Effects of speaking rate and context on the production of Mandarin tone

Joan Sereno[1], Hyunjung Lee[2], and Allard Jongman[1]

1 KU Phonetics and Psycholinguistics Lab, Department of Linguistics, University of Kansas, USA
2 Department of English Education, Kyungnam University, South Korea
sereno@ku.edu, hyunjunglee123@gmail.com, jongman@ku.edu

ABSTRACT

Acoustic variability is ubiquitous in speech. To understand how listeners cope with this variability, the present study focuses on Mandarin tones. Since the major cues to tone are both spectral (F0 height and direction) and temporal (turning point, TP), tones provide a good test case to evaluate effects of context. The present study examined production variation due to changes in speaking rate and tonal context. Twelve speakers (6M, 6F) produced target syllables with the four different Mandarin tones at fast, normal and slow speaking rates. These syllables appeared at the end of a carrier sentence, following either a high-level tone (Tone 1) or a high-falling tone (Tone 4). Speaking rate had a similar effect across all tones and consonants while preceding context resulted in changes to temporal parameters such as turning point.

Keywords: tone, Mandarin, speaking rate, context

1. INTRODUCTION

A central issue in the field of speech perception involves the notion of perceptual constancy: how do listeners derive the phonetic structure of an utterance given the pervasive acoustic variability that is present in the speech signal? The current study investigated how a major source of this acoustic variability, speaking rate, affects the acoustic properties of lexical tone. Previous studies on Mandarin tones have identified F0 and duration as primary acoustic cues to tone identification. We provide needed acoustic data documenting the range of variation that occurs in the production of utterances under varying speaking rates and coarticulatory contexts. While temporal properties of segments have been widely studied, acoustic descriptions of lexical tones have often been confined to analyses of F0 contours and F0 height. However, at different speaking rates, Mandarin tones display different phonetic characteristics, which, in turn, influence their perception.

Few studies have examined changes in tone due to speaking rate. Xu [6] compared tonal contours across three different speaking rates for four Mandarin speakers but his main goal was not to document the effect of rate on tone. Xu showed that the syllable (rather than the rhyme) is the domain for tonal alignment and that this alignment is not affected by the internal structure of the syllable or speaking rate. However, inspection of his figures revealed that the tonal contours do seem to be affected by speaking rate but these graphs do not provide the numerical detail necessary to evaluate the effect of speaking rate. Gandour et al. [3] compared tonal contours in six male speakers of Thai at different speaking rates. An increase in rate did not affect all tones uniformly, with flatter F0 slopes for the dynamic (falling and rising) tones and with less of a rate effect for stressed syllables. These findings hint at a non-negligible effect of speaking rate on tone shape although a systematic study of the acoustic properties of tones produced at different speaking rates has not been conducted.

In addition to F0 contour, *temporal properties* such as duration and Turning Point (the time interval from onset of the tone to the lowest F0 value) are important cues to signal lexical tone (e.g., [1], [4]). However, little is known about the degree to which these temporal cues are affected by changes in speaking rate.

The present study explored the extent to which variations in speaking rate affect the four tones of Mandarin. Mandarin Chinese is a language with 4 distinct lexical tones. Mandarin phonemically distinguishes a high level pitch (55, Tone 1), a high rising pitch (35, Tone 2), a falling rising pitch (214, Tone 3), and a high falling pitch (51, Tone 4). In addition, we examined stimuli produced in two different sentences to observe how coarticulatory contexts affect tonal properties under varying speaking rates. The focus was on the overall duration of the four tones as well as on the location of the Turning Point and fundamental frequency changes for Tone 2 and Tone 3.

2. METHOD

2.1. Speakers

Twelve subjects (6M, 6F) participated in the production study. All were native speakers of Beijing Mandarin Chinese from Mainland China.

2.2. Stimuli

Each subject produced fifteen monosyllables (CV) with the four Mandarin tones for a total of 60 stimuli, each produced at a fast, normal and slow speaking rate. Speaking rate was controlled by manipulating the presentation of stimuli on a computer monitor with different SOAs (stimulus onset asynchronies). SOA was 3500ms for a slow rate, 2500ms for a normal rate and 1500ms for a fast rate. Subjects produced the target stimuli embedded in two precursor sentence frames in which the tone preceding the target word was either Tone 1 (high level) or Tone 4 (high-falling). The two contrasting precursor frames used in the production study were 'zhe4 ge0 zi4 de0 fa1 yin1 *fa1* ____' ("This measured character's pronounciation is pronounced _____") and 'zhe4 ge0zi4 de0 fa1 yin1 *shi4* _____' ("This measured character's pronounciation is _____"). All reading materials were presented to subjects in Chinese characters. Pinyin with a relevant tone number was presented next to the target Chinese character (e.g., 鼻 bí2 '*nose*' or 笔 bǐ3 '*pen*'). Overall, a total of 4320 tokens were obtained [60 target words (15 words × 4 tones) × 3 rates × 2 precursors × 12 speakers].

2.3. Procedure

All speakers were recorded in an anechoic chamber, using a cardioid microphone (Electrovoice-RE 20) and a digital recorder (Marantz PMD 671). The stimuli at a slow speaking rate were always recorded first. The stimuli were recorded at a sampling rate of 22050 Hz and analyzed using the software package *Praat* ([2]).

2.4. Measurements

Three durational properties were measured: vowel duration, consonant duration, and turning point. For consonants, VOT was measured for stops and the noise portion was measured for fricatives and affricates.

Turning point duration was measured for Tones 2 and 3. Turning point duration was measured from the onset of the tone to the point of change in F0 direction from falling to rising in Tone 2 and Tone 3 as indicated in the pitch track.

In addition to the duration measures, fundamental frequency (specifically, ΔF0) was measured for Tones 2 and 3. In the pitch contour, ΔF0 was defined as the change in fundamental frequency during the time interval from onset of the tone to the lowest pitch value.

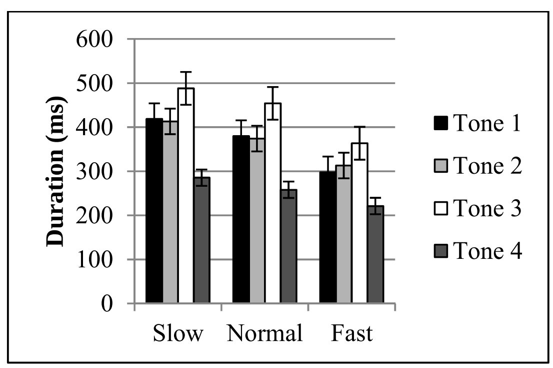
3. RESULTS

The data were evaluated based on repeated measures General Linear Model (GLM) Analyses of Variance (ANOVAs). The within-subject factors considered were Speaking Rate (slow, normal, fast); Tone (Tone1, Tone2, Tone3, Tone4); and Precursor (preceding Tone 1, preceding Tone 4). The durations of vowels and consonants were entered as dependent variables, with the Tone factor including all four tones. For the turning point analysis, the Tone factor with two tones (Tone2, Tone3) was used. In order to further explore the data within each dependent variable, pairwise comparisons were conducted ($\alpha = 0.05$) among three rates and four tones.

3.1. Tone (vowel) duration

Results of a three-way repeated measures ANOVA (Tone by Rate by Precursor) showed main effects of two within-subject factors, Tone (F (3, 9) = 37.708, $p < 0.001$) and Rate (F (2, 10) = 9.154, $p = 0.006$) on vowel duration, but no main effect of Precursor (F (1, 11) = 0.314, $p = 0.587$). Pairwise comparisons showed that the vowel duration of Tone 3 was always longer than that of any other tones, while Tone 4 was always the shortest among the four tones, with the duration of Tones 1 and 2 comparable (Tone3 > Tone1 = Tone2 > Tone4) (see Figure 1).

Figure 1: Average vowel duration (ms) for the four tones at slow, normal and fast speaking rates. Data were averaged across the two Precursors. Error bars indicate standard errors.



For the effect of Rate, the vowel duration was significantly different across all three speaking rates (slow > normal > fast), with a 19% change for fast speaking rates compared to normal rates and a 9% change for slow speaking rates as compared to normal rates. Interestingly, there was no Rate by Tone interaction, with all tones showing similar reductions/increases across speaking rates. There were no other significant two-way or three-way interactions.

3.2. Consonant duration

Results of a three-way repeated measures ANOVA (Tone by Rate by Precursor) showed main effects of two within-subject factors, Tone (F (3, 9) = 7.737, p = 0.007) and Rate (F (2, 10) = 14.789, p = 0.001) on consonant duration, but no main effect of Precursor (F (1, 11) = 0.282, p = 0.606) (see Table 1). Pairwise comparisons showed that the consonant duration of Tone 3 stimuli was similar to that for Tone 2 but significantly longer than consonant durations for Tone 1 and Tone 4 (Tone3 = Tone2 > Tone1 > Tone4).

Table 1. Average consonant duration (ms) for the four tones at slow, normal and fast speaking rates in two precursors.

Context	Rate	Tone 1	Tone 2	Tone 3	Tone 4
Preceding tone 1	Slow	86	94	93	84
	Normal	80	88	89	80
	Fast	72	77	82	69
Preceding tone 4	Slow	91	94	94	83
	Normal	81	87	89	78
	Fast	73	83	79	74

For the effect of Rate, consonant duration was significantly different across the three speaking rates (slow > normal > fast), with a 10% change for fast speaking rates compared to normal rates and a 7% change for slow speaking rates as compared to normal rates. Similar to vowel duration, consonant duration also significantly changed across speaking rates.

There were no significant two-way or three-way interactions for consonant duration.

3.3. Turning Point (TP) location

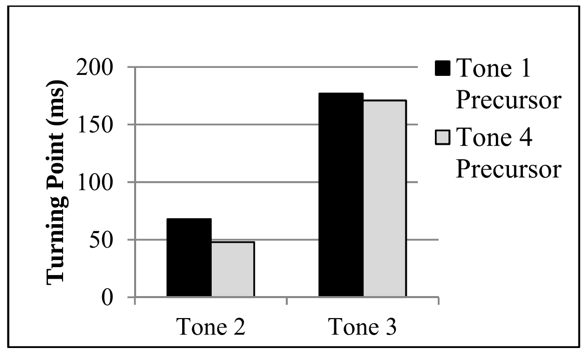
Results of a three-way repeated measures ANOVA (Tone by Rate by Precursor) showed a main effect of Tone (F (1, 4) = 33.419, p = 0.004) for turning point duration. As expected, pairwise comparisons showed that the average turning point duration is always longer for Tone 3 (174 ms) than for Tone 2 (58 ms).

Interestingly, there was no main effect of Rate for turning point (F (2, 3) = 0.747, p = 0.545) nor did Rate interact with Tone, showing similar change in turning point duration across speaking rate for Tone 2 and Tone 3.

Results also showed a main effect of Precursor (F (1, 4) = 10.38, p = 0.032) for turning point duration. For the effect of Precursor, the turning point duration was longer with a preceding Tone 1 than with a preceding Tone 4 (see Figure 2).

A significant interaction between Tone and Precursor (F (1, 4) =35.809, p=0.004) derived from the fact that for Tone 2, the turning point was 20 ms longer with a preceding Tone 1 than Tone 4, while this difference was only 6 ms for Tone 3. There were no other significant two-way or three-way interactions for turning point duration.

Figure 2: Average turning point duration (ms) for Tone 2 and Tone 3 with a Tone 1 precursor and a Tone 4 precursor.



3.4. Fundamental frequency (F0)

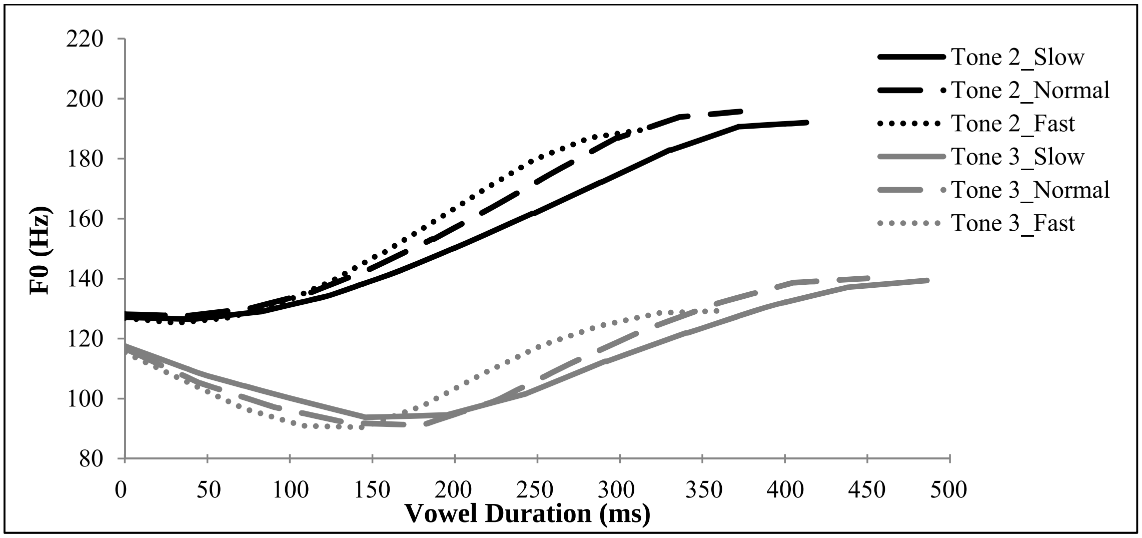
Average F0 contours for Tone 2 and Tone 3 stimuli at a slow, normal, and fast speaking rate are shown in Figure 3. The current analysis examined ΔF0. ΔF0 is defined as the change in fundamental frequency (F0) during the interval from onset of the tone to the lowest fundamental frequency value.

Results of a three-way repeated measures ANOVA (Tone by Rate by Precursor) showed a main effect of Tone (F (1, 4) = 27.354, p = 0.006). Pairwise comparison showed that the ΔF0 value of Tone 3 (37 Hz) is always greater than that of Tone 2 (6 Hz). There were no main effects of Rate (F (2, 3) = 0.415, p = 0.693) or Precursor (F (1, 4) = 4.396, p = 0.104) for ΔF0 and there were no significant two-way or three-way interactions for ΔF0.

4. DISCUSSION

The present study explored the extent to which variations in speaking rate affect the four tones of Mandarin. Twelve Mandarin speakers produced 15 monosyllables with all four tones in a sentential frame. To examine the effect of preceding context, two different sentences were used, with either a Tone 1 precursor or a Tone 4 precursor. Three different speaking rates were elicited by presenting stimuli on a monitor with different SOAs.

Figure 3: Average F0 contours for Tone 2 and Tone 3 produced at a slow, normal, and fast speaking rate.



The most interesting finding was that while there were systematic differences in tone duration observed across tone and while there were systematic differences in speaking rate, there were no interactions. In a sentence context, faster speaking rates systematically shorten vowels to a comparable degree across different tones. Tone contour (level or dynamic) or overall length of the tone does not matter. These findings agree with those for fundamental frequency ([6]) in that F0 contours of the four lexical tones of Mandarin did not vary systematically as a function of speaking rate. The present data in addition show that the temporal characteristics of tones were not differentially affected by speaking rate, with all tones similarly shortened as speaking rate was increased.

A further finding of the present set of data was that speaking rate has a similar effect on consonant durations. In a sentence context, a slower speaking rate systematically lengthened consonants as well as vowels. While the magnitude of the lengthening is much less, speaking rate does systematically change consonant duration, regardless of tone.

Turning point (TP), which is duration from the onset of a tone to the point of change in F0 direction from falling to rising, clearly distinguished Tone 2 from Tone 3 stimuli, with Tone 2 having a significantly shorter TP than Tone 3. What is interesting is that in utterances that were produced in sentence context, there is no overlap in TP between Tone 2 stimuli produced at the slowest rate and Tone 3 stimuli produced at the fastest rate. These data are in contrast to the overlap in turning point that occurs in isolated words [5]. In isolation, TP for Tone 2 decreased less across speaking rates than TP for Tone 3. Additionally, in isolation, there was a significant overlap in TP between a Tone 2 at a slow speaking rate and a Tone 3 at a fast speaking rate. Productions in sentence context did not show a similar overlap, with distinct TPs for Tone 2 and Tone 3 stimuli across all speaking rates.

An additional finding from the present data was that the context, in this case the preceding tonal context, had a substantial influence on the temporal characteristics of the tone. Turning point duration for both Tone 2 and Tone 3 was longer with a preceding Tone 1 than with a preceding Tone 4. The temporal cues in the immediately preceding context (Tone 1 versus Tone 4) were quite influential. The TP of both Tone 2 and Tone 3 stimuli were influenced by the preceding context, with an immediately preceding long duration Tone 1 context resulting in target tones with longer TPs.

The present study also examined fundamental frequency, namely change in fundamental frequency from onset to the lowest point in the tone (ΔF0). For changes in speaking rate, there were no systematic changes in the pitch contour of the tones. While speaking rate changed the turning point duration, it did not change fundamental frequency. Speakers, even at fast rates of speech, were able to maintain changes in fundamental frequency that are critical for distinguishing Tone 2 from Tone 3.

The present data were particularly significant in terms of finding no rate by tone interactions. Tone productions in sentential context appear to reduce the overlap that is often observed in isolation productions. Systematic speaking rate effects were observed but, in a sentential context, these effects are similar across tones. Instead, changes in context seem to exert a more significant and distinct effect on the timing characteristics of Mandarin tones.

5. REFERENCES

[1] Blicher, D., Diehl, R. Cohen, L. 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *J. Phon*. 18, 37-49.

[2] Boersma, P., Weenink, D. 2007. Praat: doing phonetics by computer [Computer program]. Version 4.6.21, retrieved 5 September 2007 from http://www.praat.org/.

[3] Gandour, J., Tumtavitikul, A., Satthamnuwong, N. 1999. Effects of speaking rate on Thai tones. *Phonetica* 56, 123-134.

[4] Moore, C., Jongman, A. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *J. Acoust. Soc. Am.* 102, 1864-1877.

[5] Sereno. J., Lee, H. 2010. Temporal change in Mandarin tone due to speaking rate variation. *J. Acoust. Soc. Am.* 127, 2023.

[6] Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179-203.