

**The Effects of Speaking Rate on Mandarin Tones**

BY

C2008

Yun H. Stockton

Submitted to the graduate degree program in Linguistics  
and the Graduate Faculty of the University of Kansas  
in partial fulfillment of the requirements for the degree of  
Master's of Arts.

\_\_\_\_\_  
Chairperson

Committee members \_\_\_\_\_  
\_\_\_\_\_

Date defended: \_\_\_\_\_

The Thesis Committee for Yun H. Stockton certifies  
that this is the approved Version of the following thesis:

The Effects of Speaking Rate on Mandarin Tones

Committee:

\_\_\_\_\_  
Chairperson

\_\_\_\_\_

\_\_\_\_\_

Date approved: \_\_\_\_\_

## ABSTRACT

This study examined the acoustic changes of speaking rate on Mandarin tones. 8 native Mandarin speakers produced 17 Chinese words with four tones at slow, normal and fast speaking rate. Acoustic parameters, such as consonant and vowel duration, F0 at onset and offset of vowels and the first, second and third formant frequencies were examined. The  $\Delta F0$  and the Turning Point of Tones 2 and 3 were also examined. Results showed significant changes in duration in both consonant and vowels as a function of speaking rate. The findings showed that the vowel duration of a shorter tone, such as tone 4 increases relatively more than a longer tone, such as tone 3 at a slower speaking rate. Results also found that the values of F0 at vowel onset and offset varied as a function of speaking rate. Formant frequency values showed little change when all the stimuli were examined. When the front and back vowels were analyzed separately, F2 values did show significant variations when speaking rate changed. F2 values of front vowels decreased and F2 values of back vowels increased, with an increase in speaking rate. In terms of  $\Delta F0$  and TP of Tones 2 and 3, the results showed changes across speaking rate, with a decrease in  $\Delta F0$  and an earlier TP as speaking rate increases. These effects were most prominent for Tone 3. These findings are interpreted in terms of how tones are implemented under different speaking rates in production.

## ACKNOWLEDGEMENTS

I would first like to thank my advisors Joan Sereno, Allard Jongman, and Jie Zhang for their support and dedication to my education and the completion of my thesis. From the choosing of the thesis topic to the end of the defense Joan has been a source of advice, encouragement and inspiration to me. Thank you, Joan for being a great mentor. I also would thank Allard for his being patient and great courses in phonetics which led me to the study of phonetics. Jie is also generous with his time and resource. My degree completion could not have happened without his commitment and encouragement. The entire committee, Joan, Allard and Jie, has been amazing.

The consistent caring and support of the entire Linguistics faculty and staff has been overwhelming. I'm really grateful to them for always being so helpful.

I thank my husband, Curtis, who always believes in me and encourages me. He also reminds me that there is more to life than phonetics.

## TABLE OF CONTENTS

Chapter 1		
1.1	Introduction	1
1.2	Speaking rate effects on segments and suprasegmentals	2
1.2.1	Segments	2
1.2.2	Suprasegmentals	5
Chapter 2		
2.1	Mandarin tones	15
2.2	Research Hypotheses	17
Chapter 3		
3.1	Mandarin tone production test	20
3.2	Method	20
3.2.1	Subjects	20
3.2.2	Stimuli	21
3.2.3	Procedures	23
3.2.4	Acoustic analysis	24
Chapter 4		
4.1	Results	25
4.1.1	Consonant duration	25
4.1.2	Vowel duration	26
4.1.3	Onset fundamental frequency values	29
4.1.4	Offset fundamental frequency values	30
4.1.5	F1, F2, F3 values	32
4.1.6	$\Delta F_0$ values	35
4.1.7	Turning point values	37
Chapter 5		
5.1	Discussion	40
REFERENCES		46

## Chapter 1

### 1.1 Introduction

Studies have shown that speaking rate is source of variability that affects the acoustic properties of speech, besides many others, such as phonetic context, syllable position and the speakers themselves (Miller, 1981). So far, research on the effects of speaking rate has focused mainly on segments. However, if speaking rate effects exist, they may also be observed in suprasegmentals, such as tones. It is not until recently that the question of whether speaking rate affects tones has been investigated. Evidence of this influence of speaking rate on tones has been found mainly in perception studies. Therefore, the present study extends the speaking rate research from segments to suprasegmentals by examining the effects of speaking rate on the production of lexical tones in Mandarin Chinese.

When speakers speak at different speaking rates, slow or fast, the spectral and temporal parameters of tones used by listeners as acoustic cues to distinguish tones may be modified. By looking at the speaking rate effects on the production of Mandarin tones, we can get to know the specific consequences of changes that different speaking rates bring. This, in turn, provides an idea of the kinds of perceptual compensations that listeners may have to make. This study investigates what, if any, changes in temporal and pitch properties are made when speaking rate is changed from slow to normal to fast for Mandarin tones.

The following literature review section provides an overview of speaking rate effects on the production and perception of segments and suprasegmentals. Hypotheses of this study on Mandarin tones will be established on the basis of these discussions.

## **1.2 Speaking rate effects on segments and suprasegmentals**

### **1.2.1 Segments**

Early research on speaking rate shows that the consequences of changes in speaking rate are changes in spectral and temporal properties of speech (for a review, see Miller, 1981). Lindblom (1963) studied the effect of changes in speaking rate on eight Swedish vowels /I, e, Y, æ, a, θ, ɔ, ʊ / in three CVC contexts: /b-b/, /d-d/, and /g-g/. Speaking rate was manipulated by having the speaker repeat isolated test syllables at a slow, fast and normal speaking rate. A male talker with typical Stockholm Swedish read the speech material. Lindblom measured the duration of each syllable, the F1, F2, and F3 values at the onset and offset of each syllable and at the point of each syllable at which they reached their maxima or minima. The results of this experiment showed an increasing tendency for the obtained formant frequency values to undershoot their targets as the duration of the syllable decreased (i.e., as speaking rate increased). The vowel became reduced, or more schwa-like in character.

Another study by Gay (1968) supported Lindblom's findings. The stimuli in his experiment consisted of instances of five diphthongs /ɔɪ/, /aɪ/, /aʊ/, /eɪ/, /oʊ/ in various consonantal contexts spoken by five male adults at three different rates of speech (slow, medium and fast). The word list contained fifty different CVC monosyllables with each diphthong represented ten times. The three rate conditions were dictated by the speakers' own estimation of natural productions. Ample practice time preceded each condition. Gay measured the glide duration and the onset and offset formant frequencies. The results showed that for all conditions only [ɔɪ] and [aɪ] contained relatively prominent steady states at the onset. The offset steady states were much shorter than the onset steady states. Both the onset and the offset steady states were least prominent in [eɪ] and [oʊ]. Glide duration was the longest for [eɪ] and [oʊ] and the shortest for [ɔɪ] and [aɪ]. The decrease in the glide duration was also accompanied by a decrease in F2 offset frequency. F2 offset differences between the slow and the fast speaking rate ranged from a high of 281 Hz to a low of 131 Hz. There was little consistent change in the onset frequency position.

As Gay (1968) stated, the change in rate did not affect the onset frequencies but did affect the offset frequencies. The total duration of the diphthong decreased as speaking rate increased, resulting in the termination of the formant structures before reaching the offset target. Speakers did not attain the target values at offset at the faster rate of speech. Both Lindblom's and Gay's findings suggested that the main

consequence of the increased speaking rate was simply to terminate the formant structures before the target vowel was reached.

In terms of the temporal changes at different speaking rates, Port (1976) found in his dissertation that, for some vowels, a change in speaking rate would alter not only the absolute duration of vowels but also their relative duration. For example, an increase in speaking rate reduced the duration of long /i/ relatively more than short /I/. This indicated that the absolute difference between the two vowels was reduced at a faster rate of speech. In the second experiment, Port chose four English words (*deeber, dibber and deeper, dipper*) to examine the influence of speaking rate on the duration of English vowels and stops. These sounds were each embedded in the carrier sentence *Pat tried to say \_\_\_\_\_ to his pop*. Coached individually to read the list at the slow and fast speaking rate, five male speakers of native NYC English recorded the list in a quiet room. Port (1976) found that the absolute duration of the sentences at the slow and fast speaking rate was significantly different. The mean sentence duration at the slow speaking rate was 1730 milliseconds and at the fast speaking rate 1250 milliseconds. The ratio of fast to slow was 0.72. This was also true in the change of the relative duration of /I/ and /i/. At the slow speaking rate, /I/ was 35 % shorter than /i/, a contrast that was reduced significantly to 28% at the fast speaking rate.

### 1.2.2 Suprasegmentals

Early studies on Thai tones by Abramson (1962) found that the onsets of high and falling tones were higher than low and rising tones, and the mid tone was intermediate. Fundamental frequency (F0) contours of mid and low tones fell steadily throughout, whereas those associated with high falling and rising tones changed abruptly in slope approximately halfway through their duration.

Gandour, Tumtavitikul and Satthamnuwong (1999) conducted a comprehensive study of speaking rate effects on Thai tones. In this study, Gandour et al. (1999) investigated the effects of two different speaking rates (slow and fast) on the production of five Thai lexical tones (high, mid, low, falling, rising).

The aim of the Gandour et al. (1999) study was to look for the effects of different speaking rates on the production of Thai lexical tones, specifically, the changes in the height and the slope of Thai tonal contours (high = H, mid = M, low = L, falling = F, rising = R). Each of the six Thai male speakers produced five monosyllabic target words embedded in a carrier sentence with a fixed syntactic and prosodic environment. Each utterance consisted of three syllables: *Say* \_\_\_\_, followed by two-syllable reduplicated words carrying each of the five tones. In each utterance the first and the third syllables were stressed. Height and slope characteristics of the five tone contours were compared at 11 measurement locations (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100%) in the tone contours between the fast and the slow speaking rates. Stressed (3<sup>rd</sup> syllable) and unstressed syllables (2<sup>nd</sup> syllable) were analyzed separately.

The mean difference in F0 at the fast speaking rate between the stressed and the unstressed syllables of the mid, low, falling, high and rising tones were 19, 18, 10, 9, 13 Hz, respectively. At the slow speaking rate the mean difference in the F0 of these five tones were 7, 6, 10, 1, and 2 Hz. The results of the experiments demonstrated that speaking rate affected F0 (a spectral parameter) contours of unstressed syllables both in terms of height and slope. The graphs in Figure 1 illustrated this. In the graphs, the solid lines represent the fast speaking rate. The dotted lines represent the slow speaking rate. The second syllable is where the unstressed syllables are. The third syllable is where the stressed syllables are. The height of the F0 contour was generally higher at the fast speaking rate compared with the slow speaking rate in unstressed position. In the stressed syllable positions (the 3<sup>rd</sup> syllable position), the only significant changes in height occurred in the initial epochs of Falling (0-20) and Rising (0-10), and in the final epochs (80-100) of High and Rising (see Figure 1). In most of the unstressed syllables, differences were also evident in the slope of F0 contours. The tones with substantial F0 movement, like the falling tones, high tones and rising tones, exhibited overall flatter slopes at the fast speaking rate. Those tones with lesser F0 movement, such as mid tones and low tones, showed steeper slopes at the fast speaking rate compared to the slow speaking rate. To explain the findings that only the unstressed syllables have differences in tone height and contour, Gandour et al. (1999) used Manuel's (1990) hypothesis. Namely, languages generally tend to tolerate less contextually induced changes in acoustic phonetic output if they are likely to lead to confusion of contrastive phones.

Gandour et al. (1999) further stated that the changes in the height of F0 contours might be primarily due to the preservative effects of tonal coarticulation rather than to stress itself.

Gandour et al. found that the effects of speaking rate on Thai tones were mainly on the F0 contours and the duration of the unstressed syllables. The present research will specifically examine stressed syllables and extend their investigation of speaking rate effects on tones to a different tonal language, Mandarin, by observing the changes of speaking rate affecting duration and F0 values of Mandarin monosyllabic words.

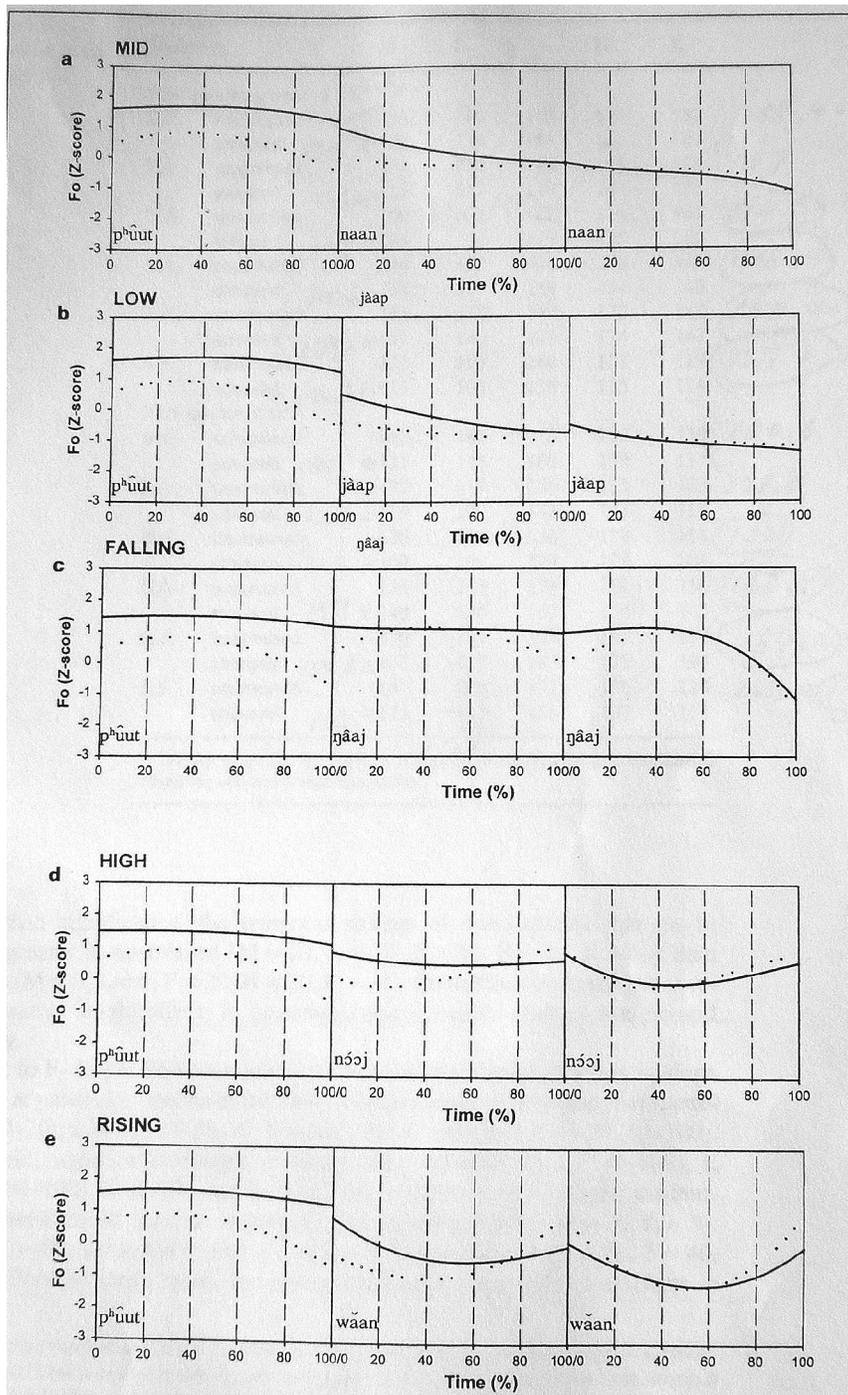


Figure 1. Average time-normalized, fitted third-order polynomial F0 curves of 6 male speakers for each of the five Thai tones (Mid, Low, Falling, High, Rising) by speaking rate. Solid lines indicate fast speaking rate. Dotted lines indicate slow speaking rate (Gandour et al., 1999).

In order to have an overall view of the behavior of Thai tones at different speaking rates, Gandour (2002) also tested the Thai emphatic tone. An emphatic tone is a tone used in emphatic reduplicatives to signal an intensification of meaning. It is an independent suprasegmental morpheme. Its segmental content is copied from the following adjective or adverb. Its meaning normally conveys the superlative degree of the adjective or adverb. The F0 contour of an emphatic tone on reduplicated forms produced in isolation is high-rising-falling (Gandour, 2002). In this companion study to his 1999 study on Thai tones, Gandour used similar stimuli (he used the same words in the same carrier sentence), similar subjects (he used the same number of speakers who were in the same age group), and similar methodology. The most important feature that distinguished the emphatic tone between slow and fast speaking rate was the overall flattening of its F0 slope at the fast speaking rate, especially in the unstressed portions. This was similar to the behavior of the regular falling, high and rising tones in his previous (1999) study.

What was not mentioned in either of the studies on Thai contour tones was the change in some temporal properties, such as the vowel duration and the Turning Point. The timing of Turning Point (TP) is defined as the duration from the onset of the tone to the point of the change in F0 direction. In Mandarin, duration is one of the important cues in recognizing the four tones. Temporal information to distinguish tone properties has been used to examine tone 2 and tone 3 because their F0 contours are similar. The overall tone duration and TP of tone 2 is shorter than of tone 3 (Blicher et al., 1990; Moore & Jongman, 1997, Jongman et al., 2006). In their study,

Blicher, Diehl and Cohen (1990) tested forty Mandarin and forty English speaking subjects with long (450 ms) and short (350 ms) Tone 2-3 continua. The results indicated that the longer syllable duration shifted the category boundary toward Tone 3 responses for both Mandarin and English speakers. Syllable lengthening seemed to enhance the F0 cues for Tone 3 category as the lengthening serves to reinforce some distinctive acoustic or auditory property of Tone 3. Blicher et al. (1989) concluded that the temporal property that was relevant in the determination of Tone 2 and 3 was the duration of the initial, non-rising part of the tone, the section from onset to TP.

In their 1997 study, Moore and Jongman systematically investigated when  $\Delta F0$  (the difference in fundamental frequency between onset and Turning Point) and TP covary, whether perception of tone 2 and 3 based on each of these parameters was equally categorical and which combinations of TP and  $\Delta F0$  triggered the shifts in identification from one tone to the other. In this study, they designed the perception experiments with stimuli varying in both  $\Delta F0$  and TP, varying in  $\Delta F0$  only or varying in TP only. The stimuli in the test varying in both  $\Delta F0$  and TP were synthesized [u] syllables that formed a continuum from tone 2 to tone 3. The  $\Delta F0$  ranged from 25 to 75 Hz, in steps of 5 Hz and the TP varied from 20 to 220 milliseconds in steps of 20 ms. The continuum varying in only  $\Delta F0$  varied in 11 steps from 5 Hz to 163 Hz with TP fixed at 120 milliseconds. The continuum varying only in TP varied from 20 to 220 milliseconds, with a constant  $\Delta F0$  of 50 Hz. Mandarin listeners were asked to identify whether they heard the Chinese word “none”, corresponding to Tone 2, or the word “dance”, corresponding to Tone 3.

These experiments used a forced-choice labeling paradigm in which subjects heard each [u] stimulus and were asked to choose from these two lexical items, “none” or “dance”. Two natural precursor phrases (“This word is \_\_\_”) were chosen from their previous production test, one was pronounced with a high-pitched voice, the other with a low-pitched voice. The two phrases differed by 39 Hz in average F0, with the high precursor spanned 192-272 Hz, as compared to 170-229 Hz for the low precursor.

The experiment examined perception of both F0 and temporal properties and only F0 of Tone 2 and 3 in high-F0 and low-F0 precursor phrases showed significant identification shifts. Thus identical stimuli were identified as tone 3 for the high precursor (high F0) condition, but as tone 2 for the low precursor (low F0) condition. These results supported their hypothesis that tone identification was influenced by changes in F0 range of the precursor. Stimuli varying in TP, however, showed no significant identification shift. These experiments suggested that normalization was triggered only when both stimuli and precursors vary along the same acoustic dimension (F0).

In pilot studies on various temporal dimensions (speaking rate), Moore (1995) found that lexical tones contrasting in temporal properties were perceived relative to speaking rate, while tones contrasting in spectral information were not affected by rate. Thus, listeners normalized for rate by shifting category boundaries for the temporal cue (turning point). Jongman, Moore and Lai (2006) further investigated the effects of speaking rate on the perception of Mandarin tones. In a series of

experiments, Jongman et al. tested twelve native Mandarin speakers on the perception of Mandarin Tone 2 (mid-rising) and Tone 3 (low-falling-rising) at different speaking rates. The aim of their study was to address the role of temporal information in the normalization of tone by manipulating temporal properties (speaking rate) rather than spectral properties of the context. They created three synthetic continua between tone 2 and 3 varying in  $\Delta F0$  only, or in the position of TP only, or in both  $\Delta F0$  and TP. They hypothesized that lexical tones, which contrast in temporal properties, will be affected by speaking rate, while tone stimuli which do not contrast in temporal information, such as  $\Delta F0$ , would not be affected by rate. Therefore, the above two continua varying in TP were predicted to show speaking rate effects, while the continuum varying only in  $\Delta F0$  was predicted not to show any effect. As Tone 2 was associated with an earlier TP than Tone 3, Jongman et al. (2006) further predicted that Tone 3 might be identified as Tone 2 at the fast speaking rate; whereas Tone 2 might be identified as Tone 3 at the slow speaking rate. Six male and six female native Mandarin Chinese participated in the perception experiments. A precursor (*Zheige zi nian* \_\_\_) with the target word [wu] (the word with a Tone 2 meaning ‘none’; with a Tone 3 meaning ‘dance’) from a Tone 2-Tone 3 continuum was played to the subjects in the context of a fast and a slow speaking rate. The subjects identified the stimuli they heard as having either a Tone 2 or a Tone 3. The results of the first test, in which Tone 2 and 3 continuums varied in both  $\Delta F0$  and TP indicated that seven out of twelve subjects had an early shift towards Tone 3 response as predicted by the hypothesis. However, this shift was not

statistically significant. The results of the second experiment, in which Tone 2 and 3 continua varied only in  $\Delta F_0$ , did support the hypothesis, which had predicted no effect of change of speaking rate in the continuum varying in  $\Delta F_0$  only. And the listeners did not show an overall earlier shift towards Tone 3 responses in the fast precursor condition while the continuum was manipulated in  $\Delta F_0$  only. The last experiment, in which Tone 2 and 3 continua varied only in TP, supported the hypothesis, exhibiting an earlier shift toward Tone 3 responses in the fast precursor condition. The findings in these experiments confirmed the hypothesis in Moore's 1995 and 1997 research, which concluded that normalization occurred when both stimuli and precursors varied in the same acoustic dimension. Thus,  $F_0$  was perceived with respect to speaker and TP was perceived with respect to speaking rate.

A study that shows little effect of speaking rate on  $F_0$  contours of Mandarin tone was conducted by Yi Xu (1998). Xu examined the consistency of tone syllable alignment across different speaking rates. His research question relating to speaking rate was whether the  $F_0$  contour associated with the rising tones, Tone 2 and 3, stretched over the entire voiced portion of the syllable as the syllable duration increased (at the slow speaking rate). Xu tested four Mandarin speakers (two males, two females) with twenty Mandarin words at a slow, normal and fast speaking rate. These words were embedded in a carrier sentence "*I say the word \_\_\_*", which was presented in Chinese on a computer. Speakers were asked to produce the sentence at a slow, normal and fast speaking rate. The signal was digitized at 22 KHz in real

time using a Sound Edit program and then stored in a computer. In his experiment Xu examined the change of F0 peak with tones in context. The results illustrated that, regardless of speaking rate, the F0 peak associated with the rising tone always occurred near the offset of the rising-tone-carrying syllable. The F0 onset of the rising tone always occurred near the center of the syllable and the slope of the rising tone did not vary systematically with speaking rate. Instead of concentrating on the effects of the context surrounding tones as Xu did, the present experiment will examine the speaking rate effect on the F0 values of all four Mandarin tones at both onset and offset of the syllable.

These experiments suggest that changes in speaking rate can affect tone perception. The present research extends these studies by carefully examining the specific acoustic changes in Mandarin tones that occur with a change in speaking rate. Given the previous research on the effects of different speaking rates on Mandarin tones, it was expected to find some evidence of a change in TP in Tones 2 and 3 in Chinese monosyllabic words at different speaking rates in production. The present research also aims at examining the changes in  $\Delta F0$ , to see whether there is any modification of  $\Delta F0$  in production when speaking rate is manipulated. In other words, the present research is a systematic acoustic analysis of the changes in Mandarin tones at different speaking rates.

## Chapter 2

### 2.1. Mandarin tones

In order to understand more about Mandarin tones, this section presents previous studies that define the acoustic characteristics of Mandarin tones.

Mandarin has four phonemic tones. When produced in isolation, their contours seem well defined and pretty stable (Xu, 1997; Jongman et al. 2006). In his 1997 study, Xu tested eight male native speakers of Beijing Mandarin. Four monosyllabic words (ma-1, ma-2, ma-3, ma-4) were produced in isolation. According to the results of this study Tone 1 starts with a high F0 value (near 130 Hz) and stays around that level throughout the syllable. Tone 2 starts with a low F0 (near 110 Hz), then falls slightly before rising (starting at 20 % into the vowel) throughout the remainder of the syllable. Tone 3 starts with an F0 value slightly lower than the onset of Tone 2, falls to the lowest F0 of all the four tones (about 90 Hz) at the vowel midpoint, and then rises sharply to the end of the syllable. Tone 4 starts with the highest F0 values of the four tones (140 Hz), continues to rise before reaching the maximum about 20% of the way into the vowel, and then falls sharply to the end of the syllable. Examples of the four tones for one speaker are shown in Figure 2.

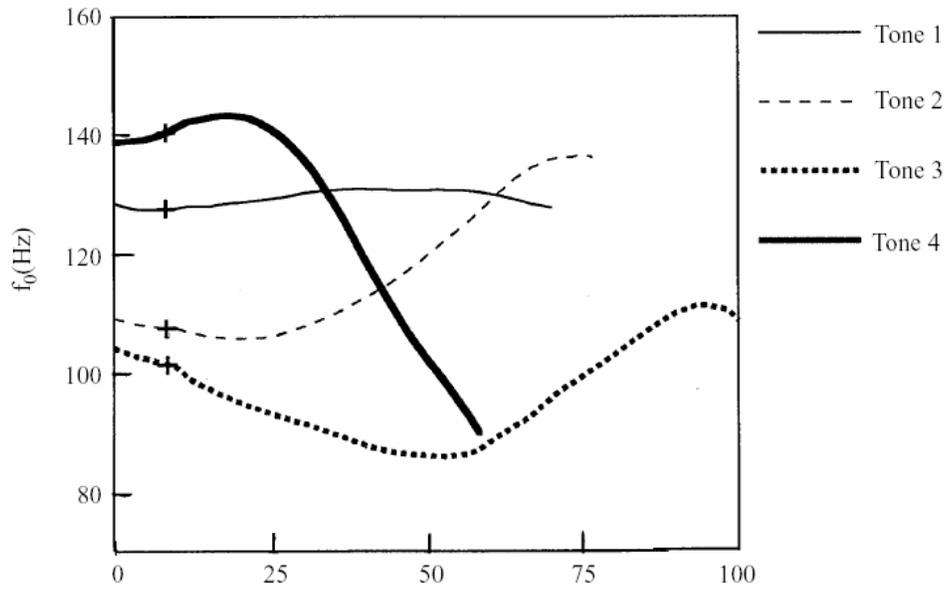


Figure 2. The four Mandarin tones, tone 1, tone 2, tone 3, tone 4 (from Xu, 1997)

Previous studies indicated that F0 was the primary cue to the perception of Mandarin tones (Howie, 1976). While F0 values represent average differences over tones, Moore and Jongman (1997) found that speakers' tones vary as a function of F0 range. In their production experiment they tested 3 males and 4 female native Chinese Mandarin speakers. They were asked to read a minimal set of the four Mandarin tones with the segmental contexts *wu*, *yi*, *bi* and *ma*. These syllables were randomized and produced in the carrier phrase *Zhe ge zi nian* \_\_\_\_ (This word is \_\_\_\_). They found that the high-pitched speaker's low region of Tone 3 could overlap with the low-pitched speaker's high region of Tone 2.

Beside the importance of the pitch properties of tones, production data have shown that tones have different temporal properties. It has been found that Mandarin tones show consistent duration differences. Tone 4 is the shortest tone (214 ms on

average). Tone 3 is the longest (349 ms), while tones 1 and 2 have intermediate durations, with Tone 2 (273 ms) longer than Tone 1 (247 ms) (Xu, 1997).

Temporal information to distinguish tone properties has also been used to examine tone 2 and tone 3 because their F0 contours are similar. Research discussed above by Blicher et al., (1990) and Moore and Jongman (1997) have shown that duration is important in identification of tones, and the TP had been demonstrated to effectively distinguish Tone 2 and 3. Shorter TP is identified as Tone 2, while longer TP is identified as Tone 3.

## **2.2 Research Hypotheses**

Previous studies on speaking rate effects on Mandarin tones have shown that the parameters that cue the production and perception of Mandarin tones include F0 and duration. At different speaking rates, Mandarin tones may display different phonetic characteristics, which, in turn, influence the perception of tones. However, a systematic study of the acoustic properties, such as the overall consonant and vowel duration, values of F0 at onset and offset, TP and  $\Delta F0$ , as well as formant frequency values of Mandarin tones produced at slow and fast speaking rates has not been conducted.

A systematic study of the effects of speaking rate on different Mandarin tones may answer the following specific questions. What are the overall F0 values and durations of each of the four tones at different speaking rates? Will they change significantly at slow and fast speaking rates? Are there any changes in the onset and

offset of F0 values of the four tones at the slow and fast speaking rate and are these changes consistent over tones? Are the Turning Point and  $\Delta F0$  values of Tone 2 and 3 at the slow and fast speaking rate different from the ones at the normal speaking rate? In order to investigate these questions, a production test is designed to test the following hypotheses:

- 1). It is hypothesized that, similar to the previous studies on segments, the overall consonant and vowel duration of the four Mandarin tones will change accordingly at the slow and fast speaking rates. The overall consonant and vowel duration of all four tones will increase at the slow speaking rate and decrease at the fast speaking rate.
- 2). It is also hypothesized that the vowel duration of the longer tones, such as Tone 3, may increase relatively more than the duration of the shorter tones, such as Tone 4, at the slow speaking rate.
- 3). Based on previous studies on the speaking rate affecting F0 values at onset and offset, it is hypothesized that speaking rate will affect F0 values at both onset and offset in Mandarin monosyllabic words. F0 values may be higher at the fast speaking rate compared to slower speaking rates.
- 4). In terms of formant frequencies it is hypothesized that formant frequency values of F1, F2 and F3 may change as a function of speaking rate. At the fast speaking rate, vowels may be reduced with an effect observed most clearly in F2 values.
- 5). It is hypothesized that different speaking rates may affect the  $\Delta F0$  values of Tone 2 and 3 such that the  $\Delta F0$  values of Tone 3 may be smaller at faster speaking rates.

6). It is also hypothesized that different speaking rates may affect the TP values of Tone 2 and 3 such that the TP values of Tone 3 may occur earlier at faster speaking rates.

## **Chapter 3**

### **3.1 Mandarin Tone Production Test**

This experiment was designed to provide spectral and temporal acoustic parameters of speakers' productions of Chinese words with four tones at different speaking rates (fast, normal and slow). The durational information of both consonants and vowels was examined. The F0 values at the onset and offset of each contour and the values of F1, F2, F3 were collected.  $\Delta F0$  and the timing of TP of tone 2 and 3 were also measured.

### **3.2 Method**

#### **3.2.1 Subjects**

Eight native speakers of Mandarin, four males and four females, aged between 21 and 32 years, participated as subjects. All subjects were undergraduate or graduate students at the University of Kansas. All spoke standard Mandarin. Four additional speakers were also recorded but were not used in the following analysis due to their use of regional dialects. The eight participants were all competent English speakers as well. The years of formal English instruction they had received before coming to the States varied from 5 years to 24 years. The number of years of their residence in the US varied from 6 months to 9 years.

All speakers reported to use at least 30 % Chinese in their daily lives. None of them reported any speech disorders. Subjects were not paid for their participation.

### **3. 2. 2 Stimuli**

Table 1 lists all the target words used in the experiment. These were seventeen syllables produced with four Mandarin tones, for a total of 68 words. Each of the eight subjects produced these 68 stimuli at a fast, normal and slow speaking rate.

Tone 1	Tone 2	Tone 3	Tone 4
bī [bi] force 逼	bí [bi] nose 鼻	bǐ [bi] pen 笔	bì [bi] must 必
bā [ba] eight 八	bá [ba] pull 拔	bǎ [ba] target 靶	bà [ba] dad 爸
fā [fa] send out 发	fá [fa] punish 罚	fǎ [fa] law 法	fà [fa] hair 发
fū [fu] husband 夫	fú [fu] clothes 服	fǔ [fu] house 府	fù [fu] father 父
dī [di] low 低	dí [di] enemy 敌	dǐ [di] bottom 底	dì [di] brother 弟
dá [da] lift 达	dá [da] answer 答	dǎ [da] hit 打	dà [da] big 大
dū [du] all 都	dú [du] poison 毒	dǔ [du] gamble 赌	dù [du] stomach 肚
zuō [zuo] suck 嘬	zuó [zuo] yesterday 昨	zuǒ [zuo] left 左	zuò [zuo] do 做
zhī [z i] branch 支	zhí [z i] straight 直	zhǐ [z i] stop 止	zhì [z i] rule 制
zhū [zu] pearl 珠	zhú [zu] chase 逐	zhǔ [zu] cook 煮	zhù [zu] live 住
shī [ʃi] teacher 师	shí [ʃi] ten 十	shǐ [ʃi] history 史	shì [ʃi] is 是
shē [ʃe] luxurious 奢	shé [ʃe] tongue 舌	shě [ʃe] give up 舍	shè [ʃe] shoot 射
shū [ʃu] book 书	shú [ʃu] ripe 熟	shù [ʃu] count 数	shù [ʃu] tree 树
gē [gɤ] brother 哥	gé [gɤ] leather 革	gě [gɤ] a surname 葛	gè [gɤ] each 个
guō [guo] pot 锅	guó [guo] country 国	guǒ [guo] fruit 果	guò [guo] pass 过
jī [tɕi] chicken 鸡	jí [tɕi] reach 及	jǐ [tɕi] squeeze 挤	jì [tɕi] remember 记
jū [tɕy] reside 居	jú [tɕy] tangerine 桔	jǔ [tɕy] lift 举	jù [tɕy] huge 巨

Table 1. 17 sets of Mandarin stimuli used in the Experiment. Minimal pairs with 4 tones (Tone1, Tone 2, Tone 3, Tone 4)

### **3. 2. 3 Procedures**

These 68 target words were presented in Chinese characters to the speakers on a Dell computer using PowerPoint. In this way, speaking rate could be manipulated. At the fast speaking rate, these words were separated by intervals of 600 milliseconds between stimuli from offset to onset; at the normal speaking rate the intervals were 1200 milliseconds and at the slow speaking rate they were 2400 milliseconds. The subjects were recorded individually in an Anechoic chamber. The subjects were seated in front of the computer monitor with a microphone approximately 5 inches away from the subjects' lips. An Electrovoice RE-20 microphone and a Marantz PMD 671 solid state recorder were used.

At the start of the session, the subjects were first given instructions on how to read these words on the screen, and then they were given twelve practice trials before starting the test trials at each speed. The subjects read aloud these 68 trials from the computer screen at the slow, normal and then the fast speaking rate. Each trial was shown in Chinese character form, which was about the size of 1 inch by 1 inch on the computer screen. The trials were presented in random order and the order of presentation was different for each subject. With eight speakers producing 68 trials at three different speaking rates (fast, normal and slow) there were all together 1632 (8 speakers x 68 trials x 3 speeds) tokens.

### **3. 2. 4 Acoustic Analysis**

The data were then analyzed with Praat. Dependent variables included consonant duration, vowel duration, onset and offset fundamental frequency values (F0 onset, F0 offset), first formant frequency (F1), second formant frequency (F2), third formant frequency (F3) values for all 4 tones as well as turning point (TP) and fundamental frequency ( $\Delta F0$ ) changes from tone onset to the lowest point of Tones 2 and 3.

Consonant and vowel duration was determined with waveform and spectrogram analysis. Consonant duration was measured from the onset of the consonant to the onset of the vowel as determined by the start of F1. Vowel duration was measured from the onset of the vowel, the start of F1, to the offset of the vowel, the end of voicing. The F0 onset and offset of the vowels were measured at vowel onset and offset, respectively. Formant values for F1, F2, and F3 were measured at the middle points of vowels. For diphthongs, the measurements of F1, F2 and F3 were done only for the first steady-state vowel.

TP, defined from the pitch track as the duration between the onset of the tone and the point at which the tone changed F0 direction from falling to rising, was determined by examining the pitch contours.  $\Delta F0$  was calculated as the F0 value at vowel onset minus the F0 value at TP.

## Chapter 4

### 4.1 Results

Two different analyses were undertaken for the duration data. The first analysis investigates the mean of the consonant and vowel duration at the fast, normal and slow speaking rates and tests whether these consonant and vowel durations undergo changes in duration as a function of rate. The findings will address the validity of the first hypothesis that overall duration of all tones increases at the slow speaking rate. The second analysis investigates the relative difference in duration across all four tones as a function of speaking rate and tests whether different tones undergo similar amounts of duration change. The findings of this analysis will address the second hypothesis, which stated that at the slow speaking rate the duration of longer tones, such as Tone 3, might increase relatively more than the shorter tones, such as Tone 4.

#### 4.1.1 Consonant Duration

A two-way ANOVA (Rate x Tone) was conducted for consonant duration. There was a main effect for Rate [ $F(2, 1144) = 112.358, p < 0.001$ ], with consonant duration varying as a function of speaking rate. Within-subjects contrast effects revealed that consonant duration at fast speaking rate (70 ms in duration) was significantly shorter than at normal speaking rate (89 ms in duration) which was significantly shorter than at slow speaking rate (104 ms in duration). Mean consonant durations across the four tones and the three speaking rates are shown in Table 2.

There was no main effect for Tone [ $F(3, 572) = .167, p = .92$ ] with consonant durations similar across tones, nor was there a significant interaction between Tone and Rate [ $F(6, 1144) = .326, p = .92$ ].

The consonant duration Ratios (see Table 2) show the significant duration change of all four tones across the three speaking rates, though there is not much difference among the four tones. For the four tones the ratio of slow-to-normal showed an increase of 22% (ratio S/N 1.22) for Tone 1, 15% (ratio S/N 1.15) for Tone 2, 16% (ratio S/N 1.16) for Tone 3 and 14% (ratio S/N 1.14) for Tone 4.

Rate	C-duration (ms)			
	(n=8) Mean (SD)			
	Tone 1	Tone 2	Tone 3	Tone 4
Fast	69 (62)	70 (61)	74 (66)	66 (61)
Normal	85 (83)	91 (81)	91 (87)	88 (83)
Slow	104 (108)	105 (99)	106 (110)	100 (103)
Ratio F-N	0.81	0.77	0.81	0.75
Ratio S-N	1.22	1.15	1.16	1.14

Table 2. Mean duration of consonants and standard deviations (SD) in parentheses of consonants of all stimuli (Tone 1, Tone 2, Tone 3, Tone 4) under fast, normal and slow speaking rate. Ratios of fast to normal and slow to normal speaking rate are also given for each tone.

#### 4. 1. 2 Vowel Duration

A two-way ANOVA (Rate x Tone) was also conducted for vowel duration.

There was a main effect for Rate [ $F(2, 1144) = 2912.023, p < .001$ ], with vowel

duration varying as a function of speaking rate. Within-subjects contrast effects revealed that vowel duration at fast speaking rate (332 ms in duration) was significantly shorter than at normal speaking rate (501 ms in duration) which, in turn, was significantly shorter than at slow speaking rate (787 ms in duration).

For vowel duration there was also a main effect of Tone [ $F(3, 572) = 88.619$ ,  $p < .001$ ]. Bonferroni posthoc tests revealed that tone 4 (449 ms) was the shortest in vowel duration and was significantly shorter than vowel duration for Tone 1 (542 ms in duration) and tone 2 (535 ms in duration) which, in turn, were significantly shorter than vowel durations for Tone 3 (633 ms). Mean vowel durations for all four tones across all three speaking rates are shown in Table 3. While consonant duration showed no difference across tones, vowel duration did.

There was also a significant interaction between Rate and Tone for vowel duration [ $F(6, 1144) = 6.296$ ,  $p < .001$ ] as shown in Figure 3. At the fast speaking rate Tone 3 had the longest duration, followed by Tone 2 and Tone 1. Tone 4 had the shortest duration. In general, for the longest tone (i.e., tone 3), slowing speech down increased vowel durations less than for the shortest tone (i.e., Tone 4), where slowing speech down increased vowel duration more.

This observation can be verified by examining vowel duration ratios (See Table 3). For tone 3, the ratio of slow to normal showed about a 40 % (ratio S/N 1.4) increase while the shortest tone (tone 4) showed the greatest increase of 74 % (ratio S/N 1.74).

Rate	V- duration (ms)			
	(n=8) Mean (SD)			
	Tone 1	Tone 2	Tone 3	Tone 4
Fast	318 (84)	334 (90)	412 (129)	260 (79)
Normal	489 (106)	497 (109)	618 (130)	397 (96)
Slow	816 (149)	774 (165)	867 (154)	688 (180)
Ratio (F-N)	0.65	0.67	0.67	0.65
Ratio (S-N)	1.67	1.56	1.4	1.74

Table 3. Mean duration of vowels and standard deviations (SD) in parentheses of all stimuli (Tone 1, Tone 2, Tone 3, Tone 4) under fast, normal and slow speaking rates. Ratios of fast to normal and slow to normal speaking rate are also given for each tone.

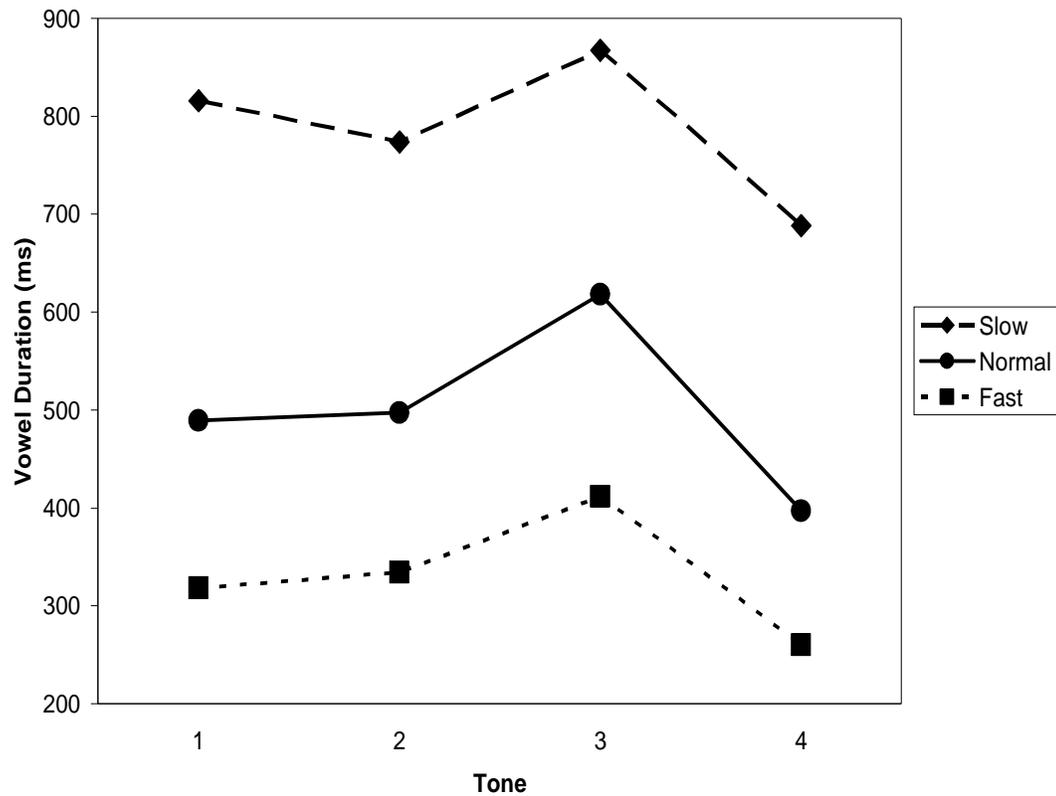


Figure 3. Vowel duration (ms) of Mandarin Tone 1 to Tone 4 at fast, normal and slow speaking rates.

#### 4. 1. 3 Onset Fundamental Frequency Values

A two-way ANOVA (Rate X Tone) was conducted for fundamental frequency (F0) values at the onset of tones. There was a main effect for Rate [ $F(2, 1144) = 30.87, p < .001$ ], with fundamental frequency varying as a function of speaking rate. Within-subjects contrast effects revealed that F0 onset values at fast speaking rate (212 Hz) were significantly higher than at both normal (206 Hz) and slow speaking rate (205 Hz). There was a pitch increase at tone onset at the fastest speaking rate.

As expected, there was also a main effect of fundamental frequency onset values for Tone [ $F(3, 572) = 133.052, p < .001$ ]. Within-subjects contrast effects revealed that, across speaking rates, F0 onset values for tone 4 (266 Hz) were the highest and were significantly different from Tone 1 (228 Hz) which, in turn, were significantly higher than Tone 2 (171 Hz) and Tone 3 (166 Hz). They were not significantly different from each other.

There was also a significant interaction between Rate and Tone for fundamental frequency onset values [ $F(6, 1144) = 6.73, p < .001$ ]. As shown in Figure 4, while fundamental frequency was the highest at the fast speaking rate as compared to normal and slow speaking rates, this pattern is most evident in tones with higher onset fundamental frequency values such as tones 1 and 4. For Tones 2 and 3, there was less difference in onset fundamental frequency across all three speaking rates.

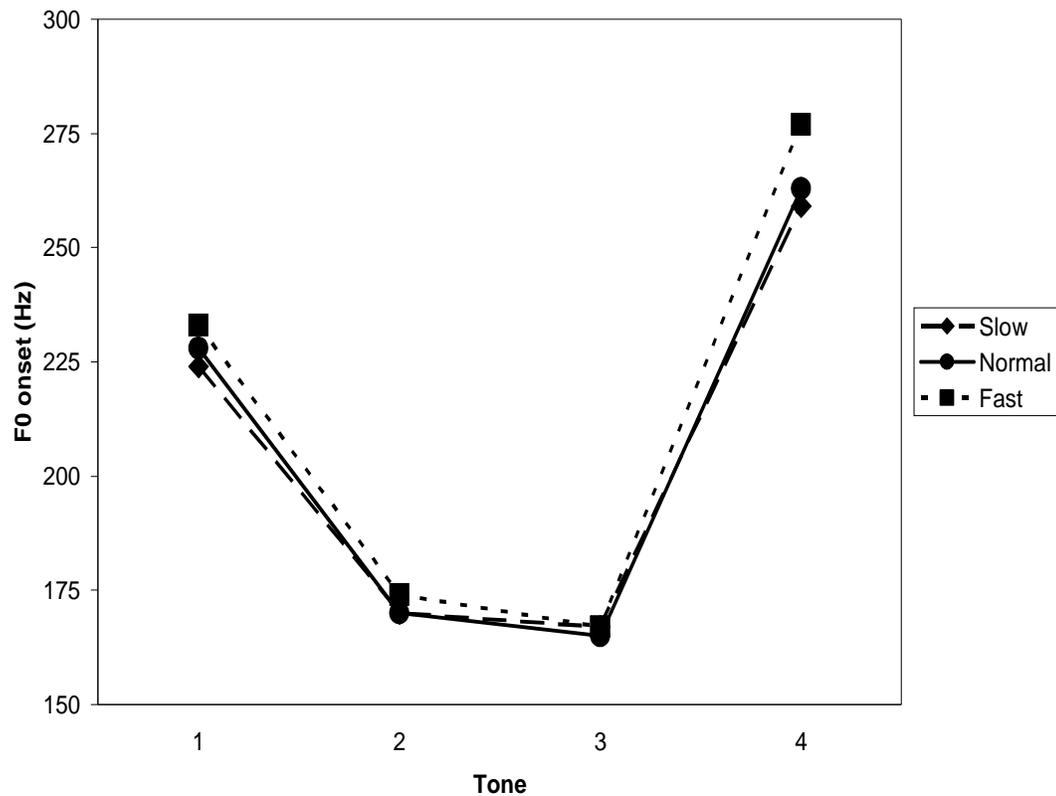


Figure 4. Fundamental frequency values at onset of Mandarin Tone 1 to Tone 4 at fast, normal and slow speaking rates.

#### 4. 1. 4 Offset Fundamental Frequency Values

To check the fundamental frequency (F0) values at the offset of tones, a two-way ANOVA (Rate X Tone) was conducted. There was a main effect for Rate [ $F(2, 1144) = 15.254, p < .001$ ], with fundamental frequency varying as a function of speaking rate. Similar to onset F0 values, within-subjects contrast effects revealed that F0 offset values at fast speaking rate (190 Hz) were significantly higher than they were at normal speaking rate (184 Hz), which, in turn, were significantly higher than they were at slow speaking rate (181 Hz).

As expected, there was also a main effect of fundamental frequency offset values for tones [ $F(3, 572) = 165.51, p < .001$ ]. A Bonferroni posthoc test revealed that, across speaking rates, F0 offset values for tone 2 (227 Hz) and tone 1 (218 Hz) were the highest and were significantly higher than tone 3 (177 Hz) which in turn was significantly higher than tone 4 (117 Hz).

There was also a significant interaction between Rate and Tone for fundamental frequency offset values [ $F(6, 1144) = 9.788, p < .001$ ]. As shown in Figure 5, while fundamental frequency was the highest at fast speaking rate as compared to normal and slow speaking rates, this pattern doesn't hold for Tone 2 and 3. For Tone 3, there was little difference in offset fundamental frequency across speaking rates. For Tone 2 there was a higher offset fundamental frequency at the slowest speaking rate.

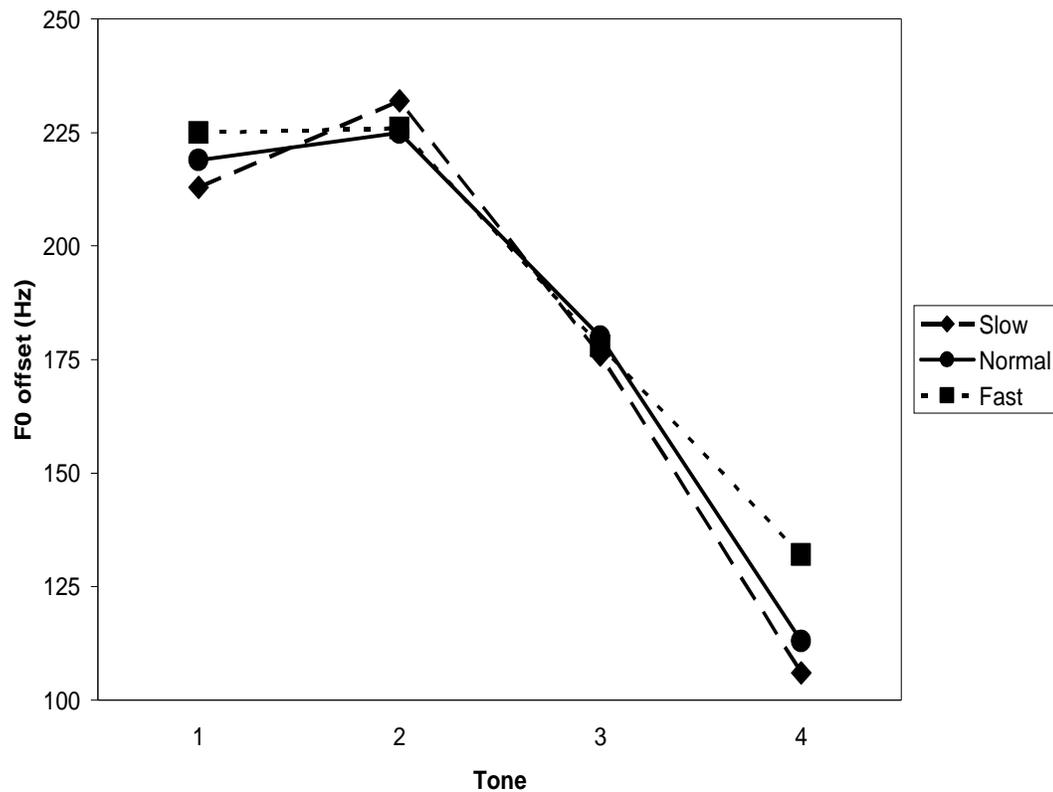


Figure 5. Fundamental frequency values at offset of Mandarin Tone 1 to Tone 4 at fast, normal and slow speaking rates.

#### 4. 1. 5 F1, F2, F3 Values

To evaluate the magnitude of changes in F1, F2, and F3 as a function of speaking rate, a two-way ANOVA (Rate X Tone) was used. Unexpectedly, none of the changes in F1, F2 and F3 for all four tones under different speaking rates were significant except for F1. There was a main effect for Rate [ $F(2, 1144) = 5.990, p < .003$ ], with F1 formant frequency values varying as a function of speaking rate. Within-subjects contrast effects revealed that F1 formant values at the fast speaking rate (508 Hz) were significantly higher than those at both normal (496 Hz) and slow (492 Hz) speaking rates.

There was no main effect for Rate of F2 [ $(2, 1144) = 2.386, p = 0.092$ ] and F3 [ $(2, 1144) = 1.289, p = 0.276$ ]. There was also no main effect for Tone, F1 [F (3, 572) = 0.354,  $p = 0.786$ ]; F2 [F (3, 572) = 0.297,  $p = 0.827$ ]; F3 [F (3, 572) = 0.540,  $p = 0.655$ ], nor was there a significant interaction of Rate and Tone for F1 [F (6, 1144) = 1.747,  $p = 0.107$ ] or F2 [F (6, 1144) = 0.180,  $p = 0.982$ ]. However, the interaction for F3 was significant, F3 [F (6, 1144) = 2.334,  $p = 0.030$ ].

Neutralization of vowel quality changes F2 differently depending on whether the vowel is a front or back vowel. If vowels are reduced, the F2 of front vowels may be lowered while the F2 of back vowels may be raised. Therefore, additional analyses were conducted, in which front and back vowels were examined separately. All diphthongs were excluded.

In analyzing the front and the back vowels, a two-way ANOVA (Rate X Vowel Quality) was conducted to determine if front and back vowels were differently affected by rate. No diphthongs were included. There was a main effect of Rate [F (2, 946) = 4.261,  $p < .04$ ], with F2 varying as function of speaking rate. Within-subjects contrast effects revealed that F2 values at the fast speaking rate (1525 Hz) and at the normal speaking rate (1529 Hz) were higher than at the slow speaking rate (1502 Hz). There was also a main effect of Vowel Quality [F (1, 473) = 514.796,  $p < .0001$ ]. A Bonferroni Posthoc test revealed that F2 values of front vowels (1994 Hz) were significantly higher than the F2 values of back vowels (1042 Hz). Between Rate and Vowel Quality there was also an interaction for F2 values [F (2, 946) = 5.394,  $p < .005$ ]. As shown in Figure 6, as speaking rate increased, the F2

of back vowels raised while the F2 of front vowels lowered. This is most clear comparing fast and normal speaking rates.

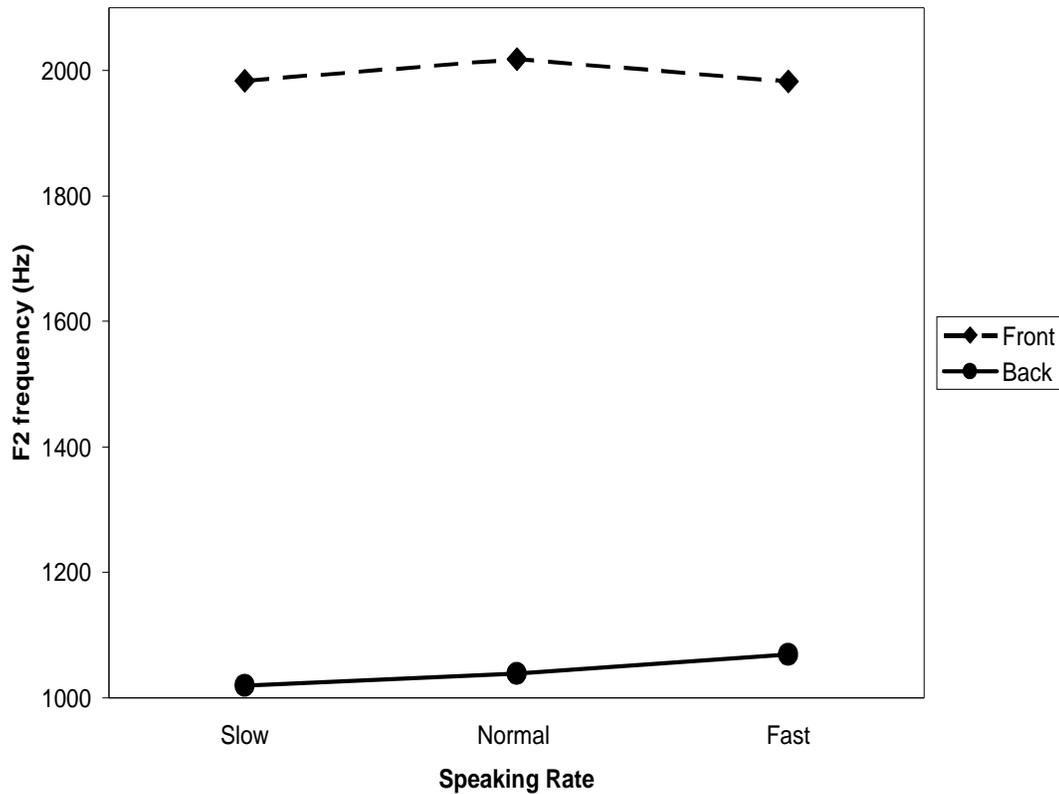


Figure 6. F2 values of the front vowels and the back vowels of across the 4 Mandarin tones at fast, normal and slow speaking rates.

Additional analyses were conducted for the front and the back vowels separately, examining the 4 tones. For front vowels, a two-way ANOVA (Rate X Tone) was used. A strong trend for Rate was found [ $F(2, 502) = 2.674, p < .07$ ]. Overall, both the fast and the slow speaking rate seemed to result in lowered F2 values (see Table 4). For the back vowels, there was a significant main effect for Rate [ $F(2, 432) = 13.559, p < .0001$ ]. Within-subject contrast effects revealed that

F2 values increased as the speaking rate increased for the back vowels (see Table 4), indicating that vowels were neutralized while speaking rate increased. This was observed for all 4 tones.

Rate	F2 Front values (Hz)				F2 Back values (Hz)			
	(n=8)				(n=8)			
	Tone 1 (Hz)	Tone 2 (Hz)	Tone 3 (Hz)	Tone 4 (Hz)	Tone 1 (Hz)	Tone 2 (Hz)	Tone 3 (Hz)	Tone 4 (Hz)
Fast	1992	1967	1950	2017	1070	1081	1031	1092
Normal	2041	2008	1967	2057	1053	1034	1020	1050
Slow	1980	2004	1924	2026	1022	1029	1009	1019

Table 4. F2 values of the front vowels and the back vowels of Mandarin Tone 1, Tone 2, Tone 3, and Tone4 at fast, normal and slow speaking rate.

#### 4. 1. 6 $\Delta F0$ Values

Two additional analyses were conducted specifically for the two contour tones of Tones 2 and 3. Recall that  $\Delta F0$  is the change in F0 from onset to the point where the contour begins to rise. With respect to  $\Delta F0$  values, a two-way ANOVA (Rate x Tone) was conducted. There was a main effect for Rate [ $F(2, 572) = 17.329, p < 0.001$ ], with  $\Delta F0$  varying as function of speaking rate. Within-subjects contrast effects revealed that  $\Delta F0$  values at a fast speaking rate (31 Hz) were significantly less than the  $\Delta F0$  value at both normal (39 Hz) and slow speaking rate (38 Hz). As expected, there was also a main effect of  $\Delta F0$  for Tone [ $F(1, 286) = 185.025, p < 0.001$ ]. A Bonferroni posthoc test revealed that  $\Delta F0$  values of Tone 2 (15 Hz) were

significantly less than  $\Delta F0$  values of Tone 3 (57 Hz). Between Rate and Tone there was a significant interaction for  $\Delta F0$  values [ $F(2, 572) = 8.085, p < 0.001$ ]. As shown in Figure 7, there was little change in  $\Delta F0$  for Tone 2 across speaking rates while there was a much greater change for Tone 3 from the fastest to the slowest speaking rate.

This observation can be verified by examining  $\Delta F0$  ratios (See Table 5). For Tone 2, the ratio of fast-to-slow showed about a 10% (ratio F/S 0.89) increase while for Tone 3 the ratio showed a larger increase over 20% (ratio F/S 0.79). The table also showed little change in  $\Delta F0$  for both Tones 2 and 3 from normal to slow speaking rate.

Rate	$\Delta F0$ values (Hz)	
	(n = 8)	
	Tone 2 (Hz)	Tone 3 (Hz)
Fast	13	49
Normal	17	63
Slow	15	62
Ratio (F-N)	0.81	0.78
Ratio (S-N)	0.91	0.97
Ratio (F-S)	0.89	0.79

Table 5.  $\Delta F0$  values of Mandarin Tones 2 and 3 at fast, normal and slow speaking rates. Ratios of fast to normal and slow to normal speaking rate are also given for each tone.

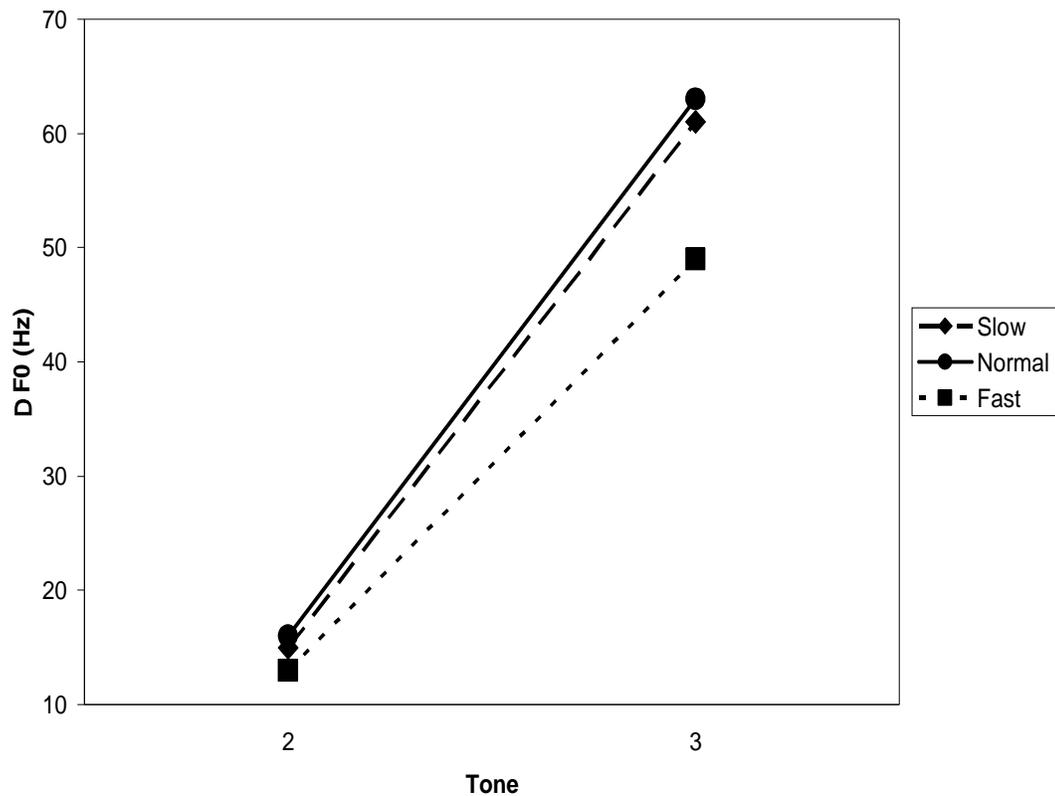


Figure 7.  $\Delta F_0$  of Mandarin tones 2 and 3 at fast, normal and slow speaking rates.

#### 4. 1. 7 Turning Point Values

To evaluate the impact of rate-induced changes in the TP of Tone 2 to 3, a two-way ANOVA (Rate x Tone) was conducted. There was a main effect for Rate [ $F(2, 572) = 201.385, p < 0.001$ ], with TP values varying as a function of speaking rate. Within-subjects contrast effects revealed that TP values at the fast speaking rate (116 ms) were significantly smaller than they were at the normal (185 ms) speaking rate which, in turn, were significantly shorter than they were at the slow speaking rate (232 ms).

As expected, there was also a main effect for Tone [ $F(1, 286) = 448.220, p < 0.001$ ]. A Bonferroni posthoc test revealed that TP values of Tone 2 (96 ms) were significantly smaller than TP values of Tone 3 (259 ms).

Between Rate and Tone there was a significant interaction for TP values [ $F(2,572) = 51.837, p < 0.001$ ] with TP values of Tone 2 to Tone 3 significantly changing from fast to normal to slow speaking rate (see Figure 8). For Tone 2 the ratio of normal-to-slow is a smaller increase (12%), while for Tone 3 the ratio of normal-to-slow is much greater (32%).

The Rate and Tone interaction can also be seen from TP values expressed as a percentage of vowel duration for Tone 2 and 3, shown in Table 6. For Tone 2 the percentage of TP values in the three speaking rates (F=18, N=22, S=16) was much lower than they were for Tone 3 (F=41, N=42, S=40). Interestingly, the TP percentage remained quite stable across different speaking rates, especially for Tone 3 (ranging from 40 % of the vowel duration at the slow rate and 42 % of the vowel duration at the fast rate). More variability in TP duration was seen for Tone 2 (ranging from 22 % of the vowel duration at the normal rate and 16 % of the vowel duration at the slow rate).

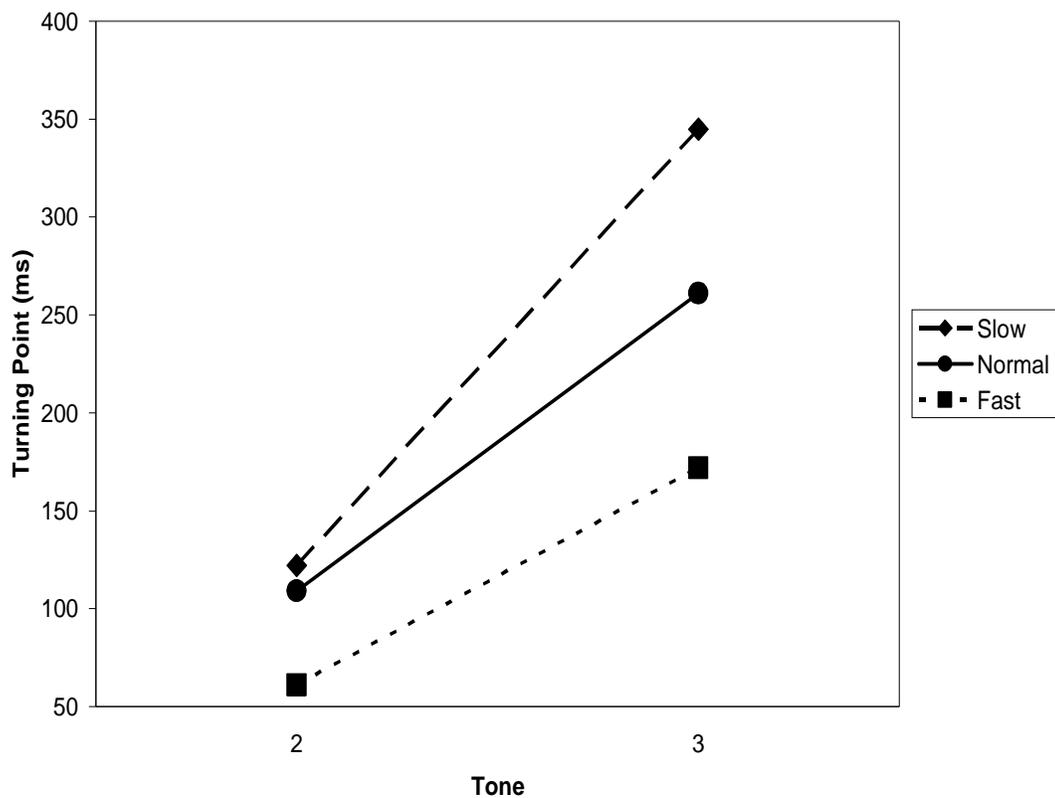


Figure 8. Turning Point of Mandarin tones 2 and 3 at fast, normal and slow speaking rates.

Rate	TP Values (ms)		TP Percentage of vowel	
	Tone 2	Tone 3	Tone 2	Tone 3
Fast	62	172	18	42
Normal	110	261	22	42
Slow	123	346	16	40
Ratio (F-N)	0.56	0.66		
Ratio (S-N)	1.12	1.32		

Table 6. Mean TP values and ratios of Tones 2 and 3 at three speaking rates. The right side columns show the TP values for Tone 2 and 3 expressed as a percentage of vowel duration for the slow, normal, and fast speaking rate.

## Chapter 5

### 5.1 Discussion

The present study examined the acoustic effects of speaking rate on Mandarin tones. Eight native Mandarin speakers participated in this research. They produced 68 stimuli at a fast, normal and slow speaking rate. A total of 1632 tokens were collected. At the fast speaking rate, the stimuli were separated by intervals of 600 milliseconds from offset to onset; at the normal speaking rate the intervals were 1200 milliseconds and at the slow speaking rate they were 2400 milliseconds.

The data were examined and analyzed for the consonant and vowel duration, for the fundamental frequency values at the onset and the offset of vowels as well as for the vowel F1, F2 and F3 values. Turning Point values and the  $\Delta F0$  values of Tone 2 and 3 were also examined.

While measuring the consonant and vowel duration, it was found that both consonant and vowel duration decreased as the speaking rate increased from slow to normal and to fast. Also, the vowel duration of a longer tone, such as Tone 3 increased less than a shorter tone, such as Tone 4 at slower speaking rates. This is contrary to what was expected.

F0 values at the onset and the offset and the formant frequency values of F2 showed significant changes at different speaking rates. Across increasing speaking rates, F0 onset values were increasingly higher with the high onset tones, Tone 1 and 4, than with Tone 2 and 3. At offset, F0 values were the highest only with Tone 1 and 4. In terms of the vowel F1, F2 and F3 values, few significant differences were

observed, except for the rate effect on F1 formant values. Since neutralization of vowels changes F2 values depending on whether the vowel is a front or a back vowel, front and back vowels were further evaluated separately. F2 of front vowels was found to be lowered, while the back vowels were raised when speaking rate increased. The present research also examined the contours of Tone 2 and 3 for the Turning Point values and the  $\Delta F0$  values at different speaking rates. Significant changes were found in both TP and  $\Delta F0$  values.  $\Delta F0$  values of Tone 2 and Tone 3 changed significantly when speaking rate changed from fast to normal speaking rate. The mean  $\Delta F0$  values of Tone 2 were significantly less than  $\Delta F0$  values of Tone 3 at all speaking rates. There is no overlap in  $\Delta F0$  between Tones 2 and 3. TP occurs earlier in Tone 3 at the fast speaking rate, whereas, TP occurs later in Tone 3 at the slow speaking rate.

The results of the present study have implications related to the research hypotheses mentioned in the Introduction. First, it was hypothesized that the overall consonant and vowel duration of the four Mandarin syllables would change at different speaking rates. In agreement with the previous research, such as that by Port (1976), the absolute duration at the slow and fast speaking rate was significantly different. There was a decrease in both consonant and vowel duration when speaking rate changed from slow to normal and then to fast. At the fast speaking rate, the consonant and vowel durations of all four tones were significantly shorter than they were at the normal and at the slow speaking rate. The experiment showed that the rate was successfully manipulated.

According to the results of Port's (1976) experiments for individual vowels, it was also hypothesized that the duration of the longer tones, such as Tone 3, would increase relatively more than the duration of the shorter tones, such as Tone 4, at the slow speaking rate. In his experiments, Port (1976) found that the duration of /l/ was relatively longer than the duration of /i/ at the fast speaking rate as compared to the slow speaking rate. For the present experiments, the interaction between Tone and Rate indicated that at different speaking rates the durations of the four different tones were also significantly different from each other. However, our results patterned opposite to Port's data predictions. For tone 3, the longest tone, the ratio of normal-to-slow showed about a 40 % (ratio S/N 1.4) increase while the shortest tone (tone 4) showed the greatest increase of 74 % (ratio S/N 1.74). The results showed that the shortest tone (tone 4) had the most change in duration across all speaking rates. This result did not support the hypothesis that the duration of the longer tones, such as Tone 3, would increase relatively more than the duration of the shorter tones, such as Tone 4, at the slow speaking rate. At the slow speaking rate, tones get lengthened to a different extent, with Tone 4 showing the greatest increase and Tone 3 the least.

Second, the results of this study demonstrated that a spectral property, F0 values at onset and offset, was also significantly affected by speaking rate. This result supported some findings of Gandour et al. (1999) for Thai tones, showing F0 values changing as a function of speaking rate. In the Gandour et al. study the F0 values of the unstressed syllables of the five Thai tones were significantly changed at different speaking rates. The F0 contour was generally higher at the fast speaking

rate compared to the slow speaking rate. Interestingly, Gandour et al. did not find consistent F0 differences in stressed syllables. In the current research, F0 at both onset and offset had their highest values at the fast speaking rate. This means that when speakers spoke at a fast speaking rate they raised their fundamental frequencies much more than they did when they spoke at the normal and slow speaking rate. Also, across speaking rates F0 onset values were much higher with the high onset tones, Tone 1 and 4, than with Tone 2 and 3. At offset, F0 values were the highest only with Tone 1 and 4. Given that Tone 1 is a high level tone and tone 4 is a falling tone there is more leeway to change F0 at onset and offset due to speaking rate. Across onset and offset, Tones 1 and 4, therefore, showed more systematic changes in F0 height across speaking rates, with higher F0 values at both onset and offset at the faster speaking rates. Tone 2 and 3 did not show this pattern of higher F0 at onset and offset at the fast speaking rate. For tones 2 and 3, at both onset and offset, there was less change across speaking rates, with both F0 onset and F0 offset values showing little difference at fast, normal or slow speaking rates. In fact, at the slowest rate, tone 2 showed the highest offset F0 value. This finding corresponds with Xu's (1998) finding that the F0 peak associated with the rising tone (Tone 2) always occurred near the offset of the Rising-tone-carrying syllable. These results indicate that F0 offset carries more information about the rising tone, and this information is clearly evident at the slowest speaking rate.

The present results also indicated that at different speaking rates,  $\Delta F0$  values changed significantly. Overall, the mean  $\Delta F0$  values of Tone 2 were significantly

less than  $\Delta F0$  values of Tone 3 at all speaking rates.  $\Delta F0$  values of Tone 2 and Tone 3 changed significantly when speaking rate changed from fast to normal. There was little difference in  $\Delta F0$  between the normal and the slow speaking rates. Across all speaking rates, there is no overlap in  $\Delta F0$  in Tones 2 and 3. While  $\Delta F0$  for Tone 2 was always less than that for Tone 3,  $\Delta F0$  did vary across speaking rates for Tone 2 and Tone 3, but that change was the greatest for Tone 3. Compared to Tone 2, Tone 3 had a greater change in  $\Delta F0$  across speaking rates.

The findings of  $\Delta F0$  of Tones 2 and 3 showed that  $\Delta F0$  did change across speaking rates. These findings can be considered with respect to the results in previous experiments, such as Jongman et al. (2006), which found no effects of a change in speaking rate in the continuum varying in  $\Delta F0$  only. In their perception experiments, the listeners did not show an overall earlier shift towards Tone 3 responses in the fast precursor condition while the continuum was manipulated in  $\Delta F0$  only. The present production data do show changes in  $\Delta F0$  at different speaking rates. Speaking rate effects showed up the most for Tone 3. In production, both  $\Delta F0$  and TP vary across speaking rate.

Regarding the hypothesis about the temporal properties (TP) of Tone 2 and 3, the present results support the hypothesis that speaking rate greatly affects the TP of Tone 2 and 3. TP occurs earlier in Tone 3 at a fast speaking rate, whereas, TP occurs later in Tone 3 at a slow speaking rate. These results are in accord with the findings of Moore and Jongman (1997) and Jongman et al. (2006) on the temporal property of Tone 2 and 3 that TP was perceived with respect to speaking rate.

F1, F2 and F3 did not vary significantly with speaking rate. F2 and F3 did not show a significant change as a function of speaking rate. All formant frequency values were quite constant except for some change in F1. While observing sizeable changes in duration and F0 at different speaking rates, vowel quality did not seem to change that much. Therefore, further analyses were conducted

Neutralization of vowel quality changes F2 differently depending on whether the vowel is a front or a back vowel. The analysis of F2 of the front and the back vowels separately showed that F2 values did change significantly at different speaking rates. F2 values of front vowels were lowered while the F2 values of back vowels were raised when speaking rate increased. This finding indicated that vowels were more neutralized when speaking rate increased.

Finally, while the findings of the present study are about how tones are implemented under different speaking rates in production, they also have certain implications for understanding how listeners perceive tones. Future research on this topic may contribute to modifying the design so that the stimuli collected are closer to natural speech in terms of different speaking rates. In the present study, the speakers spoke at either a fast or a slow speaking rate in conjunction with the pre-programmed signals on the computer. With training and a greater range of timing, the fast speaking rate might be able to be made even faster, and the slow speaking rate slower. This kind of speech may result in an overlap in  $\Delta F0$  and Turning Point for Tone 2 and 3.

## References

- Abramson, A. (1962). "The vowels and tones of Thai: acoustical measurements and experiments," *Bloomington: Indiana University Research Center in Anthropology, Folklore and Linguistics*. Publication No. 20.
- Blicher, D. L., Diehl, R. and Cohen, L. (1990). "The 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.
- Gandour, J., Tumtavitikul, A. and Satthamnuwong, N. (1999). "Effects of speaking rate on Thai Tones," *Phonetica*. 56, 123-134.
- Gandour, J., Tumtavitikul, A. and Satthamnuwong, N. (2002). "Effects of speaking rate on Thai emphatic tone," *Asia Pacific Journal of Speech, Language and Hearing*. 7, 101-110.
- Gay, T. (1968). "Effect of speaking rate on diphthong formant movements," *Journal of the Acoustical Society of America*. 44. 1570-1573.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Jongman, A., Moore, C. B. and Lai, Y. (2006). "The role of language experience in normalization: Speaker and rate normalization in the perception of lexical tone by Mandarin and English listeners," Linguistics Department, University of Kansas.

- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *Journal of the Acoustical Society of America*. 35, 1773-1781.
- Manuel, S. Y. (1990). "The role of contrast in limiting vowel-to-vowel coarticulation in different languages." *Journal of the Acoustical Society of America*. 88, 1286-1298.
- Miller, J. L. (1981). "Effects of Speaking Rate on Segmental Distinctions," in *Perspectives on the Study of Speech*, edited by Eimas, P. D. & Miller, J. L. (Lawrence Erlbaum Ass. Publishers), pp. 39-74.
- Moore, C. B. (1995). "*Speaker and rate normalization in the perception of lexical tone by Mandarin and English learners,*" Unpublished doctoral dissertation, Cornell University.
- Moore, C. B. and Jongman, A. (1997). "Speaker normalization in the perception of Mandarin Chinese tones," *Journal of the Acoustical Society of America*. 102, 1864-1877.
- Port, R. F. (1976). "*The influence of speaking tempo on the duration of stressed vowel and medial stop in English Trochee words,*" Unpublished doctoral dissertation, University of Connecticut.
- Xu, Y. (1997). "Contextual tonal variations in Mandarin," *Journal of Phonetics*. 25, 61-83.
- Xu, Y. (1998). "Consistency of tone-syllable alignment across different syllable structures and speaking rates," *Phonetica*. 55, 179-203.