Tones in Whispered Chinese: Articulatory Features and Perceptual Cues

by

Man Gao
B.A., Zhongshan University, 1999

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

MASTER OF ARTS

in the Department of Linguistics

We accept this thesis as conforming
to the required standard

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

____
Dr. Hua Lin, Departmental Member (Department of Linguistics)

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

____
Prof. Jimmy G. Harris, External Examiner (Adjunct, Department of Linguistics,
University of Victoria)

Supervisor: Dr. John H. Esling

ABSTRACT

Whisper is a common modality of speech communication, regardless of language. In most phonetic studies, descriptions of the larynx during whispering concentrate on the horizontal plane of the glottis. A review of linguistic studies on whispered speech in tone languages shows that there is more concern about people's ability to perceive tones in whispered speech than about how they perceive them, given that the most important perceptual cue, F0, is absent. In order to explain how pitch information is conveyed in the absence of F0, therefore, this study investigates how whispered tones are produced from the perspective of a broader view of the larynx and what contributes to Mandarin Chinese tone recognition in whisper.

Three experiments were conducted in this study. Laryngoscopic endoscopy was used in the first experiment to observe the pharynx and larynx when Mandarin tones are produced in phonated and whispered speech. The second experiment involved data collection for acoustic analysis. Two male and two female native speakers of Mandarin produced three sets of syllables with four contrasting tones in isolation and in sentential environments. The data collected were analyzed acoustically with speech-analysis software and played as stimuli to ten native speakers of Mandarin in the third experiment, a perception test.

A number of conclusions were drawn based on the results of these experiments: (1) the laryngeal sphincter mechanism is found to be a principal contributing physiological maneuver in the production of whisper, emphasizing the vertical rather than

the horizontal component of the laryngeal source; (2) two special behavioral maneuvers are also used in whisper: male speakers tend to lengthen vocalic duration and female speakers tend to exaggerate the amplitude contours of Tone 3 and Tone 4; (3) these two special behavioral maneuvers and two temporal envelope parameters contribute to tone recognition in whisper, but the phonetic context is shown to be a distraction; (4) the environments of the target tones cause perceptual differences, and the ranking of these environments in order of increasing degree of difficulty is: isolation, sentence-final, sentence-medial and sentence-initial; (5) the ranking of the four tones in isolation, in order of increasing degree of perceptual difficulty is: Tone 3, Tone 4, Tone 1 and Tone 2.

This is the first time that empirical evidence has been gathered to test the notion of ‘special maneuver’ proposed by linguists in the 1970s; it is also a first look in depth at the articulatory production mechanism for whisper. This study contributes to our knowledge of the phonetic production and linguistic realization of whispered tones in Mandarin Chinese by providing a complete description of the articulatory mechanism and a comprehensive report on how they are perceived in various environments.

Examiners:

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

Dr. Hua Lin, Departmental Member (Department of Linguistics)

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

Prof. Jimmy G. Harris, External Examiner (Adjunct, Department of Linguistics, University of Victoria)

ACKNOWLEDGEMENTS

It would be a mission impossible to finish this M. A. thesis without the supports from all my professors, colleagues, friends and family.

My first and deepest gratitude goes to my supervisor, Dr. John Esling, who kindles my interests in phonetics, who inspires me with this wonderful topic, and whose participation in this project composes the most valuable part of this work. His versatility, precision and patience are strong drives in my every step forward, and will always impel me in my future studies. I am also indebted to Professor Jimmy Harris, for his precious comments and insightful advice. My big thanks go to Dr. Hua Lin, for offering her constant support and having confidence in me; and to Dr. Daniel Bryant for his helpful suggestions on my work. I must also acknowledge Professor Li Shaoni for his valuable Bai speech data.

I am thankful for the Department of Linguistics for not only offering me a comprehensive graduate program, but also a warm home on the other side of the Pacific. I am obliged to Dr. Leslie Saxon and Dr. Suzanne Urbanczyk for their understanding, encouragement and support in many ways throughout the two-year program. I would like to thank Dr. Tom Hukari, Barry Carlson, Tadao Miyamoto, to name a few, for their academic help. Technical support from Greg Newton, and secretarial assistance from Darlene Wallace, Gretchen Moyer and Murren, are much appreciated.

I am thankful for all my fellow colleagues. Lei, Jingbo, Joon and Xi: your priceless friendship makes my study towards M.A. much easier; Maria, Tessie, Allison, Aliki, Chris, Heather, Mavis, Ryan and Yuko: I enjoying studying and partying with you all. I am also very grateful to Sarah and Sylvia for their fast but yet careful proofreading of this thesis.

My family and friends back in China are and will be always my ultimate source of support and motivation. Thank you my dear parents for anything and everything; thank you my best friends Yan Chang, Lei Chen, Yan Kang, Liu Liu, Zhi Liu, Kai Xiao, Huan Yang, Fan Zhou for your spiritual support through various ways.

TABLE OF CONTENTS

FIGURES

TABLES

Chapter One

INTRODUCTION

1.1 Purpose of the study

It is known that people can understand each other when they whisper instead of speaking normally. The fact that formant frequencies carry the information of vowels explains why and how most vowels are correctly perceived in whisper, and so does the fact that the spectral envelope carries the information of consonants like fricatives and stops. However, since pitch (usually fundamental frequency) is used to differentiate the meaning of various lexical items with identical consonants and vowels in tone languages, and since F0 is absent in whisper, it may be difficult to explain how native speakers of tone languages get the full information on tones from whispered speech. There are a number of possible explanations for this phenomenon: special maneuvers of the whisperer (Abramson, 1972), substitution of temporal envelope cues or of stress features (Pike, 1948), and the aid of semantic context, to name a few. This study focuses on the first two explanations.

Researchers displayed great interest in tones in whisper in the 1950s and 1960s. However, a review of most of the studies carried out during that period showed that there was more concern about how well people can perceive tones in whisper than about why they can perceive them, given that the most important perceptual cue, F0, is absent in whisper. This, therefore, triggered my interest in identifying how whispered tones are produced and what contributes to Mandarin tone recognition in whisper. In addition, most tone perception tests used isolated tonal segments as stimuli. Few of the perception tests on Mandarin tones used stimuli in contexts such as conversation, even though whisper is almost as intelligible as normal speech in conversation as it is ‘frequently used and readily understood’ by native speakers (Miller, 1934: 2).

Therefore, this study examines whispered tones in both isolation and context through acoustic analysis and perception tests, and isolated whispered tones are also examined in laryngoscopic images. More specifically, I compare whispered and phonated speech articulatorily and also acoustically to investigate two temporal envelope parameters (duration and amplitude contour) of Mandarin tones. By identification tests of tones in isolation, sentence-initial, sentence-medial and sentence-final positions, I hope to find out if whisperers do use some special maneuvers, and what their role in tone perception is if these special maneuvers exist. Mandarin Chinese is the target language in this study for three reasons: (1) Mandarin Chinese is a tone language with four distinctive tones and a neutral tone; (2) my knowledge and intuition about Mandarin tones as a native speaker are beneficial to the study; (3) a large number of subjects are available because the majority of international students at the University of Victoria are Chinese students.

1.2 Research questions

This study is aimed at discovering what contributes to tone perception in whispered speech in different contextual environments. This problem prompts a number of research questions:

- What characterizes the production of whispered tones?
- Do the duration and amplitude contour in whispered speech differ significantly from those in phonated speech as the result of some special maneuvers of the speakers?
- If speakers' special maneuvers do exist, do they contribute to tone perception in whisper?
- Do the duration and amplitude contour in the whispered context-free stimuli (tonal monosyllables) differ significantly from those in context (speech as the stimulus)?
- To what extent can the temporal envelope cues substitute for F0 in tone perception in whisper?
- Is the existence of carrier sentences an aid to tone perception in whispered Mandarin, or a distraction?
- Does the position in which the target tonal segment stays make any perceptual difference?

Three experiments are conducted in this study: The first is the physical observation of the pharynx and larynx, looking for an answer to the first research question. The second is for data collection, acoustic analysis and comparison, seeking the answers to the next three questions. The third experiment is a perception test, used to substantiate the findings of the second experiment and to answer the last three questions.

1.3 Limitations of the study

Although this study concerns tone perception in whispered speech, it is very difficult to use spontaneous whispered speech as the stimulus in both experiments. This study uses edited speech instead. The contextual stimulus is a common mini dialogue that students use in the library or classroom. Two subjects are instructed to act it out as naturally as they can by imagining they were in the library. I hope that my investigation based on this dialogue will shed some light on the phonetic realities found in real communication. Details of this stimulus context and of data collection will be provided in the section on methodology.

1.4 Outline

This paper reports on a current study on Mandarin tones in whispered speech. It contains six parts. Chapter one is an introduction to this study. Chapter two provides the theoretical and experimental background on which this study is based. Since there is little literature on Mandarin tones in whispered speech, the literature on tones and on whisper are reviewed separately. Chapter three describes in detail three experiments conducted in this study. Acoustic analysis from experiment II, results of the perception test and further acoustic analysis from experiment III are presented in chapter four. Chapter five is the core part of this paper, presenting a conclusion based on the findings of this study and answering the research question. The last chapter summarizes the paper, makes recommendations for future studies and discusses the contributions of this study to the understanding of both whispered speech and Mandarin tones.

Chapter Two

LITERATURE REVIEW

2.1 Tone

Tone has received growing attention from linguists for its important role in human languages. It has been widely studied from physiological, perceptual, acoustical, phonological and historical perspectives. This section summarizes relevant findings on the acoustical features of tone and tone perception which are a basis for the current study. These are examined separately in section 2.1.3 and 2.1.4. Section 2.1.1 presents an overview of tone systems and tone production, and section 2.1.2 provides a brief introduction to tones in Mandarin Chinese.

2.1.1 Overview of tone

Tone is very closely related to the pitch, or fundamental frequency, of the voice. In tone systems, pitch serves to contrast meanings among words or syllables. Two types of subsystems are formed based on the different phonological units in which the domain of pitch lies: word-based tone systems and syllable-based tone systems (Laver, 1994; H. Lin, 1992, 1996). Swedish, Norwegian, Slovenian and Mende are languages with word-based tones, where the pitch is associated with the entire word. Syllable-based tones are found in predominantly languages of Africa, Asia and the Americas (Laver, 1994). Tones in Mandarin Chinese are syllable-based (H. Lin, 1992, 1996). A typological division of languages with a syllable-based tone system is suggested by Pike (1948) based on features of tone: register tone systems and contour tone systems. The contour tone system, which is usually found in Asian languages, is more pertinent to my study. This is the system of tones whose feature of pitch behavior is trajectory shape rather than the relative height of pitch as in a register tone system. Section 2.1.3 will explain this further with some specific Mandarin examples.

2.1.2 Tones in Mandarin Chinese

Mandarin Chinese is a complex term with three references. First of all, it refers to the Chinese dialects spoken North of the Yangtze River, also named the Northern Dialect (H. Lin, 2001). About two-thirds of Chinese people speak Mandarin in this sense. Narrowly speaking, Mandarin can be also used to refer to as one of the dialects within the area mentioned above, such as Tianjin Mandarin. The third and also the most widely used reference is Mandarin as Standard Chinese (H. Lin, 2001). Also known as *Pǔtōnghuà*, Mandarin is officially defined as the standard language by the government of the People's Republic of China. It 'takes the Beijing sound system as its pronunciation standards, the vocabulary from the Northern Dialect as its foundation, and the grammar from the model modern writing in the vernacular as its grammar standards' (cf. H. Lin, 2001: 2). The target language in this study is Mandarin as Standard Chinese. Native speakers of Beijing dialect, the dialect on which the sound system of Mandarin is based, produced the stimuli for the experiments.

Mandarin has four distinctive citation tones and a neutral tone. The four citation tones are high level, rising, low falling, and falling, which are traditionally labeled as Tone 1, Tone 2, Tone 3 and Tone 4 respectively. The high level tone is a level pitch contour that starts at a high pitch and remains high. There is no dramatic pitch movement except 'a slight dip in the middle of the vowel and a slight rise toward the end of the syllable' (Shih, 1988: 83). Tone 2 is a rising pitch contour that starts at mid pitch and ends slightly higher above the pitch level of Tone 1. Shih (1988) observes that during the first half of this rising tone, it either stays level or drops slightly. The falling-rising tone is an inverse circumflex pitch contour that is described as starting at mid pitch, and falling down to the lowest before rising to the mid (Chao, 1948; Shih, 1988; Shen, 1990; Tseng, 1990). However, H. Lin (2001) argues that the speakers intend to start at the lowest pitch and continue at the low pitch for a while before rising to the middle pitch when producing Tone 3, but due to physiological constraints, they are unable to reach the low target at the very beginning. Speakers start slightly higher than the target pitch. This argument is verified by the physiological evidence from Esling et al. (2002) and Carlson & Esling (2001) of the slow speed of engagement of the sphincter, as I will describe later in this paper, the production of Mandarin Tone 3 involves the laryngeal sphincter mechanism. The falling tone covers the entire pitch range, starting at the highest pitch and dropping to the lowest. Compared with the four basic tones in Mandarin, neutral tone is a more controversial issue among linguists. There is no clear-cut definition of the neutral tone yet. What is recognized and accepted by most linguists is that the neutral tone is short and weak, and that it has more than one pitch value depending on its surrounding context. For more discussion on the neutral tone, the reader is referred to H. Lin (1994, 1996 and

1999). Since neutral tone is not a concern in this study, I will not elaborate this issue further.

In addition to the descriptive names of Mandarin tones such as rising and falling, and the traditional names such as Tone 2 and Tone 4, many linguists have tried to develop more accurate or theoretical ways to describe them. Among them, the Scale of Five Pitch Levels developed by Chao Y. R. (1948) is widely used in the current literature. This method divides the pitch range into five levels, represented by numbers from 1 to 5 (Level 1 represents the lowest pitch and level 5 the highest). Table 2-1 summarizes the descriptions of four tones by Chao as well as their various corresponding names.

Table 2-1 Tones in Mandarin

	Tone	Chinese name	Descriptive name	Chao's description
Mā 'mother'	Tone 1	yīnpíng	High level	55
Má 'hemp'	Tone 2	yángpíng	Rising	35
Mǎ 'horse'	Tone 3	shàng shēng	Falling rising	214 (113)
Mà 'scold'	Tone 4	qù shēng	Falling	51

Tone 3 has two values, 214 and 113, according to two different opinions on this tone. However, since 113 is arguably the intended pitch value for Tone 3 and 214 is what people actually produce, we can ignore the value 113 given that this study deals with the acoustic characteristics of speech production rather than the underlying representation.

A recent phonological description of Mandarin tones is developed by H. Lin (1992, 1996 & 1998). In these studies, H. Lin gave the following representation of the five Mandarin tones including the neutral tone. Using the symbols 'H, M, L' representing high level, mid level and low level tonemes, each of the four Mandarin tones is a sequence of three tonemes: HHH (Tone 1), MHH (Tone 2), LLM (Tone 3) and HML (Tone 4) (Lin, 1998). As to the neutral tone, it is basically a low tone represented as L. These representations of Mandarin tones, different from what other linguists have proposed, take timing into account. Thus, they are able to provide a unified account of tones in various environments.

Tone in Mandarin is carried by the vowel(s) in each syllable. According to the official figure from the People's Republic of China, there are a total of 405 syllables in Mandarin (cf. H. Lin, 2001). If the tonal difference is counted as the subtype syllable, there are about 1200 syllables (H. Lin, 2001). However, not every syllable has four distinctive tones. L. Lin and L. Wang's (1991) statistics shows that there are about 160 syllables having all four tonal subtypes. About 130 syllables have three tonal subtypes; about 70 have two tonal subtypes; and 40 have only one tonal subtype.

2.1.3 Acoustic characteristics

Pitch or fundamental frequency is the most important acoustic cue to tone, but it is not the only one. Every tone in Mandarin has its own concomitant features besides F0. They are duration, amplitude contour, and probably formant information of the vowel. This section reviews previous findings on those features.

2.1.3.1 Fundamental frequency

Fundamental frequency is a term from the study of sound physics, measuring the frequency or rate of vocal fold vibration (cycles per second, or Hz) (Laver, 1994). 'Fundamental frequency' and its abbreviated form 'F0' are terms used in acoustic phonetics. Pitch or perceived pitch, corresponding approximately to the same measurement as F0, is a perceptual concept (Laver, 1994) and is used in auditory phonetics. The frequency variations of voice occur within various linguistic units. For tone languages like Chinese, Thai, and some African languages, systematic pitch variations and pitch patterns are related to small grammatical units such as words or morphemes. For intonation languages, the pitch variations occur within larger units such as clauses and sentences (Catford, 1977). Pitch patterns used in these two different ways are called tones and intonation respectively. However, even at the level of sentences for tone languages, fundamental frequency declines gradually. This pattern is often described as a universal property of spoken languages (Kent & Read, 1992; Cruttenden, 1997). Thus, for tone languages like Chinese, the study of fundamental frequency as one of the tone features is always intertwined with the issue of intonation in the sentential context. Nonetheless, this study will be involved with other acoustic features than just with the fact that there is no vocal fold vibration in whisper. A brief summary of F0 in relation to tone in isolation is presented below.

Zhao uses five pitch levels to represent Mandarin tones. How these levels correspond to the F0 measurements is still controversial. Based on the measurements of fundamental frequency for all four tones, L. Lin and L. Wang (1991) used a mathematical logarithm to calculate the frequency range for each level. Assuming that the lowest pitch is 100 Hz and the highest is 300, level 5 corresponds to 245-300 Hz, level 4 to 195-245 Hz, level 3 to 155-195 Hz, level 2 to 125-155 Hz, and level 1 to 100-125 Hz. This is not an accurate way of calculating the frequency value for tones, although it is an insightful approach for combining the description of five pitch levels with the exact frequency of the tones.

The generalized average curve of each of the four tones in Mandarin was derived by Howie (1976) in his study of Mandarin vowels and tones. Two male native speakers produced 15 syllables per tone in all nine syllable types as the stimuli. Table 2.2 shows the detail of the frequency value at the starting and ending points, as well as at the mid point for Tone 3. For comparison, different results achieved by Tseng (1990) are also listed in the same table. Tseng's experiment on acoustic correlates of tone (1990) used a female native speaker as the subject producing 24 tokens (6 syllables per tone) in citation. Thus, it is not surprising that his result has higher pitch than Howie's experiment. Moreover, Tseng's result has an overall wider range than Howie's, especially the pitch range for Tone 2 and Tone 3. There is a 100 Hz difference for Tone 2 in Tseng's result compared to 35 Hz in Howie's result. Tseng attributed this difference to 'individual styles of speech' (1990: 18).

Finally, scholars have also computed the frequency range for four different tones in Mandarin. L. Lin and L. Wang's (1991) results, however, differ radically from both Howie's and Tseng's. Tone 1 in their study has a frequency value between 340-350 Hz, Tone 2 between 270-340 Hz, Tone 3 between 230-170-230 Hz and Tone 4 between 360-280 Hz. All four tones in their studies start at a much higher frequency than in either study in Table 2-2, but the pitch ranges they came up with are similar to those of Tseng. L. Lin and L. Wang (1991) attributed the wide range to the female subject who produced the stimuli. My impression is that the higher frequency value might relate to the way they calculate the F0 by dividing the 10^{th} harmonic by 10.

Table 2-2 Frequency values for tones in Mandarin (adapted from Tseng (1990: 18))

Tone	Sources	F0 pattern			
		beginning	F0 drop	F0 rise	ending
Tone 1 High level	Howie (1976)	150 Hz	-	-	150 Hz
	Tseng (1990)	215 Hz	-	-	215 Hz
Tone 2 Rising	Howie (1976)	115 Hz	-	35 Hz	150 Hz
	Tseng (1990)	150 Hz	-	100 Hz	250 Hz
Tone 3 Falling-rising	Howie (1976)	113 Hz	40 Hz	40 Hz	113 Hz
	Tseng (1990)	135 Hz	45 Hz	90 Hz	180 Hz
Tone 4 Falling	Howie (1976)	157 Hz	52 Hz	-	105 Hz
	Tseng (1990)	245 Hz	95 Hz	-	150 Hz

2.1.3.2 Duration

Duration is one of the temporal envelope cues for tone recognition. The duration of the vocalic part of the tonal syllable is counted as the duration of the tone. Taking the intrinsic vowel duration into consideration, Howie (1976) measured the duration of the four tones in all nine syllable types with various vocalic parts. His result is shown in Table 2.3, together with that of Tseng (1990) and Fu & Zeng (2000), again for comparison. The general pattern according to Howie is: Tone 3 has the longest duration, Tone 1 has the shortest, and Tone 2 is longer than Tone 4. This result corresponds with the claim made by Chao (1968) that the more complicated the tonal contour is, the longer duration it has. As the most complicated tonal contour among all four tones, Tone 3 is proven to have the longest duration elsewhere as well. Tseng's (1990) study presents the same result with Tone 3 as the longest tone. Tone 3 has a mean duration of 457.04 milliseconds (ms hereafter) based of durations on 6 different vowels in the tonal syllables with falling-rising tone. However, Tone 4 is found to have the shortest duration according to Tseng (1990). As for the remaining two tones, Tone 2 lasts longer than Tone 1 for four vowels (66.67%), while the reverse order is found for the other two vowels (33.33%). In addition, durations of four tones in Tseng's study are longer overall than Howie's, which is explained by the difference of the context. I will discuss the role of context further below. Recent studies on the duration patterns of Mandarin tones yield similar results. Fu and Zeng (2000) also find that in isolation, Tone 3 and Tone 4 are the longest and shortest respectively and Tone 2 is significantly longer than Tone 1.

Table 2-3 Average duration of Mandarin tones

Tone	Howie (1976)	Tseng (1990)	Fu & Zeng (2000)
Tone 1 High level	237 ms	351.76 ms	339.5 ms
Tone 2 Rising	271 ms	376.28 ms	374.7 ms
Tone 3 Falling-rising	277 ms	457.04 ms	464.3 ms
Tone 4 Falling	251 ms	236.12 ms	334.4 ms

Howie (1976) made his measurement from stimuli with a total of 136 tonal syllables. Each syllable was pronounced in the carrier sentence, *Zhèige* ____ *zì, shì Lǎo Lǐ xiě de.* 'This word, is written by Mr. Li.'. Thus the tone was examined in edited speech rather than citation (or isolation), whereas all tonal syllables were produced in isolation in Tseng's experiment (1990). The different context in these two studies shows that tones produced in isolation have a longer duration than those in sentential context, as Tseng claimed that 'the occurrence of the target token regarding its position in an utterance has a definite effect on the duration' (1990: 24). In order to further verify his claim, Tseng (1990) compared the vowel duration in citation form with that in spontaneous speech. The data from his experiment proved that Tone 3, the longest one in isolation, still had the longest duration in spontaneous speech of 221 ms, although it was much shorter than that in isolation, 457.04 ms. The shortest tone in context is Tone 2;

nevertheless, Tone 4 is the shortest tone in a context-free environment. This finding still differs from that of Howie (1976), who found that Tone 1 had the shortest duration in embedded sentences. This discrepancy is due to the difference between edited speech and spontaneous speech.

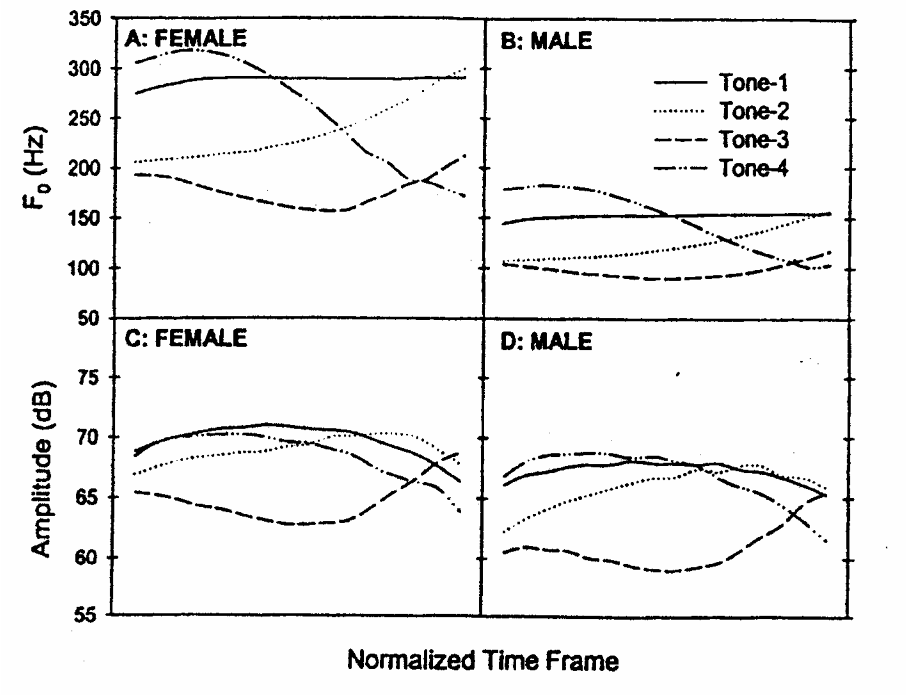
Both Ting (1971) and Ho (1976) conducted studies investigating the variation of the acoustic parameters of the four tones in isolation and in various sentence environments. Their findings on duration were similar: tones in isolated words have the longest duration; the next longest duration was in sentence-medial position (the end of a clause). For sentence-initial and sentence-final positions, Tone 1 and Tone 4 last longer in initial positions than in final positions, while the order reverses for Tone 2 and Tone 3 (Ho, 1976). Their finding reaffirms that environment does make a difference for the duration of tones, and that tone itself has an effect on the duration of syllable nuclei. Contradictorily, based on the statistics from Howie and Woo's studies, H. Lin (1996) argued that all Mandarin tones are longer in sentence-final position than elsewhere. More importantly, neither of them tried to examine the interrelation among tones in other sentence environments. Would the duration of tones in various environments follow the same order? Or would Tone 3 simply have the longest duration in any context? The current study will measure the duration of tones in isolation, and in sentence-initial, -medial and -final positions, as well as comparing those in phonated speech with those in whispered speech.

2.1.3.3 Amplitude contour

'Amplitude is the amount of deviation of the oscillating molecules from the rest position: in other words, it is the positive or negative pressure attained by a sound wave.' (Catford, 1977: 50) Amplitude is calculated with the average sound-pressure, or RMS (root-mean-square) pressure, which is related to the acoustic term 'intensity'. Intensity is measured with decibels (dB hereafter). The fact that 0 dB corresponds to the sound-pressure level of a reference sound shows that it is a relative scale (Laver, 1994). Both amplitude and intensity are related to the loudness of a sound. Loudness is a perceptual concept. Although there is no direct or parallel relation between amplitude or intensity and loudness (Crystal, 1992), amplitude and intensity are the two major things to measure and analyze when the loudness of a certain sound is studied. There are a small number of acoustic studies on amplitude contour as one of the concomitant features of tone. Among these studies, various ways of measuring loudness were used.

Fu and Zeng (2000: 46) in their experiment defined the amplitude contour as the 'fluctuations in the overall amplitude at a rate between 50 Hz and 500 Hz'. Figure 2-1 shows the mean amplitude contour of the four Mandarin tones on single-vowel syllables in citation, and the F0 contour for comparison. Panels on the left are the results from female data, while results from the male speaker data are on the right. Similar to the F0 contour, the amplitude contour for female speakers are higher than those for male speakers, especially for Tone 1, 2 and 3. The patterns of the amplitude contour, flat for Tone 1, rising for Tone 2, falling and rising for Tone 3, and falling for Tone 4 demonstrate the similarity between F0 and amplitude contour. Thus F0 and amplitude contour are highly related.

Fig. 2-1 Amplitude of Mandarin tones (Fu & Zeng, 2000)



Ho (1976) made comparisons of the highest and lowest amplitude peaks among seven various contextual environments. His observations were: the amplitude curves for all tones in isolation and sentence-initial environment are higher than those in sentence-final and sentence-medial environments; and tones in citation have the highest peak. Like Fu and Zeng, Ho concluded that amplitude and F0 are highly correlated, as exemplified by the fact that high amplitude peaks and high fundamental frequency co-exist in the same environment, as do low peaks and low F0 (Ho, 1976). However, Ho did not provide any definition of the amplitude contour or a description of the technical steps of measuring it. But it is very likely that Ho adopted a measuring method different from that of Fu and Zeng.

2.1.4 Tone perception

In this section, two different perspectives on the issue of tone perception are discussed separately with selected studies. They are the cues in tone perception and tone perception in different environments, both of which are relevant to this study.

2.1.4.1 Perceptual cues

F0, duration and amplitude contour, the acoustic parameters for tones in Mandarin studied in section 2.1.3, serve as cues in tone perception. It is agreed among almost all linguists that F0 is the most prominent acoustic element distinguishing tones, as it carries sufficient information for tone contrasts. This has been widely proven by a number of identification experiments (Jensen, 1958; Abramson, 1972, 1974; H-B. Lin & Repp, 1989; etc.) on various tonal languages as well. This conclusion has also been proven in studies on the tones of Mandarin Chinese by Howie (1976), Tseng (1990), and Fu and Zeng (2000). However, the roles of other perceptual cues such as duration and amplitude contour in tone identification have also received considerable interest as discussed below.

Based on the acoustic analysis of Mandarin tones in isolation, Tseng (1990) concluded that duration plays a secondary role in tone perception. Similarly, Lin and Repp (1989) found that duration played a secondary role for the falling-nonfalling tone distinction in their studies on the cues to Taiwanese tones. Lin and Repp found that nonfalling tones which include high-level (Taiwanese Tone 1), low-rising (Taiwanese Tone 4) and mid-level (Taiwanese Tone 5) were relatively long; while high-falling (Taiwanese Tone 2) and mid-falling (Taiwanese Tone 3) tones were comparatively short. When F0 provides ambiguous information, duration is an aid to tone perception between these two tonal categories. Blicher et al. (1990) summarized the duration patterns among tones on the basis of some cross-linguistic data in this way: with all other variables controlled, high tones tend to be shorter than low tones, and falling tones shorter than rising tones. This conforms with Cruttenden's (1986) physiological explanation for the theory of declination: since a rise in fundamental frequency takes a longer time to achieve than a fall, an upward shift of pitch involves a more difficult physical adjustment than a downward shift of pitch. The experiments (Blicher et al., 1990) on the effect of duration in Mandarin Tone 2 and Tone 3 distinction provided empirical evidence that the length effect did help in enhancing the tonal perception for both native and non-native speakers of Mandarin. Fu and Zeng (2000) used processed stimuli in their study to investigate the effect of individual temporal envelope cues. Perception tests of the duration-stimuli with the naturally varying vowel duration preserved yielded the lowest tone recognition score (35.6%). This indicated that the role of the duration cue in tone identification is minor compared with those of other cues.

Abramson, in his study of Thai (1974), suggested that the efficacy of the F0 in tone perception was enhanced by the addition of amplitude information. In his perception test, overall identifiability was improved from stimuli with F0 alone to synthetic stimuli with F0 plus amplitude. Whalen and Xu's study (1992) on Mandarin tone perception tested the perception of manipulated stimuli with the F0 and formant structure cast aside and duration controlled. They concluded with the result of this experiment that Mandarin listeners can identify the tone with 'reasonable accuracy solely from the amplitude contour' (Whalen and Xu, 1992). A comparatively high perceptual score of 58% was achieved by listeners, for the Amplitude-stimuli in Fu and Zeng's research (2000).

With both duration and amplitude, or all perceptual cues except F0, retained in the stimuli, the perceptual score would improve to a certain extent. In the same research by Fu and Zeng (2000), the accuracy of the perception test increased to 69.3% -- the highest tone recognition score in the study when the APD (Amplitude-Periodicity-Duration)-stimuli (only F0 absent) were presented to the listeners. Whalen and Xu (1992) also got similar results when they played a signal that consisted solely of an amplitude contour and duration to Mandarin speakers. Those listeners were able to identify Tone 2, Tone 3 and Tone 4 quite well, but they were unable to identify Tone 1 correctly besides telling that Tone 1 is not Tone 3. Although these perception results are not as good as when F0 is present, they imply that speech without F0 is intelligible most of the time. Whispering is a perfect example. Since fundamental frequency is totally absent in whispering, listeners must use other perceptual cues such as duration, amplitude or other spectral frequency information to substitute for F0 to perceive tones. The acoustic study of tones in whisper, therefore, is very important in understanding the role of temporal envelope cues in tone perception. In addition, all the studies discussed above used either synthetic speech or manipulated speech to generate stimuli without fundamental frequency. These stimuli are unnatural and might deviate from speech in reality. Since whisper is perfectly natural speech produced by human beings, it is an ideal stimulus in this sense, as it permits us to discover facts about perceptual cues other than F0 in tone perception. Finally, since people are able to understand each other in whisper as well as in normal speech, do people consciously or unconsciously ‘exaggerate’ the use of temporal envelope parameters to produce the tone for the sake of better understanding? This is also an interesting issue to look into in this study.

2.1.4.2 Contextual variation

Contextual difference may also alter the perception of tones. Mandarin tones in four different contexts are tested in this study: tones in isolation, and tones in sentence-initial position, sentence-medial position and sentence-final position. Semantics is controlled in these sentential contexts, thus the role of phonetic context in tone perception is studied. There are a few studies that compare tone perception in two contexts, but not in all of the contexts. Howie (1976) did perception experiments on both citation syllables and syllables in the carrier sentence *Zhèige ____ zì, shi Lǎolǐ xiě de.* The results of his perception tests of tones in isolation and sentence-medial position are: intelligibility of 95.4% for isolated syllables and 81.3% for sentence-medial syllables. Vance (1976) concluded from his listening experiment results that identification of tones is somewhat more difficult in final context than in medial contexts. Assuming that both studies are correct, the order of these contexts from the most accurate tone perception to the least is: isolation, sentence-medial, sentence-final. Studies comparing tone perception in all three or four environments including sentence-initial are needed. This study will verify this ranking in whispered speech.

2.2 Whisper

Whispering is a natural way of communication. People whisper in certain circumstances to reduce the perceptibility, such as whispering into someone's ear in the library, which is labeled as soft whisper (Weitzman et al., 1976). Soft whisper is also referred to as a quiet whisper produced 'in a relaxed, comfortable, low effort manner' in Solomon et al.'s study (1989: 161). Stage whisper, on the other hand, is a combined kind of whisper one would use if the listener is quite a distance away from the speaker (Weitzman etc., 1976). Stage whisper is actually whispery voice in phonetic terms, which is a type of phonation that involves vocal fold vibration (Laver, 1980; Esling, 1984; Esling, 1994b). This research will focus on soft whisper, which is produced without vocal fold vibration and is more commonly used in daily life.

In section 2.2.1, previous studies on the production of whisper are summarized; the physiological features of whisper are further elaborated in section 2.2.2 with Bai speech data*; section 2.2.3 deals with the acoustic characteristics of whisper as well as the acoustic differences between whisper and normal speech; previous studies on perception of whisper are reviewed in section 2.2.4.

2.2.1 What is whisper

The essential physiological characteristic of whisper lies in the configuration of the glottis and epilarynx. The vocal folds remain opening for whispering, thus friction is produced when the airstream from the lungs is forced through the passage. In normal speech, however, the vocal folds have opening and closing that produces quasiperiodic pulses of airflow. The source of whisper, therefore, is the aperiodic noise rather than quasiperiodic voice due to the lack of quasiperiodic airflow. There are different descriptions at the glottal level for whisper: Pressman & Kelemen (1955), Catford (1977) and Kallail & Emanuel (1984) describe the vocal folds as “narrowing”, “slit-like” or “slightly more adducted”; according to Tartter (1989), “whispering speech is produced with a more open glottis than is normal voicing”; Weitzman et al. (1976: 62) define the

---

* Bai speech data were produced by Li Shaoni (Central University of Nationalities, P.R.China) at the University of Victoria, and are provided by Dr. John H. Esling,

whispered vowels as produced "with a narrowing (or even closing) of the membranous glottis while the cartilaginous glottis is open. Based on the observations of phonation with photographs taken with a fibre-optic bundle inserted through the mouth, Buuren (1983: 13) claims that we would expect to see the 'familiar triangular opening between the arytenoids'. Solomon et al. (1989) studied the laryngeal configuration and constriction during whispering of 10 subjects from videotapes of the larynx. Three types of vocal-fold configuration are identified (Solomon et al. 1989): (1) the shape of an inverted 'V' or narrow slit, (2) the shape of an inverted 'Y', (3) bowing of the anterior glottis. A description scheme consisting of these three shapes by three sizes (small, medium and large) was established for judgments. Their study concluded that soft whisper has the dominant pattern of medium inverted 'V'. Fibre-optic laryngoscopic photographs in research by Esling (2002) show that the vocal folds are parted for whisper, but that the laryngeal aryepiglottic laryngeal sphincter is engaged during whispering. This will be discussed in detail in the next section.

Schwartz (1971) compared the air consumption values for voiced and whispered syllables. His study revealed that female speakers had greater air consumption for whispered syllables than for voiced ones. Similarly, Hoit & Hixon (1987) and Smitheran & Hixon (1981) compared the respiratory and laryngeal functions between whispering and normal speaking with various observations. They found that whispering has lower lung volumes, lower tracheal pressures and lower laryngeal airway resistances than speaking, but that whispering has higher translaryngeal flows.

Other comparisons between whisper and normal speech have raised the issue of articulatory reorganization in whisper. Based on two experiments in Danish and English, Mansell (1973) concluded that articulatory behavior is different in whisper and in normal speech, i.e., articulatory reorganization in whisper exists. Peterson (1961: 23) suggested whisperers' conscious efforts to produce vowels equivalent to voiced ones may result in 'minor articulatory adjustments'. Based on the findings of speakers' vocal efforts in whispered speech, Traunmüller and Eriksson in their recent study (2000) implied that speakers attempt to modify their articulation when whispering in order to increase the audibility of their whispers. In addition, tonal differences also contribute to these adjustments. Because the lack of vocal fold vibration causes the absence of pitch in whisper, listeners do not have access to the most significant perceptual cue to tone in whispered speech. With evidence of the mixed results from tone perception tasks in whispered Mandarin, Abramson (1972) argued that some speakers make use of special maneuvers to compensate for the missing pitch information, while others just use the usual instructions to their speech production mechanisms minus fundamental frequency control. From his own experiments on whispered Thai, he further proposed that 'the speaker replicates 'instructions' that he sends in phonated speech to the extrinsic muscles of his larynx and, perhaps, muscles of the tongue to make adjustments in vocal tract configuration to facilitate large pitch changes' (Abramson, 1972: 43). Given the importance of tone in this issue, I hope that this research is able to provide some insightful findings into articulatory reorganization in whisper.

2.2.2 Physiological features

Most descriptions of the larynx during whispering found in the literature were made at the cross-sectional or horizontal level of the glottis. Inquiring from the perspective of the vertical plane into the larynx might provide more insightful information on how whisper is produced. This section examines the laryngoscopic video of Bai speech data containing both phonated and whispered vowels to account for the role of the laryngeal sphincter mechanism in the production of whisper (Esling 1994b, 1995, 1996; Edmondson et al., 2000; Esling & Edmondson, 2002). Description of the data is made after a brief summary of the laryngeal mechanism and the Bai language.

Esling (1996, 1999) found that the aryepiglottic sphincter mechanism is responsible for the sound category 'pharyngeal' based on his observations from fibreoptic laryngoscopy. The aryepiglottic sphincter is engaged when 'the aryepiglottic folds (linking the adducted arytenoids with the margins of the epiglottis) press up under the body of the epiglottis as the tongue retracts' (Esling & Clayards, 1999: 28). This mechanism is also known as the 'laryngeal sphincter' (Esling & Clayards, 1999). The aryepiglottic folds 'press against the base of epiglottis', and form an angle reported within the range of 140° to 90° degrees with the arytenoids; the concomitant features are the retracting tongue root towards the back wall of the pharynx, and 'raising and fronting of the epilaryngeal cartilages which form the sphincter' (Esling & Clayards, 1999: 28). The aryepiglottic folds obscure the vocal folds and form a triangular shape epilaryngeally. In addition, it is argued that this aryepiglottic mechanism is also related to tone in tone languages such as Tibeto-Burman Mpi (Esling, 1996). Some Asian languages have tones that involve some degree of pharyngealization, in the sense of sphinteric narrowing, such as Yi and Bai.

Bai belongs to the Sino-Tibetan language family, and is the language of the Bai minority living in Southwest China (Sichuan Province and Yunnan Province). Bai has tense and lax vowels; in addition, it also has harsh voice and breathy voice (Edmondson et al., 2000). The speech data produced by Li Shaoni at the University of Victoria includes both phonated and whispered versions of the tonal paradigm in Bai, which makes the comparison of the productions of the two types of phonation (modal and whisper) possible. Given that there is no Mandarin speech data available, and Bai is also a tone language, I believe descriptions based on the Bai data are able to provide a good indication of how whispered speech in a tone language is produced.

Comparing the video files for phonated and whispered lax vowels in Bai, I observed some slight laryngeal sphinctering for the whispered vowels. The aryepiglottic folds narrow to reduce the epilaryngeal tube area to a small triangular opening. No vibration is observed between the vocal folds anteriorly, and they close towards each other with an opening in the shape of a 'Y'. The ventricular folds come together over the vocal folds and obscure the vocal folds but not the small opening. However, different from the laryngeal sphincter mechanism observed in pharyngeal sounds, the apexes of arytenoids are always separate in the whispered vowels and so are the aryepiglottic folds. The phonated versions of the lax vowels, on the other hand, do not involve narrowing of the aryepiglottic sphincter mechanism. Thus, the whispered versions of the Bai lax vowels can be associated with the laryngeal sphincter mechanism for their production.

Tense vowels in Bai are observed by Edmondson et al. (2000: 50) to be 'articulated with the laryngeal aryepiglottic sphincter mechanism'. From the video files of tense vowels, the apexes of the arytenoids and the aryepiglottic folds join together and draw close to the middle of the epilaryngeal tube area (Edmondson et al., 2000). This action of the sphincter mechanism occurs in the whispered tense vowels of Bai as well.

Similar to the phenomenon observed in whispered lax vowels, the vocal folds are not vibrating and are not completely closed, the ventricular folds impinge on the vocal folds, and the aryepiglottic folds do not close completely. The differences found between these two types of whispered vowels are: (1) the shape of the opening between the vocal folds is unable to be observed, because the whispered tense vowels involve more activation of the laryngeal sphincter mechanism than the lax vowels: that the airway is reduced to a much smaller area in the tense vowels so that the ventricular folds and the small opening between vocal folds are both obscured; (2) larynx raising is more evident in whispered tense vowels than lax vowels; and (3) the tongue root is retracted in the tense vowels also. These differences are presumed to reflect register contrast, since the tense vowels in Bai use the laryngeal sphincter mechanism to articulate while the lax vowels do not. Given that whispered phonation also involves this mechanism to some degree, the whispered tense vowels need to incorporate more activation of the sphincter mechanism to maintain the contrast with the lax vowels.

The laryngeal sphincter mechanism is active in all whispered vowels in Bai including high, mid, breathy, harsh and rising; the contrast between lax and tense vowels is also observed in whispered high mid and breathy vowels. To summarize the laryngeal sphincter as a physiological function of whispering: no vocal fold vibration is found in whisper, and there is a narrow ‘Y’-shape opening between the aryepiglottic folds in whispered lax vowels. The ventricular folds come over and obscure the vocal folds. The aryepiglottic folds press up and draw close to the middle of the airway, reducing the epilaryngeal tube area to a small triangular opening. They are observed in contact posteriorly, as are the apexes of arytenoids, and in contact laterally with the epiglottis. A slight larynx raising but no tongue root retraction is found in lax vowels.

2.2.3 Acoustic characteristics

Corresponding to the essential physiological characteristics of whisper is the lack of vocal fold vibration, the most significant acoustic characteristic of whisper is the absence of fundamental frequency and consequent harmonic relationships (Tartter, 1989). Turbulent aperiodic airflow is the only source of sound for whisper, and it is a strong, 'rich' and hushing sound (Catford, 1977: 96). Whispered pitch was explained scientifically early by Helmholtz (republished in 1954), who assigned the pitch of whispered vowels to typical values of the first formant (F1 hereafter) for the back vowels and to typical values of the second formant (F2 hereafter) for the front vowels on the basis of comparison between whispered and phonated vowels with some standard frequency source. Later researchers, however, suggested with collected data that the perceived pitch of either whispered or phonated vowels is more directly related to F2 than to the fundamental frequency or to F1 (Peterson & Barney, 1952; Thomas, 1969; McGlone & Manning, 1979).

Vowels are the focus of most studies on whispered speech, especially the first three formants of whispered vowels. Comparing them spectrographically, both Peterson (1961) and Lehiste (1970) found that the first three formants of whispered vowels are higher than those of phonated vowels. More specifically, Lehiste (1970) reported that F1 is approximately 200-250 Hz higher, whereas F2 and F3 are approximately 100-150 Hz higher in whispered vowels. Smith (1973: 82) came to the same conclusion based on his dissertation that 'formants for whispered vowels tend to be higher in frequency than those for the same vowels produced with voicing'. Although recent studies done by Kallail and Emanuel (1984a, 1984b) revealed a similar trend for whispered vowel formants, the results show that this trend was only strongly evident for F1. By comparing the formant-frequency differences between isolated whispered and phonated vowels produced by adult male and female subjects, Kallail and Emanuel (1984a, 1984b) found that the differences for F2 and F3 are smaller than for F1 between the two types of phonation. Both studies also show that only the differences for F1 were strongly evident.

Thomas (1969) made some comments on the formant bandwidths and amplitudes of whispered vowels in his study. A wide bandwidth of F1 was found for some vowels, and a narrow bandwidth (around 100-200Hz) of F3 was found for some vowels as well. As for formant amplitudes, those of F1 and F2 were approximately equal among all vowels examined, while the F3 amplitude for the back vowels was extremely low. Smith (1973: 83) compared the formant bandwidths and amplitudes between whispered and phonated vowels. His findings on the formant bandwidths showed that bandwidths were slightly wider for whispered vowels than for phonated vowels. All four test vowels (/i/, /æ/, /ɔ/, /ʊ/) showed little difference between the peak amplitudes of the three formants, and the biggest difference among them was found for the back vowel /ɔ/. Traunmüller and Eriksson (2000) found from adults' speech that the duration of all segments in whisper tended to be longer than those in phonated speech; and their analysis of children's speech showed that whispered vowels are also slightly longer, although consonants in whisper are shorter.

Of all speech sounds, vowels have been studied more thoroughly than consonants, and primarily in English. In addition to the fact that vowels are "associated with a steady-state articulatory configuration and a steady-state acoustic pattern" (Kent and Read, 1992: 87), the vowel is also the indispensable part of a syllable in most languages in the world. What is more important to this study is that vowels are the "carrier" of tones in Mandarin Chinese. However, whispered vowels in previous studies were almost all in a context-free phonetic environment, either in sustained isolated condition or in monosyllables. There have been hardly any studies that examine vowels in whispered speech, not to mention the comparison between vowels in whispered speech and phonated speech. This is one of the issues that the present study is looking into.

2.2.4 Previous studies on perception

There have been a number of studies on whisper examining the perception of various phonetic or phonological units such as vowels, consonants and tones. Gender and speaker identifiability have also been studied. This section reviews and comments on some previous research in four different categories: vowel perception, consonant perception, gender perception, and tone perception. The last section presents a brief summary of these categories.

2.2.4.1 Vowel perception

Kallail and Emanuel (1984) designed two experiments studying and comparing whispered and phonated vowel perception. Sustained vowel samples produced by male and female adult subjects respectively were examined in each experiment. They found that phonated vowels were generally identified correctly more often than whispered vowels: listeners got 85% and 80% correct for phonated vowels in each experiment but only 63% and 65% for whispered vowels. For the phonated vowels, /æ/, /i/ and /u/ were identified better than vowels /ɑ/ and /ʌ/. For the whispered vowels, vowel /i/ was identified the best with 73% correct in both experiments; vowels /ʌ/, /æ/ and /ɑ/ were identified correctly at the range of 61% to 68%; while listeners found vowel /u/ the most difficult to identify with only 54% correct judgments in both experiments.

Tartter (1991) reassessed whispered vowel identification by using a larger range of English vowels and a more natural CVC context for the vowel productions. The vowels examined were /i/, /ɪ/, /u/, /ʊ/, /ə/, /ɛ/, /ɔ/, /ʌ/, /æ/ and /a/. All 10 vowels were pronounced in a [hVd] context by both male and female speakers in both normal and whispered registers. Results of this research showed considerably better whispered vowel identification with 82% average accuracy compared with the 65% found by Kallail and Emanuel (1984), and only a 10% falloff in identification accuracy from phonated speech (Tartter, 1991). The CVC context is one of their explanations for the improvement in the accuracy percentage they obtained.

Both studies revealed that vowels are better identified in phonated speech than in whispered speech, but they differed in the accuracy percentage for whispered vowel perception. The existence of consonant conditions in the latter study might contribute to the improvement in vowel identification. However, vowels in syllables are still in a context-free condition, and whether the existence of context would further improve the judgment accuracy remains unexplored in the studies of whisper.

2.2.4.2 Consonant perception

Dannenbring (1980) studied consonant perception in whispered speech. The voicing feature was examined for 12 different consonants: /b, p, d, t, g, k, z, s, v, f, χ, θ/. These consonants are in syllable-initial position followed by four different vowels. The most important result of this experiment is that 'subjects are able to discriminate between whispered phonemes even though the most obvious feature that differentiates them in normal speech (VOT) is missing' (Dannenbring, 1980: 982). Duration of fricative noise and the amplitude difference between the noise and the following vowel were explained as the cues available instead of VOT (voice onset time) to differentiate the voice feature in whisper for phonated speech.

Similarly, Tartter (1989) conducted an experiment in which listeners were asked to identify 18 different whispered initial consonants. Besides stops and some fricatives listed for Dannenbring's research, nasals, glides and affricates were also examined in his research. The listeners achieved better than chance accuracy (64%) in the nonsense syllables (C/a/). The conclusion that consonant information is normally well conveyed in whispered speech was made.

The two studies above provide evidence for the high identifiability of consonants in whispered speech. All the studies in these two sections reveal the efficacy of whisper in transmitting perceptual cues for phonetic units, vowels and consonants. None of the above studies ever used any contexts as the stimuli, and very few studies have so far. The role of the phonetic context in whispered speech identification is still unclear.

2.2.4.3 Gender identification

Schwartz and Rine (1969) found that listeners were able to identify speakers' gender from vowels /i/ and /a/ sustained in a whisper for 3 seconds. Based on their spectrographic analysis of the stimuli, they concluded that the upward frequency displacement of the resonance peaks of female vowels was the primary acoustic cue for the listeners, with 97.5% accuracy. However, Lass et al. (1976) found poorer gender identification accuracy (75%) for whisper in their study. Sustained isolated vowels were also used as the stimuli for whisper; the basic difference lies in the timing factor, as in the latter study each sample was only sustained for 0.5 second.

It seems from the two studies above that duration plays an important role in gender identification and most likely in other kinds of perception tests: the longer the stimulus lasts, the higher the accuracy percentage. In spontaneous speech, the duration for each individual phonetic unit is shorter than in sustained or context-free conditions. Would the phonetic context compensate for shorter duration of tones in context than tones in citation, and as a result listeners achieve similar or even higher perception percentage? This study attempts to address this complex question.

2.2.4.4 Tone perception

The intelligibility of tones in tone languages in whisper has attracted many researchers' attention, since the dominant perceptual cue, fundamental frequency, is missing in whisper. Abramson (1972) summarized two distinctive opinions on this issue: the first one, represented by Panconcelli-Calzia, claimed that context makes comprehension of longer whispered utterances possible in tonal languages. The comprehension of isolated words was ruled out. In response to the claims that whispered tones are indeed audible, he advanced two thoughts: either they were not dealing with a genuinely voiceless whisper, or the semantic function of tones had been exaggerated. Giet stated the opposite view, that tonal distinctions are maintained in whispering and rendered perceptible through the substitution of other phonetic features for pitch, e.g. in real Chinese whispering, the two substitutes are changes in vowel color as well as increases in air flow for high tones and decreases in air flow for low tones (cf. Abramson, 1972).

Jensen (1958) did a series of experiments studying tone perception in four tonal languages with tonal characteristics: Norwegian, Swedish, Slovenian, and Mandarin. Contrastive word pairs were uttered in whisper in all four experiments, with the only contrast between words being the word tone. Results of his experiments showed an average identification from 66% to 81% for all four languages, and 53% to 73% for Norwegian, 100% for Swedish, 71% to 85% for Slovenian, 73% to 88% for Mandarin. The conclusion that he reached was that word tones were transmitted better in whisper in some languages (Swedish, Mandarin) than in the other two languages (Jensen, 1958). Replicating the procedures in Jensen's experiments, Miller found some 'striking contrasts' when comparing results of his perception test of Vietnamese with those of Jensen's. He concluded that very little word tone is transmitted in Vietnamese whispered speech; obtaining an overall accuracy score of 42% for Vietnamese tone recognition in words.

In addition to the findings that tone perception in whisper differs in languages, we must be aware of the context in which whispered stimuli are examined. The stimuli in both studies were monosyllabic or disyllabic (Norwegian) words, i.e. tones were perceived in a context-free environment. As discussed earlier, context as one of the possible perceptual cues was not taken into account in these two experiments.

Wise and Chong (1957) reported a recognition test that used sentences as stimuli for whispered Mandarin. Thirty pairs of sentences were composed and read aloud by one of the authors in both phonated and whispered speech. A typical pair of sentences is:

(1) Wǒ	yào	**tāng**.	/	Wǒ	yào	**táng**.
I	want	soup.	/	I	want	candy.

Six competent native speakers of Mandarin were asked to repeat what they heard from the tape which randomly played one sentence out of a total of 638. They counted the correct tones of the critical words that each subject repeated and got the accuracy of this test as low as 62%. It was claimed then by the researchers that a tone language is not consistently intelligible when whispered (Wise and Chong, 1957).

The result of this study seems to indicate that context is not a helpful cue for listeners to perceive tones in a tonal language; rather it is more like a distraction as we compare the low accuracy percentage 62% with an accuracy range between 73% to 88% that Jensen got. Interestingly enough, Jensen (1958) made some brief comments on the role of context in his experiments on whisper: ‘context of course always helps one to understand, and the same may be true of stress patterns which, for all we know, may make tone distinctions more or less superfluous’ (Jensen, 1958: 187). Thus, in order to elucidate the role of context in whispered speech, we need to carry out an experiment in which whispered tones in both context-free and context environments are produced and perceived by subjects. An acoustic analysis of both conditions will also be necessary for us to explore what specifically in the context either helps or blocks the perception of tones.

One of Abramson's (1972) tonal experiments with whispered Thai tested the accuracy of tone perception of certain words which were embedded in sentences and compared the result with the context-free condition. It seemed from the comparison that the sentence environment induced a slight improvement in perception: an accuracy range from 21.3% to 46.2% in isolated monosyllabic words increased to a range of from 48.8% to 67.5% in sentences. However, the critical tonal words for perception tests in a context-free condition were different from those in context condition; furthermore, only words with two distinctive tones were embedded in sentences as stimuli, while all five tones in Thai were used in the context-free stimuli. These factors may contribute to the improvement of context that makes the result of this experiment less convincing in support of the view that the existence of context helps people to perceive tones in whispered speech.

2.2.4.5 Summary

Jensen (1958) stated in his paper that Olof Gjerdman was the first person to raise the issue of whether pitch is heard in whispered Chinese. This issue then caught the interest of a number of linguists. Charles Boardman Miller (1934) found from an investigation of Mandarin Chinese that the four tonal phonemes are identified in whispered speech, and he came to the conclusion that this was accomplished with the help of context and energy changes indicating variations in pitch. By obtaining an affirmative answer from a Chinese speaker on whether whispering is intelligible between two persons, Gjerdman corroborated his theory that ‘a whispered language has as many distinct musical accents as the same language when voiced’ (Jensen, 1958: 187). Giet, according to Jensen (1958), had the same opinion that Chinese tones are expressed and preserved in whisper. Nevertheless, Panconcelli-Calzia, according to Abramson (1972), held the opposite opinion on the intelligibility of whispered tones as stated earlier. However, no literature I know directly addresses this conflict of opinions on whispered speech. By examining Mandarin tones in both isolation and context in this study, I hope to contribute to this debate with empirical evidence.

In summary, the selected studies we examined above have presented some good models of research on whispered speech; and their results and conclusions have provided an insightful idea of the indentifiablity of gender, and of the phonetic and phonological units in whispered speech. The current study, based on previous experiments, will examine further the role context plays in the perception of tone in whispered Mandarin. Differing from previous studies, this study will take advantage of updated equipment to collect physiological data and to search for empirical evidence that can be correlated with the acoustical analysis; specifically, the production of the four Mandarin tones is described with laryngoscopic photographs of the larynx and pharynx taken for both modal and whispered registers.

Chapter Three

METHODOLOGY

Three experiments were designed for this study: the first experiment is a physiological observation on the pharynx and larynx; the second one involves stimulus collection and acoustic analysis; and the third one is a perception test. These three experiments are described separately in Sections 3.1, 3.2 and 3.3.

3.1 Experiment I

A Kay Elemetrics Rhino-Laryngeal Stroboscope 9100 was used to observe articulations of four Mandarin tones in phonation and whisper. This is a computer-controlled system with a halogen light source, a Mitsubishi S-VHS video cassette recorder BV-2000 and an oral endoscope. The rigid endoscope was put into the mouth of the subject over the tongue so as to have a clear view of the pharynx and larynx on the monitor. Using the oral technique, only a vowel sound similar to [ɨ] as in *sī* 'to think' in Mandarin Chinese can be produced by the subject of this observation, Dr. John H. Esling, a second-language speaker whose Chinese was acquired at the University of Michigan in 1971-73. He is also the Secretary of the International Phonetic Association. Articulation of four Mandarin tones in both phonated and whispered speech is videotaped at 30 frames/sec. Videotapes were edited into clips containing one tone in both phonation type each for observation and evaluation. Observations of the pharynx and larynx in the production of whispered Mandarin tones will be described in the next chapter.

3.2 Experiment II

3.2.1 Subjects

The participants in this experiment are two male and two female students. They are all native speakers of Beijing Mandarin; they were born in Beijing, and had never stayed in any place other than Beijing for more than 3 months before they came to Canada. Their length of stay in North America is between 19 and 36 months. The age range of the participants is between 18 and 27. Three of them are from the University of Victoria, and the other male subject is in grade 12. From their self-report, all subjects are free from any speech and hearing problems.

3.2.2 Speech material

3.2.2.1 Syllables in isolation

There are three sets of stimuli under this category, with each set having four tonally contrasting monosyllables. They are: 1) *bā* ‘eight’, *bá* ‘pull out’, *bǎ* ‘handle, dispose’, *bà* ‘father’; 2) *fā* ‘distribute’, *fá* ‘penalize’, *fǎ* ‘law’, *fà* ‘hair’; 3) *mā* ‘mother’, *má* ‘hemp’, *mǎ* ‘horse’, *mà* ‘scold’. The reasons for choosing these three sets of syllables are: first, all four distinctive tonal subtypes for each syllable have meanings so that subjects can produce all these tonal subtypes naturally; second, all three syllables have the same vowel to avoid intrinsic vowel differences of duration and amplitude; and finally the variation of the consonant part is also controlled for place of articulation as all of them are labials.

Every subject was asked to read aloud every tonal syllable three times consecutively before moving on to the next one in the above order. Then they were asked to whisper them three times in the same order as the phonated stimuli.

3.2.2.2 Mini-dialogue

All three sets of the target syllables are embedded into a mini-dialogue. The basic structure of the dialogue is:

A: _____ zì nǐ zhīdào zěnme xiě ma? (___ 字你知道怎么写吗?)

'How would you write the character _____?'

B: Wǒ bù zhīdào zěnme xiě_____. (我不知道怎么写 ___.)

'I do not know how to write _____."

A: Nǐ zhīdào _______ zì yīngyǔ zěnme shuō ma? (你知道____ 字英语怎么说吗?)

'Do you know how to say the word ______ in English?'

B: Bù zhīdào. (不知道.)

'I do not know.'

Therefore, each target tonal syllable appears three times in each mini dialogue. It appears in sentence-initial position, in sentence-final and in sentence-medial positions respectively.

The subjects were divided into two groups; each group had a male and a female subject. Each group was asked to perform the dialogue in a normal way first. Then they were asked to do it in whisper, and they were instructed to act naturally as if they were in the library whispering.

3.2.3 Recording procedure

Recording of the stimuli productions took place in the Phonetics Laboratory of the Linguistics Department at the University of Victoria. Subjects sat in the sound-treated room with a condenser microphone (AKG C1000S) at an angle of 45° in front of them. The distance between the speaker's mouth and the microphone was adjusted for optimal output (about 8 inches). Software 'Cool Edit Pro LE version 1.2' developed by Syntrillium Software Corporation was used to capture all productions as waveforms on the hard drive; a digital audio tape was also used to record the productions on a DAT deck for backup.

As mentioned earlier, every subject was first asked to read aloud every tonal syllable three times consecutively before moving on to the next one in order from Tone 1 to Tone 4. Then they were asked to whisper them three times in the same order as the phonated stimuli. Great care was taken to insure that the whispered samples were true whisper. Laryngeal pulses should not be detected auditorily by the investigator, and a few defective trials were repeated until acceptable test samples were produced and recorded. A total of 144 (12 x 3 x 4) phonated syllable samples and a total of 144 (12 x 3 x 4) whispered syllable samples were collected.

A similar recording procedure was used when two subjects produced the mini-dialogue. Each pair was asked to take its time preparing and practising the dialogue until they felt comfortable with the speech material; the investigator participated in their practice so that feedback and suggestions were provided to help them produce it as naturally as possible. Voicing was auditorily monitored, and some defective trials were detected and the subjects asked to repeat them until acceptable test samples were produced and recorded. A total of 12 pieces of (4 x 3 x 4) mini-dialogue in phonated speech were recorded since every subject rotated their roles as the questioner and the answerer. For the whispered version, every mini-dialogue was repeated once resulting in a total of 24 (2 x 4 x 3 x 4) pieces of mini-dialogue recorded.

3.2.4 Acoustic analysis one

Digital techniques were used to measure the duration and intensity of the/a/ vowels for each of the target words. All the stimuli are digitized at a sampling rate of 20,000 Hz by a 16-bit A/D converter using the Computerized Speech Lab (Kay/STR) analysis software. The Praat program produced by the Institute of Phonetic Sciences of the University of Amsterdam was used to get the values for this analysis.

3.2.4.1 Speech sample organization

Every syllable produced in isolation was put in an individual sound file; so that there is a total of 288 sound files, consisting of 288 syllables produced, among which half are in voiced phonation and the other half in whispered speech. Every mini-dialogue was divided into four separate sentences, of which the last sentence was discarded since it does not contain any target syllables. Every sentence was put into an individual file, resulting in a total of 144 phonated sentences (48 per sentential position) and a total of 288 whispered sentences (96 per sentential position).

3.2.4.2 Duration

The duration of the vocalic part of every syllable was measured. The vocalic part was isolated manually on the broadband spectrogram linked with the waveform. The first set is the syllable *ba /pa/,* where the vocalic part starts from the 'center of the releasing spike' (Peterson & Lehiste, 1960: 694); the second set is *fa /fa/*, where the fricative is easily recognized at the onset of the vowel formants; and the last set is *ma /ma/,* where the nasal ends as the steady formant pattern shifts abruptly to movement (Peterson & Lehiste, 1960).

3.2.4.3 Amplitude contour

Once the vocalic part of each syllable was identified and selected, the software automatically shows the mean intensity of the chosen part. Then the maximum intensity and the minimum intensity of syllables with Tone 1, Tone 2 and Tone 4 were determined. For syllables with Tone 3, maximum intensity at the beginning of the vocalic part, minimum at the middle part and maximum at the final part were calculated separately.

3.2.4.4 Statistical analysis

SPSS 10.0 software was used for further statistical analysis, mainly for the calculation of means, but the Independent-Samples T test was used for computing the significant level of differences between the normal and whispered speech samples. The probability (p) value for this study is set at 0.05 and 0.1: any p-value smaller than 0.05 is considered significant at the $p<0.05$ level, and any p-value between 0.05 and 0.10 is considered less significant, but still significant at the $p<0.1$ level.

3.2.5 Speech sample reorganization

The purpose of this procedure is to manipulate the recorded speech data and reorganize them for the perception test -- experiment III. All the whispered productions were manipulated in this section while only a small number of the phonated productions were manipulated.

First of all, the beginning and the ending parts of every syllable in isolation were trimmed, and a 1.5-second length of silence was inserted at both ends. This is to make sure that there is an equal length of silence (3 seconds) between every two syllables when played to the listeners. Then all 144 isolated whispered syllables/files were put in random order into one category, and a total of 21 randomly chosen phonated syllables were put into another category.

Similarly, the beginning and the ending part of every whispered sentence was trimmed, and a 2.5-second length of silence was inserted at both ends so that there were 5 seconds between every two sentences. Then they were put into 3 separate categories according to the position of the target syllable, with 96 sentences in every category. Sentences within each category were randomized.

Five categories containing five types of stimuli were prepared for the following experiment; they were stored on the hard drive separate from the raw data, and a copy was burned to a CD as backup.

3.3 Experiment III

3.3.1 Subjects

Participants in this experiment are 10 females recruited among Chinese students in Victoria. They are all native speakers of Chinese. Two of them are native speakers of Beijing Mandarin, three of them have stayed in Beijing for three to six years, and the other five subjects are native speakers of northern dialects. Their length of stay in North America is between two weeks and 32 months. The age range of the participants is between 18 and 36. They are free from any speech and hearing problems from their self-report.

3.3.2 Procedure

Stimuli for the perception test are taken from the second experiment and are described in section 3.2.5. In summary, the listening test is divided into 5 sessions in which listeners listen to five groups of stimuli including phonated syllables, whispered syllables, whispered sentences I (target syllable in sentence-initial position), whispered sentences II (target syllable in sentence-final position) and whispered sentences III (target syllable in sentence-medial position). There are 10-minute intervals after each session for rest and feedback, during which the investigator would ask for comments on what they heard in the form of informal chat.

Like the production task, the perception test also took place in the Phonetics Laboratory of the Linguistics Department at the University of Victoria. Two subjects at a time sat in the sound-treated room, with two headphones hooked up through the Samson Q5 Headphone Amplifier to the PC outside the test room. The stimuli were played by the Windows Media Player in the same random order for all 10 subjects. Before the perception test started, each subject was asked to read the handout with instructions for every session. The listeners were asked to use the standard University of Victoria answer sheet and had to choose from five choices for each syllable or sentence they heard to judge the tone of the target syllable. The first session was a pretest on their capability of identifying Mandarin tones in normal speech and also a warm-up for them to get used to the environment and the speed of the test. After the first session, the investigator collected their answer sheets and marked them. Only subjects who got an accuracy score of over 90% would continue with the other four sessions. None of the subjects from this study got an overall accuracy lower than 90%. The answer sheets for the remaining 4 sections were corrected automatically by the University of Victoria Computing Center.

3.3.3 Acoustic analysis two

Isolated syllables that were correctly perceived by over 8 listeners were regrouped and analyzed acoustically for a second run. The difference between the two analyses is that the timing factor was taken into consideration in the second run.

3.3.3.1 Duration

The duration of the vocalic part of every syllable was measured again. Since the selection of the vocalic part was made manually, the difference was controlled under 5 milliseconds. The consonant part was trimmed and the starting and ending time of the vocalic part were recorded. The starting time of the vocalic part is between 1.49 to 1.50 seconds.

3.3.3.2 Intensity

Syllables with Tone 3 and Tone 4 were the focus of the intensity analysis for the second run. The maximum and minimum intensity of the vocalic part of syllables with Tone 4 were measured again, and the time of the maximum and minimum intensity were extracted and recorded this time. The same was done for syllables with Tone 3 except that two maximum intensities at both ends were measured. Since a mathematical algorithm is used in the Praat software to get maximum and minimum intensity, taking sample period into consideration but not the visually determined maximum and minimum point on the intensity tier, there is a fine difference between the two analyses. However, this difference was controlled within 0.2 dB so that it would not affect the overall outcome of this analysis.

Chapter Four

RESULTS AND ANALYSIS

This chapter reports the results of the three experiments. Section 4.1 describes the pharynx and larynx based on observations of Mandarin tones produced in both phonations in experiment I. Section 4.2 reports the results from experiment II: the duration of four Mandarin tones in isolation, sentence-initial, sentence-medial and sentence-final positions, and the amplitude contours of Tone 3 and Tone 4 in these four environments. Results are also compared between phonated syllables and whispered ones. Section 4.3 deals with the results from experiment III: the accuracy percentage the listeners achieved. Results of the second acoustic measurements are analyzed in the last section, Section 4.4.

4.1 Observation of the pharynx and larynx

Tone differences are not as articulatorily obvious (by observation) as they are auditorily. It is difficult to view directly the cricothyroid mechanism, but its effect on vocal fold length can be seen. For phonated tones, the observed length of the vocal folds decreases from Tone 1 to Tone 2, and to Tone 3. The length of the vocal folds is taken as a primary indication of changing pitch. The higher the pitch (Tone 1), the longer the vocal folds. Larynx raising is observed in the later part of Tone 2 articulation, which might be related to the rising fundamental frequency of Tone 2; and it is not observed in Tone 1. Larynx raising is observed in Tone 4 at the beginning of the articulation, followed by lowering. The raising of the larynx found in Mandarin tones, therefore, is a secondary feature of pitch shifting, which provides some evidence for the assumption that larynx-height is related to pitch. It is also possible that both larynx raising and tongue-root retraction are related to increases in pitch, but we will consider these two mechanisms primarily in connection with the laryngeal sphincter, which is not traditionally thought of as having a role in pitch change. The only phonated tone that involves the aryepiglottic laryngeal sphincter mechanism is Tone 3. Figure 4-2 shows the pharynx and larynx in the production of phonated Tone 3. Figure 4-1 shows the state of quiet breathing for identification and comparison. In Tone 3, the aryepiglottic folds narrow to reduce the epilaryngeal tube area to a small triangular opening to the extent that part of the vocal folds is obscured. The aryepiglottic folds form an angle of about 150° with the arytenoids (see Figure 4-2-middle) as compared to almost a straight 180° line found in other tones in modal phonation and in the state of breath.

Fig. 4-1 Pharynx and larynx in breath (audible breathing out)

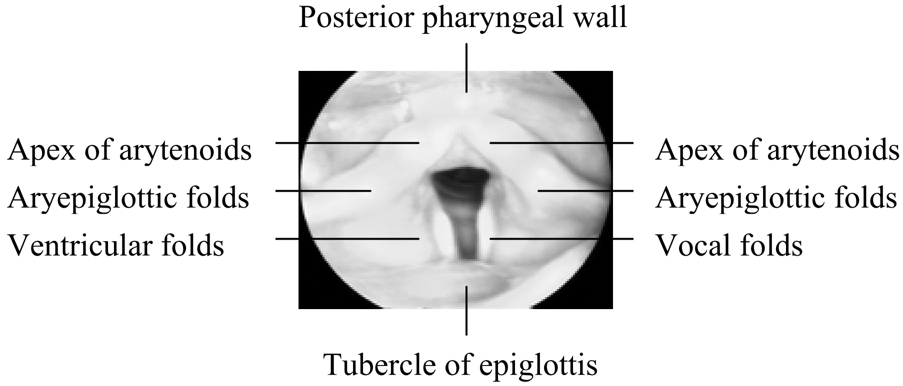
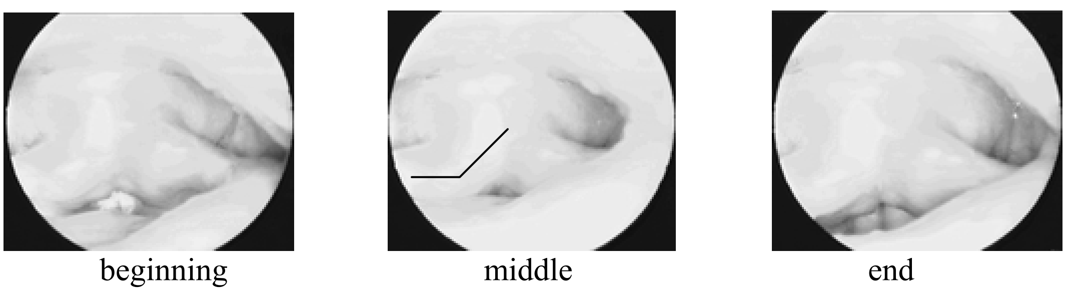


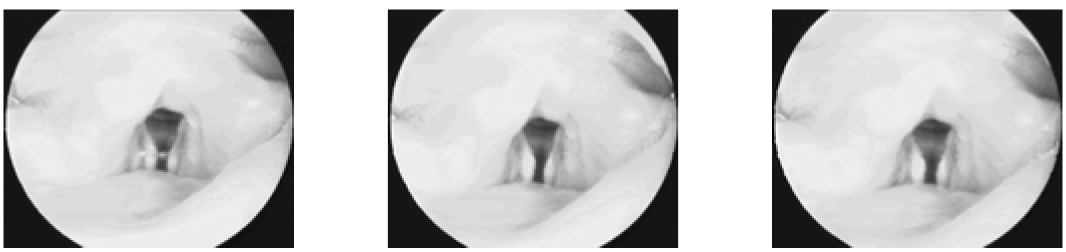
Fig. 4-2 Pharynx and larynx in phonated Tone 3



All the whispered tones are found to involve this aryepiglottic laryngeal sphincter mechanism (Esling, 1996, 1999). No vibration is observed between the vocal fold, which adduct slightly anteriorly to produce a 'Y' shape. The aryepiglottic folds adduct anteriorly and press up under the body of the epiglottis to reduce the epilaryngeal tube area to a narrow channel above the triangular posterior glottal opening. Larynx raising and tongue root retraction as concomitant features are observed. In addition, it is found that all whispered tones in these careful productions are preceded by an epiglottal stop. Because the tones were produced in citation, the speaker held his breath before each production in whisper, which would probably not occur in running speech.

Some differences are observed among the four tones. Tone 1 has a relatively larger triangular shaped lumen resulting from sphinctering, the open vocal folds are not obscured, and the aryepiglottic angle is about 120° as shown in Figure 4-3 (cf. Esling & Clayards, 1999). The larynx rises at the beginning of Tone 1, and both vocal fold opening and larynx height remain stable until the end of the syllable.

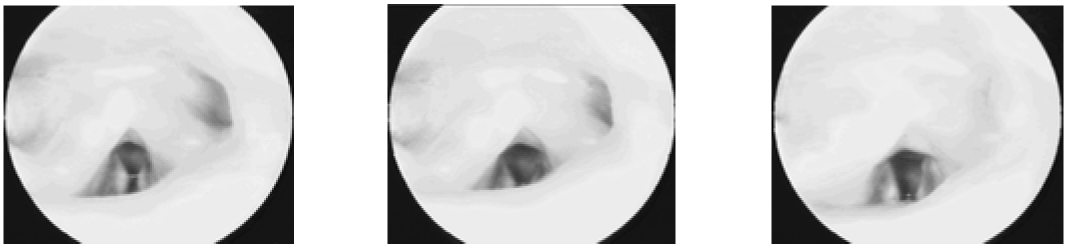
Fig. 4-3 Pharynx and larynx in whispered Tone 1



beginning middle end

Tone 2 has a smaller triangular opening in the area of the epilaryngeal tube than Tone 1, which is interpreted as a feature of larynx lowering, but the narrowing is still not small enough to obscure the vocal folds. The aryepiglottic angle seen in Figure 4-4 is also about 120° at syllable onset, similar to that of Tone 1. In addition to larynx raising as one of the concomitant features of the sphincter, the larynx is observed to rise after the beginning part of Tone 2, as can be observed in Figure 4-4 (middle and final). This corresponds to what is found in the phonated version.

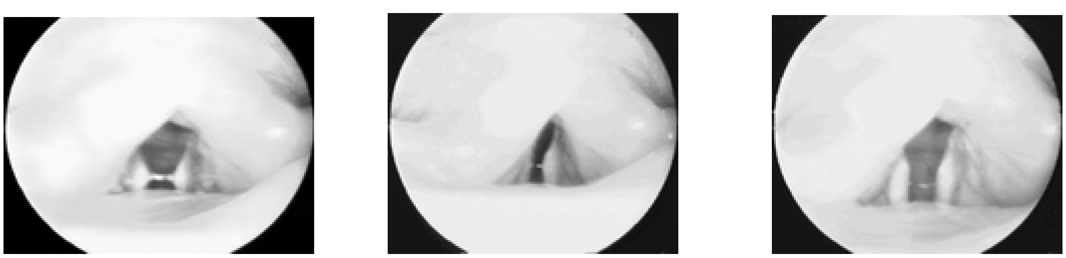
Fig. 4-4 Pharynx and larynx in whispered Tone 2



beginning middle end

Figure 4-5 shows the pharynx and larynx of whispered Tone 3, where more activation of the sphincter mechanism is incorporated, and both glottal and sphinteric changes take place as the pitch contour proceeds. Tone 3 has a sharper aryepiglottic angle of about 90°. The glottal airway is gradually reduced to a smaller opening, where the opening in the middle is smaller than Tone 2, but the vocal folds are still visible. The larynx is lowered towards the middle of Tone 3, which relates to the falling contour. In the later half (rising), the larynx rises as the opening between the vocal folds is increased. This observation of Tone 3 reaffirms findings from Tone 2 that larynx height is related to pitch; the higher the pitch, the higher the larynx. Thus, in the absence of vocal fold vibration, larynx height appears to be the primary feature for pitch shifting in whispered tones.

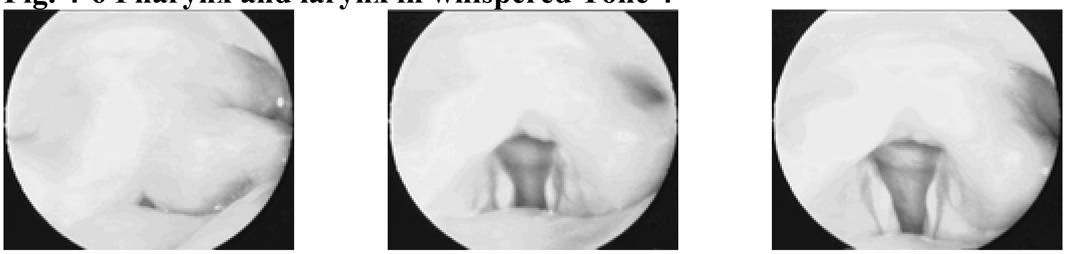
Fig. 4-5 Pharynx and larynx in whispered Tone 3



beginning middle end

Tone 4 has a dynamic articulation given that it has the greatest pitch range and is reported as the shortest tone in most of the literature. At the beginning, following full sphincteric occlusion, the vocal folds adopt a wide opening initially, accompanied by a further elevation of the larynx after the epiglottal stop. Then the larynx lowers abruptly and at the same time the glottal opening decreases in size until the end. The aryepiglottic angle remains around 100°. Figure 4-6 shows the state of the pharynx and larynx starting from the epiglottal stop for the articulation of Tone 4.

Fig. 4-6 Pharynx and larynx in whispered Tone 4



epiglottal stop [ʡ] beginning end

As discussed in the literature review, F2 (the second formant of vowel /a/) is retained in whisper and is related to the height of larynx. When the larynx is raised, F2 goes up; while when the larynx is lowered, F2 goes down. In company of the sphincter mechanism, however, the effect of sphinctering on F2 would be the same as for pharyngealization if larynx height does not change, i.e. F1 increases and F2 decreases. This complicates the evaluation of pitch in whisper using acoustic spectral means.

4.2 Acoustic analysis one

Experiment II is for data collection and acoustic analysis. Two acoustic parameters were examined: duration and amplitude contour. They are reported in sections 4.2.1 and 4.2.2 respectively.

4.2.1 Duration

Duration is one of the temporal envelope cues for tone recognition. The duration of the vocalic part of the tonal syllable is counted as the duration of the tone. A number of studies have been done on measuring the duration of Mandarin tones in normal speech. The finding that Tone 3 has the longest duration in isolation is obtained by most of these studies. However, according to Howie (1976), Tone 1 is the shortest; whereas Tseng (1990) and Fu & Zeng (2000) found that Tone 4 the shortest. These findings were based on data in various environments: Howie (1976) measured tonal syllables produced in a carrier sentence; whereas Tseng (1990) and Fu & Zeng (2000) measured isolated syllables. The discrepancy between their findings might result from the variation in contexts. The stimuli analyzed in this study include syllables in isolation and syllables in various sentence environments. The contextual variable is controlled in this analysis so that they are examined separately. Section 4.2.1.1 deals with the duration of syllables in isolation, section 4.2.1.2, section 4.2.1.3 and section 4.2.1.4 deal with the duration of syllables in three different sentential environments separately. Section 4.2.1.5 summarizes the findings on duration and compares these findings among four environments.

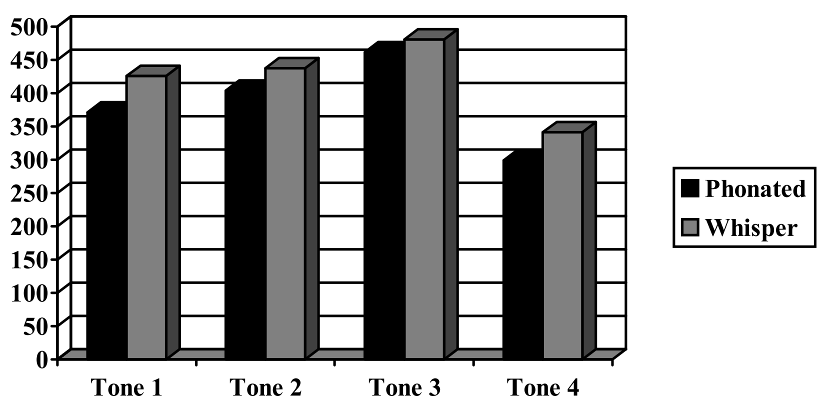
4.2.1.1 Syllables in isolation

Results from this section confirm the findings of Tseng (1990) that the longest tone is Tone 3, which has an average duration of 461.44 ms; and the shortest is Tone 4, with an average of 299.08 ms. Tone 2 (403.64 ms) is a bit longer than Tone 1 (371.40

ms). The order of duration for the four tones in this study is Tone 3, Tone 2, Tone 1 and Tone 4 from the longest to the shortest for both phonated and whispered data, the same as the ranking from Tseng's study. This result shows that the length pattern among the four tones in Mandarin is retained in whisper as well.

Comparison between normal and whispered samples reveals that all four tones are lengthened in whisper. Figure 4-7 displays this trend. The Independent-Samples T test was used to compute the duration differences between these two types of phonation, and the result is that there is significant difference for Tone 1, Tone 2 and Tone 4 ($p<0.05$), and that the difference for Tone 3 is not significant. This result provides evidence for the assumption that speakers lengthen the vocalic part in whisper as one of the special maneuvers for whispered tone production.

Fig. 4-7 Mean duration of tones in isolation (ms)



In order to control the differences that might result from the speakers' gender, the duration of all four tones is compared separately for male speakers and female speakers. Table 4-1 lists the mean duration for all four tones produced by male and female speakers and the differences between phonated and whispered speech.

Table 4-1 Average duration of Mandarin tones (ms)

	Male speakers

Tone	Phonated	Whisper	Duration Range
Tone 1	389.83	454.00	64.17*
Tone 2	425.33	472.11	46.78*
Tone 3	488.89	512.45	23.56**
Tone 4	329.61	372.22	42.61*
		Female speakers	
Tone	Phonated	Whisper	Duration Range
Tone 1	353.00	395.41	42.41*
Tone 2	381.94	402.89	20.95
Tone 3	434.00	448.61	14.61
Tone 4	268.55	309.78	41.23*
		Overall	
Tone	Phonated	Whisper	Duration Range
Tone 1	371.40	425.54	54.14*
Tone 2	403.64	437.50	33.86*
Tone 3	461.44	480.53	19.09
Tone 4	299.08	341.00	41.92*

*Difference is significant at $p<.05$
**Difference is significant at $p<.10$

The ranking of duration among four tones remains the same for both female and male speakers. Most of the duration differences between whisper and normal speech in Table 4-1 are significant. The finding that the duration of the vowels in whisper is longer than in normal speech agrees with the findings of Traunmüller et al. (2000). They found that the durations of all segments in Swedish produced by adults are longer in whispered speech than in corresponding phonated speech, but for speech produced by children, this lengthening is only found for vowels, while consonants in whisper are shorter than in phonated speech.

4.2.1.2 Target syllable in sentence-initial position

The lengthening of tones in whispered speech is verified again by the sentence-initial syllables. The durations of target syllables in whispered speech are 21.5ms to 29.5ms longer than those in phonation, and the differences for all four tones are statistically significant at the level of $p<0.05$. When the variable of the speaker's gender is controlled, most of the differences are still significant statistically except for Tone 4 produced by female speakers. The rankings of duration in the sentence-initial environment are different from the rankings among tones in citation. The pattern is shown in Figure 4-8 and 4-9.

Fig. 4-8 Duration of tones in sentence-initial position produced by male speakers (ms)

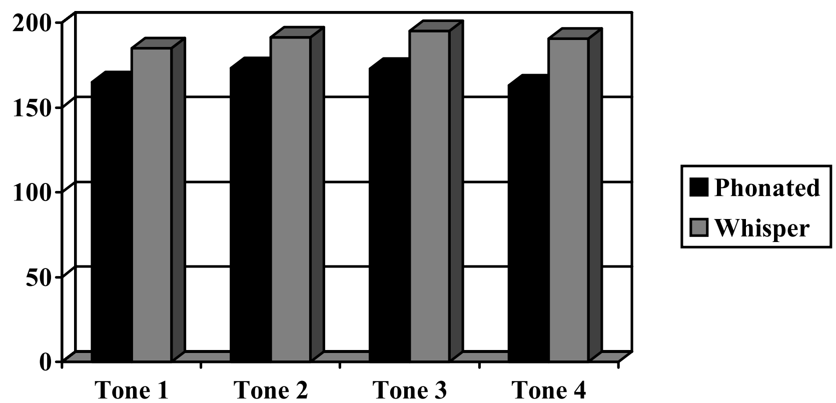
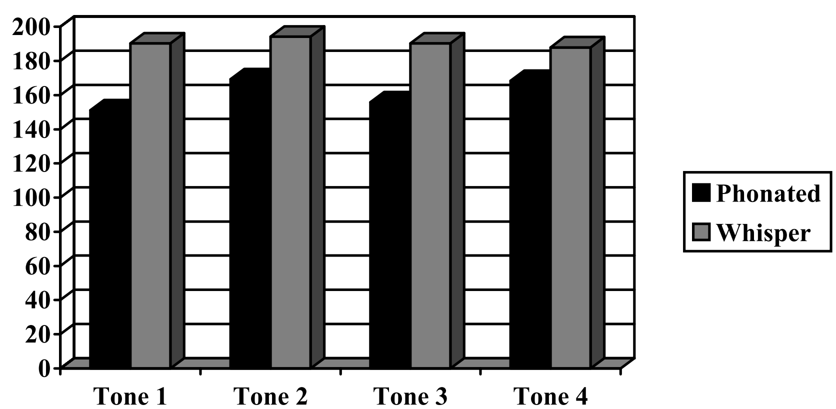


Fig. 4-9 Duration of tones in sentence-initial position by female speakers (ms)



There is no fixed order for the duration ranking of four Mandarin tones in sentence-initial position for either phonated and whispered speech. What is clear is that Tone 3 is no longer the longest tone, but ranks third in phonated speech. All four tones have similar length in this context, with the difference between the longest and shortest being around 10 ms. This tendency is more obvious in whispered speech. Given that the duration is generally longer to a large extent, the difference between the longest and shortest is only 5.21ms, less than that of phonated speech.

4.2.1.3 Target syllable in sentence-medial position

Howie (1973) made his measurement of duration of tones in the carrier sentence, *Zhèige ____ zì, shì Lǎo Lǐ xiě de.* 'This character is written by Mr. Li.' Although the sentence-medial context in this study *Nǐ zhīdào _______ zì yīngyǔ zěnme shuō ma?* Do you know how to say this word in English?' resembles it, the results of the two studies vary. Table 4-2 lists the results from the two studies for comparison.

Table 4-2. Mean duration of tones in sentence-medial position produced by male speakers

Tone	Howie (1976)	Phonated	Whisper
Tone 1	237 ms	133.17ms	152.58ms
Tone 2	271 ms	145.83ms	163.08ms
Tone 3	277 ms	139.00ms	162.25ms
Tone 4	251 ms	138.83ms	153.83ms

Since Howie used only male speakers to collect data, his results are compared with the results from male speakers in this study. Tones in whispered speech in both studies are again much longer than those in normal speech in my study, but even so, the duration of whispered tones measured in Howie's study (1976) is still about 100 ms longer than the whispered tones from this study. Two things can account for this huge difference: first, the carrier sentence in Howie's study has a pause in the middle that may have had the effect of prolonging the last stressed word before it; second, the sentences in my study were produced in the form of dialogue which resembles spontaneous speech, while there was only one carrier sentence in Howie's study, and it was presumably produced by speakers in a more formal way.

Whispered syllables produced by female speakers are also longer than phonated ones, and the differences found between these two phonations are mostly significant at either the $p<0.05$ or $p<0.1$ level.

Fig. 4-10 Duration of tones in sentence-medial position by male speakers (ms)

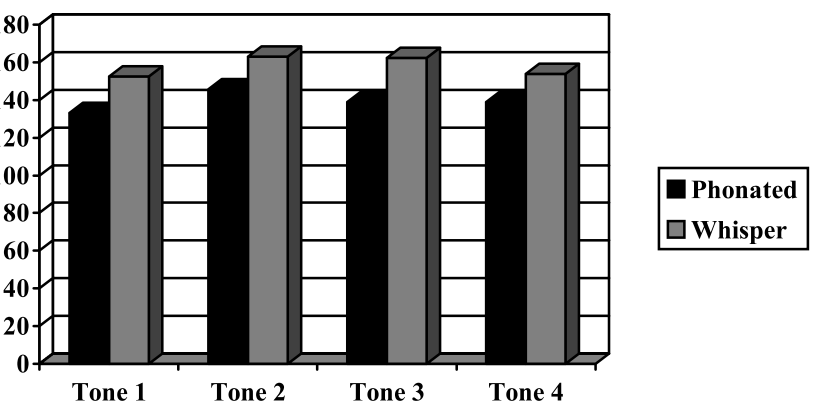
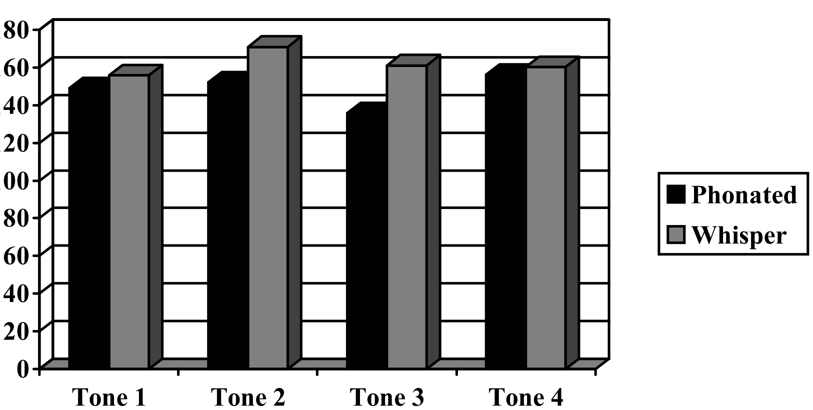


Fig. 4-11 Duration of tones in sentence-medial position by female speakers (ms)



Similar to tones in sentence-initial position, the duration ranking among the four tones in sentence-medial position has no clear pattern. Tone 3 is not the longest tone, and the duration for all tones is fairly similar. The differences between the longest and the shortest tones for two phonations are about 5 to 10 ms. Figures 4-10 and 4-11 display this clearly.

4.2.1.4 Target syllable in sentence-final position

Whispered tones in sentence-final position are again longer than their phonated counterparts. The differences for Tone 1 and Tone 4 are statistically significant, but not for Tone 2 and Tone 3. When I separate the data by speakers' gender, it can be observed in Table 4-3 that duration differences for male speakers are larger than for female speakers. In addition, whispered Tone 2 and Tone 3 produced by female speakers are shorter than the phonated versions. It is interesting that male speakers are more consistent in prolonging the length of tone in whispered speech than female speakers and do so to a greater extent. This will be discussed in the next chapter.

Table 4-3. Duration range between whispered and phonated speech in sentence-final position

Tone	Male speakers	Female speakers	Overall
Tone 1	61.01ms*	22.25ms	41.54ms*
Tone 2	35.92ms*	-9.75ms	13.11ms
Tone 3	33.00 ms	-2.83ms	15.09ms
Tone 4	61.09ms*	35.25ms	48.17ms*

*Difference is significant at p<.05

The ranking of duration for tones in sentence-final position has the same pattern as for tones in isolation. The order is Tone 3, Tone 2, Tone 1 and Tone 4 from the longest to the shortest for both phonated and whispered speech as shown in Figure 4-12, 4-13 and 4-14. Although Tone 2 is longer than Tone 1 in phonated stimuli produced by female speakers, the duration difference of these two tones is only 0.34 dB, and the order is observed in all other stimuli.

Fig. 4-12 Duration of tones in sentence-final position by male speakers (ms)

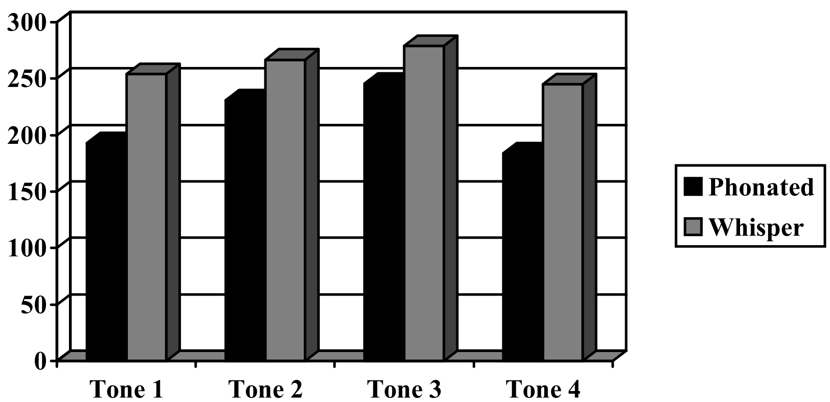


Fig. 4-13 Duration of tones in sentence-final position by female speakers (ms)

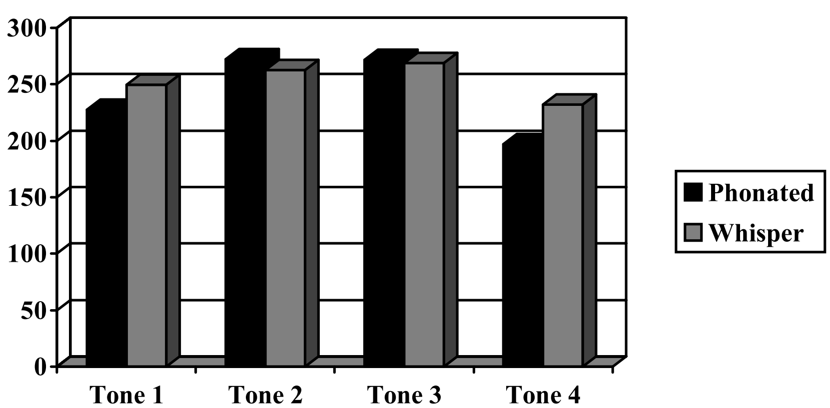
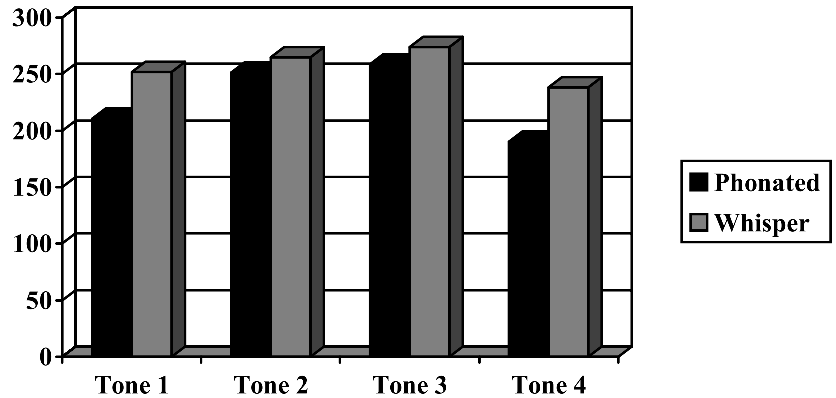


Fig. 4-14 Overall duration of tones in sentence-final position (ms)



4.2.1.5 Summary

Before proceeding to the summary of the duration section, we need to make a comparison of duration among four different environments. Duration among the four tones in Mandarin is compared in every section above, respectively, but what is the order of certain tones in different contexts? Previous studies (Ho,1971, 1976) on the influence of sentence environment on duration in normal speech revealed that the ranking in decreasing order is isolation, sentence-medial, sentence-final and sentence-initial. The ranking found in the current study is similar, except that tones in sentence-medial position are the shortest rather than the second longest as in Ho's studies (1971, 1976), shown in Figures 4-15 to 4-18 for the four different tones respectively. This inconsistency between findings from previous studies and this study can be accounted for by the difference in definition of sentence-medial position. Sentence-medial position in previous studies is the middle of a sentence but at the end of a clause and the carrier sentence is *Rú guǒ zhè gè zì zhēn niàn ______, nǐ jiù diǎndiǎn tóu ba!* 'If this character is really pronounced as _____, please nod your head.' But the target syllable in this study is in the middle of a sentence without clause or pause. Though the target syllables are both in a similar position, the one at the end of the clause resembles sentence-final position, which makes a big difference in duration.

Fig. 4-15 Mean duration for Tone 1 in various contexts (ms)

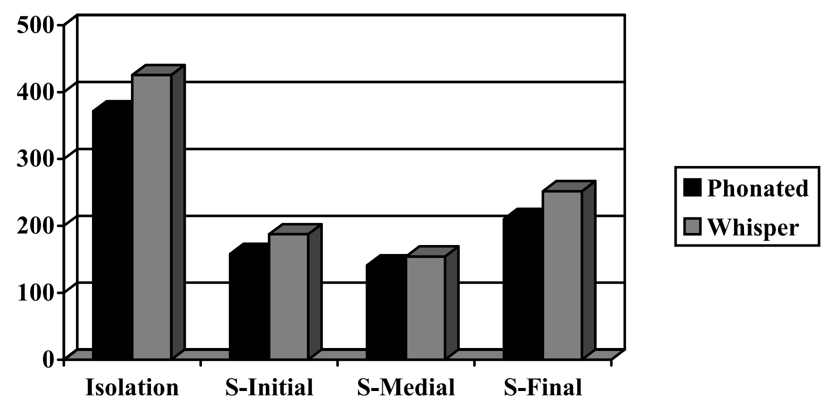


Fig. 4-16 Mean duration for Tone 2 in various contexts (ms)

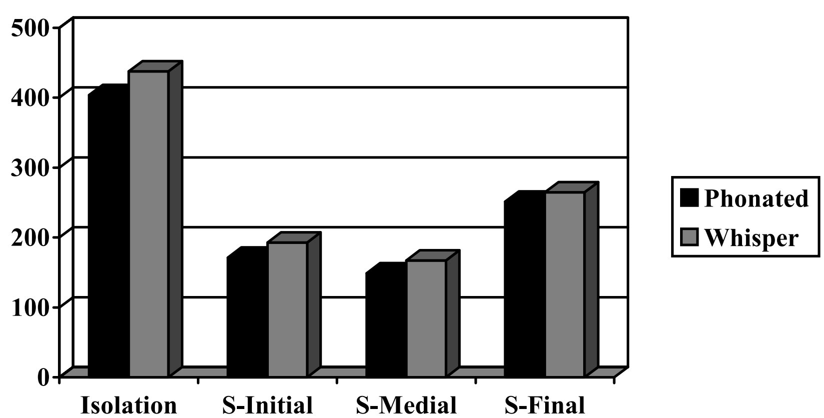


Fig. 4-17 Mean duration for Tone 3 in various contexts (ms)

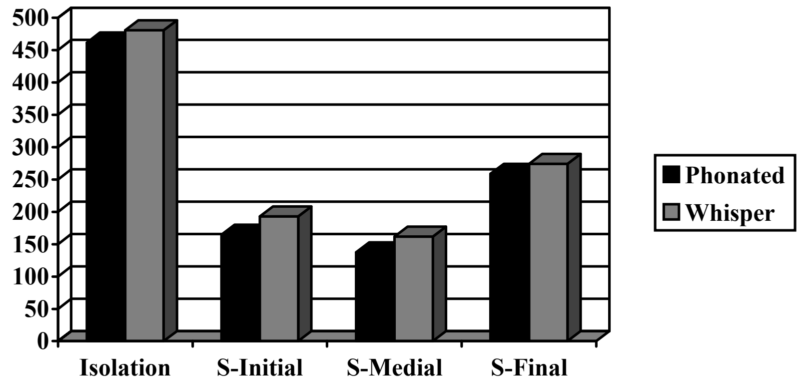
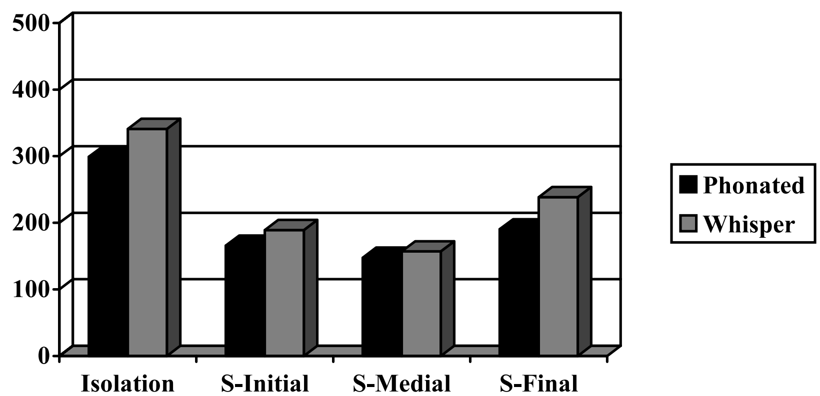


Fig. 4-18 Mean duration for Tone 4 in various contexts (ms)



The above section reports the findings on durations for four Mandarin tones in four environments produced in phonation and whisper. To sum up: (1) whispered tones are generally longer than their phonated counterparts in all four contexts, and the differences found between them are mostly significant at the $p<0.05$ or $p<0.1$ level; (2) the duration ranking among four Mandarin tones is Tone 3, Tone 2, Tone 1 and Tone 4 in isolation and sentence-final environments for both phonated and whispered speech; there is no clear ranking order for tones in sentence-initial and sentence-medial position; (3) the duration ranking within four different environments is isolation, sentence-final, sentence-initial and sentence-medial for all four tones in both phonated and whispered speech. Finally, given that all four tones in isolation have much longer duration than in other environments, and given that duration plays a role in tone perception as discussed in the literature review chapter, it is predicted that tones in isolation would be identified at a higher percentage rate than those in a sentential environment.

4.2.2 Amplitude contour

Amplitude contour as one of the concomitant features of tone has been studied by Fu and Zeng (2000), and they find that the shape of the amplitude contour of Mandarin tones is very similar to that of the F0 contour. Ho (1976) also claimed that amplitude is highly correlated with fundamental frequency, namely, the higher the F0, the higher the amplitude. Assuming they are correct, listeners would still get information about the tone from the amplitude contour in phonation even with the lack of fundamental frequency as in whisper.

Comparing the mean intensity between two phonations in the current study confirms Ho's (1976) conclusion that amplitude in whispered speech is generally lower than in normal speech. My analysis shows that most of the mean intensities of all four tones in whisper are significantly lower than those of normal speech.

The amplitude difference observed in this study is not surprising since the purpose of whispering is to reduce perceptibility. What is surprising, by comparing the mean of the intensity, is that this difference is observed to be much greater for female than for male speakers; the intensity range between phonated and whispered speech for female speakers is twice or even triple that of male speakers.

However, the analysis of the amplitude contour is more relevant than intensity to the purpose of this study, because the decrease of mean amplitude as a general trend does not necessarily mean that the speakers use any special maneuvers while whispering. But the amplitude contour, on the other hand, may provide information. Take Tone 4 as an example. Assuming that whisperers do use some special maneuvers to increase or decrease the airflow to make the falling tone more salient, then we might expect to find that the amplitude contour is more sloping in whispered speech than normal speech. Tone 3 and Tone 4 in Mandarin have clear amplitude contour patterns to examine, but Tone 1 and Tone 2 are relatively difficult to follow given that Tone 1 is a level tone and Tone 2 is a rising tone at the mid-high level. Therefore, Tone 3 and Tone 4 are the focus of the amplitude contour analysis. Section 4.2.2.1 and 4.2.2.2 examine Tone 4 and Tone 3 respectively, and section 4.2.2.3 summarizes the findings on amplitude contour.

4.2.2.1 Amplitude contour for Tone 4

As mentioned above, Tone 4 is a falling tone in Mandarin. The amplitude contour of Tone 4 decreases from the highest intensity to the lowest. The difference between the maximum and minimum intensity would provide some information on the slope of the amplitude contour. Female speakers have a significantly larger intensity scope, whereas it is the opposite for male speakers. Overall, the range between the maximum and minimum intensity for Tone 4 in whispered speech is not significantly different from that in normal speech (see Table 4-4).

Table 4-4. Mean intensity ranges of Tone 4 in isolation (dB)

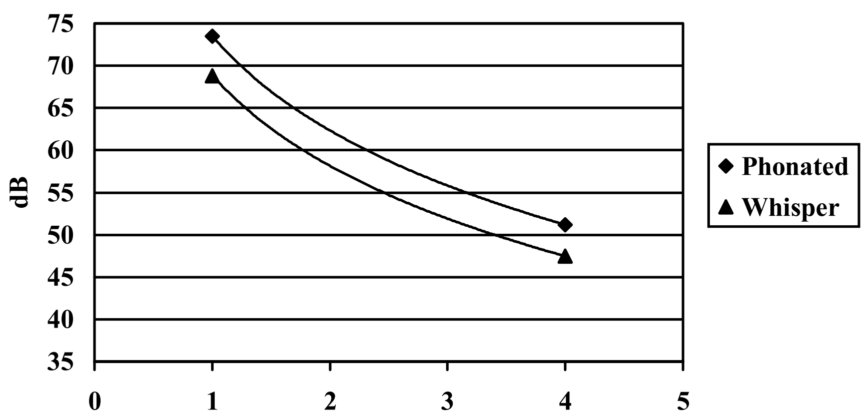
Speakers	Male		Female		Overall	
Phonation	Phonated	Whisper	Phonated	Whisper	Phonated	Whisper

Intensity Range Max.-Min. (dB)	22.29	18.78**	20.77	24.55**	21.53	21.67

**Significantly different from the correspondent under Normal category at p<.10 level

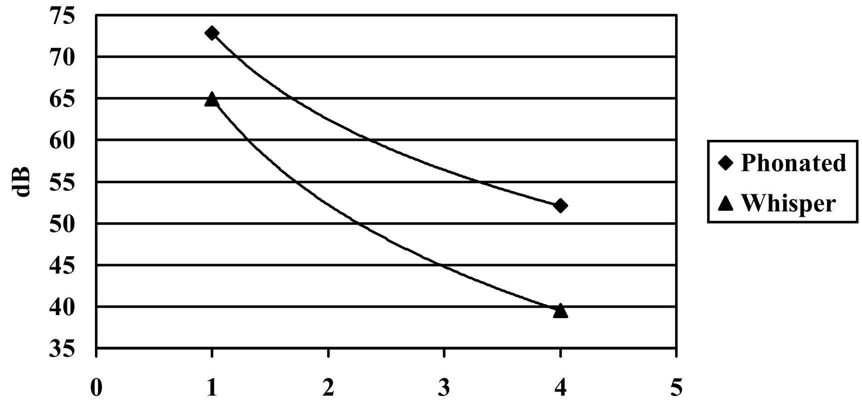
This finding was more straightforward when I computed the mean maximal and minimal intensity and put them on a scatter chart and connected them with trend lines. Figure 4-19 and 4-20 represent amplitude contour trends of Tone 4 for male and female speakers respectively.

Fig. 4-19 Amplitude contour of Tone 4 in isolation produced by male speakers



The amplitude contour lines in Figure 4-19 for male speakers are parallel, but for female speakers, the amplitude contour for whisper is more sloping than the normal one. This demonstrates the tendency for female whisperers to ‘exaggerate’ the amplitude contour when producing Tone 4.

Fig. 4-20 Amplitude contour of Tone 4 in isolation produced by female speakers



The amplitude contour of Tone 4 in three different sentential contexts has also been studied. The result conforms to the finding that female speakers tend to make the amplitude contour more salient in whisper, especially Tone 4 in sentence-initial and sentence-medial positions. Figures 4-21 to 4-24 outline this tendency with trend lines.

Fig. 4-21 Amplitude contour of Tone 4 in sentence-initial position for male speakers

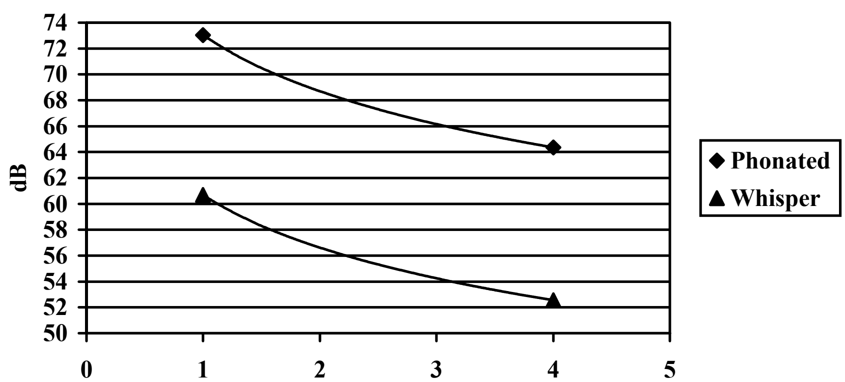


Fig. 4-22 Amplitude contour of Tone 4 in sentence-initial position for female speakers

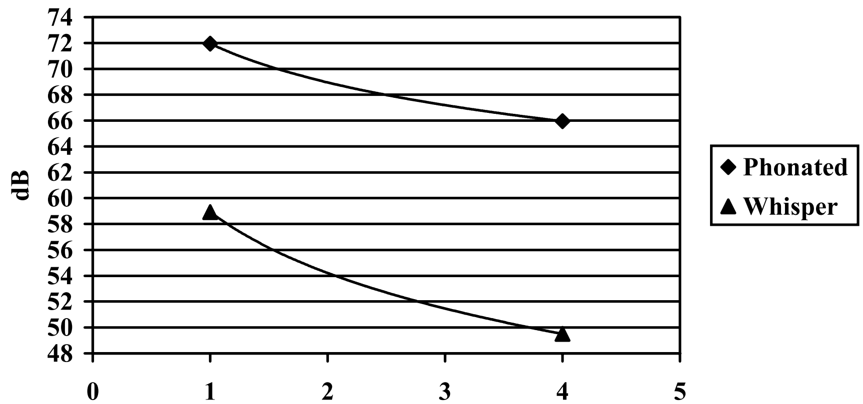


Fig. 4-23 Amplitude contour of Tone 4 in sentence-medial position for male speakers

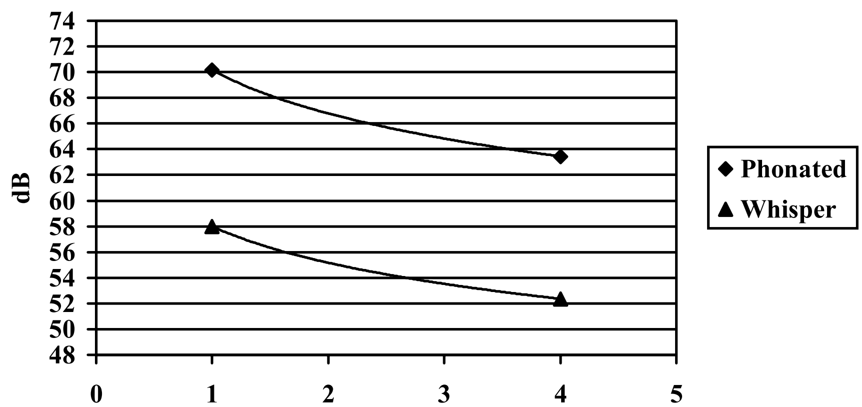
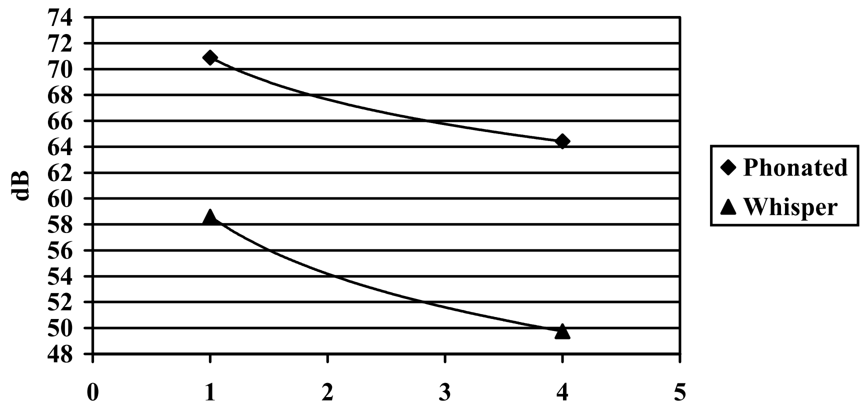


Fig. 4-24 Amplitude contour of Tone 4 in sentence-medial position for female speakers



The above figures show that the trend lines representing the amplitude contour of Tone 4 produced by male speakers in phonated and whispered speech are parallel, but those for female speakers reveal a tendency for whispered Tone 4 to have a more skewed trend line for amplitude. However, this trend is not found for Tone 4 in sentence-final positions. Although the difference is subtle, the fact that this ‘exaggerated’ amplitude contour of whispered tones by female speakers appears in various environments does provide some information on amplitude contour of tones produced by female speakers during whispering. This assumption will be strengthened later in the section on Tone 3.

Finally, the Tone 4 data in this study in a way verify the Mandarin Fourth Tone Sandhi first observed by Y. R. Chao (1968). This sandhi rule is: Tone 4 (51) → Tone 4 (53)/ ____Tone 4 (51). Although tone sandhi patterns are related to fundamental frequency, amplitude contours are highly correlated as stated earlier. Statistics show that the intensity range of phonated Tone 4 in sentence-final position (12.5 dB) is significantly higher than in sentence-initial (7.34 dB) and sentence-medial (6.59 dB) positions since another Tone 4 follows the target sample in sentence-initial and medial positions. The same pattern is also found among whispered samples. Thus, Tone 4, whether phonated or whispered, undergoes tone sandhi to become a falling tone from high level to mid level when followed by another Tone 4.

4.2.2.2 Amplitude contour for Tone 3

Tone 3 is a falling-rising tone in Mandarin, as discussed in the Literature Review chapter. Its amplitude contour first falls from maximal intensity on the left-hand side to minimal in the middle, and then rises from the minimal to the maximal intensity on the right-hand side. In order to compare the gradient of the falling and rising of Tone 3 in whispered and phonated speech, I separately measured the highest intensity on the left side, on the right side and for the lowest intensity in between. The differences between the left-hand maximum and the minimum, and between the minimum and the right-hand maximum were obtained from all samples. Table 4-5 provides detailed information on these differences and the comparison between the two types of phonations.

Table 4-5. Mean intensity ranges of Tone 3 in isolation (dB)

	Male		Female		Overall	
	Phonated	Whisper	Phonated	Whisper	Phonated	Whisper
Maximum(L)-Minimum(dB)	7.10	9.69	8.98	15.87*	8.04	12.78*
Maximum(R)-Minimum(dB)	8.38	7.94	7.59	10.04**	7.98	8.99

*Difference is significant at p<.05
**Difference is significant at p<.10

Statistics show that male speakers in this experiment did not use any special maneuvers for increasing loudness or decreasing it dramatically when whispering; rather, a smaller loudness increase occurred by comparison. The intensity range between maximal and minimal intensity for female speakers shows that their intensity scopes are significantly greater in whisper than in normal speech. Thus, this finding again confirms the general trend of the female speakers' intention to 'exaggerate' the intensity changes in whisper. Observe the following figures.

Fig. 4-25 Amplitude contour of Tone 3 in isolation produced by male speakers

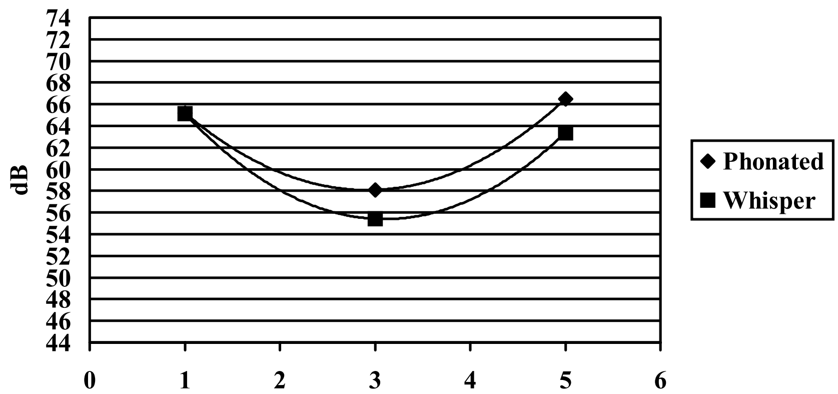
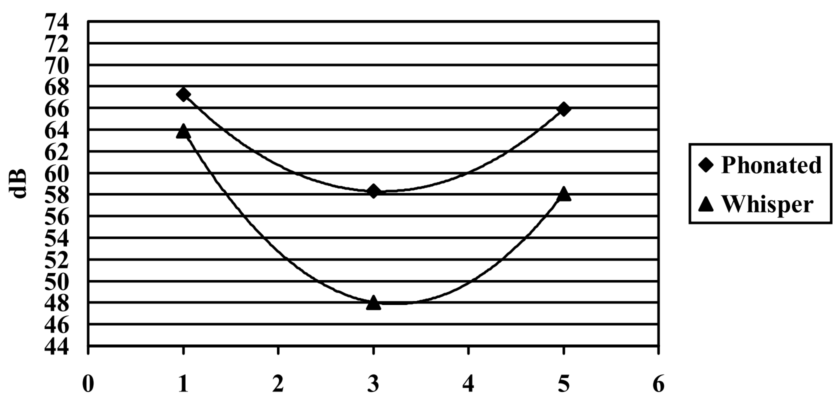


Fig. 4-26 Amplitude contour of Tone 3 in isolation produced by female speakers

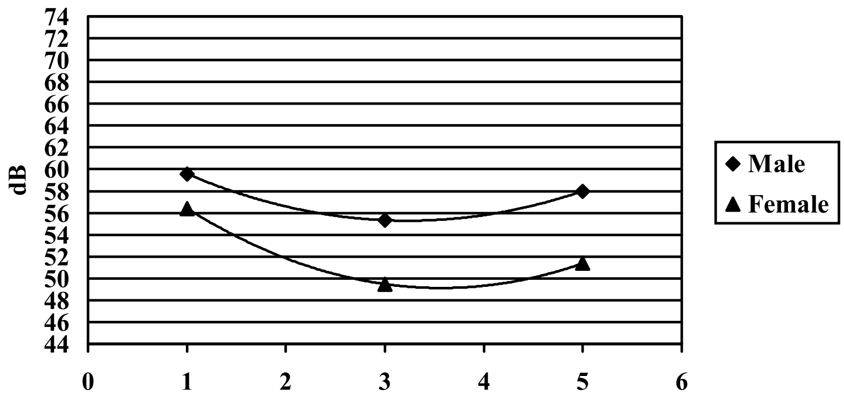


Similar to Tone 4, the average maximum on both sides and the minimum in the middle are computed, and these three values are plotted on the chart with trend lines used to reconstruct the amplitude contour for Tone 3. Figure 4-25 is for male speakers and Figure 4-26 is for female speakers. In Figure 4-25, the left maximum for whisper is more sloping than for normal phonation, but the lines on the right are parallel; thus it is not surprising to see that the amplitude trend lines of Tone 3 in two phonations are similar in shape and close to each other. The picture in Figure 4-26 differs and provides further evidence for the hypothesis that amplitude contour for whisper is more dramatic than for normal speech, resulting from the special maneuvers of female whisperers. The trend line for female whisperers is lower than the normal one in position, and its shape is more inclined.

Analysis of the amplitude contour for Tone 3 in three sentence environments is more complicated. According to H. Lin's (1996) non-linear phonological analysis on Mandarin tones, each of the four Mandarin tones consists of three level tonemes: Tone 3 is represented as LLM. If the tone is in non-final position, the last toneme is deleted (H. Lin, 1996). Since Tone 3 is a falling-rising tone, it would be difficult to keep the complete falling-rising amplitude contour when it is in sentence-initial and sentence-medial positions. Statistics from this study provide empirical evidence to support this analysis. Of all 36 Tone 3 samples from sentential environments produced in normal speech, only ten are observed to have the falling-rising contour from the intensity tier on the spectrogram, nine of which are from the sentence-final position. For whispered Tone 3, 24 out of 72 are observed to have the falling-rising contour. Comparison of the amplitude contour between phonated and whispered Tone 3 is made based on the ten phonated and 24 whispered ones.

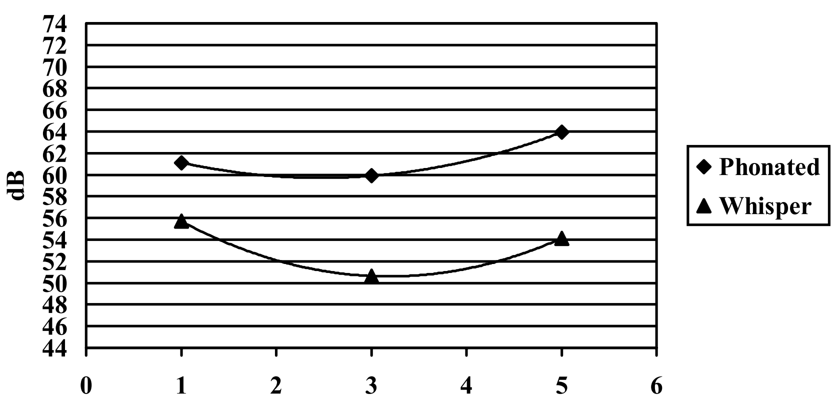
For the sentence-initial environment, a total of seven samples produced by male speakers and a total of four produced by female speakers are observed in whispered speech. It is clear from Figure 4-27 that the line representing amplitude contour for female speakers is more slanted in the falling part (left-hand) than that of male speakers.

Fig. 4-27 Amplitude contour of whispered Tone 3 in sentence-initial position for male and female speakers



Only one sample from normal speech and two from whisper are found in sentence-medial position, and they are all produced by female speakers. Again the tendency is shown in Figure 4-28 with the trend line representing the amplitude contour. The falling part of the whispered trend line is obviously more sloping than the normal one.

Fig. 4-28 Amplitude contour of whispered Tone 3 in sentence-medial position for female speakers



Samples from both male and female speakers and both normal and whispered speech in sentence-final position are found, for the reasons discussed above. Scatter charts (see Figure 4-29 and 4-30) are made for male speakers and female speakers separately. The fact that trend lines from male speakers for the two types of phonation almost overlap, whereas those from female speakers differ, adds another piece of evidence for the tendency of female speakers to try to make the amplitude contour more salient during whispering.

Fig. 4-29 Amplitude contour of whispered Tone 3 in sentence-final position for male speakers

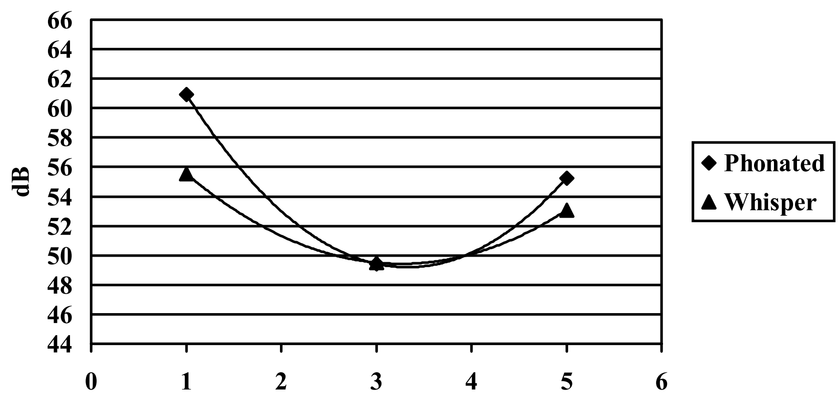
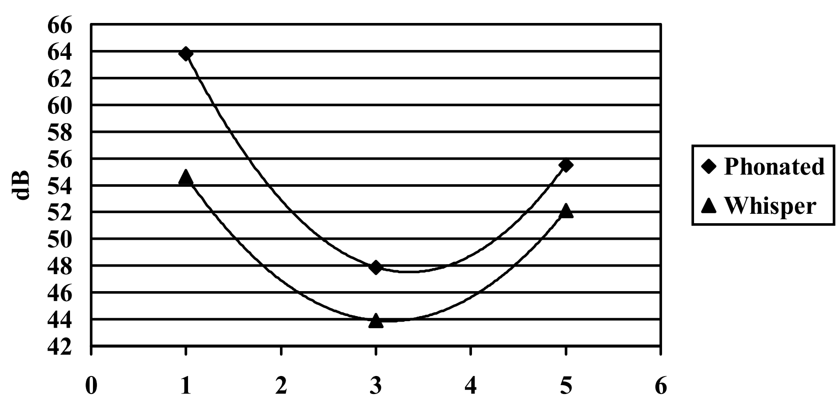


Fig. 4-30 Amplitude contour of whispered Tone 3 in sentence-final position for female speakers



Findings from Tone 3 are similar to those from Tone 4, in that male speakers tend to retain the same amplitude contour in whisper in regard to shape or slope; while female speakers tend to 'exaggerate' the amplitude contour, resulting in greater difference between the highest and the lowest intensity points.

4.2.2.3 Summary

This section examines the mean intensity difference between phonation and whisper for all four Mandarin tones first, but the focus is on the amplitude contours of Tone 3 and Tone 4. Major findings from this part are: (1) the mean intensity of all four tones in normal speech is significantly higher than in whisper, which is obvious in the sense that people want to reduce the perceptibility of speech when they whisper; (2) the pattern of amplitude contour difference found in both Tone 3 and Tone 4 matches in that male speakers retain a similarly slanted contour in whisper; while female speakers tend to 'exaggerate' the slope. However, the second finding is only preliminary, due to the fact that the duration factor was not taken into consideration in the analysis of the amplitude contour. Given that there is a significant difference in duration between phonated and whispered speech, and that the time when the maximum and minimum intensity occur is ignored, this result just sketches the possibility of further analysis. In the last section of this chapter, I will come back to this issue.

4.3 Tone identification

This section reports the results from the perception test in experiment III. Ten subjects listened to the stimuli and made judgments on the tones of the target syllable. The results are analyzed in sections 4.3.2, 4.3.3, 4.3.4 and 4.3.5 for four different environments, respectively; the effect of context on tone identification is examined separately in section 4.3.6. Before proceeding to the results of the test, I will summarize in section 4.3.1 the feedback collected from the subjects during the perception test.

4.3.1 Feedback from listeners

After listening to the group of whispered tones in isolation, the comment by most subjects is that they had never thought it was difficult to distinguish tones in whispered speech, and they found it to be hard and confusing. They had had a lot of experience communicating in whisper in Mandarin, and had seldom encountered any misunderstanding of tones. But when they went on to listen to the sentence environment groups, most of them found it even harder to distinguish tones than in citation. Six subjects thought that whispered tones in isolation were easier to hear than in sentences, and two thought that it was easier to perceive some tones in citation than in sentences, although they did not specify which tones. One of those subjects specified that tones in citation were easier to perceive than in sentence environments because they have longer duration. The other two subjects thought that tones in sentence environments are easier than those in citation, especially tones in sentence-final position. The reason is that tones before the target one provide some information on the speaker's pronunciation to aid the judgment. Within the three various sentence environments, eight subjects believed that target tones in sentence-final position were clearer than those in the other two positions. Among these eight subjects, some thought sentence-medial and sentence-initial were somewhat difficult to identify, while others favored one over another.

Nine subjects thought that it was more difficult to perceive Tone 1 than Tone 3 and Tone 4, and two of them thought that Tone 1 was the toughest tone to perceive of all four tones. Some subjects said that they always confused Tone 1 with Tone 2. Only one subject thought the other way, that Tone 1 and Tone 3 were easier to perceive than the other two tones. Two subjects told me that they hardly heard any Tone 1 in the sentence-initial group, although the number of each tone was the same in every group. Similarly, Tone 2 was not easy to perceive especially when compared with Tone 3 and Tone 4. Two subjects thought that Tone 2 was the most difficult to perceive of all four tones. All subjects agreed that Tone 3 was easier to perceive, at least compared to Tone 1 and Tone 2. One subject commented that the rising part in the end of Tone 3 was the key to perception. And another subject commented that when Tone 3 was not long enough, it sounded like Tone 2. As to tone 4, most subjects grouped it with Tone 3 as the easiest tones to perceive. In sentence-initial position, Tone 4 was not as obvious as in other environments, so that three subjects could hardly perceive any Tone 4 in this group.

Finally, these ten listeners also made comments on the four speakers that show the idiosyncrasies of the speakers as well as of the listeners. Seven listeners felt that the stimuli produced by male speakers were easier to perceive than those produced by female speakers. Both of the two male speakers were favored by some of these listeners according to their description of the sound. One of the female speakers was ranked as the least clear by two listeners, in that all the stimuli produced by her sounded like Tone 1 or neutral tone. Two listeners gave more credit to the stimuli produced by the female speakers than the male speakers, and the other subjects did not think there was much difference in the difficulty of perception of the stimuli produced by male and female speakers.

4.3.2 Results of the stimuli in isolation

The mean score for the perception test on isolated tones is 86.5 out of 144, and the average accuracy percentage is 60.1%. Table 4-6 lists the accuracy percentage for every subject, coded from 001 to 010.

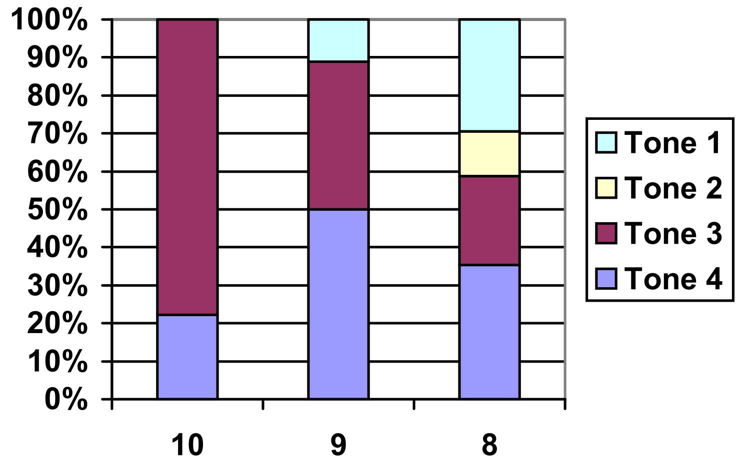
Table 4-6. Perception result for whispered tones in isolation

Subject Code	001	002	003	004	005	006	007	008	009	010
Correct No.	83	92	85	87	81	96	92	85	89	75
Percentage (%)	57.6	63.9	59.0	60.4	56.3	66.7	63.9	59.0	61.8	52.1

Comparing results from this study with previous studies, an average of 60.1% is much lower than what the listeners in Jensen's (1958) study achieved. The highest percentage, shown in the above table, 66.7% (subject 006), is still much lower than the lowest accuracy percentage of 73% in Jensen's study and the highest percentage of 88%. However, this discrepancy results from the different stimuli used in the two studies. Jensen used one minimal set *lā* (to pull, Tone 1) and *lá* (to cut, Tone 2) for two sessions, while in this study all four tones for the same phonetic form are tested. Although six minimal sets of words with all four tones were used in the other two sessions, only two tones were chosen for each minimal set. What is more important, the fact that 'the subjects were informed about the tone contrast applied to each of the minimal sets' (Jensen, 1958: 193) clearly helped the subjects to perceive tones at a higher accuracy rate.

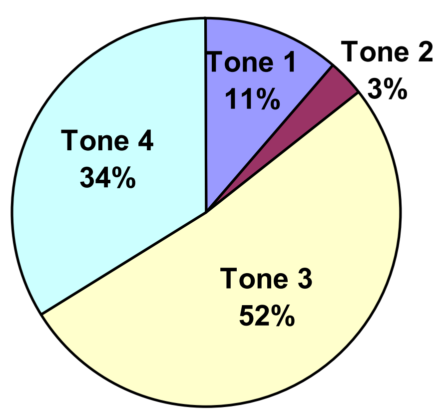
Out of 144 items in this test, 27 tones were correctly perceived by all ten listeners, 18 tones were correctly perceived by nine listeners, and 17 by eight listeners. Figure 4-31 shows how the four tones are distributed in these small groups.

Fig. 4-31 Tones in isolation correctly perceived



The number under each bar represents the number of subjects that made the correct judgment. It is clear from this chart that Tone 3 and Tone 4 are the main items that were identified as the intended tone by over eight listeners, while there are only a few instances of Tone 1 and Tone 2. This trend will be more evident if we add them up. Figure 4-32 reveals that over half of those samples are Tone 3, and one third of them are Tone 4.

Fig. 4-32 Distribution of four tones in isolated samples correctly perceived by the majority



Therefore, the ranking among the four tones in this environment in decreasing order of difficulty is Tone 2, Tone 1, Tone 4 and Tone 3. Of the nine items that none of the subjects perceived correctly, seven are Tone 2 and the remaining two are Tone 1, which adds another important piece of evidence to the ranking.

4.3.3 Results of the stimuli in sentence-initial position

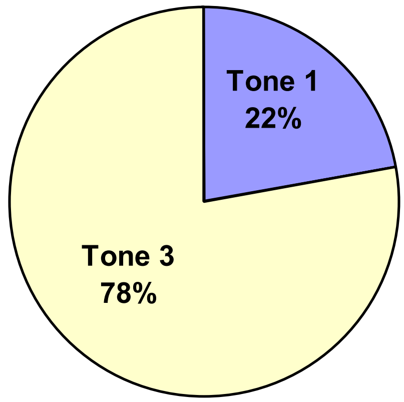
The mean perception score from this group of stimuli was much lower than that from the isolated stimuli, only 35 out of 96 were perceived correctly (36.5%). Table 4-7 lists the score and percentage for each subject.

Table 4-7. Perception result for whispered tones in sentence-initial position

Code	001	002	003	004	005	006	007	008	009	010
Correct No.	45	38	27	35	38	34	30	37	44	22
Percentage (%)	46.9	39.6	28.1	36.5	39.6	35.4	31.3	38.5	45.8	22.9

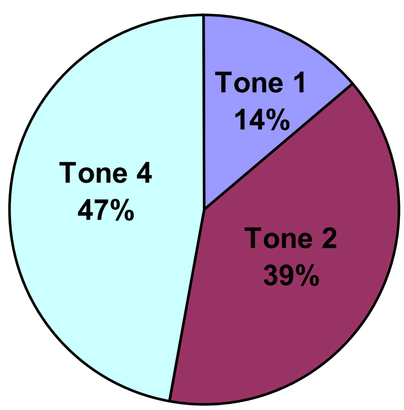
The fact that the highest percentage the subjects obtained is less than half (46.9%) provides evidence that tones in sentence-initial position are more difficult to perceive than tones in isolation. The number of samples that were perceived as the intended tone was far less than that of isolated stimuli. There is only one sample, Tone 3, was correctly perceived by all subjects; six samples by nine subjects, including five Tone 3 and one Tone 1; three samples by eight subjects, including two Tone 3 and one Tone 1. Put them all together in a pie chart in Figure 4-33, and the trend is obvious that Tone 3 keeps its dominant position as the easiest tone to perceive. Tone 1 becomes the second easiest tone to perceive over Tone 4.

Fig. 4-33 Distribution of four tones in sentence-initial samples correctly perceived by the majority



There are 36 out of 96 samples in this test that were incorrectly perceived by the majority of subjects (eight or over). It is surprising to find that about half of them are Tone 4, which turns out to be the most difficult tone to perceive; Tone 2 is the second most difficult, and some Tone 1 as well, shown in Figure 4-34.

Fig. 4-34 Distribution of four tones in sentence-initial samples incorrectly perceived by the majority



Thus, the ranking for this group of stimuli is Tone 4, Tone 2, Tone 1 and Tone 3 as the degree of difficulty decreases. This ranking is verified further when we look at the nine samples that none of the subjects perceived correctly, eight of which are Tone 4.

4.3.4 Results of the stimuli in sentence-medial position

The mean accuracy score from this group of stimuli is similar to the sentence-initial stimuli; an average of 36.3 out of 96 were perceived correctly (37.8%). The highest percentage achieved is 49% and the lowest is 22.9% as listed in Table 4-8.

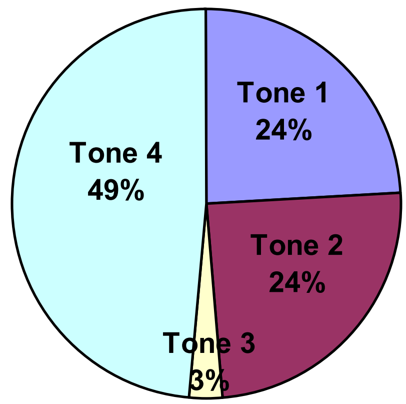
Table 4-8. Perception result for whispered tones in sentence-medial position

Code	001	002	003	004	005	006	007	008	009	010
Correct No.	40	34	35	27	39	48	34	41	43	22
Percentage (%)	41.7	35.4	36.5	28.1	40.6	49.0	35.4	42.7	44.8	22.9

Only two samples out of 96 were correctly identified as intended by all subjects, and they are both Tone 3; three samples by nine subjects, and two of them are Tone 3, one is Tone 1; and six Tone 3 by eight subjects. However, since there is only one Tone 1 that was correctly perceived by the majority, and the rest are all Tone 3, it is only certain that Tone 3 again is the easiest tone to perceive in sentence-medial position. Samples that were misjudged by the majority of subjects need to be analyzed before making the ranking.

There are a total of 17 samples that confused all listeners in this test, 11 samples misled nine of them, and nine samples trapped eight subjects. Figure 4-35 displays the distribution of the four tones in these samples. Tone 4 again takes the lead among all four tones, and Tone 1 and Tone 2 share the same number in these samples. In addition, one Tone 3 was not correctly judged by the majority of subjects.

Fig. 4-35 Distribution of four tones in sentence-medial samples incorrectly perceived by the majority



The ranking in this group of stimuli is then Tone 4, Tone 2, Tone 1 and Tone 3 as the degree of difficulty decreases, which is the same as the ranking for stimuli in sentence-initial position. Although Tone 2 and Tone 1 seem to be difficult to the same extent, given that eight Tone 2 were misjudged by either all listeners or nine of them, Tone 2 is ranked over Tone 1 in this part.

4.3.5 Results of the stimuli in sentence-final position

Wise and Chong (1957), in their perception test of the whispered tone of Mandarin Chinese, also used stimuli in sentence-final position. Their result is 2391 out of 3824 times (62%) correct (Wise & Chong, 1957). The accuracy percentage obtained from this study is about 10% less than theirs, average 51.4% correct (49.3 out of 96 samples). This difference can also be accounted for by the stimuli used. They composed thirty pairs of sentences, with the two sentences differing only in the tone of the target syllable in every pair. Thus, two tones were differentiated at a time, rather than four tones; in addition, there was some semantic context in the sentence that aided comprehension.

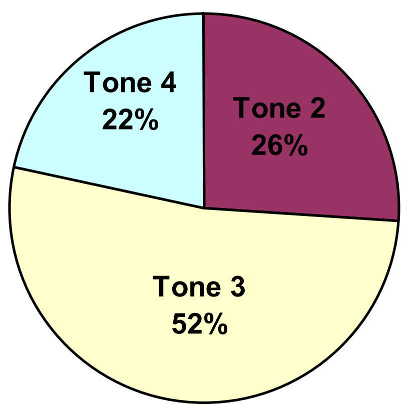
Therefore, the result we obtained in this study is generally consistent with the results from previous studies. Obviously, subjects were doing much better with stimuli in sentence-final position. The detail of each subject is in Table 4-9.

Table 4-9. Perception result for whispered tones in sentence-final position

Code	001	002	003	004	005	006	007	008	009	010
Correct No.	51	51	45	41	52	58	48	55	55	37
Percentage (%)	53.1	53.1	46.9	42.7	54.2	60.4	50.0	57.3	57.3	38.5

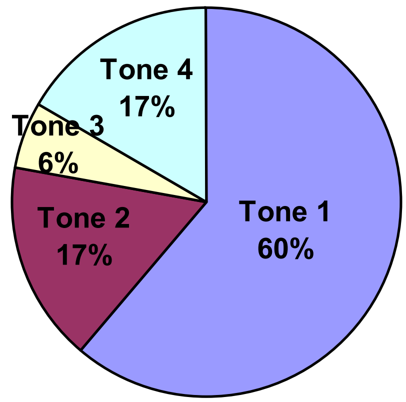
Subjects generally perform better in this part, and 23 samples out of 96 were perceived correctly by the majority of subjects. Figure 4-36 displays the perception pattern among four tones.

Fig. 4-36 Distribution of four tones in sentence-initial samples correctly perceived by the majority



Tone 3 keeps a leading role in the samples that received accurate judgments by eight listeners or over. Tone 4 and Tone 2 play a part as well. Tone 1 is not well identified in sentence-final position. Examining the samples that are not correctly identified by the majority of listeners reaffirms the observation. Over half of those samples are Tone 1, and Tone 2 and Tone 4 share the same number, as displayed in Figure 4-37.

Fig. 4-37 Distribution of four tones in sentence-initial samples incorrectly perceived by the majority



To summarize briefly, the ranking in this part is Tone 1, Tone 4, Tone 2 and Tone 3 as the degree of difficulty decreases.

4.3.6 Summary

The findings from experiment III, the perception test in this study, will be summarized here by comparing the results reported in the above section. In addition to learning how well listeners did in perceiving whispered tones, I also want to discover the effects that various environments have on the identification of whispered tones.

Comparing mean accuracy percentages that the subjects obtained with four types of stimuli, the percentage achieved with stimuli in isolation ranks first, followed by the stimuli in sentence-final, sentence-medial and sentence-initial position, as in Figure 4-38.

Fig. 4-38 Ranking of mean accuracy percentage for stimuli in four environments

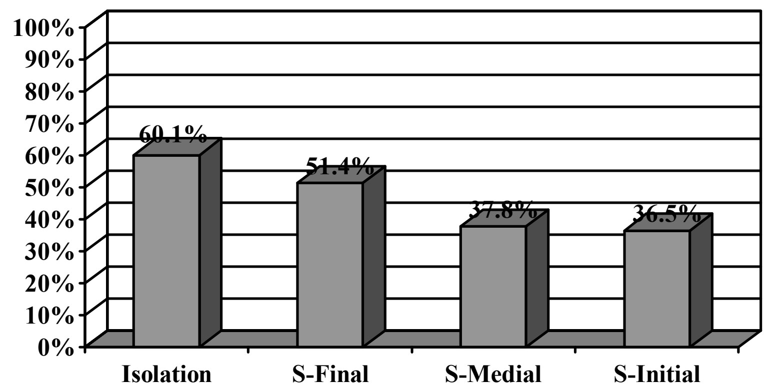
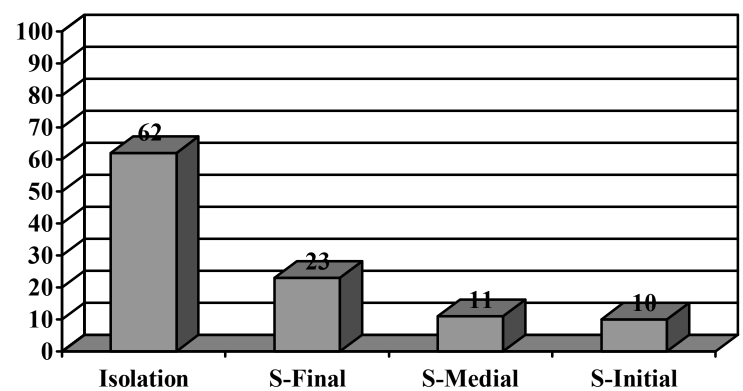


Fig. 4-39 Ranking of the number of samples correctly perceived by the majority



This ranking is further verified by the number of samples that were judged correctly by the majority of subjects. The big gap between the numbers for stimuli in isolation and in sentence environments, shown in Figure 4-39, substantiates the prediction made in experiment II that stimuli in isolation would be better identified. However, the order between stimuli in sentence-medial and sentence-initial positions is not clear; what we know for sure is that they are much more poorly identified than the other two types of stimuli.

Specific tones were perceived quite differently in the four contexts environments, especially Tone 4, Tone 2 and Tone 1. In addition to their all being incorrectly judged at a higher rate in sentence environments than in isolation, their rankings in different positions change, too. Table 4-10 outlines the four rankings among tones in four environments. The ranking as discussed at the end of each section is based on samples that were correctly and incorrectly perceived by the majority of subjects.

Table 4-10 Ranking of tones with respect to the degree of difficulty in perception test

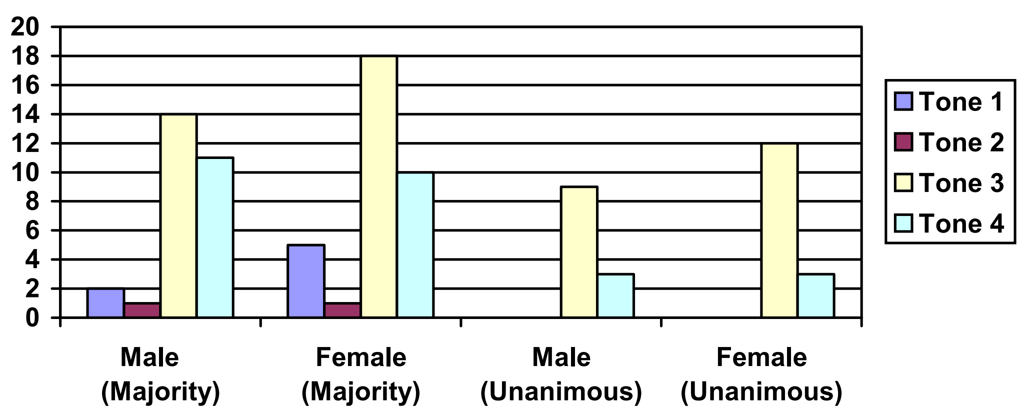
Contexts	Ranking
Citation	Tone 2 >> Tone 1 >> Tone 4 >> Tone 3
Sentence-initial	Tone 4 >> Tone 2 >> Tone 1 >> Tone 3
Sentence-medial	Tone 4 >> Tone 2 >> Tone 1 >> Tone 3
Sentence-final	Tone 1 >> Tone 4 >> Tone 2 >> Tone 3

As the degree of difficulty decreases in the rankings, Tone 3 remains the easiest tone to identify in all four environments. Although Tone 4 is the second easiest to perceive in isolation, listeners had a hard time judging it correctly in sentential environments, especially in sentence-initial and sentence-medial ones. Tone 2 was always difficult to perceive except in sentence-final context, and Tone 1 was easier than Tone 2, except in sentence-final context.

Finally, some brief comments are made in response to the listeners' feedback. First of all, that tones in isolation are perceived at a higher percentage rate than tones in other contexts confirms their feelings about the tones in different contexts. Tone 3 is unanimously agreed upon as the easiest tone to perceive in all contexts and was verified so from the results of the perception test. Their comments on other tones generally agreed with the results. Nevertheless, it is interesting to note the discrepancy between listeners' comments on speakers' performance and the listeners' actual performance in the test.

Seven listeners felt that stimuli produced by male speakers were easier to perceive than those produced by female speakers, two felt the opposite way, and one had no preference. The result, based on the analysis of the speech samples correctly perceived by the majority of listeners, is that the listeners did better with stimuli produced by female speakers. Figure 4-40 compares the number of the Majority samples (samples perceived as intended by the majority of subjects) and Unanimous samples (samples perceived as intended by all subjects) produced by female and male speakers separately. It is obvious that the majority of listeners were able to identify more tones produced by female speakers, especially Tone 1 and Tone 3. Although this discrepancy between listeners' intuitions and actual judgments in this study was unexpected and irrelevant to the topic, it would be a very interesting topic for future study.

Fig. 4-40 Distribution of the samples correctly identified by listeners



4.4 Acoustic analysis two

As described in chapter three, isolated samples that were perceived correctly by the majority of listeners were analyzed acoustically for a second time. The purpose of this analysis was to validate the findings and conclusions made in the first acoustic analysis, specifically the tendency of female speakers to exaggerate the amplitude contour when whispering. This analysis is based on a much smaller amount of data: 62 whispered tones in isolation, which include seven Tone 1, two Tone 2, 32 Tone 3, and 21 Tone 4.

This section is divided into three subsections dealing with duration, amplitude contour, and summary of findings, respectively.

4.4.1 Duration

The duration of these samples was measured for a second time. The duration for the whispered tones is still much longer than for phonated tones. Table 4-11 lists the overall duration of tones, and durations for male speakers and female speakers separately.

Table 4-11. Average duration of Mandarin tones (ms)

		Male speakers		
Tone	Phonated	Whisper	Whisper (Majority)	Whisper (Unanimous)
Tone 1	389.83	454.00*	479.5*(2)	N.A
Tone 2	425.33	472.11*	493**(1)	N.A
Tone 3	488.89	512.45**	520.3*(14)	535*(9)
Tone 4	329.61	372.22*	371.45*(11)	372.33(3)
		Female speakers		
Tone	Phonated	Whisper	Whisper (Majority)	Whisper (Unanimous)
Tone 1	353.00	395.41*	388(5)	N.A
Tone 2	381.94	402.89	398(1)	N.A
Tone 3	434.00	448.61	449.28(18)	445.83(12)
Tone 4	268.55	309.78*	297.2**(10)	306.33 (3)
		Overall		
Tone	Phonated	Whisper	Whisper (Majority)	Whisper (Unanimous)
Tone 1	372.00	425.54*	414.3**(7)	N.A
Tone 2	403.64	437.50*	445.5(2)	N.A
Tone 3	461.44	480.53	480.34(32)	484.05(21)
Tone 4	299.08	341.00*	336.1*(21)	339.33 (6)

*Significantly larger than the correspondent under Phonated category at $p<.05$ level
** Significantly larger than the correspondent under Phonated category at $p<.10$ level

Durations under the ‘Majority’ category refer to the duration of the tones perceived as intended by the majority of subjects (eight and over), while those under the ‘Unanimous’ category are those perceived as intended by all subjects. The number in the brackets on the right of duration means the number of samples that the mean duration is based on. The mean durations for all four tones under different whispering categories in this analysis are longer than their normal counterparts, and over half of them are statistically significantly longer. Thus, the finding that whispered tones are longer than phonated ones has been confirmed. However, it is interesting to note that the mean durations for tones produced by male speakers are almost all significantly longer, but not for female speakers. This issue will be further addressed in the discussion chapter.

Comparing the durations of the Majority samples with those of all whispered samples, there is no clear pattern found. Except that some tones under Majority produced by male speakers are longer, all the other durations under Majority are either shorter or about the same as those under Whisper. The difference between the durations under Unanimous and those under Whisper is not significant either.

4.4.2 Amplitude contour

Re-examining the amplitude contour for Tone 3 and Tone 4 is the key issue. As mentioned earlier, the tendency of female speakers to ‘exaggerate’ the amplitude contour is only an assumption, as the factor of timing is not taken into consideration in the first analysis. In order to verify this assumption, the amplitude contours of Tone 3 and Tone 4 in this small database were measured again, and so were the phonated counterparts matching the numbers (32 Tone 3 and 21 Tone 4) for later comparison. All these samples measured have about a 1.5-second silence at the beginning, and the consonant part in each sample is trimmed off so that the vocalic part of all the samples starts at about the same time, around 1.5 seconds. The starting time, time of maximum and minimum, and ending time of each sample were measured and recorded, and the mean values were computed. Figures 4-41 and 4-42 reconstruct the amplitude contour for Tone 4 produced by male and female speakers.

Fig. 4-41 Amplitude contour of Tone 4 (Majority) in isolation produced by male speakers

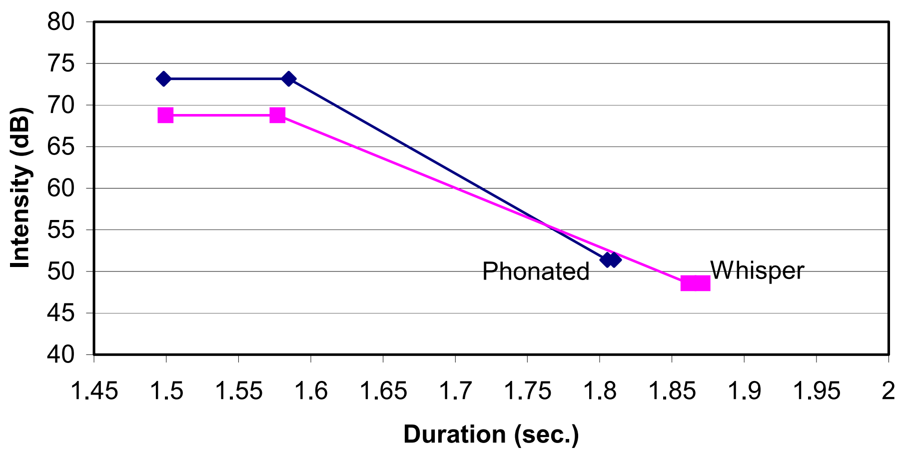
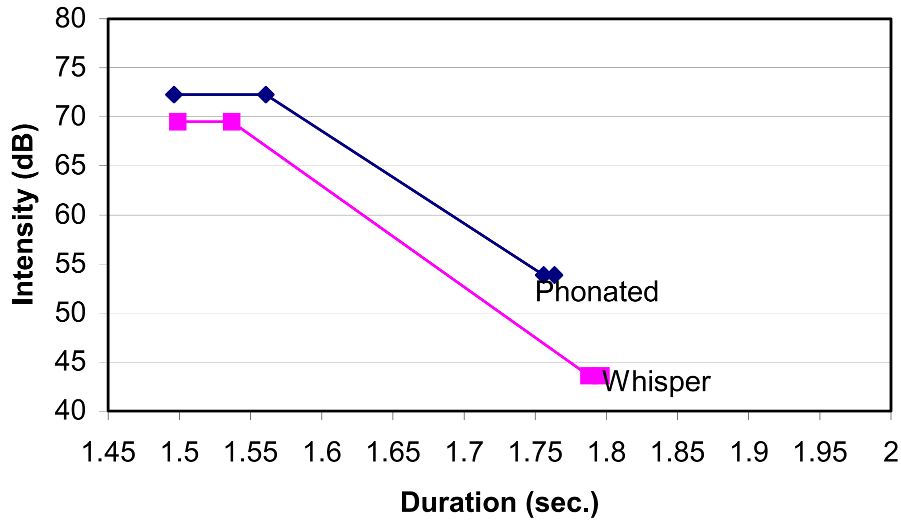
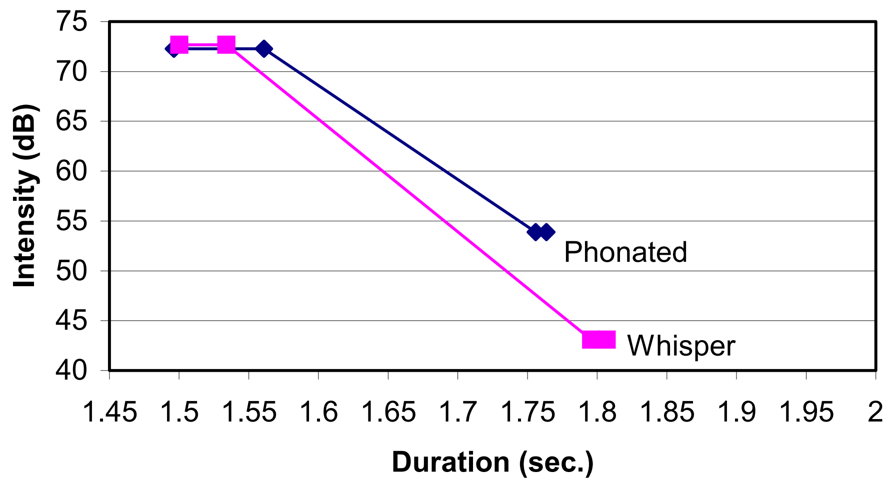


Fig. 4-42 Amplitude contour of Tone 4 (Majority) in isolation produced by female speakers



The mean starting time for Tone 4 is around 1.5 seconds for both phonated and whispered samples, but the time of maximal intensity for female speakers is earlier in whispered samples than in phonated ones. Obviously, the mean ending time of whispered samples is later than phonated ones, and so is the time of minimal intensity. Comparing the trend line of amplitude contour for male speakers, it is not surprising that the line for whispered speech is not parallel, and there is even less slope than the line for phonated speech. Male speakers made no attempt to ‘exaggerate’ the amplitude line. Female speakers, on the other hand, tended to make the amplitude contour more salient. Even though the two lines are visually parallel, the intensity ranges between the maximum and minimum intensity are significantly different. This tendency to ‘exaggerate’ in whisper by female speakers is strengthened by the analysis of the Unanimous samples, shown in Figure 4-43.

Fig. 4-43. Amplitude contour of Tone 4 (Unanimous) in isolation produced by female speakers



For all Tone 4 that have been accurately perceived by all listeners, there is a greater slope in amplitude contour for whisper than that of the phonated counterpart, and a greater intensity range. The maximum intensity occurs at about the same time for phonated and whispered samples, but the minimum intensity for whispered samples reaches a lower level. The amplitude contour hypothesis is substantiated with the visual observations of Tone 4 in this analysis. But for stimuli produced by male speakers, the amplitude contour lines overlap, and the phonated one is visually more sloping than the whispered one. Table 4-12 lists the intensity range for phonated samples, Majority samples and Unanimous samples, showing the tendency observed among female speakers in whispering.

Table 4-12. Intensity range (maximum – minimum) of Tone 4

Data Type	Phonated	Whisper-Majority	Whisper-Unanimous
Male speakers	21.78 dB	20.18 dB	26.57 dB
Female speakers	18.39 dB	25.89 dB*	29.59 dB*

* Difference is significant at p<.05

Similar analysis was applied to Tone 3 samples, except that Tone 3 has one more time of maximum intensity to observe. The trend lines for amplitude contour are sketched in Figure 4-44 and 4-45.

Fig. 4-44 Amplitude contour of Tone 3 (Majority) in isolation produced by male speakers

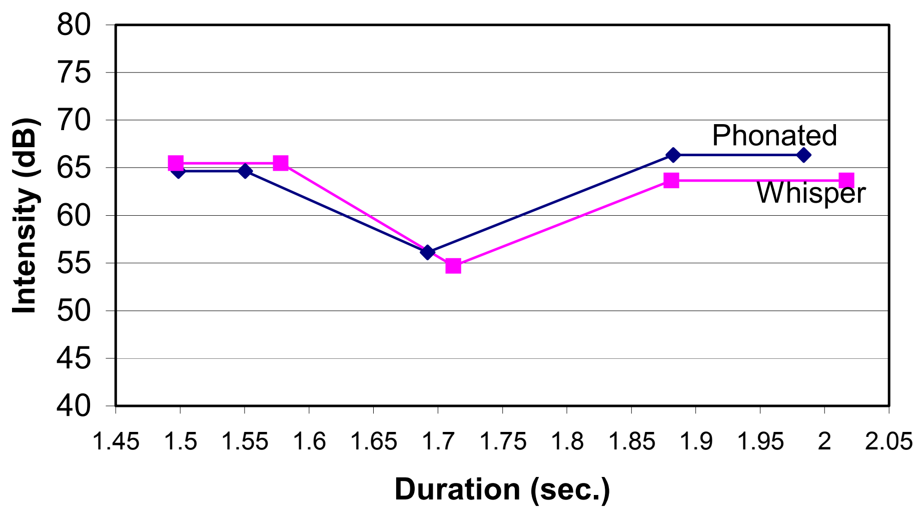
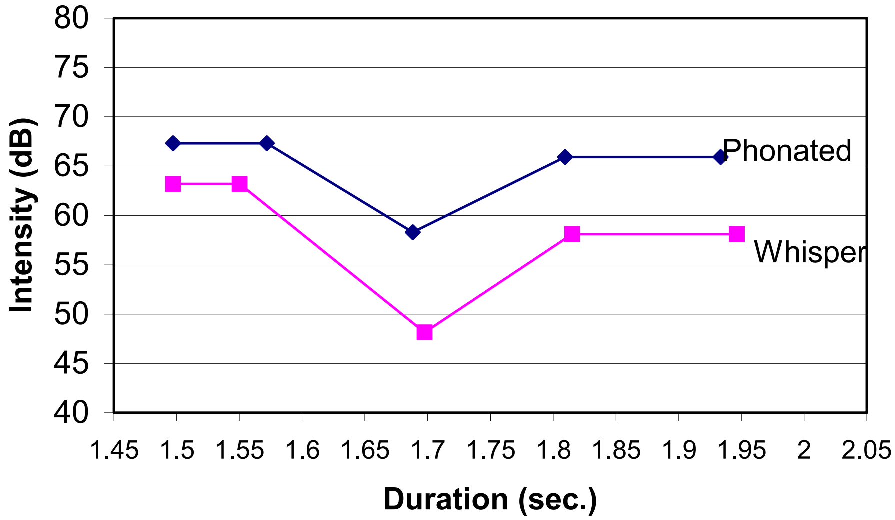
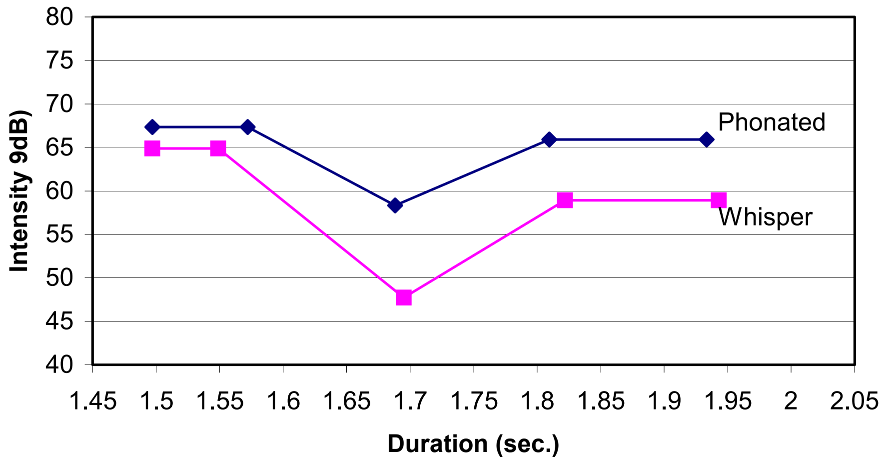


Fig. 4-45 Amplitude contour of Tone 3 (Majority) in isolation produced by female speakers



The amplitude contour for male speakers in whispered speech is parallel to that in phonated speech as to the falling part on the left, while the rising part on the right for whispered speech is less slanted than the part for phonated speech. But for female speakers, there is a subtle difference between the two lines representing the amplitude contour. The line for whispered speech is slightly more sloping than the one for normal speech on both falling and rising parts. This tendency is more evident when comparing the unanimous data with the phonated data in Figure 4-46. The amplitude contour for whispered speech is clearly steeper than the phonated one, especially the falling part on the left side. Therefore, the tendency of female speakers to increase the range and the slope of intensity for Tone 3 and Tone 4 during whispering is demonstrated by the analysis of Tone 3.

Fig. 4-46 Amplitude contour of Tone 3 (Unanimous) in isolation produced by female speakers



The intensity range of both the rising and falling part is listed in Table 4-13 for comparison between three types of data. Statistically, intensity ranges of samples produced by female speakers are all significantly greater than those of the normal samples, but only the range difference for the falling part of Unanimous samples produced by males is significant.

Table 4-13. Intensity range (maximum – minimum) of Tone 3

Data Type	Phonated		Whisper-Majority		Whisper-Unanimous	
	Max(L)-Min	Max(R)-Min	Max(L)-Min	Max(R)-Min	Max(L)-Min	Max(R)-Min
Male speakers	8.52dB	10.22dB	10.79dB	8.97dB	13.59dB*	10.36dB
Female speakers	9.02dB	7.6dB	15.03dB*	9.94dB**	17.15dB*	11.18dB*

* Difference is significant at $p<.05$
** Difference is significant at $p<.10$

4.4.3 Summary

This section reports the results of the second acoustic analysis based on a relatively small set of data. Nevertheless, the data are of importance to this study in that at least eight of the listeners out of ten gave the right answers to these samples, given that the perception accuracy is at a lower level (60.1%). Information provided by these samples is twofold: about what helps the listeners make correct judgments, and about the whisperers' ability to increase the identifiability of the tones. This analysis focuses on the latter one. Duration and amplitude contour were re-measured and analyzed, and the results substantiated the findings and assumptions made in the first acoustic analysis. In summary, the findings are: (1) durations of whispered tones are longer than phonated tones; (2) the durations of the samples that were identified as the intended tones by the majority are not necessarily longer than the mean duration of all whispered samples; in particular the samples produced by female speakers are generally shorter than the overall mean duration; (3) the reconstructed amplitude contour for Tone 4 in this small database is the proof for the assumption that female speakers tend to exaggerate the amplitude contour during whispering while male speakers have no such tendency; (4) the finding from Tone 4 is reduplicated in the analysis of the amplitude contour for Tone 3; so the special maneuver of exaggerating the amplitude contour observed among female speakers is proven with both visual and statistical evidence. This section concludes the experiments carried out in this study. Discussion of the research is presented in the next chapter on the basis of the findings reported in this chapter.

Chapter Five

DISCUSSION

This study is aimed at determining what contributes to the perception of tone in whispered speech. Three experiments were designed and carried out in this study. This chapter discusses the research questions in view of the findings and results of the experiments. Section 5.1 discusses the special maneuvers that speakers may use during whispering; section 5.2 deals with the temporal envelope parameters that listeners may use as a perceptual cue to whispered tones; section 5.3 comments on the effect of sentential environments in tone perception. The last section wraps up the discussion by providing a tentative answer to the main research question.

5.1 Speakers' special maneuvers

The notion of 'special maneuvers' was suggested by Abramson (1972) when he tried to explain the mixed perception results of whispered Mandarin tones. He argued that some speakers might 'go through special maneuvers to compensate for the missing pitch information' (Abramson, 1972: 33). A similar idea is also seen in other studies (Meyer-Eppler, 1957; Peterson, 1961; Mansell, 1973), but they used various terms like 'minor adjustment' (Peterson, 1961: 23) and 'reorganization' (Mansell, 1973: 2). A number of studies have been done on this issue at the articulatory level, that is on the reorganization of the articulatory mechanism in whisper. But few studies have investigated the nature of special maneuvers from the perspective of whisperers' behavior. This study attempts to investigate if speakers do use some special behavioral maneuvers during whispering. Based on the acoustic analysis of two temporal parameters, duration and amplitude contour, in the vocalic part of the target syllable in Mandarin Chinese, this study is able to come to a tentative conclusion.

Generally speaking, the duration of tones in whispered speech is considerably longer than that in phonated speech. In most cases, this difference between whispered and phonated speech is statistically significant. This finding is substantiated when speaker gender and speech context are controlled. For male speakers, the differences for all tones are found to be statistically significant, except for Tone 3 in sentence-final position. For female speakers, the duration of whispered speech is much longer than that of phonated speech, except for Tone 2 and Tone 3 in sentence-final position where whispered samples are slightly shorter than phonated ones. The pattern of tone lengthening found in whispered speech is definitely a special maneuver resulting from the phonation type. And it is reasonable and logical that under the circumstance of no voicing being allowed, whisperers should prolong the articulation time to provide more information to the listeners. As the statistics show, male speakers use this special maneuver to a greater extent than female speakers.

From the listeners' point of view, do they benefit from the lengthening of the vocalic part? Of the 62 samples collected which were correctly perceived by eight or more subjects in this study, the overall mean duration is about the same as that of all 144 whispered samples. The mean duration of the 62 samples is longer for male speakers, but not for female speakers. One possible explanation is that this special behavior increases the intelligibility of the speech by providing more information about the sound when the duration is prolonged to a certain extent. The lengthening strategy ceases to be of help to the listeners beyond that point, and listeners have to rely on information provided by other parameters.

The amplitude contours of Tone 3 and Tone 4 in Mandarin were studied. Scatter charts were used to plot the mean maximal and minimal intensity of Tone 3 and Tone 4. Then trend lines were used to connect these dots to reconstruct the amplitude contours. The trend lines for whispered samples produced by male speakers are either parallel or less sloping than the trend lines for phonated samples in all four contexts. However, the trend lines for whispered samples by female speakers are more sloping than those of phonated samples. In another words, female speakers use the special maneuver of increasing the falling slope for Tone 4 and Tone 3, and also the rising slope for Tone 3 to aid the identifiability of whispered tones. This finding is only preliminary, as duration and the time when maximum and minimum intensity occur were not taken into account. Nevertheless, the tendency based on this preliminary analysis was verified in the second acoustic analysis, in which the timing factor was taken into consideration. The trend line representing the amplitude contour for Tone 3 and Tone 4 is more sloping for whispered samples than for phonated samples. Statistical analysis also provides empirical support for this finding, that the range between the maximal and minimal intensity of whispered samples produced by female speakers is statistically greater than that of phonated ones; but the reverse pattern applies for samples produced by male speakers. Thus, this special maneuver of increasing the intensity range and slanting the shape of the amplitude contour to a greater degree is only found among female speakers. Female speakers exaggerate by producing a ‘dramatic’ intensity range to help listeners to perceive Tone 3 and Tone 4.

This special maneuver does aid the speaker to perceive Tone 3 and Tone 4 to a higher degree of accuracy. This finding comes from the comparison between samples perceived as the intended tone by the majority subjects and by all subjects. The intensity range for the Unanimous samples is significantly greater than for the majority. It is interesting that the same pattern was also found among samples produced by male speakers. Take Tone 4 for example: the intensity range of the Unanimous samples is over 6 dB greater than that of the Majority samples. It might be explained that male speakers use this special maneuver inconsistently when whispering or, similar to what Abramson (1972) suggested, that some of the speakers use this special maneuver but not all of them. However, since only two subjects of each gender were recruited in this study, further experimentation with a larger number of subjects is needed to validate this possible explanation.

The last but not least finding from this study on special maneuvers is unexpected but interesting and might be important to studies of gender differences in speech. Male and female speakers are found to use or prefer different speaking strategies during whispering to provide perceptual aids to listeners. As mentioned before, male speakers made use of lengthening, while female speakers increased the intensity range and slanted the amplitude contour. Although whispered stimuli produced by female speakers were also found to be longer than the phonated stimuli, most tones were not statistically significantly longer than their phonated counterparts. In addition, Tone 2 and Tone 3 in sentence-final position are even shorter than the phonated ones. Results of the perception test also support this finding. As discussed earlier, for male speakers, the mean duration of the Majority samples produced is longer than the mean duration of all samples; furthermore, the mean duration of the Unanimous samples is even longer than the Majority ones. This progression suggests that duration played an important role when listeners made correct judgments of the stimuli produced by male speakers: The longer the stimuli the higher the proportion of listeners who perceived them correctly. Since male speakers generally do not make use of the special maneuver for amplitude contour, lengthening becomes the only way that listeners can get extra information on tones in whispered speech. For the stimuli produced by female speakers, on the other hand, the mean duration shows the opposite pattern: that the Majority and Unanimous samples are in most cases shorter than the mean duration of all whispered samples. This is because listeners rely more on the special maneuver for amplitude contour to get the extra information on tones which female speakers use. This is symmetrical in the sense that male and female speakers adopt different speaking strategies in parallel with their gender difference. I will leave this issue to phoneticians who study gender differences in speech for further study.

5.2 Temporal envelope cues

This section again deals with duration and amplitude contour of tones in Mandarin Chinese. Discussion in this part differs from the last section in that the focus is on the intrinsic differences of these two parameters among the four tones, and how these differences contribute to tone perception in whispered speech. As reported in the literature review chapter, a number of experiments (Lin & Repp, 1989; Blicher et al., 1990; Tseng, 1990; Whalen & Xu, 1992; Fu & Zeng, 2000) have been conducted on the roles of duration and amplitude contour in tone identification. In summary, their findings are: (1) duration plays a minor role in tone identification compared with other cues; (2) duration helps to enhance the perception of the difference between Tone 2 and Tone 3; (3) listeners are able to perceive tones at a reasonable percentage solely from the amplitude; and (4) speech without F0 is intelligible most of the time. However, these findings were based on studies with synthetic or manipulated speech, which might deviate from speech in reality. Thus, one of the research goals of this study is to provide some insightful information on this topic.

Comparing the duration of tones in phonated and whispered speech, I find that the intrinsic tonal differences are well maintained in whispered speech, and that the ranking from the shortest to the longest tone is the same as that in phonated speech: Tone 4, Tone 2, Tone 1, Tone 3. The result from the perception test shows that Tone 3, being the longest in duration, is also the easiest to perceive. Is it because its length provides more information to the listeners? The result of the perception test for Tone 4 does not support this explanation. Being the shortest tone among all four tones, Tone 4 in isolation is the second easiest tone to perceive. Parameters other than the duration of Tone 4 must provide information to the listeners. Although duration does not play the key role in tone identification, we should nevertheless not overlook its function as a cue to tone perception. One listener from the perception test thought that when Tone 3 is not long enough, it sounds like Tone 2. Another listener made a comment that ‘some of the stimuli are too short to tell, I need a longer time to judge’. The same listener also explained why she thought tones in sentences are more difficult to perceive than tones in isolation: 'single syllables (tones in citation) are easier to perceive than in sentence environment, because they have a longer duration.' Indeed, the great difference in the duration between isolated stimuli and stimuli in sentences reveals that tone is definitely related to the fact that the perception scores from stimuli in various sentence environments are much lower than those from the stimuli in isolation. The gap, which can be as big as over 300 ms between stimuli in isolation and stimuli in sentence-medial position, would certainly reduce the identifiability of tones in sentence-medial position as well as in other sentential environments.

Amplitude contour, or intensity, is considered to play a more important role in tone perception by researchers. Pike (1948: 34) supposed that under circumstances where the toneme is either ambiguous or indistinguishable such as in whisper, intensity may help the perception, and may 'substitute for pitch and partly preserve the contrasts'. Analysis of the amplitude contour for Tone 3 and Tone 4 in this study provides some evidence for this suggestion. The reconstructed amplitude contours for Tone 3 and Tone 4 in both phonations resembles the shape of the pitch contour studied by a vast body of literature: falling and rising for Tone 3 and falling for Tone 4. Tone 4 is the second easiest tone to perceive is because of its huge intensity range from the beginning to the end of the vocalic part, even though it has the shortest duration. Feedback from the listeners also tells us that the amplitude contour aids them to perceive; one listener made her judgment of Tone 3 based on the rising part at the end. Like Tone 3 and Tone 4, Tone 1 and Tone 2 also have amplitude contour. Though they are not examined in this study, a flat shape for Tone 1 and a flat shape with a bit rising at the end for Tone 2 are supposed from their pitch contour. However, their amplitude contours do not seem to assist the listeners' judgments. The majority of listeners felt that Tone 1 and Tone 2 were more difficult to discern than the other two tones and in fact they were ranked much higher than Tone 3 and Tone 4 as the difficulty level decreased based on the perception results. Therefore, what really helps perception of Tone 3 and Tone 4 is not the amplitude contour, but the changing or moving intensity over time. The reason that Tone 1 and Tone 2 were poorly recognized is their lack of a salient amplitude contour to catch listeners' attention. Comparing Tone 3 with Tone 4, Tone 3 has a more prominent amplitude contour than Tone 4, characterized as first falling then rising; whereas the perception of Tone 4 depends on the huge range between maximal and minimal intensity. If the intensity range is not big enough, Tone 4 has no advantage for better perception than Tone 1 and Tone 2. As such, Tone 4 in sentence-initial and sentence-medial environments is ranked as the most difficult tone to perceive, and it is the second most difficult in sentence-final position. Of all the 62 samples that were correctly perceived by the majority of subjects, 32 are Tone 3, 21 are Tone 4, seven are Tone 1 and only two are Tone 2. In other words, about 90% of Tone 3 and about 60% of Tone 4 in the listening material could be identified by the majority of subjects. Therefore, the amplitude contour with distinctive characteristics plays a very important role in the perception of tone in whispered speech, and for Tone 3, intensity almost takes the place of pitch as the key perception cue.

With the absence of fundamental frequency in whispered speech, duration and amplitude contour clearly help people to identify tones to a certain degree, as the mean accuracy percentage shows. But they are unable to substitute for fundamental frequency completely, as the percentage score is much lower for whispered speech than for normal speech. The role of duration and amplitude contour in tone perception overall can also be proven by the comparison of percentage scores in isolation and in sentence environments. With the length of tone declining dramatically and the intensity range being unable to display in sentential environments, people could hardly get any information from the duration and amplitude contour to make correct judgments, and they did poorly on the test.

5.3 Contexts

The effects of context on tone identification have been studied for both phonated and whispered speech. Context in most of the studies reported here refers to phonetic context but not semantic context. Howie (1976) did his perception test on phonated stimuli in isolation and in a carrier sentence. The result is that the intelligibility is 95.4% for isolated stimuli, and 81.3% for sentence-medial stimuli. His study shows that the existence of phonetic context is actually a distraction to the tone perception in normal speech. Wise and Chong (1957) embedded target syllables from Mandarin Chinese into sentences and did a perception test on six subjects. The conclusion they came to with the result of the perception test is that a tone language is not consistently intelligible when whispered (Wise & Chong, 1957), indicating that context is of no help for listeners in making correct judgments. However, Jensen (1958) made his positive comments on the basis of context, although he did not specify the type of context.

Three different phonetic contexts are used in this study to determine whether the existence of contexts is a distraction or a helpful cue to tone perception in whispered speech. With the semantic context controlled, the target syllable is embedded in sentence-initial, sentence-medial and sentence-final positions. As summarized in previous chapters, the existence of context overall is a distraction for tone perception in whispered speech: the accuracy percentage for stimuli in sentential environments is significantly lower than that for stimuli in isolation. From the listeners' feedback, most of them felt the same way, that generally speaking, sentential contexts do not help at all. The reasons they gave are that syllables in sentences are too short to judge, and that there is no semantic context. We have discussed the first reason in the last section: because syllables are much shorter in sentential environments, especially in sentence-initial and sentence-medial positions, speakers are unable to fully 'express' the amplitude contour of a certain tone in a sentence. As a result, listeners are unable to get the same amount of information from duration and amplitude as they do from isolated syllables, and they are not likely to make judgments at higher or even the same accuracy rates with stimuli in sentences as with stimuli in isolation. The other factor, the semantic cue, is an important perception cue that this study does not cover. But undoubtedly, semantic context is always playing a crucial role in the intelligibility of speech especially speech under special circumstances such as whispering. Although Wise and Chong (1957) claimed that Mandarin tones are not consistently intelligible in whispered speech, they obtained a much higher average perception score (62%) with their tonal syllables embedded in sentences than the outcome of this perception test (51.4%). The semantic context in their carrier sentence could account for this discrepancy. The example they gave of a carrier sentence is:

(1) Wǒ yào **tāng**. / Wǒ yào **táng**.
I want soup. / I want candy.

The target syllables that these sentences carry are Tone 1 and Tone 2. Corresponding Tone 3 and Tone 4 syllables would not fit syntactically and semantically into the carrier sentence as they are both verbs, *tǎng* ‘lie down’ and *tàng* ‘burn’. Therefore, listeners would eliminate Tone 3 and Tone 4 as possible choices when they heard this sentence, realizing that only syllables with Tone 1 or Tone 2 can fit. Their chances of getting the right answer were doubled by the aid of this semantic context. Everything considered, this study reaffirms the significant role of semantic context in communication with whisper.

In addition to finding out the effects of context overall, I also wanted to find out the different effects that these three environments may bring to listeners’ perception. The ranking I found in decreasing order of perception difficulty was: sentence-initial, sentence-medial and sentence-final. Tones in sentence-final positions are generally about 100 ms. longer than those in other sentence positions, which accounts for the different outcomes of the perception tests. One of the listeners also mentioned that the tones before the target tone provide some information on the speaker’s pronunciation to aid judgment. This is also a possible explanation for the higher accuracy percentage for sentence-final stimuli. As to the other two sentence environments, the order between them is only tentative. What we know for sure is that listeners did worse with stimuli in sentence-initial and sentence-medial environments than in sentence-final environments and in isolation.

5.4 Summary

Having examined the three major possible explanations for tone perception in whispered speech that were raised by different linguists, we are able to provide a tentative answer to the main research question of what contributes to tone perception in whispered speech in different contextual environments.

Speakers' special maneuvers during whispering provide extra information on tones that listeners do not usually get from normal conversation. In particular, male speakers lengthen the vocalic part that carries the tone to provide this additional information, while female speakers prefer to 'exaggerate' the amplitude contour to do so. However, this finding is based only on the analysis of Tone 3 and Tone 4. This extra information is actually of some help to listeners since the samples that were correctly perceived by the majority of listeners have longer duration if they were produced by male speakers, or have more sloping amplitude contour if they were Tone 3 or Tone 4 produced by female speakers. Therefore, whisperers' special maneuvers observed in this study contribute to tone perception in whispered speech. However, they could only contribute to a certain extent: listeners were unable to have consistent access to full information on tones from these special maneuvers as seen by the relatively low average accuracy percentage for whispered samples (60.1%) compared to over 90% for the phonated samples.

Two major temporal envelope cues were examined in this study: duration and amplitude contour. Duration is found to play a smaller role in tone perception in this study. Although Tone 4 is the shortest tone, it is ranked as the second easiest tone to perceive in whisper. But the contribution of duration to tone perception is indisputable, and the evidence comes from the comparison of perception scores between stimuli in isolation and in sentential environments. Amplitude contour plays an important role especially for Tone 3 and Tone 4. As discussed in the last chapter, the distinguishing amplitude contour of Tone 3 and Tone 4 makes them more salient and easier to recognize than the other two tones. Taking the total number of isolated samples played in the perception test into account, the accuracy percentage for Tone 3 is 327 out of 360 (36 x 10) = 90.8%. This result is comparable to that of phonated stimuli, which means Tone 3 is almost as identifiable in whisper as in normal speech. The amplitude contour is able to substitute for fundamental frequency as the main perception cue to Tone 3 to some extent. The accuracy percentage for Tone 4 is found to be 74.7% (269 out of 360), so amplitude contour is an important cue to Tone 4 perception in whisper but not the only cue. F2, as discussed earlier, could also convey pitch information, but further study would have to take individual vowel quality into account. Unfortunately, little can be said from this study on the contribution of amplitude contour to Tone 1 and Tone 2 perception. Thus, we can say in conclusion that amplitude contour plays an important role in Tone 3 and Tone 4 perception in whispered speech.

As discussed, extended phonetic context does not contribute to tone perception in whisper; rather it is a distraction to listeners. But semantic context is important in whispered speech as in other special circumstances. It is particularly important for Tone 1 and Tone 2, and important to Tone 4 as well, according to the findings of this study.

In summary, three things are proven to be important to tone perception in whispered speech. They are: whisperers' special maneuvers, duration and amplitude contour, and semantic context. Although there might be other things that also help listeners to make correct judgments, such as the tracking of F2 spectral energy, these three elements are considered the most consistent and influential in the literature and gain further support from this study.

Chapter Six

CONCLUSION

I conclude as follows: section 6.1 summarizes this thesis, section 6.2 makes recommendations for future studies based on the limitations and findings of the current work, and section 6.3 presents the contributions of this study.

6.1 Summary

This study examines whispered tones in citation and sentence environments with acoustic analysis and perception tests to investigate what contributes to tone perception in whispered speech. As reported in the literature review chapter, there were a number of studies on whispered speech in the 1950s and 1960s. However, more emphasis was given to how well people perceive whispered tones rather than how to produce them and what helps listeners to perceive. A few acoustic studies have been done more recently on the formant features of vowels in whispered speech but not on the temporal envelope parameters that are more related to tones. Therefore, Mandarin Chinese as the native language of the investigator was chosen as the target language in the present study to examine whispered speech in tone languages. In addition, stimuli in four environments were examined to determine the effects of context on tone perception.

Three experiments were conducted in this study. Laryngoscopic endoscopy was used in the first experiment to observe the pharynx and larynx when Mandarin tones were produced in phonated and whispered speech. The second experiment was for data collection and acoustic analysis. Three sets of syllables with four contrasting tones were used as the stimuli. Two male and two female native speakers of Mandarin produced these target syllables in isolation and in sentential environments, and in both phonated and whispered speech. The data collected were analyzed acoustically with speech-analysis software, and were used as the stimuli for the perception testing in the third experiment. Ten native speakers of Mandarin participated in this experiment. They listened to the whispered stimuli in both isolation and in context and made judgments of the tones they perceived. A second acoustic analysis was carried out based on the result of the perception test. Only the isolated stimuli perceived correctly by the majority listeners were analyzed again to collect further information on how people perceive whispered tones. The summary of the findings is reported in response to the seven research questions listed in the introduction chapter that comprised the main research goal.

The laryngeal sphincter mechanism as the primary physiological characteristic of whispering is observed and described in this study. No vocal fold vibration is found in whisper, and there is a ‘Y’ shaped glottal opening observed in all whispered tones, with the vocal folds parted anteriorly. The aryepiglottic folds press up and forward to draw close to the middle of the airway over the glottis, thereby reducing the epilaryngeal tube area to a small channel which differs among tones. Larynx raising and tongue root retraction occur as concomitant features of sphinctering, as in phonated speech, but in whisper, larynx height is primarily related to shifting of pitch, substituting for the length and vibratory frequency of the vocal folds in phonation. In other words, the larynx rises in rising contours and lowers in falling contours during whisper. This is suspected to influence F2 frequency by changing the shape of the pharyngeal resonating cavity. As to the degree of sphincteric narrowing, it might primarily relate to the noise production in whisper that controls the intensity of tones. Kratochvíl (1968) accounts for the loudness of phonated tones in citation in such a way that loudness increases in rising contours and decreases in falling contours. The tendency is observed from the laryngoscopic images in this study that the sphincteric opening enlarges in the rising contour while reducing in the falling contour, showing that the degree of sphincteric narrowing may relate to the intensity of tones.

Speakers in this study were found to use two special behavioral maneuvers during whispering. Specifically, male speakers used the maneuver of lengthening vocalic duration so that whispered vowels were significantly longer than phonated vowels in all four environments. Female speakers also lengthened the vocalic part of whispered syllables in most cases, but the duration differences between whispered and phonated speech were mostly not statistically significant. Rather, they prefered the special maneuver of increasing the intensity range for Tone 3 and Tone 4 and making a more salient amplitude contour in whisper than in normal speech. Based on the acoustic analysis of Tone 3 and Tone 4 syllables produced by female speakers, the intensity ranges between the maximal and minimal intensity were all significantly greater in whispered speech than in phonation. The reconstructed amplitude contours with trend lines also display the tendency for whisper to be more slanted than for phonation.

However, findings on stimuli produced by male speakers do not reveal this pattern. The second acoustic analysis validates the findings on these two maneuvers with supplementary evidence.

The two behavioral special maneuvers found in this study are proven to contribute to tone perception to a certain extent. The second acoustic analysis found that mean durations of the Majority agreement samples produced by male speakers are longer than all samples, and mean durations of the Unanimous agreement samples are even longer than the Majority ones. This indicates that listeners make use of the extra information provided by male speakers through lengthening to make correct judgments. However, this pattern is not found among Majority and Unanimous samples produced by female speakers. Listeners make use of the greater intensity range of the Majority and Unanimous samples produced by female speakers to make correct judgments for Tone 3 and Tone 4.

Duration in the context-free condition is found to differ significantly from that in contextual conditions. Stimuli produced by both male and female speakers in citation are significantly longer than stimuli in all three sentential environments. Among these three environments, durations of stimuli in sentence-final position are significantly longer than in the other two positions. Durations in the other two environments are about the same, although stimuli in sentence-initial position are a bit longer. No comparison was made for amplitude contour among the sentential environments.

Temporal envelope cues can substitute for F0 to various extents depending on different tones. Generally speaking, duration plays a smaller role in whispered tone identification. Amplitude contour plays a more important role for Tone 3 and Tone 4 but not for the other two tones. Because Tone 3 and Tone 4 have a more salient and active amplitude contour, it is very easy for people to perceive them with only the amplitude contour, especially Tone 3. Even though fundamental frequency is absent in whisper, an accuracy percentage of 90.8% for whispered Tone 3 in isolation is achieved in this study. The falling-rising amplitude contour of Tone 3 can almost totally substitute intensity for fundamental frequency as the major perception cue to Tone 3 in whispered speech. The accuracy percentage of 74.7% for whispered Tone 4 in isolation obtained in this study reveals that amplitude contour can substitute for pitch for Tone 4 in whisper to a great extent, but other perception cues are also needed to get a perception score equal to that of normal speech. Amplitude contours for Tone 1 and Tone 2 are not analyzed in this study, but it is known that amplitude contour does not play an important role for these two tones; otherwise Tone 1 and Tone 2 in isolation would not be ranked as the most difficult tones to perceive.

The carrier sentence in this study is shown to be a distraction to tone perception in whispered Mandarin. Comparing the accuracy percentages for stimuli in four different environments, the perception score for stimuli in isolation is much higher than that for stimuli in the other three environments. This is because the carrier sentence in this study does not include any semantics at all. The main reason, as we have discussed, is that durations of tones in sentential environments are significantly shorter than in isolation, which affects the amplitude contour as well. Given that these two temporal envelope parameters are both perceptual cues to whispered tones, the perception score for stimuli with envelope cues carrying less information is definitely lower than that for stimuli with more information from these two cues. Although the semantic context is controlled and is not the target in the present study, other findings indicate that semantic context plays a very important role for tone perception in whisper, in particular for Tone 1 and Tone 2. Among the other main possible perceptual cues studied, the absence of semantic context is the most likely reason they were identified with a very low accuracy percentage.

Finally, the position of the target tonal segment makes some perceptual difference. Stimuli in sentence-final position were judged as intended at a much higher percentage than stimuli in sentence-medial and -final positions. The perception scores that the subjects got for these two environments were similar. An acoustic explanation is that mean durations of stimuli in sentence-final position are much longer than in sentence-initial and sentence-medial positions. As we have just discussed, more information from the envelope cues is provided to the listeners if the stimuli are in sentence-final positions; thus listeners were able to make more correct judgments.

6.2 Future studies

The present study is the first to analyze and compare the temporal parameters of tones in whispered and normal speech, and the first to compare them among stimuli in four different environments. Furthermore, it is also the first to look in depth at the articulatory production mechanisms for whisper. However, it is important that future studies should look into this issue more comprehensively and include other concerns that this study did not address.

Two special maneuvers were observed in this study for male and female speakers respectively. However, only two subjects for each gender were included in this study, which makes this result less convincing. Future investigators are recommended to recruit more subjects to investigate the observed special maneuvers and gender differences, and the number of the subjects should be large enough to make the result representative of the population. In addition, there might be other special maneuvers that whisperers use during whisper, such as larynx height and degree of sphinteric narrowing that we observed in this study. Since the observation was made on one speaker producing only one possible vowel, we need to investigate these two articulatory parameters further with a larger group of subjects producing complete Mandarin syllables in contrasting vocalic environments. An instrumental approach might be to use a nasendoscopic laryngoscope rather than the oral scope. Moreover, there might be other special behavioral maneuvers given that only two acoustical parameters were examined in this study.

While the amplitude contours of Tone 3 and Tone 4 are a focus of this study, further studies need to consider the amplitude contours of Tone 1 and Tone 2. Besides, mathematical algorithms might be necessary in future studies to compare the degree of sloping to verify the finding that the amplitude contour produced by female speakers in whisper is more slanted than in phonation. Since perceptual cues other than duration and amplitude contour, such as formant information of the vowels, interacting with larynx height, might also contribute to whispered tone perception. These should be analyzed and accounted for in future work. With respect to the physiology of amplitude, verification of the relation between the dynamic effect of airflow and the degree of sphincteric narrowing should be made in future studies as well.

Although stimuli in sentential environments are included in this study, this is done mainly for the purpose of examining context effects on whispered tone perception. Only preliminary acoustic analysis was done on them. It is still a mystery why Tone 4 was not well perceived in all these sentential environments as was Tone 3, since they both have more salient amplitude contours than the other two tones.

Vowel /a/ of Mandarin Chinese was chosen as the target vowel carrying tones in this study. It will be very interesting to study the possible effects that other vowels bring to the perception test. Since Kallail and Emanuel's (1984) experiments on whispered vowel perception found that different vowels are perceived with different accuracy percentages, would certain tones on different vowels be easier or more difficult to perceive than on /a/? These are all interesting questions that could be pursued in future studies. Similarly, findings from a parallel study on tone languages other than Mandarin will be very important for understanding laryngeal control in tone languages as a whole.

Two final very intriguing topics for future studies arise. The first one is inspired by the very informative and insightful data from the Bai language (Esling & Edmondson, 2002). A close examination of the laryngeal features of native Mandarin speakers for both phonation and whisper will make the study of Mandarin tones in whispered speech more complete. The second one, mentioned in the discussion chapter, is that we need to further analyze from the psycho-physical point of view the different strategies that male and female speakers choose during whisper in order to express differences in gender behavior.

6.3 Contributions

The notion of 'special maneuvers' by whisperers was proposed as early as the 1950s (Meyer-Eppler, 1957). Besides a few studies on articulatory mechanism reorganization in the 1970s (Mansell, 1973), no one really knows what these 'special maneuvers' are. This is the first study that brings empirical evidence to this phenomenon. This study finds two specific special maneuvers from the whisperer's behavior perspective: one is lengthening vocalic duration and the other is increasing the amplitude contour. Articulatorily, some possible special maneuvers are observed in this study: the mechanical adjustment of the larynx sphincter that is found in the production of all four whispered tones, the relationship between the sphincter and larynx height, and the control of airflow. However, further investigations are needed to substantiate and specify these articulatory adjustments in whisper since larynx height was not isolated empirically and spectral frequencies in contrasting vocalic environments were not studied.

This study provides a complete description of Mandarin tones in various environments. The findings of the perception test offer a comprehensive report on how whispered Mandarin tones are produced in isolation and how they are perceived in isolation and in three sentential environments.

The findings of the present study may also contribute to research on synthesized speech in speech technology applications. The significant differences between the temporal envelope parameters of whispered and phonated tones found in this study suggest that duration and amplitude contours for phonated tones would not be appropriate to use when synthesizing tones in whisper. Researchers should also take the speaking strategies adopted by different genders into consideration when synthesizing whispered tones.

REFERENCES

Abramson, Arthur S. (1972). Tonal experiments with whispered Thai. In Albert Valdman (Ed). *Papers in linguistics and phonetics to the memory of Pierre Delattre*, 31-44. Paris: The Hague.

Abramson, Arthur S. (1974). The tones of Central Thai: some perceptual experiments. *Haskins Laboratories Status Report on Speech Research*, 39-40, 91-104.

Allen, J., & Haggard, M. (1977). Perception of voicing and place features in whispered speech: A dichotic choice analysis. *Perception & Psychophysics*, 21, 315-322.

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

Buuren, L. V., (1983). Observations on phonation. *Journal of the International Phonetic Association*, 13(1), 13-23.

Carlson, B. F., Esling, J. H. (2001). The shift of uvulars to pharyngeals in Nootka. Paper presented at the SSILA conference, July 2001, Santa Barbara.

Catford, J. C. (1977). *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

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

Crystal, David. (1991). *A Dictionary of Linguistics and Phonetics*. Blackwell Publishers.

Cruttenden, A. (1986). *Intonation*. New York: Cambridge University Press.

Cruttenden, A. (1997). *Intonation* (2nd edition). New York: Cambridge University Press.

Dannenbring, G. L. (1980). Perceptual discrimination of whispered phoneme pairs. *Perceptual & Motor Skills*, 51, 979-985.

Dongilli, Paul A. (1994). Semantic context and speech intelligibility. In James A. Till, Kathryn M. Yorkston, & David R. Beukelman (eds.), *Motor Speech Disorders: Advances in Assessment and Treatment*. Baltimore: Paul H. Brookes Publishing Co.

Dow, Francis D. M. (1972). *An Outline of Mandarin Phonetics*. Canberra: Australian National University Press.

Edmondson, J. A., Esling, J. H., Li Shaoni, Harris, J. G. & Lama Ziwo. (2000). The aryepiglottic folds and voice quality in Yi and Bai languages: Laryngoscopic case studies. *Minzu Yuwen*, 6, 47-53. (in Chinese).

Eklund, I., & Traunmüller, H. (1996). A comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Quarterly Progress and Status Report – Royal Institute of Technology*, 2, 131-134.

Endres, W., Bambach, W. & Flosser, G. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America*, 49(6), 1842-1848.

Esling, J. H. (1984). Laryngographic study of phonation type and laryngeal configuration. Journal of the International Phonetic Association, 14, 56-73.

Esling, J. H. (1994a). Some perspectives on accent: Range of voice quality variation, the periphery, and focusing. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New views, new directions* (pp. 49-63). Alexandria, VA: TESOL.

Esling, J. H. (1994b). Voice quality. In R.E. Asher & J.M.Y. Simpson (Eds.), *The Encyclopedia of language and linguistics* (pp. 4950-4953). Oxford: Pergamon Press.

Esling, J. H. (1995). Pharyngeal phonetics: Larynx height, tongue root, and pitch dependence. *Chicago Linguistic Society, 31*, 143-152.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association, 26*, 65-88.

Esling, J. H. (1998). Laryngoscopic analysis of pharyngeal articulations and larynx-height voice quality settings. *Proceedings of the 5th International Conference on Spoken Language Processing*, vol. 2 (pp. 543-546). Sydney: ASSTA.

Esling, J. H. (1999a). The IPA categories "pharyngeal" and "epiglottal": Laryngoscopic observations of pharyngeal articulations and larynx height. *Language & Speech*, 42*,* 349-372.

Esling, J. H. (1999b). *University of Victoria Phonetic Database* (version 4.0). Victoria:Speech Technology Research Ltd.; and Lincoln Park, NJ: Kay Elemetrics Corp.

Esling, J. H. (1999c). *Voice quality settings of the pharynx. Proceedings of the 14th International Congress of Phonetic Sciences,* vol. 3 (pp. 2449-2452). Berkeley: University of California.

Esling, J. H. (2000). Crosslinguistic aspects of voice quality. In R.D. Kent & M.J. Ball (Eds.), *Voice quality measurement* (pp. 25-35). San Diego: Singular Publishing Group.

Esling, J. H. (2002). The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai. Paper presented at the LSA annual meeting, January 2002, San Francisco.

Esling, J. H., Ball, M. J. & Dickson, B.C. (2000). The transcription of voice quality. In R.D. Kent & M.J. Ball (Eds.), *Voice quality measurement* (pp. 49-58). San Diego: Singular Publishing Group.

Esling, J. H., Carlson, B. F. & Harris, J. G. (2002). A laryngoscopic phonetic study of Nootka and Salish glottal stop, glottalized resonants, and pharyngeals. Paper presented at the SSILA conference, January 2002, San Francisco.

Esling, J. H., Clayards, J. A. W., Edmondson, J. A., Qiu, F-Y. & Harris, J. G. (1998). Quantification of pharyngeal articulations using aryepiglottic angle measurements from laryngoscopic images. *Proceedings of the 5th International Conference on Spoken Language Processing*, vol. 7 (pp. 3091-3094). Sydney: ASSTA.

Esling, J. H., & Clayards, J. A. W. (1999). Laryngoscopic analysis of pharyngeal articulations and larynx-height voice quality settings. In A. Braun (Ed.), *Advances in Phonetics (Zeitschrift für Dialektologie und Linguistik, Beiheft 106)* (pp. 22-33). Stuttgart: Franz Steiner Verlag.

Esling, J. H., Dickson, B. C. & Snell, R. C. (1994). Real-time processing of electroglottographic waveforms for the evaluation of phonation types. *ICSLP 94* (pp. 647-650). Yokohama: Acoustical Society of Japan.

Esling, J. H., & Edmondson, Jerold A. (2002). The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai. In Angelika Braun & Herbert R. Masthoff (eds.), Phonetics and its applications: Festschrift for Jens-Peter Köster on the occasion of his 60th birthday. Stuttgart: Franz Steiner Verlag.

Esling, J. H., Heap, L. M., Snell, R. C. & Dickson, B. C. (1994). Analysis of pitch dependence of pharyngeal, faucal, and larynx-height voice quality settings. *ICSLP 94* (pp. 1475-1478). Yokohama: Acoustical Society of Japan.

Fletcher, H. B. (1995). The speaking mechanism. In ASA (ed.), Speech and Hearing in Communication. New York: Acoustical Society of America.

Fromkin, Victoria A. (1978). *Tone: A Linguistic Survey*. San Francisco: Academic Press.

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

Fu, Q., Zeng, F., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America*, 104(1), 505-510.

Gårding, Eva, Kratochvíl, Paul, Svantesson, Jan-Olof, & Zhang, Jialu. (1986). Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech*, 29(3), 281-293.

Giet, Franz. (1956). Kann man in einer Tonsprache flüstern? *Lingua*, 5, 372-81.

Harbold, Beorge J. (1958). Pitch ratings of voiced and whispered vowels. *Journal of the Acoustical Society of America*, 30(7), 600-601.

Helmholtz, H. L. F. von. (1954). *On the sensations of tone*. New York: Dover Publications.

Higashikawa M. (1999). Acoustical-perceptual correlates of 'whisper pitch' in synthetically generated vowels. *Journal of Speech, Language & Hearing Research*, 42(3), 583-596.

Higashikawa, M., Nakai, K., Sakakura, A., & Takahashi, H. (1996). Perceived pitch of whispered vowels-relationship with formant frequencies: A preliminary study. *Journal of Voice*, 10(2), 155-158.

Ho, Aichen T. (1971). *Mandarin tones in selected sentence environments: an acoustic study*. Ph.D. dissertation, University of Wisconsin.

Ho, Aichen T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33, 353-367.

Hoit, J. & Hixon, T. (1987). Age and speech breathing. *Journal of Speech and Hearing Research*, 32, 351-366.

Howie, John Marshall. (1976). *Acoustical Studies of Mandarin Vowels and Tones.* Cambridge: Cambridge University Press.

IPA. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Jensen, M. K. (1985). Recognition of word tones in whispered speech. *Word*, 14, 187-196.

Jovičić, S. T. (1998). Formant feature differences between whispered and voice sustained vowels. *Acustica*, 84(4), 739-743.

Kallail, K.J. (1981). *An Acoustic Comparison of Isolated Whispered and Phonated Vowel Productions*. Ph.D. dissertation, University of Oklahoma.

Kallail, K.J., & Emanuel, F. W. (1984a). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *Journal of Speech and Hearing Research*, 27, 245-251.

Kallail, K.J., & Emanuel, F. W. (1984b). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, 12, 175-186.

Kallail, K.J., & Emanuel, F. W. (1985). The identifiability of isolated whispered and phonated vowel samples. *Journal of Phonetics*, 13, 11-17.

Kent, Ray D., & Read, Charles. (1992). *The acoustic analysis of speech*. San Diego: Singular Publishing Group.

Klich, R. J. (1982). Effects of speech level and vowel context on intraoral air pressure in vocal and whispered speech. *Folia Phoniatrica*, 34(1), 33-40.
Lashley, C., Hicks, D. M. (1984). Vibratory action of the vocal folds during whisper. *The IASCP Bulletin*, 2, 32-35.

Laver, John. (1980). *Phonetic Description of Voice Quality*. New York : Cambridge University Press

Laver, John. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lin, Hua. (1994). Mandarin stress revisited. In Jose Camacho & Lina Chouerri (Eds.), *Proceedings of the Sixth North American Conference on Chinese Linguistics.* Los Angeles: University of Southern California Press.

Lin, Hua. (1996a). *Mandarin Tonology*. Taipei: Pyramid Press.

Lin, Hua. (1996b). Phonetics or phonology: Mandarin neutral tone revisited. In Tsaifa Cheng, Hongming Zhang & Yafei Li (Eds.), *Proceedings of the Fourth International Conference on Chinese Linguistics*, 2, 219-233. Los Angeles: GSIL Publications, University of Southern California Press.

Lin, Hua. (1998). Diaosu lun ji Putonghua liandu biandiao [Theory of "Tonal Elements" and Tone Sandhi in Putonghua]. *Zhongguo Yuwen [Chinese Linguistics]*, 1998(1), 31-39.

Lin, Hua. (1999). Putonghua qingsheng diaozhi de zonghe fenxi [A unified analysis of the neutral tone values in Mandarin]. In Lu, Shinan, *Xiandai Yuyinxue Lunwenji [Papers on Modern Phonetics and Phonology]* (pp. 175-183). Beijing: Golden City Press.

Lin, Hua. (2001a). *A Grammar of Mandarin Chinese*. Munich: Lincom Europa.

Lin, Hua. (2001b). Stress and the distribution of the neutral tone in Mandarin. In De Bao Xu (Ed.), *Chinese Phonology in Generative Grammar* (pp. 139-161). San Diego: Academic Press.

Lin, H-B., & Repp, Bruno H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, 32(1), 25-44.

Lin, T., & Wang, L. J. (1991). *Yuyinxue Jiaocheng* [Course on Phonetics]. Beijing: Beijing University Press.

Mansell, P. (1973). An experimental investigation of articulatory reorganization in whispered speech. *Working Papers, Institut für Phonetik und sprachliche Kommunikation der Universität München Forschungsberichte*, 2, 201-253.

McGlone, R. E., & Manning, W. H. (1979). Role of the second formant in pitch perception of whispered and voiced vowels. *Folia Phoniatrica*, 31(1), 9-14.

Meyer-Eppler, W. (1957). Realization of prosodic features in whispered speech. *Journal of the Acoustical Society of America*, 29, 104-106.

Miller, J. D. (1961). Word tone recognition in Vietnamese whispered speech. *Word*, 17, 11-15.

Miller, C. B. (1934). An Experimental-Phonetic Investigation of Whispered conversation, Considered from the Linguistic Point of View. Ph.D. dissertation, University of Hamburg.

Moore, Crinne B. (1993). Phonetic observation on stress and tones in Mandarin Chinese. *Working Papers of the Cornell Phonetics Laboratory,* 8, 87-117.

Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, 4, 10-29.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America*, 24, 175-184.

Peterson, G. E. & Lehiste, Ilse. (1960). Duration of Syllable Nuclei in English. *Journal of the Acoustical Society of America*, 32(6), 693-703.

Pike, Kenneth L. (1948). *Tone Languages*. Ann Arbor: University of Michigan Press.

Preminger, Jill E., Van Tasell, Dianne J. (1995). Quantifying the relation between speech quality and speech intelligibility. *Journal of Speech and Hearing Research*, 38(3), 714-726.

Remez, Robert E., Rubin, Philip E. (1990). On the perception of speech from time-varying acoustic information: contributions of amplitude variation. *Perception and Psychophysics*, 48(4), 313-325.

Ross, Elliott D., Edmondson, J.A. & Seibert, G. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis. *Journal of Phonetics*, 14, 283-302.

Schwartz, M. F. (1971). Air consumption values for oral and whispered plosive-vowel syllables. *Journal of the Acoustical Society of America*, 49(2), 610.

Schwartz, M. F., & Rine, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustical Society of America*, 44(6), 1736-1737.

Shen, Xiaonan S. (1989). *The Prosody of Mandarin Chinese*. Los Angeles: University of California Press.

Shen, Xiaonan S. (1990). On Mandarin Tone 4. *Australian Journal of Linguistics*, 10(1), 41-59.

Shigeno, S., & Hiroya, F. (1980). Context effects in phonetic and non-phonetic vowel judgments. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, 14, 217-224.

Shih, Chi-lin. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory,* 3, 83-109.

Smith, W. F. (1973). *A formant study of whispered vowels*. Ph.D. dissertation, University of Oklahoma Health Sciences Center.

Smitheran, J. & Hixon, T. (1981). A clinical method for estimating laryngeal airway resistance during vowel production. *Journal of Speech and Hearing Disorders,* 46, 138-146.

Solomon, N. P., McCall, G. N., Trosset, M. W. & Gray, W. (1989). Laryngeal configuration and constriction during two types of whispering. *Journal of Speech and Hearing Research*, 32, 161-174.

Stathopoulos, Elain T., Hoint, Jeannette D., Hixon, Thomas J., Watson, Peter J., & Solomon, Nancy Pearl. (1991). Respiratory and laryngeal function during whispering. *Journal of Speech and Hearing Research*, 34, 761-767.

Schwartz, M. (1971). Air consumption values for oral whispered plosive-vowel syllables. *Journal of the Acoustical Society of America*, 49, 610.

Tartter, V. C. (1989). What's in whisper? *Journal of the Acoustical Society of America*, 86, 1678-1683.

Tartter, V. C. (1991). Identifiability of vowels and speakers from whispered syllables. *Perception & Psychophysics*, 49, 365-372.

Thomas, I. B. (1969). Perceived pitch of whispered vowels. *Journal of the Acoustical Society of America*, 46, 468-470.

Thomas, I.B. (1969). Perceived pitch of whispered vowels. *Journal of the Acoustical Society of America*, 46(2), 468-470.

Traunmüller, Hartmut, & Eriksson, Anders. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America*, 107(6), 3438-3451.

Tseng, Chiu-yu. (1990). *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Taipei: Institute of History & Philology Academia Sinica.
Tsunoda, K., Niimi, S., Hirose, H., Harris, K. S., & Baer, T. (1991). An electromyographic study on laryngeal adjustment for whispering. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, 25, 33-38.

Vance, Timothy J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, 33(5), 368-392.

Venkatagiri, Horabail S. (1981). Reaction time for voiced and whispered /a/ in stutterers and nonstutterers. *Journal of Fluency Disorders*, 6(3), 265-271.

Vilkman, E., Aaltonen, O., Raimo, I., Ignatius, J., & Koi, P. V. (1987). On stress production in whispered Finnish. *Journal of Phonetics*, 15, 157-168.

Warren, Richard M., & Bashford, James A. (1978). Production of white tone from white noise and voiced speech from whisper. *The Bulletin of the Psychonomic Society*, 11(5), 327-329.

Watterson, Thomas, & Emanuel, Floyd. (1981). Effects of oral-nasal coupling on whispered vowel spectra. *The Cleft Palate Journal*, 18(1), 24-38.

Weitzman, R. S., Sawashima, M. & Hirose, H. (1976). Devoiced and whispered vowels in Japanese. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, 10, 61-79.

Wise, C. M., & Chong, Lily Pao-Hu. (1957). Intelligibility of whispering in a tone language. *Journal of Speech and Hearing Disorders*, 22(3), 335-338.

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

Xu, Yi. (1995). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America,* 95(4), 2240-2253.

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

Yip, Michael C. W. (2000). Spoken word recognition of Chinese homophones: The role of tone neighbors. *Psychologia*, 43(2), 135-143.

APPENDIX I

Transcript for the Recording session

WORD LIST

❶	bā	bá	bǎ	bà
	八	拔	靶	爸
❷	fā	fá	fǎ	fà
	发	罚	法	发
❸	mā	má	mǎ	mà
	妈	麻	马	骂
❹	bī	bí	bǐ	bì
	逼	鼻	比	必

DIALOGUE

_____ zì nǐ zhīdào zěnme xiě ma?
A:_____ 字你 知道 怎么 写 吗?

bā	bá	bǎ	bà
1. 八	2.拔	3.靶	4.爸
fā	fá	fǎ	fà
5.发	6.罚	7.法	8.发
mā	má	mǎ	mà
9.妈	10.麻	11.马	12.骂

Wǒ bù zhīdào zěnme xiě_____.
B:我 不 知道 怎么 写 _____.

bā	bá	bǎ	bà
1. 八	2.拔	3.靶	4.爸
fā	fá	fǎ	fà
5.发	6.罚	7.法	8.发
mā	má	mǎ	mà
9.妈	10.麻	11.马	12.骂

Nǐ zhīdào _______ zì yīngyǔ zěnme shuō ma?
A:你 知道 _______ 字 英语 怎么 说 吗?

bā	bá	bǎ	bà
1. 八	2.拔	3.靶	4.爸
fā	fá	fǎ	fà
5.发	6.罚	7.法	8.发
mā	má	mǎ	mà
9.妈	10.麻	11.马	12.骂

Bù zhīdào.
B:不 知道 .

APPENDIX II

Instruction for the perception session

Date: __________________

Age: ___________________

Education: ______________

Native Dialect: ________________

Foreign Language(s): _______________

How long have you been in North America: __________________

Please give yourself a code (e.g. GM) : __________________

How long have you been staying in Beijing? (If applicable): _______________

Section 1: Please listen to the word carefully and choose one of the five choices on the answer sheet. You are going to hear a total of 144 whispered words.

A: Tone 1 (一声)

B: Tone 2 (二声)

C: Tone 3 (三声)

D: Tone 4 (四声)

E: Do not know. (不清楚)

Section 2: Please listen to the sentence carefully and choose one of the five choices on the answer sheet.

A: Tone 1 (一声)

B: Tone 2 (二声)

C: Tone 3 (三声)

D: Tone 4 (四声)

E: Do not know. (不清楚)

◆**For the first 96 sentences, you are going to hear a sentence within the frame:**

_____ zì **n**ǐ zhīdào zěnme xiě ma? **(___ 字你知道怎么写吗?)**

Please identify the tone of the first word and choose on the answer sheet.

◆**For the second 96 sentences, you are going to hear a sentence within the frame:**

Wǒ bù zhīdào zěnme xiě_____. **(我不知道怎么写 ___.)**

Please identify the tone of the last word and choose on the answer sheet.

◆**For the second last 96 sentences, you are going to hear a sentence within the frame:**

Nǐ zhīdào _____ zì yīngyǔ zěnme shuō ma? **(你知道____ 字英语怎么说吗?)**

Please identify the tone of the word in the middle of the sentence and choose on the answer sheet.

VITA

Surname: Gao Given Name: Man

Place of Birth: Wuhan, Hubei Province, P. R. China

Educational Institutions Attended:

University of Victoria	2000 to 2002
Zhongshan University	1995 to 1999

Degrees Awarded:

B.A.	Zhongshan University	1999

Honors and Awards:

University of Victoria Dean's Scholarship	2000 to 2002
Faculty of Graduate Studies Teaching and Research Fellowship	2000 to 2002
The Henry and Michiko Warkentyne Graduate Scholarship	2001 to 2002

Publications:

2002 An OT Analysis of Chinese Abbreviatory Compounds. In (eds.), Working Papers of the Linguistic Circle of the University of Victoria 16 (in press)

2002 Chinese Ba-structure in HPSG. In (eds.), Working Papers of the Linguistic Circle of the University of Victoria 16 (in press)

Title of thesis:

Mandarin Tones in Whisper: An Investigation of the Perceptual Cues

Author ___________________

Man Gao

August 15, 2002