TONAL VARIATIONS IN MANDARIN: DATA FROM SPONTANEOUS AND READ SPEECH

Li-chiung Yang[a] *& Richard Esposito*[b]

[a]Tunghai University, Taiwan; [b]Spoken Language Research, VA, USA
lyang@sprynet.com; esposito_r@bls.gov

ABSTRACT

It is well-known that tone modification occurs in read speech due to a number of different contextual effects on tonal targets, most prominently due to local tone sequencing and effects of downstepping, and that in spontaneous speech, a greater variability of tone shape has been observed as arising from a number of global discourse contextual effects. In this paper we show comparative results on the variability of tonal target realization in read and spontaneous speech. Specifically, our study confirms the effects of tonal sequencing on pitch, and further shows that tonal shape conformity to the defined lexical shapes is greater in both spontaneous speech and read speech when amplitude is higher.

Keywords: tones, Mandarin, amplitude, variability

1. INTRODUCTION

Previous research on read speech and spontaneous speech in Mandarin has demonstrated the wider variability of pitch movement in the latter, and researchers have attributed the greater variability to a number of different factors. Research on read speech [5, 8] has predominantly focused on production and perception of tones in read or experimentally elicited speech. In these studies, a prominent result has been the elucidation of rules for tone sandhi, and the modification of tonal shapes under differing aspects such as questioning, focus, or emphasis.

In research on the relationship between tone and intonation, researchers [4, 6] have pointed out two theories of intonational effect, the first in which intensification occurs when tonal and intonational values correspond, while tone neutralization occurs when tonal and intonational values diverge; the second theory derives from [1] in which tonal value may be changed by intonation such that their tonal values may become unrecognizable, and as such involves not only production, but perception.

Research on speech prosody in both tone and non-tone languages [2, 3] has shown that there are a number of important influences on the prosody of natural speech, including interactive monitoring activity, emotional state, and the level of uncertainty, as well as topic organization and phrasal position. In the current paper, we study how tonal sequences and amplitude affect the realization of lexical tone shape, and compare results obtained from a read speech corpus to results found in spontaneous Mandarin speech.

2. DATA AND METHODOLOGY

The data utilized in this study are part of a larger project of Mandarin speech, totaling over 20 hours of speech. For this study, a subset of continuous speech of approximately 20 minutes in duration for spontaneous speech and 10 minutes of read speech were analyzed. The spontaneous speech data were collected in informal home settings, and the read speech corpus was collected in a laboratory setting.

Data were manually segmented to the phrase, word, and syllable level, and time, amplitude, and pitch (f_0) values were extracted from the speech files using speech software. f_0 values were automatically corrected for formant outliers via programming, and slight errors in segmentation among the different label files were also adjusted automatically.

A total of 5,755 lexical syllables from the 20 continuous minutes of segmented spontaneous speech, which amounted to 4,737 syllables after eliminating overlapped syllables for which $f_0$values were ambiguous between speakers. For read speech, there was a single speaker, and a total of 2,074 syllables of speech.

Table 1: Number of syllables by tone, spontaneous speech.

Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	Total
689	543	731	1425	1349	4737

Table 2: Number of syllables by tone, read speech.

Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	Total
356	300	318	502	598	2074

3. VARIABILITY OF TONAL SHAPES

The defined lexical tones in Mandarin are commonly recognized as Tone 1: high pitch level and flat (55); Tone 2: start at mid to low pitch level, and rising (35); Tone 3: starts mid to high level, falls, then rises (214); Tone 4: high and falling (51), and a neutral tone, Tone 0 (variable). Tone 3 has two recognized variants in spontaneous speech, a high to low fall with no final rise, and a short duration, low pitch value (21). With these defined shapes, we used the slope of the f_0 as a simple shape measure of the f_0 values. The linear slopes and intercepts of each syllable were calculated using S-plus. Since a quadratic would be more applicable for Tones 2 and 3, a quadratic fit was also produced for each syllable, and used in measuring the effects of amplitude in Section 3.2.

3.1. Influence of preceding and following lexical tone

Local tone sequences can theoretically affect the realized target in two primary ways, through anticipatory effects of the upcoming syllable, or through carry-over effects of the preceding syllable. For example, [8] found greater evidence for the strength of carry-over effects than for anticipatory using balanced sequences in experimental speech. In the current study, we were interested in investigating if the same effects on the realization of target tones hold in our spontaneous speech data.

In Tables 3 through Table 6, we compare the average slope of each lexical syllable in the data, grouped by immediately preceding and succeeding syllable tone, by calculating the linear regression slopes for two speakers in spontaneous conversation for all tones that could reliably isolated to a given speaker, as grouped by the subsequent lexical tone in Tables 3 and 4, and by the preceding tone for Tables 5 and 6.

According to the anticipatory theory, high succeeding onset tones 1 and 4 should show a relatively positive slope, and conversely for low and low onset tones 2 and 3. In forward assimilation not based on the succeeding onset, a succeeding rising tone 2 can cause a rising effect. We can see that there is considerable variability in the slopes of each tone depending on the tonal sequence; we can also see that this is in part speaker dependent.

The data suggest that different tone combinations merge contextually in different ways. For both speakers, the anticipatory condition is true for Tone 4 followed by all tones, and both show the effects of the 3rd tone sandhi (33—>23). When Tone 3 is followed by Tone 1 though, Tone 3 remains low, rather than assimilating to high Tone 1, but succeeding Tone 4 has a relative positive effect on Tone 3 for both speakers.

Table 3: Mean slope in Hz per second, of each tone, by tone of following syllable, Speaker P.

	When followed by:				
Slope of	Tone 1	Tone 2	Tone 3	Tone 4	Neutral
Tone 1	-78.4	-76.2	-125.3	-35.6	-119.5
Tone 2	-66.1	24.7	-115.8	-16.0	-53.2
Tone 3	-185.4	-76.7	-0.4	-149.5	-97.8
Tone 4	-133.6	-254.7	-248.8	-187.1	-202.9

Table 4: Mean slope in Hz per second, of each tone, by tone of following syllable, Speaker S.

	When followed by:				
Slope of	Tone 1	Tone 2	Tone 3	Tone 4	Neutral
Tone 1	71.1	-114.3	-51.3	-53.4	-81.3
Tone 2	-23.7	-48.1	-58.7	52.6	-83.9
Tone 3	-211.1	-145.4	-87.8	-98.0	-205.5
Tone 4	-35.6	-233.8	-177.5	-139.9	-119.9

Table 5: Mean slope in Hz per second, of each tone, by tone of previous syllable, Speaker P.

	When preceded by:				
Slope of	Tone 1	Tone 2	Tone 3	Tone 4	Neutral
Tone 1	-56.4	-35.2	-67.4	-75.2	-102.6
Tone 2	-160.0	-130.0	30.3	-18.5	17.0
Tone 3	-211.3	-163.6	-118.5	-106.2	-26.7
Tone 4	-225.5	-249.7	-122.9	-201.4	-180.1

Table 6: Mean slope in Hz per second, of each tone, by tone of previous syllable, Speaker S.

	When preceded by:				
Slope of	Tone 1	Tone 2	Tone 3	Tone 4	Neutral
Tone 1	-116.5	-35.3	57.4	-104.3	-32.1
Tone 2	-124.7	-61.5	34.6	-12.3	-1.7
Tone 3	-332.8	-87.3	-149.9	-163.8	-114.4
Tone 4	-212.2	-127.7	-69.4	-151.9	-27.5

When looking at the influence of the preceding tone in Tables 5 and 6, we see that a preceding Tone 1 has a negative effect on the subsequent tonal slope. When preceded by Tones 2 or Tone 3, tonal slope becomes relatively higher for Tone 1 and Tone 2 for both speakers, and these results are in accordance with the carry-over predictions, and preceding Tone 1 has a negative effect on all tones for speaker S, and for all tones except for Tone 1 for speaker P.

Both anticipatory and lag effects are found in the above tables, and the differences evident between the speakers suggest that the influence of sequencing in speech on tonal targets may be conditioned by a number of speaker, speech rate,

and speaker style factors. The results given confirm the results of prior researchers who have found substantial effects of preceding and subsequent tones on the realized tonal pitch values in read and experimental speech.

3.2. Influence of amplitude

A range of phonological and linguistic phenomena has been cited as affecting the pitch values of syllables. Among the factors affecting pitch are position in phrase, phonemic identity, emotional and interactive effects, and prosodic environment.

The results of the current study indicate that amplitude is also correlated with changes in tonal shape. Prior research using experimental speech has found that the distribution of energy in syllables is correlated with tone identity in Mandarin tones [7]. A key finding in the current study is that in both spontaneous and read speech, higher amplitude is associated with a greater realized shape of the defined lexical tone shape. There are some interesting differences by lexical tone, and differences between read speech and spontaneous speech as well, as seen in Figures 1-2 for Tones 1 and 4 in spontaneous speech, and Figures 3-4 for read speech. Comparisons of the spontaneous to read speech in the figures show that spontaneous speech is more widely dispersed around the linear regression line, indicating the presence and greater influence of contextual variables other than amplitude in the determination of pitch shape, with many more slope outliers in the spontaneous speech, and a greater proportion of syllables that diverge from their defined values than in read speech.

For both read and spontaneous, the average slopes for Tones 3 and 4 fall, in line with their lexical definition. This is especially clear in the case of Tone 3 and Tone 4, where the calculated slopes in read speech deviate from having a downward slope in only a few cases which appear above the zero slope line. For Tone 1, the read speech slope values cluster relatively symmetrically around zero, in accordance with the defined flat slope of Tone 1, although for both read and spontaneous, the overall slope average was slightly below zero, as shown in Table 7.

Table 7: Mean slope of syllables for four tones, read speech and spontaneous speech.

	Tone 1	Tone 2	Tone 3	Tone 4
Read	-69.59	5.14	-214.9	-301.1
Spontaneous	-48.97	-27.58	-117.09	-166.84

Figure 1: Slope of f_0 for Tone 1 by amplitude, spontaneous speech.

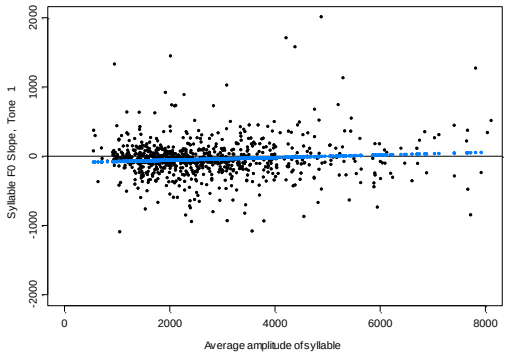


Figure 2: Slope of f_0 for Tone 4 by amplitude, spontaneous speech.

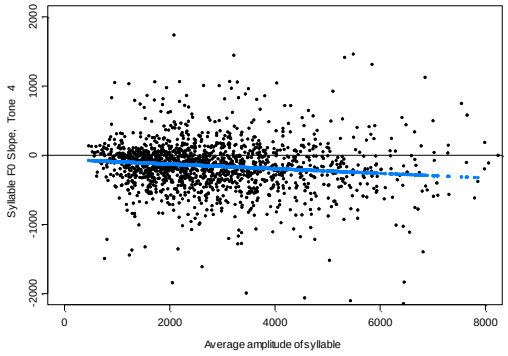


Figure 3: Slope of f_0 for Tone 1 by amplitude, read speech.

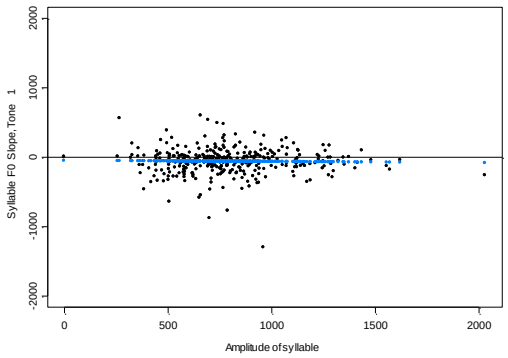
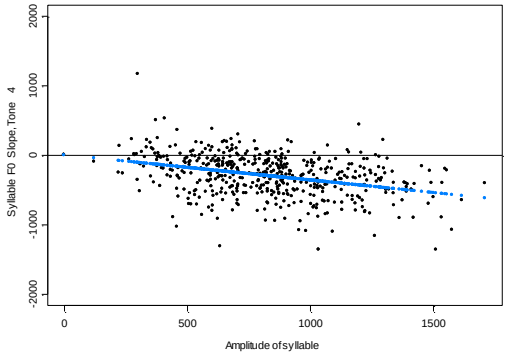
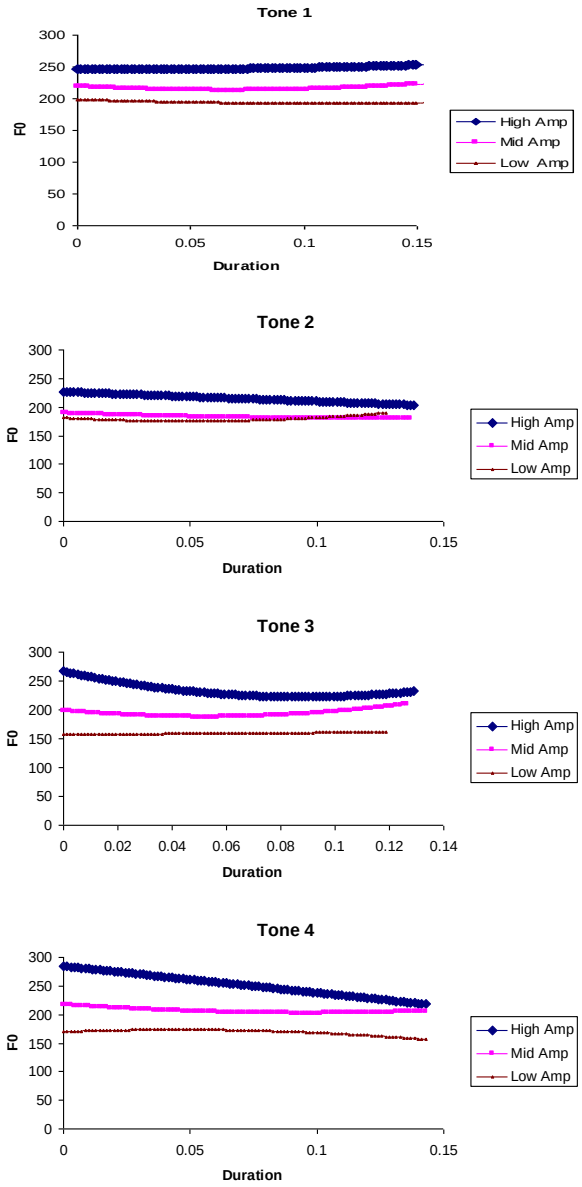


Figure 4: Slope of f_0 for Tone 4 by amplitude, read speech.



For Figure 5 a-d, we calculated an average of the quadratic fits for each of the tones, grouping by the average amplitude of each syllable, but including all speakers and all syllables with 1/3 of syllables by tone in each group, normalized by duration, and depicting the average duration within groups. These figures indicate that *higher* amplitude is associated with greater conformity to the tonal target: Tone 1 becomes higher and flatter, and Tones 3 and 4 fall more with higher amplitude. Only Tone 2 breaks this pattern: higher amplitude Tone 2 becomes higher in average pitch, but exhibits a relatively greater fall in slope.

Figure 5 a-d: (in upper left, right, lower left, right order) Slope of F0 for Tones 1-4 by amplitude, spontaneous speech.



The greater overall conformity to defined shape for read shape may arise because of the relative lack of other contextual influences. However, the greater fall conformity for greater amplitudes in both read and spontaneous speech must arise from a different cause. As is well-known, amplitude frequently is associated with *emphasis*, and these results suggest that one way to give emphasis to a lexical item in a tonal language such as Mandarin is to provide a more obvious and distinct lexical tone shape.

4. CONCLUSION

In this study, we have focused on several factors that determine the realized tonal targets. We have shown evidence for the greater variability of spontaneous speech, and have confirmed prior results on the influence on the tonal f_0 values due to tone sequencing and sandhi. We have also shown that greater amplitude is associated with a greater conformity to tonal targets, and suggest that this partially expresses the emphasis that is possible in a tone language because of the lexically defined pitch values.

5. ACKNOWLEDGMENTS

This research was supported by Taiwan National Science Council under grant no. NSC97-2410-H-029-026.

6. REFERENCES

[1] Chao, Y.R. 1968. *A Grammar of Spoken Chinese.* Berkeley, CA: University of California Press.

[2] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation.* Cambridge University Press.

[3] Hirst, D., di Cristo, A. (eds.). 1998. *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

[4] Shen, X.S. 1990. *The Prosody of Mandarin Chinese*. University of California Press.

[5] Shih, C., Sproat, R. 1992. Variations of the Mandarin rising tone. *Proceedings of the IRCS Workshop on Prosody in Natural Speech* Upenn, 193-200.

[6] Tseng, C.Y. 1981. *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Ph.D. dissertation, Brown University.

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

[8] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.