency as child learners because other factors will interfere with the development of long-term memory representations of an L2 segment that are identical to a monolingual's L1. Phonetic categories are also defined in the postulates as the specified language-specific aspects of speech sounds in long-term memory representations. The other two postulates state that L1 phonetic categories evolve over a life span, reflecting the L1 or L2 phones' properties, and that bilinguals maintain a contrast between L1 and L2 phonetic categories in a so-called "common phonological space". The two languages thus eventually influence each other.

SLM (Flege 1995, 2002) proposes some important hypotheses to account for cross-language speech perception and production, including:

- Sounds in L1 and L2 are perceptually related to each other at an allophonic level rather than at an abstract phonemic level;
- If an L2 sound is perceived as very phonetically different from the closest L1 sound, it is very likely that a new category will be established for the L2 sound;
- The likelihood for (2) to happen increases when the perceived dissimilarity increases;
- When the age of learning an L2 increases, the likelihood of discerning L1 and L2 sounds, as well as L2 sounds that are noncontrastive in the L1, will be decreased;
- SLM proposes the concept of equivalence classification, which may eventually prevent creating new categories for certain L2 segments. This means that when equivalence classification happens, a single phonetic category will be formed to process perceptually linked L1 and L2 sounds (diaphones), which in production will resemble either L1 or L2 sounds.
- A bilingual may establish a phonetic category that differs from a monolingual's if the bilingual's category falls in the common L1-L2 phonological space, or if the

bilingual's representation is based on features that are different from a monolingual's.

Production of a sound will ultimately correspond to its property presentation established in its phonetic category. This implies that without an accurate perceptual input, the production of an L2 sound will be inaccurate.

There are two specific mechanisms that SLM (Flege 1995) proposes through these hypotheses by which the phonetic categories that represent L1 and L2 phonetic systems interact: *category assimilation* and *category dissimilation*.

When a category is not formed for an L2 sound due to its great similarity to an L1 sound, then the long-term memory representation that is used for an L1 phonetic category and a similar L2 phonetic category will assimilate, gradually leading to a "merged" L1-L2 category. Take Flege's (1987) for example, in which he examined the production of /t/ in French and English words by English-French and French-English bilinguals both of whom were late L2 learners. The segment /t/ in English is usually realized with long-lag VOT values, and French /t/, on the other hand, is usually realized with short-lag VOT values. The English and French monolinguals' production data confirmed this. Interestingly, French-English bilinguals produced English /t/ with longer VOT than French monolinguals' production of French /t/; however, the VOT of their English /t/ was not as long as that of English monolinguals. In contrast, the English-French bilinguals produced French /t/ with shorter VOT than English monolinguals' production of English /t/; however, the VOT of their French /t/ was longer than that of French monolinguals. Thus, the study showed that neither of the L2 groups established a new phonetic category for the L2 /t/; instead, the two sources of input information for

/t/ from both L1 and L2 seemed to have caused them to merge their /t/ to somewhere between L1 and L2, reflecting both features of English /t/ and French /t/.

The other mechanism, category dissimilation, happens when a new category is established for an L2 sound. It may differ from neighboring L1 and/or L2 sounds that are produced by a monolingual. Flege & Eefting (1987), for example, examined the production of /p t k/ in Spanish and English words by Spanish-English bilinguals. Those segments are usually realized with short-lag VOT in Spanish and long-lag VOT in English. The results showed that both adult and child bilinguals produce Spanish /p t k/ with much shorter VOT than Spanish monolinguals. This is consistent with the hypothesis that the bilinguals seemed to modify their L1 /p t k/ to make them more distinct phonetically from those categories that they established in English.

Perceptual Assimilation Model (PAM). Another influential model in cross-language speech perception is Best's (1995) Perceptual Assimilation Model (PAM). This model focuses primarily on the perception of a non-native language by naïve learners who have not had any learning experience with the language. The fundamental premise of the model is that nonnative segments tend to be perceived "according to their similarities to, and discrepancies from, the native segmental constellations that are in closest proximity to them in native space. (Best 1995, p. 193)"

According to PAM, three patterns of perceptual assimilation of non-native segments will happen:

- Assimilation to native category. This means a non-native segment could be heard by a naïve listener as either a good exemplar of his/her native category, an acceptable but not perfect exemplar of the native category, or a notably deviant exemplar.

- Assimilation as an uncategorizable speech sound. This means that if a non-native sound cannot be categorized/ assimilated to any native categories, but still is heard as a speechlike sound, then it is likely to be assimilated within native phonological space.
- Not assimilated to speech. This happens when a non-native sound cannot be assimilated to native phonological space and is thus perceived as a nonspeech sound.

Although PAM does not directly address issues in cross-language speech learning, it can be extended to apply to the second language learning process. In order to predict how naïve listeners discriminate nonnative phonological contrasts, PAM also proposed patterns for how each phone in a contrasting nonnative pair is perceptually assimilated. For instance, very good discrimination is predicted for Two Category (TC) assimilation; poor discrimination is predicted for Single Category (SC); etc. PAM takes into account both phonological and phonetic levels when explaining how the L1 system influences the perception of nonnative sounds (Best & Tyler, 2007). One notion in which PAM differs from that of SLM is that listeners extract information about articulatory gestures from the nonnative speech whereas the learners described in SLM try to form phonetic categories based on the acoustic cues they hear. In a manner similar to SLM, PAM also agrees that listeners continue to refine their perception of speech gestures over their lifetime.

Native Language Magnet model (NLM). Kuhl and her colleagues (1992, 1994) developed the Native Language Magnet model (NLM) with the goal of characterizing the developmental changes reflecting how infants reorganize their phonetic perception during first language acquisition. The basic claim of NLM is that infants have the ability to discriminate any contrast in speech sounds in the world even if they have not heard them before. By the end of their first year, this ability is gradually lost and their brains

are more tuned to only the native sounds to which they are exposed. In their experiments, both American adults and 6-month-old infants exhibited a perceptual magnet effect, such that if two tokens were within one category, their physical phonetic distance would be perceptually shrunk or ignored by listeners, and if the two tokens were perceived as in a different category, their physical phonetic distance would be perceptually stretched.

The implications of NLM can be extended to both first language acquisition and adults learning a second language. For an infant, the shrinking and stretching of perceptual space can actually facilitate L1 acquisition since the irrelevant information will be left out in perceiving phones categorically. For adults learning a second language, NLM claims that being exposed to one language "has distorted the underlying perceptual space by reducing sensitivity near phonetic prototypes, and these perceptual effects can be difficult to alter (Kuhl & Iverson 1994, p561)." In particular, if an L2 sound is close to a native language prototype, NLM predicts that it will be very difficult for L2 learners to perceive the phonetic contrast in the second language.

How, then, can these models be extended to account for the acquisition of lexical tones by nonnative speakers and what predictions can we make? Those models deal mainly with the segmental features of speech sound perception, but variations in discriminating nonnative speech go beyond vowels and consonants in many languages, for example, the tonal contrast in tone languages.

The phonetic feature "pitch" is very different in a tone language like Mandarin Chinese, and an intonation language like English. Mandarin Chinese uses pitch information to discriminate lexical meanings, in addition to vowels and consonants, the

three of which are all integrally perceived by native speakers. On the other hand, in an intonation language, pitch is used to signal syntactic or emotional information at the phrase or sentence level. Therefore, when a non-native speaker starts to learn a tone language, great difficulty can be predicted simply because in their native phonological system, tone as an integral cue for identifying a speech sound does not exist.

Both SLM and PAM touch upon the relationship between the difficulty of perceiving an L2 sound contrast and the similarity/difference between L1 and L2 sounds.

SLM claims that the more an L2 sound is perceived differently from the closest L1, the more likely a new category for the L2 sound will be established. However, in the tone language case, this cannot be applied, because the phonetic distance here is not at the same dimension since tones are used by non-tonal speakers for pragmatic purposes, such as asking a question or expressing emotions. On the other hand, if the new cue is given much attention and can be successfully separated as another dimension of perception by an L2 learner, then the SLM model would predict that it will be more likely for the L2 learner to produce the L2 sound in an accurate way. Moreover, both SLM and PAM agree that throughout the life span, learners will continue to refine their perception of speech gestures. In fact, studies have shown that even for naïve speakers of tone languages, the perception of tones is not totally non-existent (Halle, Chang & Best, 2004).

The NLM effect predicts that if an L2 sound falls within the psychoacoustic space of an L1 sound, the L2 sound will be difficult to discriminate (Wayland, 2007). If this model is to be extended to tone acquisition, then L2 learners will face great difficulties in

extracting pitch contour information instead of only paying attention to the averaged pitch. For instance, for Tone 1 in Mandarin Chinese, the phonological category does not change when the averaged pitch height changes, which might be easier for native English speakers, since there is no nonnative contrast in perceiving Tone 1. However, for Tone 2 and Tone 3, which are acoustically very similar and usually only differ in timing of the turning point, then English speakers have to learn to be more sensitive to the pitch contour instead of the absolute pitch value. In this latter case, the nonnative contrast falls outside the phonological space of an English speaker's native system, thus it will be especially difficult to learn.

In conclusion, the three influential speech perception models have been reviewed in terms of their major claims and hypotheses. They address the issues in speech perception of segments from slightly different points of view and all have implications in cross language speech learning. In the next few sections, background on perception of tone languages, particularly Mandarin Chinese tones, and creaky voice is reviewed.

Tone Languages

According to Yip (2002), in approximately 60-70% of the world's languages, tones determine meaning at a lexical level. Tone languages can be found in Asia, West Africa, and Europe, and they are estimated to be spoken by more than half of the world's population (Fromkin, 1978). Pitch variation is used in all the world's languages to signify meaning, emotion, or intent at a sentential level (Burnham & Mattock, 2007). For instance, a rising intonation of a sentence conveys a question, as in the English sentence "Do you speak Chinese?"; or an exaggerated intonation can be used in baby talk. These are different from using pitch variation or tone at a lexical level to distinguish lexical meaning. For example, in Thai, there are five phonologically distinctive tones;

therefore, for the same syllable /k^{h}a:/ there will be different meanings depending on the tone, e.g. /k^{h}a:/ (mid) ‘to be stuck’, /k^{h}a:/ (low) ‘galangal, a rhizome’, /k^{h}a:/ (falling) ‘to kill’, /k^{h}a:/ (high) ‘to engage in trade’, /k^{h}a:/ (rising) ‘leg’.

Yip (2002) gives a clear explanation of three important terms: fundamental frequency (F_0), pitch, and tone. According to Yip (2002), F_0 is a pure phonetic term which refers to the signal and how many pulses per second the signal contains, expressed in Hertz (Hz). Each pulse, in the case of speech, is produced by a single vibration of the vocal folds. Pitch, on the other hand, is more a perceptual term, meaning what a hearer’s perception is of a signal. Very often, pitch and fundamental frequency are used interchangeably; however, there is no one-to-one correspondence between these two (Jongman et al., 2007). Pitch can also consist of non-speech signals, for example, in music that varies in pitch. The third term, tone, is a linguistic one, referring to “a phonological category that distinguishes two words or utterances (Yip, 2002).”

Tone Production and Perception

How is tone produced? As previously mentioned, the fundamental frequency of a sound, which is perceived as pitch, is a function of the rate of vocal fold vibration (Ohala, 1978). According to the description in Jongman et al. (2007), tone production is a process that “changes in fundamental frequency (or in rate of vocal fold vibration) are made by manipulating tension in the vocal folds. “

In order to perceive a tone, a hearer must depend in whole or in part on pitch, and hence on fundamental frequency (Yip, 2002). The signal must contain a F_0 that fluctuates, which also needs to be large or prominent enough to be detected in order to have pitch differences. Other factors include duration and amplitude.

Before tone perception, an important issue needs to be discussed, pitch detection. Klatt (1973)'s study reported that the minimal detectable differences for sounds with a level F_0 was about 0.3Hz (also known as "just noticeable differences, JNDs"). It went up to 2Hz if the sound had F_0 with ramp or slope. Of course in human languages, tones are much further apart in tonal space (Yip, 2002). It was also shown in Klatt's (1973) study that pitch is easier to discriminate on a steady-state vowel than a non-steady vowel, as in a diphthong. Therefore, the duration is important for pitch detection as well, especially for contours. According to Greenberg & Zee (1979), if the syllable is less than 40-65ms long, contours cannot be perceived. Instead, they are mostly perceived as level. According to them, a real robust perception of "contouricity" needs to be about 130ms for the entire signal duration, though this is longer than some stop-final syllables in many Chinese dialects (Yip 2002).

If F_0 is important for pitch detection, it is, therefore, reasonable to assume that it is also the primary cue for tone perception. This is true in some languages; however, Yip (2002) points out that in some other tone languages, tones differ not only in F_0, but also in duration, amplitude, and voice quality or phonation type.

According to Gandour (1978), for many tone languages, including Thai, Mandarin Chinese, Yoruba, and Swedish, F_0 is an indispensable cue for tone recognition. Some studies (e.g. Fork 1974, Abramson 1978) have shown that when other cues are removed from the signal except for the fundamental frequency, native speakers can still discriminate tones with a high degree of accuracy. In another study (Cao & Sarmah, 2007) investigating Mandarin Chinese Tone 2 and Tone 3, native speakers did not recognize the synthesized stimuli with only F_0 the same way as when all other cues

were present. On the other hand, if the original fundamental frequencies of stimuli were removed and only cues like duration and amplitude remained intact, tonal discrimination was greatly impaired (e.g. Whalen & Xu, 1992; Fu & Zeng, 2000).

Although fundamental frequency is very important in perception in tone languages, there are slight differences in the weighing of this cue in different cases. It has been shown that Thai listeners can easily identify the tones in monosyllabic words (Abramson, 1962). Though Thai has three level tones, high, mid, and low, the confusion of mid and low tones results in only a very small number of errors, which are eventually eliminated when the stimuli are produced by one single speaker instead of ten (Abramson, 1976). Testing fundamental frequency alone can be a cue for identifying tones. When Abramson (1962), for example, superimposed synthetic averaged F_0 contours on each of the natural speech monosyllabic words: *naa* 'field', naa (a nick name), naa 'face', naa 'aunt', and naa 'thick', the identification rate by native speakers was nearly perfect. The study suggested that fundamental frequency can override other cues like duration or amplitude that may be associated with a tone. However, the addition of amplitude can enhance the perception of tones, although this addition by itself is not sufficient enough for tone identification (Abramson, 1975).

In the case of Yoruba, a Kwa language spoken in Nigeria, there are also three contrastive level tones: high, mid, and low. There is a phenomenon in this language that on disyllabic words with high-high, mid-mid, and low-low patterns, only the final low tone is markedly lower than its preceding tone. Studies (Hombert, 1976, for example) have investigated the relative importance of fundamental frequency, amplitude, and duration in distinguishing a low tone from a mid tone in word final position. The results from that

study indicated that if the duration or amplitude increases, it does not cause a shift in identification, which suggests that fundamental frequency is the principal acoustic correlate of Yoruba tones, except that the falling contour is the primary perceptual cue of a low tone in word-final positions.

In Mandarin Chinese there are other parameters to be used as perceptual cues besides fundamental frequency, such as intensity and duration. Perception and production of Mandarin Chinese is reviewed in the next section.

Mandarin Chinese

Native Perception and Production of Chinese

Mandarin Chinese has four contrastive tones in its phonological inventory: Tone 1 (high level), Tone 2 (high rising), Tone 3 (low-dipping) and Tone 4 (high-falling) (Chao, 1948). For instance, the syllable *yi* has four different meanings depending on the tone; yi (high-level) ‘one’, yi (high rising) ‘aunt’, yi (low-dipping) ‘chair’, and yi (high-falling) ‘to recall; to remember’.

Fundamental frequency contours have been examined in many studies to show their primacy in the perception cues of Mandarin Chinese. For example, in one of Howie (1972)’s perception experiments, synthetic tones were imposed on the real-speech Tone 1 syllable *bao*, producing four different tone stimuli. Native speakers demonstrated 95% accuracy in identification tasks. In his other study, a set of stimuli was generated to form tokens that were minimally differentiated by tones taken from real-speech citation syllables and their fundamental frequencies were suppressed and replaced by a constant fundamental frequency of 128 Hz. It turned out that native speakers did very poorly in tone identification. This suggests the primacy of fundamental frequency patterns in Mandarin Chinese tone identification.

Gandour (1984) conducted a study to examine whether F_0 height or F_0 contour is more important as a perception cue. The results suggest that both of them are important, though native listeners seem to weigh F_0 contour slightly more importantly than F_0 height. In another study Massaro, Tseng & Cohen (1985) examined Tone 1 and Tone 2, and showed that both cues were used by native listeners, while neither F_0 height nor F_0 contour alone were sufficient enough for accurate identification.

In the contour tones of Mandarin Chinese, the confusion between Tone 2 and Tone 3 has long been reported and studied. In physical fundamental frequency contours, these two tones share a great similarity. What, then, are the acoustic properties that can distinguish these two tones? Shen & Lin (1991) have reported that besides the F_0 height, which contributes to distinguishing these two tones, there are two other parameters that are relevant. These include: (i) the timing of the turning point, defined as the duration from the onset of the tone to the point of change in F_0 direction, and (ii) the decrease in F_0 from the onset of the tone to the turning point, which they call the ΔF_0. Shen & Lin (1991) found that Tone 2 usually has an earlier turning point and smaller ΔF_0 than Tone 3. When ΔF_0 is set at 30 Hz, native speakers of Mandarin Chinese perceive Tone 3 when the turning point is more than 40% of the total length of the stimuli. When ΔF_0 is 15 Hz, Tone 3 is perceived when the turning point is more than 60-70% of the total length of the stimuli. However, Shen and Lin (1991) were not specific about the geographical affiliation of the Mandarin Chinese speakers who participated in their study. Marked idiosyncrasies among Mandarin Chinese speakers in producing Tone 3 can be safely considered as regional features; e.g. Yip (2002) reports

that the Tianjin variety of Mandarin Chinese differs significantly from the Beijing variety in terms of the production of Tone 3.

Other perception studies have also shown that these two parameters are very important cues to distinguishing Tone 2 and Tone 3. Moore & Jongman (1997) reported the simultaneous effect of timing of the turning point and ΔF_0 in Mandarin Chinese speakers' perception of Tone 2 and Tone 3. In one of their experiments, they manipulated the timing of the turning point and ΔF_0 systematically in isolated synthetic speech stimuli. They first created 12 stimuli having turning points at various times from 20 to 240 ms in 20 ms steps, then each stimulus was varied in terms of ΔF_0 ranging from 10 to 70 Hz in 5 Hz steps. All 156 stimuli (12 X 13) were presented to native speakers of Mandarin Chinese in a forced choice perception test where lexical entries with Tone 2 ('not' 无) and Tone 3 ('dance' 舞) were the choices. The results showed that when the turning point is earlier than 240ms and ΔF_0 is below 30 Hz, more Tone 2 is perceived. When the turning point is more than 200ms and ΔF_0 is larger than 35Hz, subjects report Tone 3 responses. They concluded that ΔF_0 becomes crucial in the perception of Tone 3 when the turning point is late; however, in that case ΔF_0 needs to be more than 35 Hz.

Cao & Sarmah (2007)'s study also investigated the turning point in perceiving Mandarin tones. The results showed that when the timing of the turning point is between 42.5% and 72.5% of the total length of the stimuli, the stimuli will be associated with the lexical meaning 'horse' (Tone 3, 马). However, if the timing of the turning point is less than 42.5% of the total duration of the stimuli, the stimuli will be associated with the

lexical meaning ‘hemp’ (Tone 2, 麻). If the turning point occurs after 72.5% of the total length of the stimuli, the stimuli will be associated with the meaning ‘scold’ (Tone 4, 骂).

In Chuang et al. (1972)’s study, most of the identification errors resulted from the confusion between Tone 2 and Tone 3, but the pattern of confusion is asymmetrical. There were many more errors resulting from misidentifying Tone 3 as Tone 2, than from misidentifying Tone 2 as Tone 3. If this cannot be easily explained by the cues discussed above, it could, perhaps, be related to a phonological rule that Mandarin Chinese has, known as tone sandhi (Chao, 1948)—a Tone 3 becomes Tone 2 when it precedes a Tone 3. For instance, the two compounds, *fen-chang* (Tone 2-Tone 3) ‘graveyard’ and fen-*chang* (Tone 3-Tone 3) ‘flour factory’ are homophonous in connected speech.

Besides the primary cue, F_0, in Mandarin Chinese tone perception, there are other cues, such as duration and amplitude. From production data, it has been shown that Mandarin tones also differ in their overall duration (e.g., Chuang et al., 1972). Tone 2 and Tone 3 seem to be longer than the other two tones, and Tone 4 is the shortest, although this may change if they are in different positions and serve different functions in a sentence. Blicher et al. (1990) reported that lengthening the duration of stimuli which are ambiguous between Tone 2 and Tone 3 will make the identification more often Tone 3. In Cao & Sarmah (2007), even though the total duration of ambiguous Tone 2 and Tone 3 was kept constant, when duration of turning point and/or the timing of turning point changes, duration alone could not be a distinctive cue for the perception of Tone 3 as other factors would still cause native speakers to make judgments favoring Tone 2.

Studies have also investigated how amplitude affects tone perception. For instance, Chuang et al. (1972) showed that overall, Tone 4 has the highest amplitude and Tone 3 has the lowest. Although amplitude is considered to have little effect on tone perception, there is some evidence that it can be used alone as a perceptual cue. When Whalen & Xu (1992), for example, removed the F_0 and formant structure from a natural speech signal and only the amplitude information remained, native listeners were successful in tone identification for all but Tone 1 tokens.

From the studies above, we have seen that the primary acoustic and perceptual cue for Mandarin tone identification is fundamental frequency. Besides this most important cue, ΔF_0, turning point, duration, and amplitude all contribute to the perception of Mandarin tones.

Non-Native Perception and Production of Chinese and its influence on L2 acquisition of Tones

We have seen that native speakers of Mandarin Chinese identify tones using various acoustic cues. Studies using neuroimaging techniques also suggest that Mandarin tones, for native speakers, are lateralized in the left hemisphere, which suggests that lexical tones are processed as linguistic elements in the same way that other segmental properties are processed (Wang, Jongman, & Sereno, 2001). Will nonnative speakers process tones the same as native tone speakers? Will they process tones linguistically or auditorily? How will the similarity and difference influence their learning of Chinese tones?

From various studies, it has been shown that since the F_0 and pitch pattern associated with the lexicon is unfamiliar to speakers whose native language is nontonal, tone is very difficult for them to learn (e.g. Shen 1989).

First, tonal errors made by nonnative speakers have been investigated in some studies. For example, in Shen (1989), American learners who had learned Chinese for four months were tested for production of tones. The results showed that the learners made errors on all tones with Tone 4 having the highest error rate (55.6%). Shen speculated that Tone 4 is more likely to be subject to first language interference, because it is prosodically less marked for English learners. Another study by Miracle (1989) also looked at tonal errors and reported an overall rate of 42.9% of errors by second-year American learners of Chinese. He classified the tone errors as tonal register, meaning either too high or too low, or tonal contour errors. What is interesting in that study is that the two types of errors were found evenly distributed among tones and tone errors. Specifically, the register errors for Tone 1 were produced when the high level tones were too low in tonal space; the contour errors for Tone 1 were realized by a falling contour instead of a level contour. The register error for Tone 2 was too high in tonal space and the contour error occurred when a level or falling contour was produced instead of a correct rising contour. The same register error occurred with Tone 3 and the contour error occurred when it was realized as a rising tone. The register error for Tone 4 was realized in the mid-low register and contour errors were mainly level ones instead of falling contours.

Secondly, there seem to be great differences between the pitch range of native Chinese speakers and English speakers, which might influence L2 learning. Chen (1974) conducted an experiment comparing pitch range between Chinese and English speakers with sentences in both languages. He found that the pitch range of Chinese speakers speaking Chinese was 1.5 times wider than for English speakers speaking

English. Even when English speakers spoke Chinese, their pitch range increased, but not to the same degree as that of native speakers. Therefore, the implication for American English learners of Chinese is to make pitch range much wider.

There are also differences in perceiving Chinese tones between native and non-native speakers. Fundamental frequency is the primary perceptual cue for native Chinese speakers. The two parameters of height and contour, however, have been shown to be weighted differently, depending on the linguistic experience of listeners (Wang 2006, Gandour 1983). In Gandour (1983)'s study, perception of tones was tested by listeners of both tonal languages and non-tonal languages. The four types of tonal language listeners were Mandarin, Cantonese, Taiwanese, and Thai, and the nontonal listeners were English. By using multidimensional scaling, Gandour found that English listeners placed much more importance on F_0 height, and less attention was given to the contour, compared to listeners of most of the tonal languages. The author speculated that since there are no contrastive tones in English and pitch variation is usually used at sentence level, English speakers only direct their attention to F_0 height.

Duration is another perceptual cue that is used by listeners. Tone 3 is usually longer than Tone 2 in citation form; however, there seems to be a difference in its role among native and non-native speakers. In Change (2011), duration-normalized and non-normalized stimuli of Tone 2 and Tone 3 were compared in a perceptual experiment by eight native Chinese and eight non-native listeners. Their results showed that for native listeners, the absence of the duration cue did not affect the accuracy of the tone recognition but longer reaction time was needed; on the other hand, non-native listeners suffered greatly in their perceptual accuracy as well as reaction time. The

author concluded that syllable duration may facilitate Tone 2- Tone 3 distinction for native speakers and serves as a primary cue besides F_0 for non-natives.

Another difference lies in the context of tone presentation and can be attributed to first language interference. Broselow et al. (1987) compared the perception of tones presented in isolation and in the context of two or three syllables by American learners of Mandarin. The interesting finding from this study is mostly on the nonnative perception of Tone 4. Different from Shen (1989)'s result, Broselow et al. (1987) found that Tone 4 was the most easily identified tone when presented in isolation, as well as in the final position of a two/three syllable word. The errors of Tone 4 only became worse when the tone was presented in non-final position. Broselow et al. (1987) discussed the possible rationale for this phenomenon and concluded that it might have resulted from the interference of English intonation patterns. In English, falling intonation occurs at the end of a declarative sentence, which is acoustically similar to that of Tone 4 in Mandarin. That being the case, it is possible to explain why English speakers identified Tone 4 more accurately in final position by virtue of the fact that they are very familiar with this pattern. In addition, the study showed that Tone 4 was more likely to be misidentified as Tone 1 when in utterance-final position. The authors also explained this as due to first language interference. Because both Tone 4 and Tone 1 start with a high pitch, English listeners are more likely to pay attention to the high part and ignore what follows, as is the case at the end of an English sentence. White (1981)'s study also showed first language influence on tone perception. He found that English listeners tend to identify the high tones as stressed and low tones, like Tone 3, as unstressed. In fact,

however, stress in Mandarin is not realized from fundamental frequency but from duration and amplitude.

In conclusion, we have discussed phonetic implementation in production and perception of lexical tones in tone languages in general and in Mandarin Chinese in particular. Native and nonnative perception and production of Mandarin Chinese were also compared, with the result that linguistic experience does influence the acquisition of tones, either positively or negatively. Though different patterns of tone processing have been seen in nonnative speakers, as well as in hemisphere lateralization (Klein et al., 2001), the performance of tone learning can be improved through training, even after a short period of time (Wang et al., 2000). Moreover, cortical involvement when processing tones can also be modified as proficiency improves (Wang et al., 2000).

Categorical Perception and Mandarin Chinese Tones

Tones are perceived linguistically by native speakers just like vowels and consonants are. This has been supported by a number of studies. One of the earliest studies (Van Lancker & Fromkin, 1973) found in native Thai speakers a right ear advantage in hearing tones in words rather than hearing “hummed” tones with no segmental information. Similarly, Wang et al. (2001) found the same patterns in Mandarin Chinese by comparing the tone errors of dichotically presented tone pairs by Chinese and English native listeners. Chinese listeners showed a significant right ear advantage whereas English speakers showed no ear preference. In recent years, studies using neuroimaging techniques have also yielded results that support this claim. For instance, Gandour et al. (2000) used PET to examine tone processing in Native Thai speakers and found that they used mainly the left hemisphere to process tones as

other linguistic elements, whereas native English speakers did not show any lateralization.

If tones are perceived as a linguistic segment, are they perceived categorically? Categorical perception, according to Liberman (1957, 1967), refers to the phenomenon "whereby small steps along an acoustic continuum will produce perceptible differences when they occur between phonetic categories, but not when they occur within a phonetic category (Gandour, 1978, pp. 58)." A few decades later, Kuhl (1991, 1994) and her colleagues proposed the concept of "perceptual magnetic effect", which also argues that speech sound perception is strongly influenced by category goodness—it is more difficult to discriminate a prototype (good exemplars) from its variants than a nonprototype from its variants. This phenomenon has been widely studied in consonants and vowels in a number of languages. However, very limited work has been done on categorical perception of the suprasegmental feature, tones.

One of the early studies by Abramson (1976) examined the native perception of three Thai tones: the three static low, mid, and high tones. A continuum of 16 level tones with constant F_0 contours was used on the syllable [k^ha:] between 92 and 152 Hz. The results showed that most Thai listeners were able to place the continuum into their corresponding three tone categories. However, their discrimination performance was high throughout the continuum and no clear peaks at the "presumed" boundaries were shown. Halle et al. (2004) speculated that the static tones might be expected to yield lower categoricity and perhaps the perception of dynamic tones is more categorical in nature. Wang (1976) also reported a study on categorical perception in Mandarin Chinese. There was an 11-step continuum of the syllable [i], starting with a 135Hz level

tone and tones that rose in a linear fashion from 105-132 Hz to 135Hz. From the perception results, native Mandarin Chinese speakers did show a pattern of categoricity, which was absent for English listeners. Based on the study, Wang (1976) concluded that tone perception in Mandarin Chinese is categorical. However, there were very limited numbers in that study, with two Chinese participants and three English speakers. Another study by Stagray & Downs (1993) also provided some evidence for this claim. They presented a continuum of level tones with pitch variations to Mandarin Chinese listeners and English listeners. It was revealed that Mandarin listeners were not as sensitive as English listeners to small F_0 contour variations when doing a same-different discrimination task. Chang & Halle (2000) claimed to have the first study using tone continua for the issue of tone categorization with Taiwanese Mandarin listeners. Their study reported a gradient in categoricity, in the sense that tones are perceived in a roughly similar degree of categoricity compared to vowels.

Chang and Halle (2000)'s claim notwithstanding, there was an earlier study in 1976 in which Zue reported a study on the categorical perception of Mandarin Chinese tones (Gandour, 1978). The study used a continuum that had nine variants with a possible intended target of between Tone 2 and Tone 3. Those nine tones were superimposed to a synthetic syllable *bao*, with a rising linear fundamental frequency contour of 100 to 160 Hz. What caused these nine tokens to differ was the duration of the portion before the point of rising (also called "turning point", according to Shen & Lin, 1991), which varied from 0- 400 ms with a 50ms step. Interestingly, both Chinese listeners and English listeners showed a sharp category boundary at Variant 5 (in which the point of rising was in the middle of the stimuli) and a peak in discrimination at the

category boundary. This suggests that both native and nonnative listeners treat Variant 1-4 as “predominantly rising” and Variant 5-9 as “predominantly level”. However, the similarity between native and nonnative perception in categorizing these variants might be due to the limited number of subjects and tonal stimuli as well. Cao and Sarmah (2007) also conducted a study, similarly but on a finer scale, to examine the role of the shape of the pitch contour in the perception of Mandarin Chinese Tone 3 and categorical perception of tones. A set of stimuli was constructed by varying a recorded Tone 3 of syllable *ma* in two conditions: (1) varying the duration of the dip (or turning point) and (2) varying the timing of the turning point (duration of the slope). There were thus 40 stimuli total in a continuum with a constant duration of 400ms. Those variations of duration of dip or timing of the turning point were made incremental at a 10ms step. The manipulated stimuli were presented to native speakers of Mandarin Chinese in two sets: (a) a set of speech stimuli and (b) a set of non-speech stimuli. The results showed that to be perceived as Tone 3, the duration of the dip should not be more than 67.5% of the total length of the stimuli. Once the dip is more than 67.5% of the total length of the stimuli, Mandarin Chinese speakers perceive the stimuli as Tone 1. In addition, a Tone 3 is perceived if the turning point occurs between 42.5% and 72.5% of the total length of the stimuli. If it occurs before 42.5%, it is most likely to be perceived as a Tone 2 and if it occurs after 72.5% of the total length of the stimuli, the stimuli is perceived as Tone 4. This study provided evidence of a clear categorical perception of tones by native speakers, though the non-speech stimuli that were devoid of consonantal and vocalic information did make it difficult for native speakers in accurate identification and thus they perceived them categorically. Halle et al. (2004) conducted a tone continuum

perception experiment on Taiwanese Mandarin listeners and French listeners. The stimuli consisted of three types of continuum: Tone 1- Tone 2, Tone 2- Tone 4, and Tone 3 –Tone 4. The Mandarin listeners showed a categorical perception at the tone boundaries, whereas the French listeners showed no increased sensitivity near category boundaries, suggesting the influence of a language experience effect on speech perception. On the other hand, the performance of nonnatives was not that bad, according to the results. The authors speculated that this was due to the sensitivity to intonation contours by the French listeners; or, according to the PAM model, tones would fall in the "uncategorized" phonetic space for French listeners and they would have fair to good performance depending on the perceived salience.

Another issue is the influence of speaker F_0 range. Tones can be perceived using acoustic cues, but in natural speech they are also perceived relative to other tones. Listeners have to pay attention to a speaker's F_0 range in order to distinguish tones that differ only in F_0 height. A few studies have examined the role of the extrinsic F_0 in tone perception. For instance, Moore and Jongman (1997) investigated speaker normalization in the perception of Tone 2 and Tone 3 in Mandarin Chinese by examining listeners' use of F_0 range as a cue to speaker identity. There were two speakers with different F_0 ranges so that Tone 2 of the low-pitched speakers and Tone 3 of the high-pitched speakers occurred approximately at the equivalent F_0 height. The three tone continua varied in either turning point, $\Delta F_{0,}$ or both, and both attached to a natural precursor phrase from each of the two speakers. The results from the study showed that identification shifted such that the same stimuli were identified as low tones in the high precursor condition, and as high tones in the low precursor condition. The

study suggested that tone identification is influenced by F_0 change and listeners use that cue as a reference to interpret ambiguous tones.

The present study focuses on one of the important perceptual cues for Mandarin tones: turning point, and particularly the timing of the turning point in the tone continuum of Tone 2 and Tone 3. We hope to discover the perceptual boundary of these two tones. In addition, we examine voice quality (creaky voice) to determine its role in Tone 2 and Tone 3 perception.

Creaky Voice and Mandarin Chinese

In the production of Mandarin tones, there is one acoustic cue that is of interest here and which has received less attention, namely phonation quality. Phonation quality, such as creaky voice, breathy voice, and modal voice, has generally been a less studied area in language research. In some languages, there is vowel or consonant phonation contrast; therefore, in those languages, phonation quality may be associated with a perception cue. In Mandarin Chinese, these phonation qualities do not form segmental contrasts.

Creaky voice is characterized by reduced intensity in waveform, lower fundamental frequency, and less frequent pitch periods with irregular duration (Gordon & Ladefoged, 2001). According to Keating (2006), creaky voice, or a constricted glottis, is associated with aperiodic glottal pulses. The irregular pitch periods can be visually reflected by “the increased distance between the vertical striations reflected pitch pulses, before the modal voicing commences…” (Gordon & Ladefoged, 2001, p.387).

Although creaky voice does not differentiate words at their semantic level, it has been studied in its association with production and perception of Mandarin Chinese tones.

Creaky voice has been reported as being associated with Tone 3 and Tone 4 in Mandarin Chinese (Belotel-Grenie, A. & Grenie, M., 1994). In that study, production data from 7 native speakers (4 M, 3 F) of 31 words were analyzed, and the vowel in those words was /a/. The results showed that for all four male speakers, creaky voice was always associated with Tone 3 (8 times for 8 words); for the three female speakers; however, the occurrence of creaky voice was 4/5, 1/8, and 3/6. In the study that the same authors conducted later (1995), similar results were reported based on data from a male speaker reading monosyllabic Chinese words with no initials but all kinds of vowels (low vowel, high front vowels, and high back vowels) were included. The data from that speaker showed the percentage of words produced with creaky voice to be 45.8% in Tone 3 and 10.5% in Tone 4, with no occurrences of creaky voice in Tone 1 and Tone 2. High front vowels were also more associated with creaky voice than high back vowels. However, because of the very limited number of participants in these production studies, it is difficult to generalize this result into a solid conclusion.

What is the role of creaky voice in the perception of Mandarin Tones then? There have been a limited number of studies in this area. In Belotel-Grenie, A. & Grenie, M (1997), it was determined that the recognition point was earlier for a creaky Tone 3 than for a non-creaky Tone 3. In the experiment reported in that study, four words containing the vowel /a/ and consonant initials--/m/, /n/, /l/, and /d/ were recorded by two speakers (one female and one male). The male speaker produced creaky voice and the female did not (it is not clear, however, whether the creaky voice was present for all or some of the tones reported therein). Those stimuli were subsequently partitioned into 30%, 40% segments, up to 100% of the duration of the tone for the perception experiment. Ten

native speakers participated in the identification task. The results indicated that a Tone 3 produced with creaky voice is recognized more quickly (at 60% of the duration) than a Tone 3 without creakiness (at 70%). The authors asserted that creaky voice is a secondary cue for Tone 3 perception, based on the results of their study.

Creaky voice has also been studied in other tonal languages and dialects. For example, creaky voice in Cantonese has been reported as being associated with the production of the lowest tone (mid-falling), Tone 4 (Vance, 1977). Yu (2010)'s study reported speech data from eight Mandarin speakers and eight Cantonese speakers which showed that 68% of Mandarin Tone 3 was found to have creaky voice, but only 25% overall in Cantonese Tone 4.

These results notwithstanding, there are some limitations in these studies. The number of participants from whom production data was collected, for example, was small. Also, the number of vowels and consonants included in the speech data was limited. Furthermore, the issue of how creaky voice interacts with categorical perception of Mandarin tones not only among native speakers but also among non-native speakers was not addressed. Therefore, the current study is an attempt to fill these gaps and provide a more comprehensive understanding of the role of creaky voice in the production and perception of Mandarin tones, specifically, Tone 2 and Tone 3.

Determination of Creaky Voice

Figure 2-1 is an example of modal, breathy, and creaky voice given in Gordon and Ladefoged (2001, p.390) from voiced vowels in Jalapa Mazatec words. In their paper, Gordon and Ladefoged discuss a number of phonetic properties that are associated with distinguishing creaky voice from breathy phonation and modal voices, including

periodicity, acoustic intensity, spectral tilt, fundamental frequency, formant frequencies, duration, and airflow. Although different languages do not have a uniform measurement for these parameters, some general contrasts in these phonetic properties can be found.

In the present study, aural and visual inspection of the waveform was the primary method used to determine the presence of creaky voice. There were two raters who listened and inspected the waveforms of the production data. There are other measurements, however, that can also be used to detect creaky voice, such as running a Fast Fourier Transform (FFT). According to Gordon and Ladefoged (2001), the spectral tilt (difference between the amplitude of the second harmonic to fundamental frequency, H2-F_0) is most steeply positive for creaky vowels (Figure 2-2.).

The snapshots of waveform (Figure 2-3, Figure 2-4) were taken from two native Chinese speakers who participated in the production experiment—pronouncing the syllable [ma] with Tone 3 and Tone 2. Creaky voice can be seen in the circled area in the top waveform (Figure 2-3), while the bottom waveform (Figure 2-4) has no creaky voice present. In the FFT spectra of a slice from the vowel [a], there is a positive slope from F_0 to H2 in the top spectrum (Figure 2-5), indicating the presence of creaky voice, which is absent in the second FFT spectrum (Figure 2-6).

If the production data from the study above was conclusive that creaky voice is highly associated with production of Tone 3 in Mandarin Chinese, then how important is it in perceiving Tone 3? Will the absence or presence of creaky voice influence the categorical perception of Tone 3 and Tone 2 among native listeners and nonnative listeners? Do native listeners and nonnative listeners pay attention to this cue in the

same way? What is the difference in the relative importance between phonation cues and pitch contour in tonal identity?

The goal of the present study is to answer those questions by exploring the categorical perception of Mandarin Tone 2 and Tone 3, and the role of creaky voice in the perception of these two tones, among native and non-native speakers. This research was guided by the following questions:

Research Question 1: Is creaky voice associated with a certain tone(s) produced by native Mandarin Chinese speakers (NC)?

Research Question 2: How does the presence of creaky voice affect the perception of tones by native Chinese speakers (NC) and L2 learners of Chinese (NE), respectively?

Research Question 3: What is the perceptual boundary of the Tone 2-Tone 3 continuum perceived by native Chinese speakers (NC)? How does it differ from L2 learners of Chinese (NE) with different proficiencies? What is the role of creaky voice in their perception?

Chapter 3 describes the methodology used in the present study. Results and a discussion are presented thereafter.

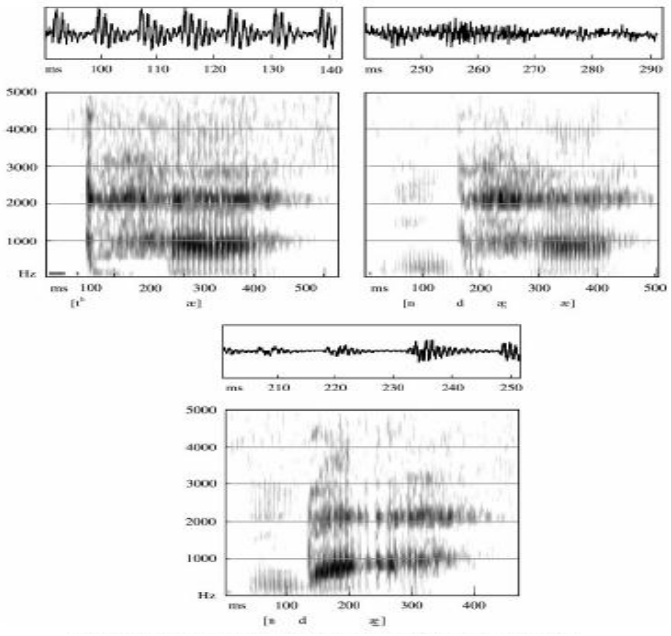

Figure 2-1. An example of modal (top-left), breathy (top-left), and creaky voice (bottom) given in Gordon and Ladefoged (2001, p.390) from voiced vowels in Jalapa Mazatec words.

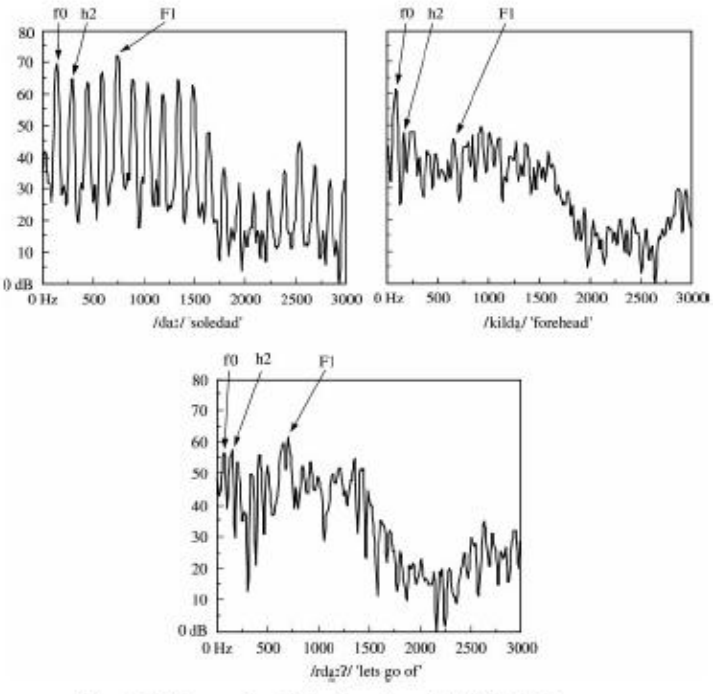

Figure 2-2. FFT spectra of modal (top-left), breathy (top-right) and creaky (bottom) vowel /a/ in three San Lucas Quizvini Zapotec words (Gordon & Ladefoged, 2001)

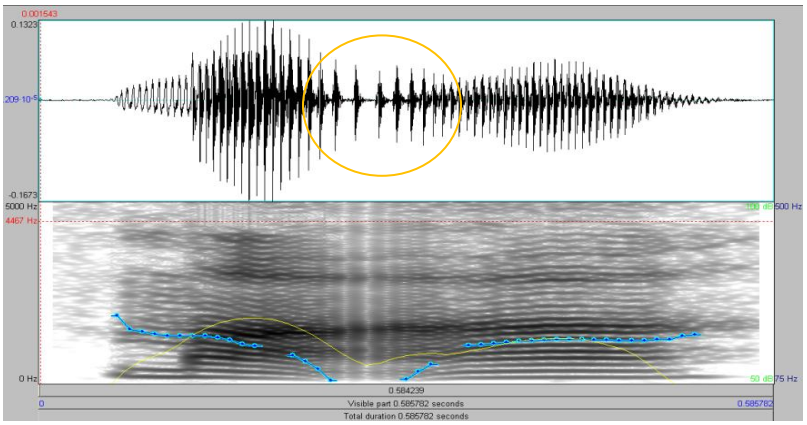

Figure 2-3. Waveform of a Tone 3 [ma] produced with creaky voice.

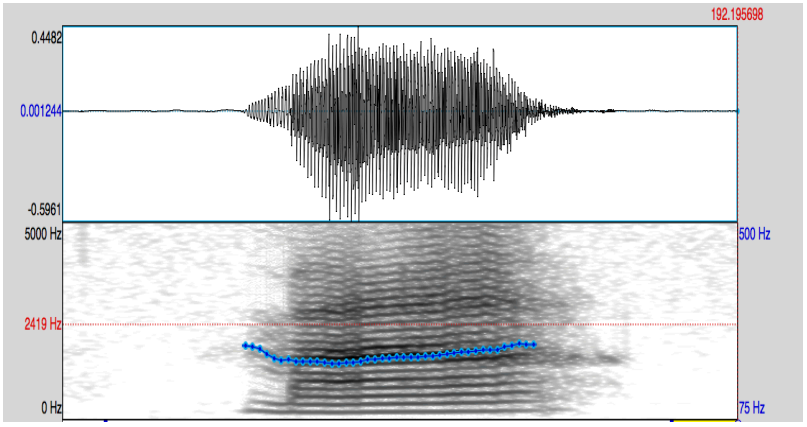

Figure 2-4. Waveform of a Tone 2 [ma] produced without creaky voice.

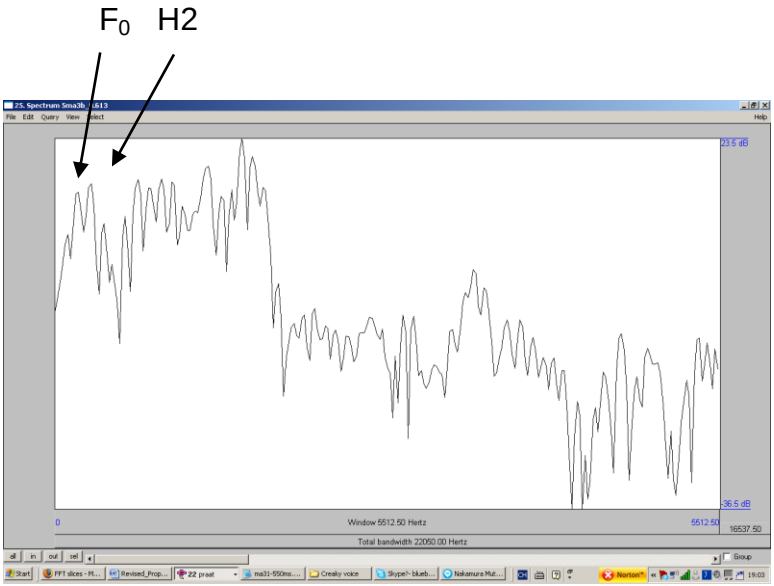

Figure 2-5. FFT spectra of a slice from the vowel [a] with creaky voice

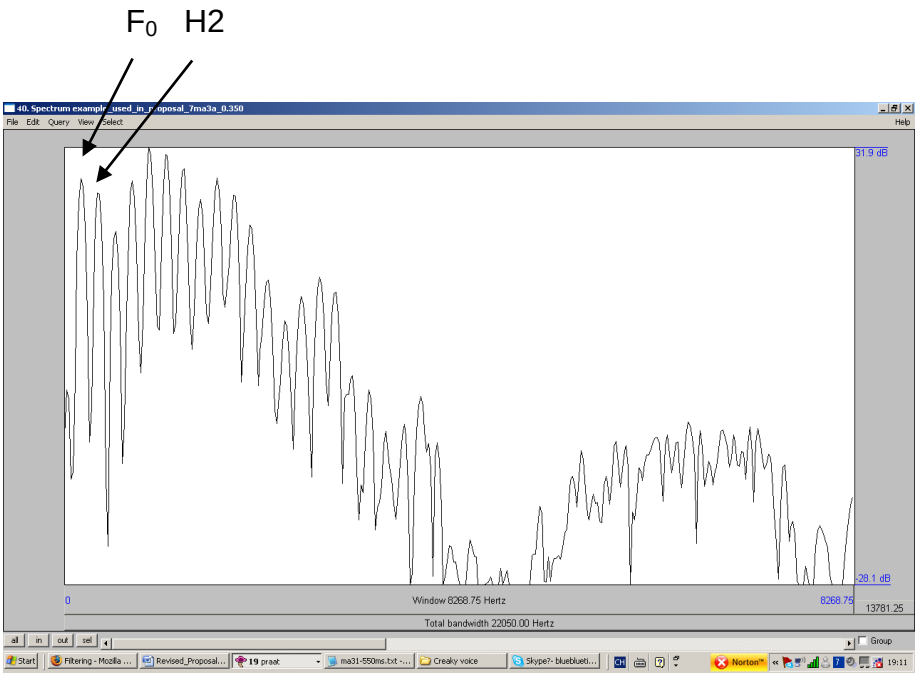

Figure 2-6. FFT spectra of a slice from the vowel [a] without creaky voice

CHAPTER 3
METHOD

The two major goals of this dissertation are to investigate the use of creaky voice phonation in native production of Mandarin Chinese tones, and to examine the effect of creaky voice in the categorization of Mandarin tones 2 and 3 among native and non-native speakers of Mandarin. The study was guided by the following research questions:

Research Question 1: Is creaky voice associated with a certain tone(s) produced by native Mandarin Chinese speakers (NC)?

Research Question 2: How does the presence of creaky voice affect the perception of Tones by native Chinese speakers (NC) and L2 learners of Chinese (NE), respectively?

Research Question 3: What is the perceptual boundary of the Tone 2-Tone 3 continuum perceived by native Chinese speakers (NC)? How does it differ from L2 learners of Chinese (NE) with different proficiencies? What is the role of creaky voice in their perception?

The experiments consisted of three parts—a tone baseline task, a tone production experiment, and a tone categorization experiment. First, native speakers of Mandarin Chinese and L2 learners of Chinese did a tone baseline task for the purpose of measuring proficiency of Mandarin tone perception, and to divide the NE up into two groups. Each of the native Chinese participants was then recorded producing utterances with target mono-syllabic Chinese words with all four tones and the presence of creaky voice was examined in the production data. In the third task, categorization of tone 2 and tone 3 among both NC and NE was examined. Two sets of tone continuum were created such that creaky phonation was present during the dip duration in the "creaky voice" set, but was absent in the "clear or modal voice" set. All other parameters were the same. Both sets of stimuli were randomly presented to both NC and NE for

tone categorization. Categorization differences, particularly for tone 2 and tone 3 among the two groups were examined. In addition, the effect of L2 proficiency on tone categorization among L2 learners of Chinese (NE) was investigated.

Participants

Group 1 (Native Chinese Speakers): Thirty-three native speakers of Chinese, age 22-40 (Mean=28.5, SD=5.3), with no reported hearing or speaking problems, participated in the experiment; however, one participant was disqualified as a result of failing the tone baseline task (less than 50% correct) and two other participants' perceptual data were lost due to a technical problem. Therefore, data from 30 participants were included in the analysis.

Among the thirty native Chinese speakers, fifteen (7 females and 8 males) were born and grew up in Beijing (capital of the People's Republic of China) until the age of at least 18. The other fifteen (9 females and 6 males) were from other areas of China, including ten provinces (*Heilongjiang, Liaoning, Shandong, Hebei, Henan, Xinjiang, Jiangsu, Anhui, Hubei, and Jiangxi*) and one municipality (*Shanghai*). All of the Chinese participants could speak standard Mandarin. Among the fifteen non-Beijingers, four (two from *Jiangsu*, one from *Jiangxi* (male), and one from *Shanghai*) also spoke a variety of *Wu* dialect, and the others were from regions where mainly Mandarin is spoken.

The thirty native speakers of Chinese were either studying or working at the University of Florida, and had been in the United States for 7 months to 12 years (Mean= 2.9 yrs, SD.=2.4). Among the Chinese participants, eleven of them reported having musical training experience from 1 to 16 years (Mean=2.5 yrs, SD= 4.4). Two are left-handed. They were all paid for participating in the study.

Group 2 (English Learners of Chinese): Forty-two native speakers of American English, age 18-23 (Mean = 20.9, SD =1.4), 14 females and 28 males, participated in the experiment. All of them except one (first year graduate student) were undergraduate students at the University of Florida. None reported any speech or hearing problems. Four indicated left-handedness. Seventeen of the English speakers were paid for participation, and the other twenty-five received extra credit for their Chinese classes.

All the native English speakers had completed, or almost completed, at least two semesters of Chinese at the University of Florida or another institution. Except for one participant, who started learning Chinese with a tutor at age 6 for four years and continued in college from age 18-21 (thus had been excluded from data analysis), all other native English speakers were late learners of Chinese (AoA later than age 16) and the length of their Chinese learning ranged from two semesters to four years. Eighteen of them had study aboard experience in China for at least three months in Beijing, Shanghai, Chengdu, or Taiwan. There is another speaker who did not pass the baseline task thus was eliminated from the study. Therefore, there are 40 NE speakers (14 Females, 26 Males) in the analysis of the current study.

Among the 40 English speakers, 28 had played a musical instrument for 2-18 years (starting age 5-17). One, who did not play any musical instrument, had been in a chorus for one year. Four English participants reported themselves to be left-handed.

Procedure and Stimuli

All experiments were conducted at the linguistics laboratory of the University of Florida. Before the experiments began, participants were briefed as to what to expect in the study and informed consent forms were signed. A questionnaire was then filled out by each of the participants, including information on their education/language

background and music background. After that, the following three experiments were conducted in the following order: 1) Tone Baseline Task, 2) Production Experiment, and 3) Categorization Experiment, followed by a debriefing session. The entire experiment took about one hour for each participant. The stimuli used and procedures in each of the tasks are explained below.

Tone Baseline Task

Stimuli

The stimuli used in the tone baseline task consisted of 40 mono-syllabic words of Mandarin Chinese, with 10 *Consonant+Vowel* combinations produced with all four tones by two native Chinese speakers from Beijing, one female and one male, resulting in 80 target words total. The 40 words were chosen from the first-year Chinese textbook (*Integrated Chinese I*) used by the Department of Languages, Literatures, and Cultures of the University of Florida, so that all L2 learners of Chinese should have encountered all the words prior to participating in the study. (See Appendix for the complete list of 40 words). Each of the target words appeared at the final position in a frame sentence (in Chinese): "*please read this word out loud* -- _____." The recording was conducted in the soundproof booth in the Linguistics lab, using a digital recorder (Marantz PMD660) and a head mounted microphone (Shure SM 10A) at a sampling rate of 44.1 kHz. The sound files were then transferred to a computer and saved as .WAV files. Each of the 40 target words from the two speakers' utterances was then segmented out from the carrier sentence and saved as 80 individual sound files (.wav) using PRAAT. All tokens were normalized at 98% peak intensity with the UAB software developed by Steve Smith at the University of Alabama in Birmingham.

Procedure

The tone baseline bask was carried out in a quiet room located in the Linguistics lab by means of a computer. The presentation of the stimuli in the tone baseline task was controlled by the UAB software. The 80 stimuli were presented in two blocks, one for the 40 words read by the female, and the other for the 40 words read by the male. The 40 target words in each block were randomized.

Presentation of the two blocks of words was counter-balanced for all participants. There was a break after the first 40 words. Participants were tested individually. After a participant entered the room and sat in front of the computer, s/he was instructed (in her/his native language) to identify the tone by clicking one of the four buttons marked with “Tone 1, Tone 2, Tone 3 and Tone 4” on the computer screen after hearing each word through a pair of headphones connected to the computer. There was no time constraint in this task; that is to say, participants were allowed to take as much time as they needed to make a choice. There was no feedback provided throughout the task, except that a red dot would appear on top of the tone square a participant clicked, as an acknowledgement of a response to a stimulus.

Tone Production Experiment

Stimuli

Materials used in the tone production task consisted of 40 different monosyllabic Chinese words, with 10 *Consonant+Vowel* combinations for each of the four tones, with each repeated three times, resulting in 120 words in total. These monosyllabic words were constructed with initials [m, n, p^h, p, k^h, k, t], followed by finals of all possible monophthongal vowels in Mandarin Chinese: front vowel /i/, central vowel /a/, and back vowels /ɤ, u/. These consonants and vowels were chosen to ensure that they could be

produced with all four Chinese tones to generate real Chinese words (Table 3-1). Notice that these words are different from those used in the baseline task. The 120 words were all randomized and each was placed at the final position of a carrier sentence in Chinese "*Please read this word out loud _____".* To make sure that all the words were pronounced with their correct tones, *pinyin* was written next to each of the target words.

Procedure

The tone production experiment was conducted in the soundproof booth in the linguistics lab at the University of Florida. A head-mounted microphone (Shure SM 10A) and a digital recorder (Marantz PMD660) were used to record the production of Chinese words. Each of the participants read the same list of 120 sentences in Chinese with target words at the end of each sentence. The same order was used for every participant. These sound files were transferred to a computer and saved as .WAV files. Each of the target words was examined for the presence of a creaky voice by two raters using audio and visual inspection.

Tone Categorization Experiment

Stimuli

There were two goals in the tone perception (categorization) experiment: one was to examine the categorial boundary in the perception of Chinese Tone2-Tone3 by native and non-native (L2) listeners, and the other was to investigate whether the presence of creaky voice affects the perceptual boundary between the two tones among both native speakers of Chinese and native speakers of American English who are learning Chinese. Therefore, two identical sets of tone stimuli were created for this experiment, except that creaky voice was present in one set, but was absent in the other.

The first set of tone stimuli consisted of 34 different tokens with no creaky voice generated from a single token of a [ma] (tone 3, 'horse', 马) syllable produced by a female native speaker of Mandarin Chinese from the Beijing area. This word was recorded using a head-mounted microphone (Shure SM 10A) and a digital recorder (Marantz PMD660) at a sampling frequency of 44.1 kHz and was saved as a .WAV sound file to a computer. An examination of the spectrogram generated by PRAAT software (Paul Boersma & David Weenink) indicated that this word was produced with no creaky voice. In addition, the total duration of this syllable was 460 milliseconds (ms) long with F_0 at syllable onset of 203.8 Hz and offset of 253.5 Hz. All 34 tokens of the non-creaky stimuli were created from this naturally produced model [ma] syllable by manipulating pitch contours using the PRAAT software. In this original iteration of [ma] (Figure 3-1), four pitch heights were observed: syllable onset F_0 = 203.8Hz, onset of vowel [a] F_0= 203.8Hz (starting at 57ms), minimal F_0 (turning point) = 170.5Hz, and the syllable offset F_0 = 253.5Hz.

To generate the 34 tokens of creaky stimuli, another [ma] syllable produced with tone 3 by the same female speaker was used as the model for examining the natural parameters of creakiness. Examination with PRAAT indicated that this utterance was produced with creakiness at the dip of the Tone3 pitch contour with duration of 68ms. Therefore, the dip of the non-creaky tokens in the first group was also set to 68ms, with the minimal observed F_0 of 170.5 Hz from the non-creaky [ma] utterance.

Stimulus group 1- non-creaky tokens: For the first group of non-creaky tokens, an empty pitch tier (length=460ms) was created in PRAAT for each token, then 5 pitch points were added to each empty pitch tier (Figure 3-2),1) syllable onset (0ms, F_0=203.8

Hz), 2) starting of vowel (57ms, F_0=203.8 Hz), 3) starting point of the dip (starting from 58ms to 388ms, in an increment of 10ms, F_0=170.5 Hz), 4) ending point of the dip (starting from 126ms up to 456ms, in an increment of 10ms, F_0=170.5 Hz), and 5) syllable offset (460ms, F_0=253.5 Hz). Therefore, without manipulating the total length of the syllable, the length of the dip was kept constant at 68ms, same as in the observed value from naturally produced syllable, only the timing of the turning point (dip) was altered; in other words, the time that the turning point (dip) occurred in each token differed by 10ms, resulting in 34 different pitch tiers.

The modal [ma] token was then synthesized with these 34 different pitch tiers, generating 34 non-creaky tone stimuli to be used in the tone perception experiment.

Stimulus group 2: creaky tokens: The second group of perception stimuli consisted of 34 tokens that were identical to the 34 tokens in the first group, except that tokens in this group were manipulated again with PRAAT so that each of them had creaky voice at the dip.

Just as in the first group, an empty pitch tier (length=460ms) was created in PRAAT for each token, and the same five pitch points were added. In addition, the 68ms dip was manipulated again to create a creaky voice effect. Recall that the prominent characteristics of creaky voice include irregular periodicity and sudden decrease in fundamental frequency and intensity. Therefore, creakiness was generated by randomly adding extremely low and irregular pitch points at a level far below the non-creaky dip to create “jitter” (Figure 3-3). The creakiness sounded quite close to a naturally produced creakiness according to two native Chinese speakers. Thus the values of each of the pitch points in the “jitter” dip were kept the same for all creaky

tokens. Each of the pitch tiers was then synthesized with the naturally produced [ma], generating 34 creaky tone tokens for the perception experiment.

Altogether, 68 different tokens were created for the tone categorization experiment. The presentation of these stimuli was again operated with UAB software. They were first normalized at 98% peak intensity with the UAB program. Then the two groups of creaky and non-creaky tokens were mixed together and randomized. The stimuli were presented to each participant in three blocks, with two breaks to reduce the fatigue effect. In each block, the mixed 68 tokens were repeated two times (68*2=136 stimuli). The order of the mixed 68 tokens was different in each block. The presentation of the three blocks was counter-balanced across the participants in the following way: 123, 132, 213, 231, 321, and 312, thus resulting in six possible orders of presentations. In total, there were 68 tokens*2 repetitions *3 blocks=408 stimuli in the tone perception experiment for each participant.

Procedure

The tone categorization experiment was also conducted in the same room where the tone baseline task was carried out. Participants were individually tested. Each participant sat in front of a computer, wearing headphones, and was instructed (in their native language) to identify the tone after hearing each stimulus by clicking one of the four squares on the computer screen with texts of “Tone1, Tone2, Tone3, and Tone4” on each of them. Although responses were expected to be mostly Tone 2 and Tone 3, the purpose of having four tone responses was to make the task more natural and to find out whether some stimuli were perceived as other tones. Responses had to be made within 3 seconds; otherwise, the following stimulus would be presented. There was no feedback provided throughout the experiment, except that a red dot would

appear on top of the button to acknowledge the response. The participants were also instructed that they should respond according to their first intuition as soon as possible and that they should take two breaks by taking off the earphones and resting for a few minutes.

Debriefing

After finishing all three tasks—tone baseline, tone production, and tone categorization experiments, each of the participants was interviewed. The interviews were conducted in each participant's native language. The main purposes of the debriefing were to find out whether any of the participants was aware of creaky voice before and during the experiments and what strategies/criteria were used in the tone categorization task.

Table 3-1. Forty tokens of mono-syllabic Chinese words used in tone production experiment

	T1	T2	T3	T4
[mi]	眯	迷	米	密
[ni]	妮	泥	你	逆
[pi]	逼	鼻	比	必
[p^hi]	批	皮	匹	屁
[ma]	妈	麻	马	骂
[pa]	扒	拔	把	爸
[ta]	搭	达	打	大
[p^hu]	扑	葡	普	瀑
[k^hɤ]	棵	壳	渴	客
[kɤ]	哥	隔	葛	个

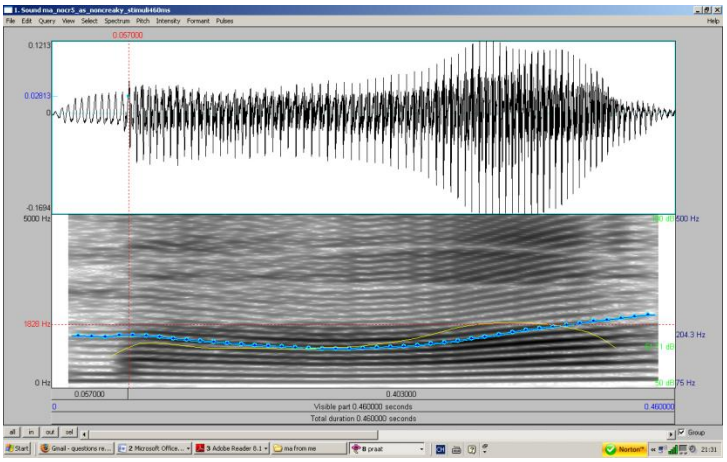

Figure 3-1. Natural utterance of [ma] with Tone 3 (non-creaky), based on which stimuli in the categorization task were created

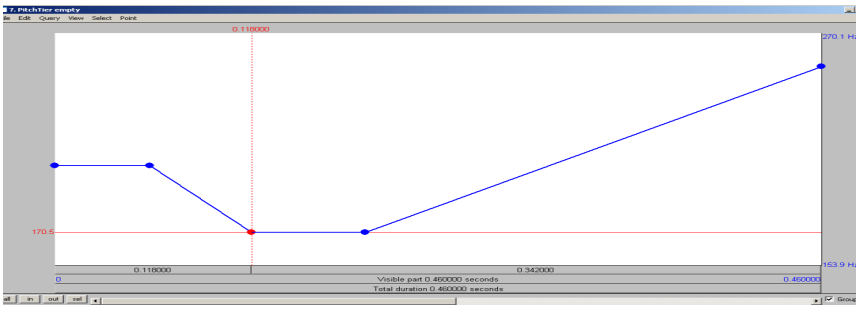

Figure 3-2. An example of an empty pitch tier in PRAAT with five pitch points added to create a non-creaky token: 1) syllable onset (0ms, F_0=203.8 Hz), 2) starting of vowel (57ms, F_0=203.8 Hz), 3) starting point of the dip (in this example: 118ms, F_0=170.5 Hz), 4) ending point of the dip (186ms, F_0=170.5 Hz), and 5) the syllable offset (460ms, F_0=253.5 Hz).

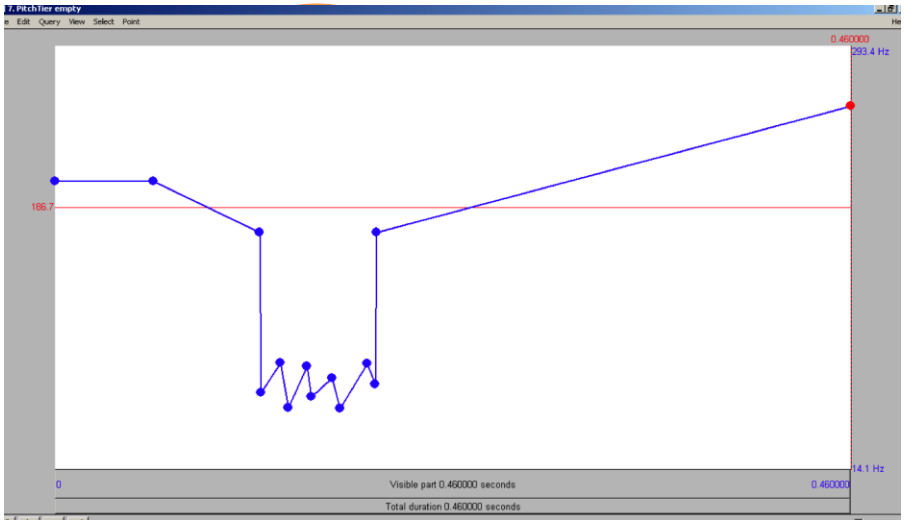

Figure 3-3. An example of an empty pitch tier in PRAAT with five pitch points added. Then randomly added extremely low and irregular pitch points below the non-creaky dip to create a “jitter”.

CHAPTER 4
RESULTS

The perception experiments examined two research questions on the categorical perceptual boundary of the Tone2-Tone3 continuum and the role of creaky voice in boundary shift by native Chinese speakers and L2 learners of Chinese. The production experiment examined the presence of creaky voice with Mandarin Chinese tones, as well as its correlation with the perception result. Results from the baseline task are discussed first in order to divide native English speaker participants into subgroups based on their Mandarin proficiency level. Then, research question 1 is answered through analysis of production data. Finally, research questions 2 and 3 are then elaborated based on the results from the perception experiments.

Research Question 1: Is creaky voice associated with a certain tone(s) produced by native Mandarin Chinese speakers (NC)?

Research Question 2: How does the presence of creaky voice affect the perception of Tones by native Chinese speakers (NC) and L2 learners of Chinese (NE), respectively?

Research Question 3: What is the perceptual boundary of the Tone 2-Tone 3 continuum perceived by native Chinese speakers (NC)? How does it differ from L2 learners of Chinese (NE) with different proficiencies? What is the role of creaky voice in their perception?

Results of the Baseline Experiment

In order to answer the first two questions, the Chinese tone proficiencies of all the participants need to be reported here from the baseline experiment.

In the baseline experiment, participants listened to 80 Chinese words in isolation followed by a judgment task of what tone they heard after each token. There were 20 words for each of the four tones. The baseline score of each participant was the percentage of the total number of correct answers divided by 80.

The range of the 30 native Chinese speakers' scores was from 91.25% to 100% (mean=98.9, SD=1.9). Among the 42 native English speakers, there were two participants who were eliminated from the study: one had an age of acquisition (AoA) of 6 (while other NEs' AoA>16) and the other one scored only 47.5%, which is more than 2 SD lower than the group mean (mean=87.1, SD=10.5). The range of the 40 remaining native English speakers' scores was from 53.75% to 100%.

Since there was noticeable variance among the scores, the native English group needed to be divided into two subgroups: NE-High and NE-Low. The two groups were divided such that the number of participants in each group was as equal as possible and that the mean scores between the two groups were significantly different. To meet these two criteria, the best cut-off point to divide the NE group was determined to be 90%. In this way, one group had a score range of 91.25% to 100%, which was the same as the native Chinese group, and the numbers in each group were as close to each other as possible, plus the means of the two groups were significantly different from each other. The NE-High group included 18 participants whose scores ranged from 91.25% to 100% (mean=94.9, SD=2.9). The NE-Low group had 22 participants and their scores ranged from 61.25% to 90% (mean=80.7, SD=10.1). T-tests were run for the two pairs: NE-High vs. NE-Low and NE-High vs. NC. The results showed that the two English groups were significantly different from each other in their mean scores (p=.00), while the NE-High and NC were not (p=.90). However, for the purpose of the current research, it is necessary to regard the NE-High as a different group from the native Chinese speakers. Therefore, for all data analyses reported in this chapter, all

participants are considered to be in one of three proficiency groups: NC, NE-High, and NE-Low.

Table 4-1 shows the error rates that participants made in each of the tone categories (number of errors in each tone category divided by the total number of the expected number of correct responses). A One-Way ANOVA was run for the error rates within each of the groups, with “tone category” being the factor. The results showed that tone category had a significant effect on error rates in both of the NE groups (NE-H: [$F(3, 68)=5.934$, $p=.001$]; NE-L: [$F(3, 84)=2.93$, $p=.038$]), but not on NC ([$F(3,116)=1.289$, $p=.282$]). The post-hoc test (Bonferroni) for the two NE groups suggested that NE-High had the two highest error rates in Tone 2 and Tone 3 (8.1% in both) and the NE-L group had the highest error rate in Tone 2 (27.5%) compared to the other tones category, which had about the same rate ($p>.05$).

In addition to calculating the mean scores for tone proficiency, Pearson Correlation tests were also performed to ascertain the relationships between the English speakers’ tone proficiency scores (total) and their age, gender, right-handedness, length of Chinese study (# of semesters), length of time studying abroad (# of months), and music experience (# of years). The results showed that none of the above-mentioned factors had any significant correlation with the English speakers’ tone proficiency scores ($p>.001$). The same test was also run for Chinese speakers and the results showed that there was no correlation between their tone scores and their age, gender, right-handedness, years in the U.S., whether or not they were from Beijing, nor their music experience ($p>.001$).

In the following sections, data from the perception and production experiments is presented. The goals of the analysis are to report 1) how creaky voice affects perceptions of tones among native and non-native speakers, 2) what the categorical perceptual boundaries for Tone 2 and Tone 3 are, and what role creaky voice plays in perception, and 3) how much creaky voice is present in native speakers' tone production.

Results from the Production Experiment

In the production part of the experiment, all native Chinese participants were recorded reading 120 target words—all monosyllabic Chinese words, with 10 Consonant+Vowel combinations for each of the 4 tones, and with each repeated three times.

The recordings of each target word from each participant were carefully examined for the presence of creaky voice by two raters. The raters listened to each of the target words and made a judgment as to the presence of creaky voice. After the first rating, the two raters compared their ratings. There were a number of words for which the raters had different judgments. A second rating was then performed by both raters until agreement was reached for each target word by both raters. Samples of the stimuli were chosen to be visually inspected. The results agreed with the rating of creakiness.

Table 4-2 summarizes the percentages of creakiness detected in all the target words in four tone categories. There were 30 native Chinese speakers and each of them read 120 target words (30 words for each tone), resulting in 900 (30*30) tokens for each tone category. A Chi-Square test was run on the numbers of words that were produced with creaky voice, and the results showed there is a strong relationship between creakiness and tone category [$X^2(3, N=1046) =1952.8, p=.00$]. As can be

clearly seen, Tone 3 had the most occurrences of creaky voice (97.6%) compared to the other three tones (Tone 1 1.9%, Tone 2 11.6%, and Tone 4 5.2%). This is in accordance with previous research suggesting that creaky voice is associated with Mandarin Tone 3. Moreover, the results of the present study suggest that creaky voice is highly associated with Mandarin Tone 3 production.

Results from Perception Experiment

Results of Tone Reponses for Each Stimulus

Percentages of four tone responses: There were 34 stimuli (differing in their turning point locations, with 6 repetitions for each), 4 types of tone responses, 3 groups (NC=2, NE-High=1, NE-Low=0), and 2 conditions (Non-creaky=0, Creaky=1). The percentages of the responses for each tone category were calculated by running crosstab and a Chi-square test of the data, and Table 4-3 was generated to show the overall picture from the three groups. Chi-Square tests on all four tone responses of the three groups in two conditions were performed. The results showed that there was a significant relationship between “group” and “tone responses” in both conditions (Non-creaky: $X^2(8, N=14280) =580$, p=.00; Creaky: $X^2(8, N=14280) =773.5$, p=.00).

In order to ascertain whether those differences in tone response percentages are significant, six Chi-Square tests with numbers of tone responses were run to compare Tone 2 vs. Tone 3 response for each of the responses within each group. The results showed that except for the NC Tone 2 vs. Tone 3 response in Creaky condition which did not exhibit a significant difference ([$X^2(2, N=5120) =3.51$, p=.06>.01), all other comparisons yielded a significant result:

Non-creaky condition

- NC Tone 2 vs. Tone 3: $X^2(1, N=5282) =896.4$, p=.00

- NE-H Tone 2 vs. Tone 3: $X^2(1, N=3010) = 603.7$, p=.00
- NE-L Tone 2 vs. Tone 3: $X^2(1, N=3428) = 142.9$, p=.00

Creaky condition

- NE-H Tone 2 vs. Tone 3: $X^2(1, N=2932) = 237.2$, p=.00
- NE-L Tone 2 vs. Tone 3: $X^2(1, N=3250) = 287.1$, p=.00

The results are summarized as follows:

1. In Non-creaky condition, all three groups had more Tone 3 responses than Tone 2 responses. Among the three groups, NE-Low had the most Tone 2 and the least Tone 3 responses.

2. In Creaky condition, both NE groups had less Tone 3 responses than Tone 2, but the NC group had the same for each tone category. Among the three groups, NC had less Tone 2 and more Tone 3 than NEs.

3. Across the two conditions, all three groups had more Tone 2 and less Tone 3 responses when there was creaky voice present in the stimuli.

The pattern seems to be very clear from the results above, namely that the presence of creakiness leads to increased perception of Tone 2 and less perception of Tone 3 regardless of language proficiency. This suggests that creakiness is not the primary cue for identification of Tone 3; otherwise, there would have been more Tone 3 responses in the Creaky condition, given that all other factors were the same.

When looking at the change in Tone 3 responses across conditions, it is found that the NC group showed the least amount of decrease from Non-creaky (60.9%) to Creaky (42.9%), when compared to both NE-High and NE-Low (59.3 % to 28.6%, 46% to 25.4%). This suggests that English speakers are more affected by the presence of creaky voice than native Mandarin listeners when it comes to Tone 3 perception. Given the fact that the two NE groups had the same number of Tone 2 responses and Tone 3 responses only in Creaky condition, we can surmise that creaky voice makes English speakers with different Tone proficiencies behave in a similar manner.

Perceptual Boundaries of Tone 2 and Tone 3

In order to investigate the questions of perceptual boundary from Tone 2 to Tone 3 and how creaky voice makes a difference, if any, two major statistical analyses were performed. First, the average turning points (ATP) were calculated from those stimuli that received only Tone 2 or Tone 3 responses. This was done in order to determine the actual turning point at which perception shifts from Tone 2 to Tone 3 under each condition across the three groups. Second, binary logistic regression tests were run for each participant's responses in order to determine the predicted 50% category-crossover point for each participant. The crossing points between two groups (NC and NE with two proficiency levels: Low and High) were then compared in both conditions (non-creaky and creaky).

Before running the two tests, Tone 1 and Tone 4 responses had to be eliminated from the data for two reasons. In the first place, the current research focuses only on Tone 2 and Tone 3. The response buttons designed for the perception experiment, which included all four tone responses, were for the purpose of making the task as natural as possible since there are four tones in Mandarin Chinese. The other reason for the elimination of the Tone 1 and 4 responses was that for the analysis of binary logistic regression, only data with a binary outcome could be used.

In order to determine whether the trials with Tone 1 and Tone 4 responses could be eliminated, a Chi-Square test was run on the combined responses of Tone 2+Tone 3 and Tone 1 +Tone 4 among the three groups, in both conditions. The results showed that for all groups and under both conditions, the Tone 2 and Tone 3 responses were significantly greater than the Tone 1 and Tone 4 responses ($p=.00$). In the following analysis, only stimuli with responses of Tone 2 and Tone 3 are included.

Average Turning Point (ATP)

In order to locate the perceptual boundary between Tone 2 and Tone 3, the average turning point (turning point refers to the timing of the start of the low dip in F_0 in milliseconds in the current study) was calculated from stimuli that received 100% Tone 2 or Tone 3 responses. This means that what was included were those stimuli that either a) received six Tone 2 or Tone 3 responses out of 6 repetitions or b) received five Tone 2 or Tone 3 responses out of five repetitions when there was a missing response. The turning point across these stimuli was then averaged for each participant.

The ATP of tone 2 responses: In the examination of Tone 2 responses, it was found that besides the expected peak of Tone 2 responses on the stimuli with turning points which occurred around the middle of a syllable, most participants showed a very consistent Tone 2 perception on stimuli for which the turning point started at or later than 348ms (75% of the syllable length). Figure 4-1 is an example of this pattern from a native Chinese participant in Creaky condition (NC24). The x-axis was the timing of the turning point, and the y-axis displays the numbers of the tone responses. As can be seen, the number of Tone 2 responses was quite high when turning points occurred earlier in a syllable, while it decreased when Tone 3 responses rose. This is not surprising since Tone 2 is associated with earlier turning points compared to Tone 3. Towards the end, however, there is a very abrupt increase of Tone 2 responses (almost exclusively) on stimuli with turning points occurring after 348ms. This phenomenon was found for 80% of NC, 62.5% of NE in Non-creaky condition and 100% of NC, plus 80% of NE in Creaky condition. This may be due to the fact that when the turning point occurs very late in a syllable, the ending slope tends to be very sharp.

Given the above, the ATPs of Tone 2 responses were divided into two parts. The first part consisted of all stimuli with turning points occurring before 348ms and receiving 100% Tone 2 response, while the second part included stimuli with turning points occurring at or later than 348 ms and receiving 100% Tone 2 responses. The ATP was then calculated by averaging the turning points from those stimuli in each part.

The average turning points for Tone 2 responses for the NC and the two NE groups (NE-High, NE-Low) are shown in Table 4-4. As can be seen in this table, the data suggests that for all three groups, the average turning points for Tone 2 responses were earlier when creaky voice was absent.

A paired-sample T-test was run on the data and the results confirm this. The NC's first average turning point was 10.2 ms earlier in Non-creaky (73.9 ms) condition compared to Creaky condition (84.1 ms, p=.003), and there is no difference in the 2nd ATP for NC (both at 368.6ms). Although descriptively, the data seem to hold true for both the NE-High and NE-Low groups, the data, nonetheless, show no significant differences in both of the ATPs within each group going from Non-creaky to Creaky. When these two groups are combined (see Table 4-5), however, the differences (9 ms and 8 ms) in both ATPs are significant ([t(42)=-.845, p=.003]; t(66)=-3.748, p=.00). In other words, both of the average turning points of those stimuli that received 100% Tone 2 responses were 9 and 8 ms later when creaky voice was present for Native English speakers. In summary, then, both native and non-native speakers of Chinese exhibit later turning points for Tone 2 responses in Creaky condition.

The fact that Tone 2 perception occurs at a later turning point for creaky stimuli suggests that the presence of creaky voice may have perceptually shortened the initial falling portion of the tonal contour before it rises (i.e., the turning point) among listeners.

The ATP of tone 3 responses: The average turning points for Tone 3 responses for the NC and the two NE groups (NE-High, NE-Low) are shown in Table 4-6. As can be seen in this table, the data suggests that for all three groups, the average turning points for Tone 3 responses are also earlier in the Non-creaky condition.

A paired-sample T-test was run for the ATPs within each group and the results were similar to the results from the ATPs of Tone 2 responses. The NC group showed a significant difference ([t(55)=-2.59, p=.012)in Tone 3 ATPs across conditions—18.2 ms earlier in Non-creaky (193.2ms) than in Creaky (211.4ms). Neither of the NE groups showed a significant difference in their Tone 3 ATPs across conditions (p=.43 for NE-High; p=.39 for NE-Low); however, when these two groups were combined (see Table 4-7), the difference (18.4 ms) in their Tone 3 ATPs was significant (Non-Creaky ATP for NE: 175.6 ms; Creaky ATP for NE: 194.0 ms. [t(59)=-1.835, p=.02]). In other words, disregarding tone proficiencies, both native and non-native speakers of Chinese show a later turning point for Tone 3 response in Creaky condition. These results, similar to those from the Tone 2 ATPs, again suggest that the perceptual boundary seems to occur at a later point when creaky voice is added.

Table 4-8 summarizes the results from this section of average turning points. Notice that the data for NE here is a combined result from the NE-High and NE-Low groups. The differences in ATPs of Tone 2 and Tone 3 responses across conditions are all significant, and they suggest that the presence of creaky voice does affect Tone 2

and Tone 3 perceptions for both native and non-native listeners. In particular, the duration of the initial falling of a tone contour seems to be perceptually shorter when there is creaky voice present. It is unclear why the differences in Tone 2 and Tone 3 ATPs between the two groups (NC and NE) are so close to each other (Tone2: NC-10.2ms, NE-9ms; Tone 3: NC-18.2ms, NE-18.4ms). One interpretation could be that the extent of how much creaky voice shortens perception is similar among both native and non-native listeners.

The analysis of the average turning point shows how creaky voice plays a role in the perception of Tone 2 and Tone 3 among both native and non-native Chinese speakers. Previous research has shown longer duration of a syllable may yield more Tone 3 identification compared to Tone 2 (Blicher et al., 1990) and smaller ΔF_0 has often been found with Tone 2 rather than Tone 3 (Shen & Lin, 1991). Since the stimuli in the current study did not vary in their duration, nor ΔF_0, what we have surmised from the results (i.e., that creaky voice makes the turning point occur at a later point) indicates that the presence of creaky voice may have perceptually shortened the initial falling portion of the tonal contour before it rises (i.e., the turning point) among both native Chinese and English listeners. Figure 4-2 describes the effect of adding creaky voice to the stimuli. The x-axis represents the timing of turning points, and the two tone contours could be either Tone 2 or Tone 3 in both conditions. Regardless of whether the tone is Tone 2 or Tone 3, the responses were triggered with a later turning point when creaky voice was present.

In the following section, a data analysis of binary logistic regression is presented. This test was used to ascertain the predicted 50% crossing point (CP) of Tone 2 and Tone 3 categorial perception among listeners.

Binary Logistic Regression for 50% Crossing Point (CP)

Binary logistic regression is used to study the probability of a certain event's occurrence when there are two outcomes (e.g. pass/fail, life/death, either ordinal or nominal responses). The purpose of the test is to determine how a response (outcome) changes according to one or more independent variables. For example, if we wanted to find out how hours of training affect test results--pass or fail, then the independent variable would be hours of training, and the two outcome values would be pass or fail. For the research reported herein, for each participant, the tone response outcome has two values, either Tone 2 or Tone 3 (after the elimination of Tone 1 and Tone 4 responses).

Logistic regression works better here because our participants only listened to the same stimulus six times. If the method of drawing lines across the 50% had been used, it would have created some uncertain situations. For instance, as shown in the graphs (Figure 4-3 and Figure 4-4), it can be very clearly seen that for participant NC15 (Figure 4-3), in Creaky condition, the 50% crossing point where Tone 2 responses switched to Tone 3 was at stimulus 3 and vice versa at stimulus 30. Those two stimuli can then be converted to the millisecond of the turning point: 78ms and 348ms. This means that for this participant, if the turning point of the dip occurs between 17% (78/460 ms) and 75.7% (348/460 ms) of the total syllable length, the stimulus is heard as Tone 3. If the turning point occurs before 17% or after 75.7%, it is heard as Tone 2. However, if we look at Figure 4-4, which shows the data table from participant NC21 in Creaky

condition, it is very difficult to decide where exactly the 50% categorial crossover line is located since the participant's responses went back and forth between Tone 2 and Tone 3 among stimuli 6 to 8. Therefore, the drawing method cannot solve the problem for the current study.

Using binary logistic regression can allow us to predict where participants would have a crossover point if they heard stimuli an indefinite number of times. Thus, binary logistic regression was run for each participant in both conditions. Since the 34 stimuli in each condition differ only in their turning points, the timing of those turning points is the predictor of responses. In this case, for the two outcomes possible, for Tone 2 it would be 0 and for Tone 3 it would be 1. Logistic regression then creates a curve for each participant, in each condition, describing the probability that Tone 3 or Tone 2 is heard at any point on the continuum.

Recall that when examining the ATP of tone responses in the last section, many participants showed a very abrupt rise in Tone 2 responses with stimuli that have very late turning points (after 348ms). For the purpose of the binary regression, those stimuli were temporarily excluded here.

After each curve was created, the predicted 50% crossing point (CP) was then calculated by using the formula $x= -\alpha/\beta$ for each participant (Agresti & Finlay 2009, p. 485). X stands for the timing of turning points (58-368 ms), α and β represent the constant and B value. Figure 4-5 is an example of a binary regression curve. The x-axis is the timing of the turning point, and the y-axis displays the two outcomes—1 (Tone 3) and 0 (Tone 2). A logit of 0 (i.e. where the y-axis is crossed) means 50%, and this is how we calculate where this point is: $0= \alpha+ \beta X$; therefore, we have the formula

mentioned above: x= -α/β. When B>0, it means that the probability of the stimulus being perceived as Tone 3 increases as the location of the turning point increases. When B<0, the probability of the stimulus being perceived as Tone 3 decreases as the location of the turning point increases (in other words, Tone 2 is more likely to be perceived on stimuli with earlier turning points).

To give an example, Table 4-9 shows the results from the logistic regression run on participant NC24 in Non-creaky condition. The significant P-value (0.003, <.05) suggests that turning point is a strong predictor for this participant in responding with either Tone 2 or Tone 3. The constant (α) here is -3.108 and β (B) is .504. Therefore, X= -α/β= -(-3.108)/.504=57.6ms. The parameter β indicates whether the probability P(y=1) increases or decreases as X increases (see Agresti & Finlay 2009, p.484). When β>0, y increases as X increases; when β<0, y decreases as X increases. This means that for participant NC24, when there is no creakiness in the stimuli, it is predicted that when the turning point occurs before 12.5% (57.6/460ms) of the total syllable length, the stimulus will most likely be heard as a Tone 3; when the turning point occurs after 12.5% of the syllable, it will be heard as Tone 2.

Let us now take a look at the same person in the creaky condition. Using the same calculation with the above, the X here = -(-5.1)/.0043=118.6 ms, which is at 25.8% of the syllable. If we compare the predicted CPs in these two conditions for this participant (12.5% in Non-creaky; 25.8% in Creaky), the results suggest that by adding creakiness, the predicted crossing point where perception shifts from Tone 2 to Tone 3 seems to come later, as compared to conditions without creakiness. In other words, it can be

inferred that, when there is creaky voice, the actual length of the falling part in the token seems shorter to the ears.

The P-value for each participant was examined and it was found that not all p-values from the binary regression are significant. In Non-creaky condition, 90% of NC, 83% of NE-High, and 50% of NE-Low had significant p-value ($P<0.05$), which means that for those participants, turning point is a predictor of their tone response to a certain stimulus. In Creaky condition, the significant portions are 100% of NC, 72.2% of NE-High, and 50% of NE-Low. The final average perceptual boundaries (X) were calculated only with those that had a significant p-value.

As previously mentioned, $\beta>0$ or <0 represents different curves that have different shapes of contours. When $\beta>0$, the probability of the stimulus being perceived as Tone 3 increases as the location of the turning point increases, which means the X is the boundary of Tone 2 shifting to Tone 3 perception. When $\beta<0$, the probability of the stimulus being perceived as Tone 3 decreases as the location of the turning point increases, which means the X is the boundary of Tone 3 shifting to Tone 2 perception. Therefore, it is necessary to analyze the binary result in two different categories according to β. Table 4-10 summarizes the results of Binary logistic regression averaged from participants in the three groups (only those using turning points as a predictor of tone responses).

In Table 4-10, T-tests and one-way ANOVA were performed to see whether all the differences were significant. The results from one-way ANOVA showed that there is no significant difference among the three groups in each of the columns; therefore, an average is calculated at the bottom of each column in Table 4-10. In Non-creaky

condition, most of the participants' (33/53=62%) perceptual boundary of Tone 2 shifting to Tone 3 occurred at 76.6ms (16.7%) and the rest (38%) of the participants shifted their perception from Tone 3 to Tone 2 at the turning point of 361.1ms (78.5%); in Creaky condition, the participants' average perceptual boundary of Tone 2 to Tone 3 was at 160.7ms (35%). Within each of the groups, T-tests showed that the differences of the Tone 2 to Tone 3 boundary in Non-creaky condition are all significant for each of the groups (NC: p=.00, NE-High: p=.039, and NE-L: p=.009).

In summary, for native and nonnative listeners, the predicted perceptual boundary from Tone 2 to Tone 3 occurs later when creaky voice is present (from 76.6ms to 160.7ms).

Since some of the stimuli that have turning points later than 348ms were excluded from the binary logistic regression analysis, it is necessary to include them now. In Non-creaky condition, we see that the predicted perceptual boundary of Tone 3 to Tone 2 for some participants occurs at 361.1ms (78.5%). It falls right in the middle of 348 ms to 388 ms, as has been shown from the ATP analysis. In addition, creaky voice also made those few participants (from all three groups) have more consistent Tone 2 responses when the turning points occurred later than 348ms.

This means that for a listener (Chinese or English), if the turning point of the syllable occurs before 76.7 ms (or 16.7% of the syllable length) when there is no creaky voice, it will most likely be perceived as a Tone 2. If the turning point occurs after 76.7 ms (16.7%) and before 361.1ms (78.5%), it will most likely be heard as a Tone 3. When the turning point occurs after 361.1ms (78.5%), it will be heard as a Tone 2 again. However, when creaky voice is present in the syllable, the perceptual boundary of Tone

2 to Tone 3 is predicted to be at 154.7 ms (or 33.6% of the syllable length). Stimuli with turning points before that will be heard as Tone 2, otherwise as Tone 3.

These results clearly show that for both native and non-native listeners, the perceptual boundaries at which Tone 2 switches to Tone 3 seem to come later when there is creaky voice, which is consistent with the findings from analysis of the average turning points of Tone 2 and Tone 3 responses from the previous section.

Binary logistic regression was also run for each participant to see whether the stimulus type, i.e. whether there is creakiness or not, is a predictor of Tone 3 responses. Tone response (Tone 3= “1”, Tone 2= “0”) was the dependent variable and stimulus type was the independent variable. Not surprisingly, the results showed a significant result for 90% of the Native Chinese and 85% of the Native English speakers ($p<.05$). This means that for 90% of Chinese and 85% of English speakers, the likelihood of a Tone 3 response decreases with the presence of creaky voice. In other words, the presence of creaky voice is a predictor of a smaller number of Tone 3 responses.

In Chapter 5, the results from production experiments and perception are reviewed in relation to previous research and findings. Limitations and future direction are also addressed.

Table 4-1. Error rates for each tone category from Baseline task and mean scores

	#of sub	Mean Score (%)	Error Rate (%) Tone 1	Tone2	Tone 3	Tone 4	sig.?
NC	30	98.9	0.7	1.8	1.3	0.7	N
NE-High	18	94.9	2.5	8.1	8.1	1.7	Y
NE-Low	22	80.7	14.1	27.5	17.7	18	Y
NE(combined)	40	87.1	8.9	18.8	13.4	10.6	Y

Table 4-2. Percentage of creaky voice present in four tones for NC

	With Creaky (out of 900)	Percentage
Tone1	17	1.90%
Tone2	104	11.60%
Tone3	878	97.60%
Tone4	47	5.20%

Table 4-3. Percentages of four tone responses received on all stimuli from 3 groups.

	Non-creaky				Creaky			
	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T1 (%)	T2 (%)	T3 (%)	T4(%)
NC	7.7	25.4	60.9	5.4	9.4	40.7	42.9	6.4
NE-H	2.4	22.6	59.3	13.4	3.1	51.3	28.6	14.9
NE-L	8.4	30.4	46	13.6	11.3	47	25.4	14.3
Average	6.6	26.2	55.8	10	8.4	45.4	33.7	11.1

Table 4-4. Two Average Turning Points (ATP) for Tone 2 responses (ms) for three proficiency groups

	1st ATP of Tone 2			2nd ATP of Tone 2		
	Non-Creaky	Creaky	Change	Non-Creaky	Creaky	Change
NC	73.9	84.1	10.2	368.6	368.6	0
NE-High	90.3	101.3	9	359.1	367.4	7.6
NE-Low	97.4	105.5	8.1	360.4	368	8.3

Table 4-5. Two Average Turning Points (ATP) for Tone 2 responses for NC and NE (combined)

	1st ATP of Tone 2			2nd ATP of Tone 2		
	Non-Creaky	Creaky	Change	Non-Creaky	Creaky	Change
NC	73.9 (16.1%)	84.1 (18.3%)	10.2 (2.2%)	368.6 (80.1%)	368.6 (80.1%)	0
NE	94.2 (20.5%)	103.2 (22.4%)	9 (2%)	359.7 (78.2%)	367.7 (79.9%)	8 (1.7%)

Table 4-6. Average Turning Point (ATP) for Tone 3 responses (ms) for three groups

	Non-creaky T3	Creaky T3
NC	193.2	211.4
NE-High	178.6	187.3
NE-Low	188.5	201.3

Table 4-7. Average Turning Point (ATP) for Tone 3 responses (ms) for NC and NE (combined)

	ATP of Tone 3		
	Non-Creaky	Creaky	Change
NC	193.2	211.4	18.2ms
NE	175.6	194	18.4ms

Table 4-8. NC and Combined NE of their Tone 2 and Tone 3 ATPs (ms)

	1st ATP of Tone 2			2nd ATP of Tone 2			ATP of Tone 3		
	Non-Creaky	Creaky	Change	Non-Creaky	Creaky	Change	Non-Creaky	Creaky	Change
NC	73.9 (16.1%)	84.1 (18.3%)	10.2 (2.2%)	368.6 (80.1%)	368.6 (80.1%)	0	193.2 (42%)	211.4 (46%)	18.2 (4%)
NE	94.2 (20.5%)	103.2 (22.4%)	9 (2%)	359.7 (78.2%)	367.7 (79.9%)	8 (1.7%)	175.6 (38.2)	194 (42.2)	18.4 (4%)

Table 4-9. SPSS output table of binary logistic regression on NC24 (Non-creaky)

		Variables in the Equation					
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	TurningPoint	0.054	0.108	8.768	1	0.003	1.056
	Constant	-3.108	1.521	4.177	1	0.041	0.045

[a.] Variable(s) entered on step 1: TurningPoint

Table 4-10. Results of Binary logistic regression for three groups (predicted crossing point of perceptual boundary of Tone 2 and Tone 3).

	Non-creaky		Creaky	
	Tone2→Tone3 (B>0)	Tone3→Tone2 (B<0)	Tone2→Tone3 (B>0)	Tone3→Tone2 (B<0)
NC (N=30)	79.4ms (17%) (N=21) (90% sig.)	362.5ms (18.8%) (N=6)	143.9ms (31.3%) (N=30) (100% sig.)	(N=0)
NE-High (N=18)	63.9ms (13.9%) (N=8) (83% sig.)	344.5ms (75%) (N=7)	160ms (34.8%) (N=13) (72.2% sig.)	(N=0)
NE-Low (N=22)	84.4ms (18.4%) (N=4) (50% sig.)	376.5ms (82%) (N=7)	178.2ms (38.7%) (N=11) (50% sig.)	(N=0)
Average	76.7ms (16.7%)	361.1ms (78.5%)	160.7ms (35%)	

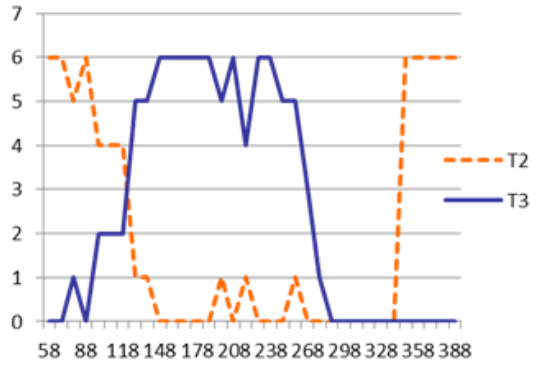

Figure 4-1. An example of having Tone 2 responses again after 348 ms (from NC24, in Creaky condition).

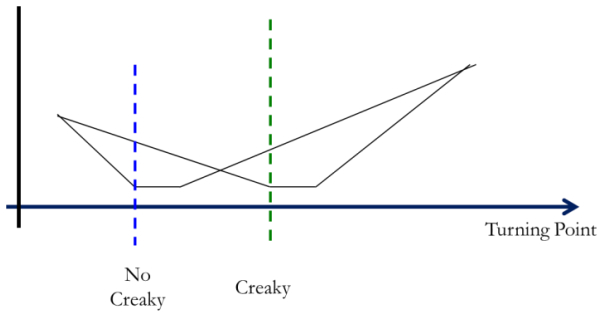

Figure 4-2. Illustration of effect of creaky voice on tone responses.

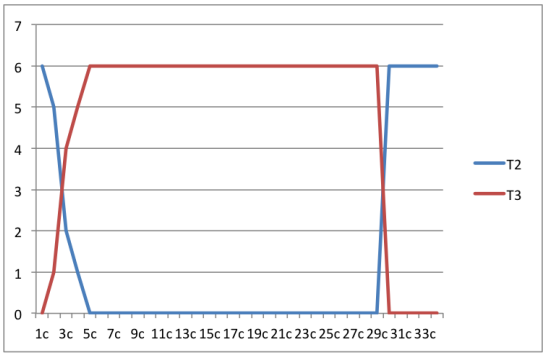

Figure 4-3. NC15 Tone 2/3 responses in Creaky condition

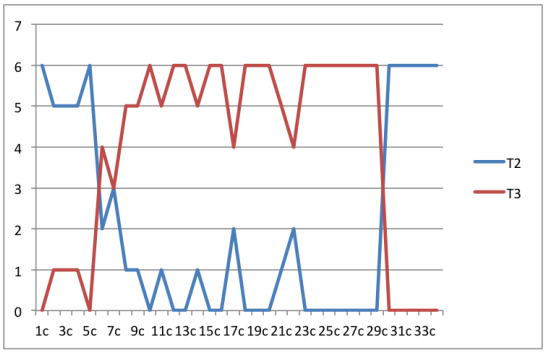

Figure 4-4. NC21 Tone 2/3 responses in Creaky condition

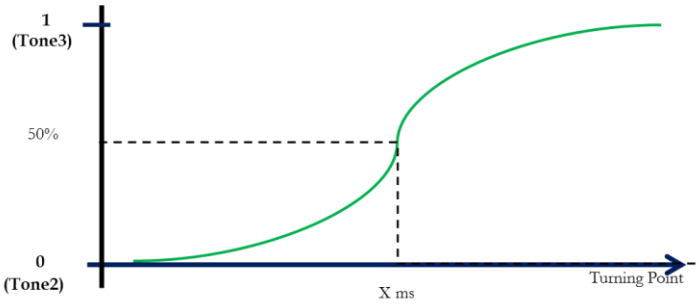

Figure 4-5. An example of probability curve of a binary regression analysis. (B>0)

CHAPTER 5
DISCUSSION AND CONCLUSION

The goal of this research has been to explore the perception of Mandarin Chinese Tone 2 and Tone 3 and the role of creaky voice among native speakers of Chinese (NC) and native speakers of English (NE) who are also Chinese language learners. In order to take L2 language proficiency into consideration, the non-native speakers were divided into two subgroups: NE-High and NE-Low. Thus, the participants were all placed in three groups: 30 native Chinese (NC) speakers, 18 high-proficiency native English speakers (NE-High), and 22 low-proficiency native English speakers (NE-Low). This chapter discusses the results of the current study and provides interpretations responding to each of the three research questions. Finally, the limitations inherent in a study like this are addressed and the future direction of research is suggested.

Research Question 1

Is creaky voice associated with a certain tone(s) produced by native Mandarin Chinese speakers (NC)? In the present study, production data from 30 native Chinese speakers (16 females and 14 males) were collected. The number of participants exceeded that of studies conducted previously (e.g. Belotel-Grenie, A & Grenie, M. 1994, 1995, Yu 2010). Each of the recordings included 120 words with 40 different words, with each word repeated three times. In the 40 target words, six initials [m, n, p^h, p, k^h, k, t] and four vowels (front vowel /i/, central vowel /a/, and back vowels /ɤ, u/) were used and there were 10 words for each tone category.

Recordings of all the target words (120*30=3600) were examined for the presence of creakiness. The results confirmed that creaky voice is extremely common in the production of Mandarin Tone 3 (97.6%). This supports previous research that creaky

voice is indeed associated with Tone 3, but we observed a much higher frequency in the present study. This is probably due to the fact that the number of the participants in the current study was much greater than previous studies and that more combinations of initials and finals were included in the target words. In addition, we determined that the second highest tone category that has creaky voice present is Tone 2—11.6%. This is different from the results of other studies such as those reported in Grenie's 1995 study in which the second highest category was Tone 4 (10.5%). However, due to the limited number of participants in Grenie's study, the results from the present study would seem to have more validity.

Research Question 2

How does the presence of creaky voice affect the perception of Tones by native Chinese speakers (NC) and L2 learners of Chinese (NE), respectively? To answer this question, the percentages of different tone responses received for each stimulus from the three groups in each condition were calculated. The results (see Table 4-4, Chapter 4) show that when there is no creaky voice, all three of the participating groups (NC, NE-High, NE-Low) had more Tone 3 than Tone 2 responses. On the other hand, when creaky voice is present, all three groups showed an increase in their Tone 2 responses and a decrease in their Tone 3 responses. This change was particularly noticeable in the two NE groups in that they had more Tone 2 responses than Tone 3 in Creaky condition. At the same time, the responses in the two categories were the same in the NC group.

When we look at the differences across conditions, within each of the three groups, the presence of creakiness resulted in more Tone 2 and fewer Tone 3 responses. This holds true for all native and non-native speakers and may indicate that

creaky voice is not a primary cue for identification of Tone 3 for either native or non-native listeners. If it were a primary cue, then the presence of creaky voice would have led to more Tone 3 responses.

Which group was affected the most by the presence of creaky voice? If we examine the percentages of tone response changes from Non-creaky to Creaky condition, we can see that the NE-high group is affected the most by the presence of creaky voice, resulting in more Tone 2 (+127%) and less Tone 3 (-30.7%) responses, and NC is the least affected (Tone 2: 15.3% increase; Tone 3: -18% decrease). This suggests that native English speakers are more affected by creaky voice than native Mandarin speakers in their perception of Tone 3. In particular, it appears that native English listeners do not use creaky voice to facilitate their identification of Tone 3.

This is, in fact, somewhat different from what previous research might have predicted. In Grenie's (1997) study, ten native Chinese speakers participated in a perception test with partially presented words with Tone 3 with and without creaky voice. The results showed that a Tone 3 produced with creaky voice was more quickly recognized (beginning at 60% of the duration of a syllable) than a Tone 3 without creakiness (70% of the duration). Their study suggested that creaky voice is a secondary indicator of Tone 3. It would thus be a reasonable prediction that creaky voice would facilitate tone 3 identification; however, this is not supported by the results from the current research.

Several reasons may contribute to this discrepancy. First, the methodology in the current study is different from that in the previous study. In the current study, the stimuli consisted of complete syllables when they were presented and they differed not only in

their creakiness but also in the timing of the turning points. On the other hand, the stimuli in Grenie's study (1997) were partially presented to see how fast an accurate Tone 3 recognition occurs. The two experimental designs were different from each other because their research goals were not the same. The current study aims to determine whether stimuli with creaky voice will generate more Tone 3 responses while the other study attempted to find out the earliest recognition point before a complete syllable unfolds itself. Secondly, the stimuli from the previous study were from naturally recorded production from native speakers. Therefore, for each stimulus (either partially or fully presented) there was a "correct" or expected tone response. In the current study, by contrast, all the stimuli were manipulated in their timing of turning point and presence of creaky voice. Thus, there were no right or wrong answers for each tone response. It appears that there might have been some interaction between the effect of phonation type and tone contour such that when the tone contour is fully presented with creakiness, the creakiness affects people's tone perception, particularly among non-native listeners. Grenie's (1997) study tested only native Mandarin speakers; therefore, whether creakiness would make the Tone 3 recognition point early remains unknown.

Jongman & Moore (2000)'s study suggested that native Mandarin listeners could use their language background to help in distinguishing tone contrasts that had variations in F_0 and speaker rate, but native English listeners still used only acoustic variation as a cue to discriminate phonemic contrast. These results seem to be supported by the current study in that the creaky voice was not used as a helpful cue but rather took on the role of a limited perceptual resource for listeners.

If creaky voice is indeed a primary perceptual cue, then it should make the Tone 3 identification easier with this extra cue and produce more Tone 3 responses. Contrary to that expectation, however, the presence of creaky voice actually reduced the total number of Tone 3 responses and more Tone 2 responses were generated. Does this mean creaky voice is a redundant cue in Tone 3 perception? Does it mean that adding creaky voice will only confuse listeners? Although this may appear to be the case, this phenomenon is not as easily explicable as it may seem. To understand the role of creaky voice, we need to take a closer look at the perceptual boundary of the tone continuum in the next question.

Research Question 3

What is the perceptual boundary of the Tone 2-Tone 3 continuum perceived by native Chinese speakers (NC)? How does it differ from L2 learners of Chinese (NE) with different proficiencies? What is the role of creaky voice in their perception? From the results obtained from calculating Average Turning Points (ATP) for both tone responses, we found that for both native and non-native speakers, the average location of the turning point is later when creaky voice is present. This holds true for both Tone 2 and Tone 3 responses. There was no significant difference in the two native English groups; therefore, their L2 proficiencies do not seem to make a difference in the average location of the turning points for Tone 2 and Tone 3 perception.

In addition, the current study also found that when the turning point occurs late in a stimulus (on or after 348ms in the current experiment), the stimulus will be identified as a robust Tone 2 from most listeners, regardless of their native languages. Since the dip in all stimuli in the current research was kept at 68ms, the real rising of those tone

contours actually occurred at 416 ms (348+68), which was about 90% of the duration. This steep rising appears to override all other parameters in a stimulus, including the presence of creakiness. This is also different from the findings of a previous study conducted by Cao & Sarmah (2007), in which stimuli with turning points later than 72.5% of the syllable were identified as Tone 4. An explanation of this would be the different parameters used in these two studies. In the 2007 study, the F_0s of stimulus onset, turning point, and offset were 186Hz, 163 Hz, and 211Hz, and the syllable duration was 400ms. In the current study, however, the corresponding F_0s are 203Hz, 170.5Hz, and 253 Hz, with the duration of syllable at 460ms. In fact, the angle at which the rising occurred in the present study was much steeper compared to the previous one. Therefore, the results do not contradict the previous finding in that study.

These data were calculated from the tokens that received 100% either Tone 2 or Tone 3 responses. Thus, these average turning points indicate the average value of the location of turning point from those stimuli that are perceived as a certain tone. For instance, the ATP for native Chinese speakers in Non-creaky condition is 16.1%. This means that when there is no creaky voice, a typical stimulus that will be perceived as a Tone 2 is the one that has its turning point at 16.1% of the syllable. However, if creaky voice is present, the turning point of this typical stimulus will move to 18.3% of the syllable. This change for a Native English listener would be from 20.5% to 22.4%. Similarly, when creaky voice is absent, a typical stimulus that will be heard as a Tone 3 by Native Chinese listeners is the one with a turning point at 42%. When there is creakiness, the turning point shifts to 46%. This shifting pattern for Native English listeners would be from 38.2% to 42.2%.

To summarize, the presence of creaky voice makes the average turning point appear later, regardless of the listeners. In other words, creaky voice makes the duration of the initial falling of a tone contour seem to be perceptually shorter to both native and non-native ears. This finding is particularly enlightening because none of the previous studies we have seen appear to make similar discoveries.

To predict the perceptual boundary of the Tone 2- Tone 3 continuum, Binary Logistic Regression was used in the present study. This method of calculating is different from that used previously in similar research (e.g. Cao & Sarmah 2007, Yang 2011). It is, in fact, a better approach. In perception experiments, the same stimulus is usually repeated only a limited number of times (6 times in the present study). As shown in Chapter 4 (figure 4-4), using the traditional 50% categorical crossover line to see the boundary might not be applicable to all participants due to the limited number of responses to one stimulus. Binary logistic regression is a more suitable method here because it provides us with a predicted value of that 50% categorical crossover line if there is an indefinite number of repetitions of a stimulus and an indefinite number of responses. This is more accurate and better applies to our data.

The results from binary logistic regression show that for both native and non-native listeners, the perceptual boundary of Tone 2-Tone 3 shifts to a later point when creaky voice is added. There is also no significant difference between native and non-native speakers on those boundaries. When creaky voice is absent, the boundary of shifting Tone 2 to Tone 3 perception happens at 16.7% of the syllable. When creaky voice is added, the perceptual boundary shifts to 35% of the syllable.

The results here seem to be different from the previous study (Cao & Sarmah, 2007) that the crossing point of the Tone 2- Tone3 perception happened at 42.5% of the syllable. This can be attributed to the following factors. First, the parameters of the stimuli are different in these two experiments. As previously mentioned, the shape of the tone contours in the stimuli used in the two studies are not the same, which would result in different angles of rising/falling contours even when the turning points are the same. Secondly, the methods for calculating perceptual boundaries are very different in these studies. The previous study used the traditional method of drawing a 50% crossing point, which could only represent the actual value collected from those participants listening to limited numbers of stimuli from a certain study. The current research, on the other hand, utilized a better approach—binary logistic regression-- to calculate the predicted boundary as if participants were listening to an unlimited number of stimuli.

In summary, the role of creaky voice in Mandarin Tone 2 and Tone 3 perception is neither critical nor redundant. Creaky voice is not a primary cue when it comes to Tone 2 and Tone 3 distinction because its presence does not predict more Tone 3 responses. However, creaky voice is not a redundant perceptual cue because it does affect the perception of certain tone contours when the turning point falls within a certain range. Non-native listeners seem to be more influenced by this perceptual cue.

In addition, the presence of creaky voice seems to shorten the perceptual duration of the initial falling of a syllable in both native and non-native listeners. The psychoacoustic effect of creaky voice has not been studied intensively in previous research; however, it has been suggested that non-modal vowels are associated with longer duration phonetically compared to their modal counterpart (Gordon 1998), such

as in Kedang and Jalapa Mazatec. This may, perhaps, help to explain the perceptual shortening found in the current research. If in some languages, a creaky vowel has to be longer than a clear vowel, it implies that listeners perceive the vowel with creakiness as the same vowel without it. Therefore, the fact that the turning point occurs at a later time when creaky voice is present does not seem surprising at all.

To ascertain whether this is also true with for Mandarin tone production, a sample of Tone 3 tokens was taken from the Native speakers' production data in the current study. Two sets of words were chosen—produced with and without creaky voice. They were compared in terms of their syllable duration and timing of the turning points. The results did not conform to previous research. The durations of the words produced with creaky voice tended to be shorter (413ms) than those that are clear (567ms, p=.00). The average timing of the turning points in those creaky words occurred at 50% of the syllable, later than that in clear words (40.5%, p=.00). This was also different from what we found in the perception experiment in the current study, where the falling part of a tone contour seems shorter when creaky voice is added. It has been established that perception precedes production and phonetic perception is not necessarily equivalent to phonological interpretation (Dinnsen 1985). The discrepancy found in the current study provides additional evidence for this. In addition, in future research, it would be more beneficial to examine native English speakers' production data and see how it relates to their perception.

The current study also has some implications for the speech models that were reviewed in Chapter 2. Those models can be extended to account for the suprasegmental features and phonation types as explored in the current research. For

example, the Speech Learning Model predicts that when an L2 sound is very different from the closest L1 sound, a new category will likely to be established. According to the SLM, if native English speakers can perceive how Mandarin Tone 3 is different from Tone 2, it is very likely that the two different tone categories will be built in their L2 phonological inventory. However, previous research (e.g. Wang 2006) and the current study both suggest that L2 learners of Mandarin tones do not use perceptual cues in the same way that native Chinese speakers do. The fact that the L2 speakers in the current study are more affected by the presence of creaky voice suggests they do not use pitch contour as the primary cue for tone recognition, or to the same extent as native speakers. At least with their current L2 proficiency levels, creaky voice seems to have a negative influence on their ability to fully perceive pitch contour. One of the predictions by the Perceptual Assimilation Model is that a non-native segment could be assimilated to the native category if it is heard as a good, acceptable, or deviant exemplar from the native category. In our L2 speakers' native language, English, there is no lexical tone category. The closest thing that tone can be related to is the intonation that is used at sentence level. Lexical tones would fall out of their native phonological space and thus they are predicted to pose potential difficulty for English speakers. In this case, degree of perceptual difficulty is predicted by PAM to depend on the phonetic saliency between the two tones. As both SLM and PAM agree, L2 listeners will continue to refine their perception, and when their L2 proficiency is more native like, the differences between Mandarin Tone 2 and Tone 3 will be more salient. PAM would then predict that an L2 learner will perceives tones with a moderate to good level eventually. The Native Language Magnet model predicts that if an L2 sound is very close to a native language

prototype, it will be difficult for L2 learners to perceive the phonetic contrast. Since our L2 learners do not have lexical tone prototypes in their native language, it is difficult for them to extract pitch contour information to perceive tones. Depending on the psycho-acoustic distances between tones, the NLM would predict that Tone 2 and Tone 3 are difficult to distinguish, since they are acoustically very similar to each other. Tones have to be mapped differently from their current phonological space in order to be perceived in the same way as native speakers of Mandarin.

Conclusions, Limitations, and Future Directions

The present study explored Mandarin Tone 2 and Tone 3 perception among native and non-native speakers and the role of creaky voice in their perception of these tones. The location of the turning points that mark the perceptual boundaries between Tone 2 and Tone 3 are 16.7% (no creaky) and 35 % (with creaky) for native Chinese and English listeners. When the turning point occurs after 90% of the syllable, perception reverts to Tone 2. Creaky voice is not a primary cue for Tone 3 perception (although it is frequently found in production of Tone 3), but it is not redundant. It shortens the perceptual length of the initial falling of a tone contour, making the perceptual boundary shift earlier for both native and non-native listeners.

There are some limitations in the present study that are relevant for future research. In the perception experiment, for example, stimuli with only vowel /a/ are included. Future studies could consider having more vowels in the perceptual stimuli to see if they affect perceptual boundaries. In addition, a secondary confirmation of the creaky voice inspection in the production data, such as examining the FFT of each word, would allow for stronger claims.

In conclusion, our study not only explored the perception of Mandarin Tone 2 and Tone 3 to a finer extent, but also examined the role of creaky voice in the perception of tones among both native and non-native listeners. The results provided some interesting findings in the psychoacoustic domain showing that voice quality could interact with perception of certain sounds. It is hoped that the research reported herein contributes to a better understanding of the perceptual cues used in determining Mandarin tones for both native speakers and L2 learners. With regard to the teaching of Chinese to L2 learners, perhaps the findings here will be of use in the development of teaching plans or chapters in Chinese L2 textbooks that focus on how to listen more effectively for Chinese tones.

APPENDIX
FORTY WORDS USED IN TONE BASELINE TASK (WITH PINYIN +TONE MARKS)

八ba1
杯bei1
车che1
吃chi1
东dong1
多duo1
分fen1
喝he1
家jia1
开kai1
来lai2
门men2
钱qian2
人ren2
十shi2
台tai2
玩wan2
学xue2
谁shei2
别bie2
比bi3
我wo3
打da3
懂dong3
给gei3
好hao3
几ji3
可ke3
李li3
请qing3
四si4
问wen4
下xia4
姓xing4
又you4
在zai4
这zhe4
课ke4
近jin4
会hui4

LIST OF REFERENCES

Abramson, A. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. International Journal of American Linguistics, 28 (2). Bloomington: Indiana University Research Center in Anthropology, Folklore and Linguistics.

Abramson, A. (1975). The tones of central Thai: Some perceptual experiments. In J.G. Harris & J. Chamberlain (Eds). *Studies in Thai linguistics.* Bangkok: Central Institute of English Language. pp. 1-16.

Abramson, A. (1976). Thai tones as reference system. In *T. Gething, J. Harris, and P. Kullavanijaya (Eds.), Tai linguistics in honor of Fang-Kuei Li*. Bangkok: Chulalongkorn University Press, pp, 1-12.

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech 21*: 319-25.

Belotel-Grenie, A. & Grenie, M.. (1994). Phonation types analysis in Standard Chinese. *Proceedings of ICSLP'94, Yokohama,* Japan, pp. 343-346.

Belotel-Grenie, A. & Grenie, M. (1995). Consonants and Vowels Influence on Phonation types in isolated words in Standard Chinese*. Proceedings of XIIIth ICPhS, Stockholm,* Sweden, pp. 400-403.

Belotel-Grenie, A. & Grenie, M.. (1997). Types de phonation et tons en chinois standard. *Cahiers de Linguistique- Asie Orientale, 26(2):* 249-279.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language research*, York Press: Baltimore, pp. 171-203.

Best, C. T. and Tyler, M. D. (2007) Nonnative and second-language speech perception: commonalities and complementarities. In the *Language Experience in Second Language Speech Learning: In honor of James Emil Flege (*edited by Bohn, O-S and Munro, M), pp. 213-234.

Blicher, D. L., Diehl, R., and Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics, 18,* 37–49.

Broselow, E. Hurtig, R.R. and Ringen, C. (1987). The perception of second language prosody . In G. Ioup and S.H. Weinberger (eds.). *Inter-language Phonology, the Acquisition of Second language Sound System.* Cambridge: Newbury House Publishers, pp. 350-361.

Burnham, D. & Mattock, K. (2007) The perception of tones and phones. In *the Language Experience in Second Language Speech Learning: In honor of James Emil Flege (*edited by Bohn**,** O-S and Munro,M**),** pp. 259-279.

Cao, R. & Sarmah, P. (2007). A Perception Study on the Third Tone in Mandarin Chinese. *UTA Working Papers in Linguistics (2007),* 2, 50-66.

Chao, Y. R. (1948). Mandarin primer. Cambridge: Harvard University Press.

Chen, G. T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics, 2,* 159-171.

Chuang et al. (1972). The acoustic features and perceptual cues of the four tones of standar colloquial Chinese*. In Proceedings of the 7th International Congress of Acoustics (Vol. 3).* Budapest: Akademial Kiado. pp. 297-300.

Chang, Y.S. (2011). Distinction between Mandarin Tone 2 and Tone 3 for L1 and L2 listeners. *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23), 1*, 84-96.

Davison, D.S. (1986). An acoustic study of so-called creaky voice in Tianjin Mandarin. *Working Papers in Phonetics, UCLA, 78,* 50-57.

Dinnsen, D. A. (1985). A re-examination of phonological neutralization. *Journal of Linguistics,* 21, 265-279.

Flege J. & Eetfing W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics* 15: 67-83.

Flege J. & Eetfing W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category information. *Journal of the Acoustic Society of America 83*: 729-740.

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*, York Press: Baltimore, pp. 233-277.

Flege, J. E. (2002). Interaction between the native and second-language phonetic systems. In *An integrated view of language development.* (edited by Burmeister, P. et al.) pp. 217-243.

Fork, C.Y.-Y. (1974). A perceptual study of tones in Cantonese. Centre of Asian Studies, University of Hong Kong.

Fromkin, V.A. (1978). Tone: a linguistic survey. NY: Academic Press. Inc.

Fu, Q.J. and Zeng, F.G.(2000). Identification of temporal envelop cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing 5,* 45-57.

Gandour, J. (1978). The perception of tone. In *Tone: a linguistic survey (Fromkin, 1978)*. NY: Academic Press. Inc.

Gandour, J. T. (1983). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics, 12,* 235-261.

Gandour, J. et al. (2000). A cross-linguistic PET study of tone perception. *Journal of Cognitive Neuroscience, 12(1),* 207-222.

Gordon, Matthew. (1998). The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. *Berkeley Linguistics Society 24.* 93-105

Hombert, J. M. (1976). Consonant types, vowel weight and tone in Yoruba. *UCLA Working Papers in Phonetics, 33,* pp. 40-54.

Joogman, A. and Moore, C. (2000). The role of language experience in speaker and rate normalization process. *Proceedings of the 6th international conference on Spoken Language Processing,* I. 62-65.

Jongman, A., Wang, Y., Moore, C.B., & Sereno, J. A. (2007). Perception and production of Mandarin Chinese tones. In the *Language Experience in Second Language Speech Learning: In honor of James Emil Flege (*edited by Bohn, O-S and Munro,M**),** pp. 209-217.

Halle, P.A, Chang, Y.C, Best, C.T. (2003) Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics 32,* 395-421.

Halle, P.A, Chang, Y.C, Best, C.T. (2004) Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners**.** *Journal of Phonetics* Volume 32, Issue 3, July 2004, pp.395-421.

Keating, P.A. & Esposito, C. (2006) Linguistic Voice Quality. *UCLA Working Papers in Phonetics, No. 105*, 85-91.

Klatt, D. (1973). Discrimination of fundamental frequency contours in synthetic speech duplications for models of pitch perception*. Journal of the Acoustical Society of America, 53,* 8-16.

Klein, D. et al. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage 13*, 646-653.

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

Kuhl, Patricia K. (1992). "Infants' perception and representation of speech: development of a new theory", In *ICSLP*-1992, 449-456.

Kuhl, P. K. and Iverson, P. (1994). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Acoustic Society of America,* 553- 562.

Massaro, D.W. et al. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13,* 267-289.

Miracle, W.C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of Chinese language teachers association*, *24,* 49-65.

Moore, C.B., and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America, 102,* 1864-1877.

Liberman, A.M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117-123.

Liberman, A.M. Cooper, F.S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Ohala, J. J. (1978). Production of tone. In *Tone: a linguistic survey (Fromkin, 1978)*. NY: Academic Press. Inc.

Sebastian-Galles N. (2005). Cross-language speech perception. In *the Handbook of Speech Perception (Pisoni, D. & Remez, R. editors)*, pp. 546-566. London: Blackwell Publishing.

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

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

Strange, W. (2007). Cross-language phonetic similarity of vowels: theoretical and methodological issues. In the *Language Experience in Second Language Speech Learning: In honor of James Emil Flege (*edited by Bohn**,** O-S and Munro,M**),**

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

Van Lancker, D., & Fromkin, V. (1973). Hemispheric specialization for pitch and “tone”: Evidence from Thai. *Journal of Phonetics, 1,* 101-109.

Vance, T.J. (1977). Tonal distinctions in Cantonese. *Phonetica*, 34, 93-107.

Wang, W. S. –Y. (1976). Language change. *Annals of the New York Academy of Sciences, 28,* 61-72.

Wang, Y., Sereno, J. A., Jongman, A., and Hirsch, J. (2000). Coritcal reorganization associated with the acquisition of Mandarin tones by American learners: An fMRI study. *Proceedings of the 6th International conference on Spoken Language Processing II.* 511-514.

Wang, Y., Jongman, A. and Sereno, J. A. (2006). L2 acquisition and Processing of Mandarin Tone. In Li,P. et al. (eds.) *Handbook of Chinese Psycholinguistics.* Cambridge University Press.

Wang, Y., Jongman, A., and Sereno, J.A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language 78,* pp.332-348.

Wayland, P. R. (2007). The relationship between identification and discrimination in cross-language perception. In the *Language Experience in Second Language Speech Learning: In honor of James Emil Flege (*edited by Bohn**,** O-S and Munro,M), pp. 201-217.

Whalen, D. H. and Y. Xu. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49,* 25-47.

White, C. M. (1981). Tonal perception errors and interference from English intonation. *Journal of Chinese Language Teachers Association,* 16, 27-56.

Yang, R. X. (2011). The phonation factor in the categorical perception of Mandarin Tones. *ICPhS XVII*, 2204-2207.

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

Yu, K. M. (2010). Laryngealization and features for Chinese tonal recognition*. In Interspeech-2010*,1529-1532.

Zue, V. (1976). Some perceptual experiments on the Mandarin tones. *Paper presented at the 29th Meeting of the Acoustical Society of America,* San Diego, California.

BIOGRAPHICAL SKETCH

Rui Cao was born and grew up in Tianjin, China. She went to NanKai University in her hometown and received her BA in English for foreign trade in 2001. She worked at Tianjin IEILTS Training Center for a year as an English teacher after graduation from college. Rui came to the US in 2002. She earned her MA in English (Teaching English as a Second Language) from the University of Toledo, where she also worked as a Teaching Assistant teaching college ESL courses. In 2004, Rui started her PhD study in linguistics at the University of Florida (UF). Her interests include phonetics, tones, second language acquisition, foreign language teaching and learning, and neurolinguisitcs. During her PhD study, Rui taught English at the English Language Institute (ELI) at the University of Florida from 2005-2007, 2011-2012, and undergrad course of Mandarin Chinese at UF from 2008-2009. She then worked as an Assistant Professor in the Chinese Department of the Defense Language Institute (DLI) for two years in Monterey, California. In spring 2012, she was awarded the CLAS Dissertation Fellowship from the University of Florida. She received her PhD in linguistics from the University of Florida in summer 2012. After graduation, Rui is going to move to Texas to join her husband.