A Corpus Study of the 3rd Tone Sandhi in Standard Chinese

Yiya Chen [1], *Jiahong Yuan* [2]

[1] Department of Linguistics, Radboud University Nijmegen
[2] Department of Linguistics, University of Pennsylvania
yiya.chen@let.ru.nl, jiahong@ling.upenn.edu

Abstract

In Standard Chinese, a Low tone (Tone3) is often realized with a rising F_0 contour before another Low tone, known as the 3rd tone Sandhi. This study investigates the acoustic characteristics of the 3rd tone Sandhi in Standard Chinese using a large telephone conversation speech corpus. Sandhi Rising was found to be different from the underlying Rising tone (Tone2) in bi-syllabic words in two measures: the magnitude of the F_0 rising and the time span of the F_0 rising. We also found different effects of word frequency on Sandhi Rising and the underlying Rising tones. Finally, for tri-syllabic constituents with Low tone only, constituent boundary showed interesting but puzzling effects on the 3rd tone Sandhi.

Index Terms: Tone, Tone Sandhi, Conversation, Corpus

1. Introduction

In languages where fundamental frequency changes (F_0) indicate lexical tonal contrasts, it is often observed that tones undergo sandhi changes when words are combined in speech, and surface with F_0 contours that are different from the basic tonal shapes produced in isolation. Such a tonal change process is commonly referred to as tone sandhi. During the last two decades, much work has been done on tone sandhi in various Chinese dialects, which culminated in the seminal work by M. Chen (2000). While previous studies have greatly improved our understanding of the tone sandhi phenomena in general, the drawback in most, if not all studies, is that the generalizations are based on introspective judgments or laboratory speech of limited number of speakers. It is thus desirable to complement the existing literature by examining the realization of tone sandhi in a large data corpus with naturally occurring speech. The specific sandhi phenomena that we focus on in this paper is the 3rd (Low) Tone Sandhi in Standard Chinese, where in a sequence of two Low tones, the first surfaces with a rising F_0, comparable to a Rising tone in the language.

Previous studies on the 3rd tone sandhi have mainly concerned with the formation of tone sandhi domain (Shih 1986, Zhang 1988, M. Chen 2000, among others). Few studies have examined the acoustic realization of the Sandhi Rising tone (hereafter SR), in particular, as compared to the underlying Rising tone (hereafter R), and how the Sandhi Rising tone is represented and processed (Speer, Shih, & Slowiaczek 1989, Peng 2000, Kuo, Xu, & Yip, to appear). One consensus that emerges from the literature is that a bi-syllabic word with two Low tones forms a 3rd tone sandhi domain where the first Low tone changes to a Sandhi Rising tone. The application of the 3rd tone sandhi across linguistic boundaries above the level of word, however, is determined by a number of other factors such as syntactic structure, information structure, speaking rate, prosodic weight, etc

(Zhang 1988, Shih 1997, M. Chen 2000, Chen 2003, Chen 2004). Speer et al. (1989) show that listeners are indeed sensitive to a constituent's phrasal structure in judging the application of the 3rd tone Sandhi to constituents which could be ambiguous between an underlying Rising tone and a Sandhi Rising tone. Their results suggest the possibility that the higher linguistic boundary it is between two Low tones, the less likely the 3rd tone sandhi rule is applied. With regard to the difference between the underlying Rising tone and the Sandhi Rising tone, Peng (2000) show that the F_0 maximum of SR is lower than R. Furthermore, in fast speech, a Sandhi Rising tone may flatten and show no apparent F_0 rise (Kuo, Xu, and Yip, to appear). Despite the reported acoustic differences, thus far, however, there is no conclusive evidence that speakers are able to exploit the subtle acoustic differences in distinguishing the two tones.

In the light of these previous results, as a preliminary study, the specific goals of our paper are:

1) To confirm, with corpus data, the lower-level acoustic differences between an underlying Rising tone and a Low tone within a 3rd tone sandhi domain.
2) To explore, given the benefit of large corpus, whether other factors such as word frequency may have any effect on the acoustic realization of the Low tone and the underlying Rising tones within a 3rd tone sandhi domain;
3) To seek acoustic evidence that higher level linguistic organization such as constituent structure may affect the Low tone realization, as compared to the realization of underlying Rising Tones.

2. Method

2.1. Corpus

Data were taken from the HKUST Mandarin Chinese corpus of telephone speech (LDC2005S15) and its transcripts (LDC2005T32). Conversations not from speakers of Standard Chinese (as stated in the document file) were excluded. Syllable boundaries were obtained through forced alignment using the SONIC speech recognition system (Pellom & Sonic 2001), with an acoustic model trained on the CALLHOME (LDC96S34) and CALLFRIEND (LDC96S55) Mandarin Chinese corpora and the CALLHOME Mandarin Chinese lexicon (LDC96L15). We also used the CALLHOME Mandarin Chinese Lexicon to select tonal sequences from the corpus.

2.2. Data

We included: 1) bi-syllabic words with four possible tonal sequences: Low-Low (T3+T3), Low-Rising (T3+T2), Rising-Low (T2+T3), and Rising-Rising (T2+T2); 2) tri-syllabic constituents with Low tone only which include three

constituent types: 1) three mono-syllabic words (hereafter referred to as the 1+1+1 pattern); 2) a bi-syllabic word followed by a mono-syllabic word (hereafter referred to as the 2+1 pattern); and 3) a mono-syllabic word followed by a bi-syllabic word (hereafter referred to as the 1+2 pattern). Tri-syllabic words were excluded because there are only 12 items in the corpus. Table1 lists the total number of tonal sequences with different constituent structures that are used in this study.

Table1. Total number of cases for tonal sequences of different structures.

Tonal sequence	Number of tokens
$(T2+T3)_{word}$	8,113
$(T3+T3)_{word}$	3,938
$(T2+T2)_{word}$	6,515
$(T3+T2)_{word}$	8,112
$T3_w+T3_w+T3_w$ (1+1+1)	2524
$(T3+T3)_w+T3$ (2+1)	636
$T3+(T3+T3)_w$ (1+2)	582

2.3. Acoustic Measurements

To quantify how a tone is realized, we first extracted the F_0 contour of the target tone, located its minimum F_0 and the F_0 at the offset of the tone-bearing syllable, and then calculated two measurements. One is the LogRange of F_0 rise, which is the log of the ratio between the minimum F0 and the F0 at the syllable offset. The other is the percentage of rise time derived by calculating the percentage of duration between minimum F0 and syllable offset over the duration of the tone-bearing syllable. All measurements were automatically extracted using Praat and Python.

3. Results and Discussion

3.1. Acoustic Realizations

For bi-syllabic words, we examined the acoustic realization of the first syllable. As shown in Table 2, both Tone of the target syllable (i.e. T2, Rising vs. T3, Low) and Following Tone, which the target tone precedes (i.e. X+T2 vs. X+T3), had a significant effect on the LogRange of the F_0 rise as well as on the percentage of F_0 rise duration of the target tone. There was also a significant ordinal interaction of these two factors.

Table 2. *Results of statistical analyses*

	df	F value	P value
LogRange of F_0 rise			
Tone	(1, 266674)	397.2	< .0001
Following Tone	(1, 266674)	188	< .0001
Tone*Following Tone	(1, 266674)	211.1	< .0001
Percentage of rise duration			
Tone	(1, 266674)	2101.7	< .0001
Following Tone	(1, 266674)	875.3	< .0001
Tone*Following Tone	(1, 266674)	579	< .0001

Figure 1 shows that when the Following Tone is Rising (i.e. in X+T2), X differs significantly in the magnitude of F_0 rise between the Rising and the Low tone. When the Following Tone is Low (i.e. in X+T3), X remains significantly different in the magnitude of F_0 rise but the Low tone shows much more F_0 rise compared to its rise in the X+T2 context. Results from our corpus study are thus consistent with Peng (2000) which shows that the F_0 peak of a Sandhi Rising tone is lower than that of an underlying Rising tone.
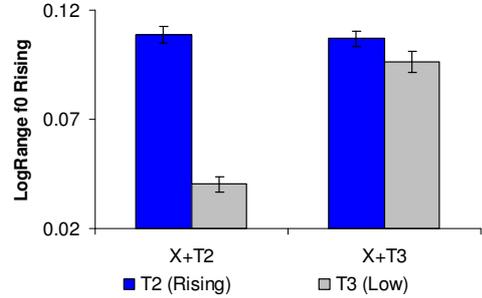
Figure 1: *Means (and ± two standard errors) of the LogRange F_0 rise within Rising vs. Low tones when the tone-bearing syllable either precedes a Low tone or a Rising tone.*

We further observe that there was a significant difference between a Sandhi Rising and an underlying Rising tone with regard to the percentage of F_0 rise duration (i.e. the distance from the F_0 minimum to the end of the tone-bearing syllable as percentage of the total duration of the syllable). Figure 2 shows that when the Following Tone is Rising (i.e. X+T2), the target Rising tone is in general aligned earlier within the tone-bearing syllable (i.e. further away from the syllable offset) than that of a target Low tone. When the Following Tone is a Low tone (i.e. X+T3), the 3rd tone sandhi applies. The percentage of rise duration of an underlying Rising tone is only slightly greater, suggesting that the Sandhi Rising tone becomes more like a Rising tone although they still differ significantly.
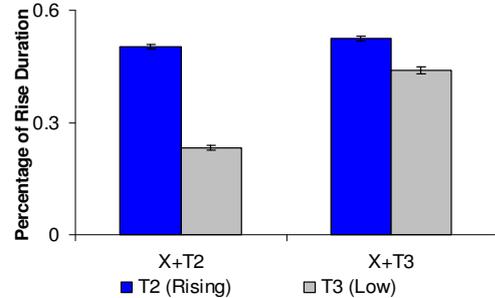
Figure 2: *Means (and ± two standard errors) of the percentage of rise time over the tone-bearing unit within Rising vs. Low tones when the tone-bearing syllables either precede a Low tone or a Rising tone.*

To summarize, both LogRange F_0 rise and percentage of F_0 rise duration suggest that despite the great similarity between an underlying Rising tone and a Sandhi Rising tone, they are indeed different. Thus, results from both laboratory speech and corpus data conjointly suggest that the 3rd tone sandhi is not the change of one toneme (i.e. the Low tone) to another toneme (i.e. the Rising tone) in Standard Chinese, as most phonological accounts have suggested.

3.2. Frequency Effect

We further examined whether frequency has any effect on the realization of the Low vs. the Rising tone. We focused on two tonal sequences where the 3rd tone sandhi is expected to apply: Low-Low (i.e. T3+T3) and Rising-Low (i.e. T2+T3). For each tonal sequence, we separated the bi-syllabic words into four frequency bins: 0-10, 10-100, 100-1000, and above 1000, based on the frequency counts of 3,431,707 words in Xinhua newswire contained in the CALLHOME Mandarin Chinese Lexicon. Figure 3 shows that for the tonal sequence of Low-Low, there was a significant drop in the LogRange F_0 rise of the first Low tone for words with high frequency (i.e. above 1000). As a contrast, Figure 4 shows that for the tonal sequence of Rising-Low, such an effect of frequency does not hold for the Rising tone.
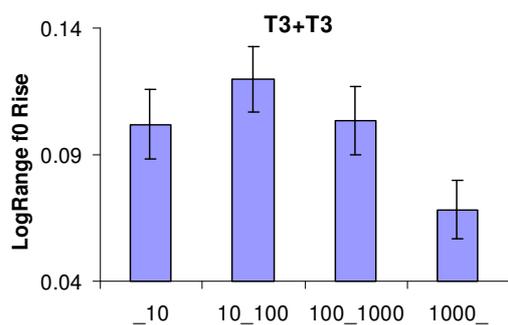
Figure 3: Means (and ± two standard errors) of the LogrRange F_0 rise of the Low tone within different frequency ranges
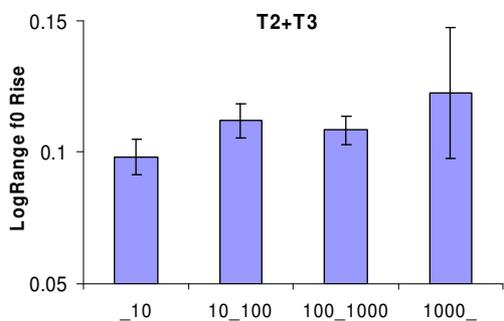
Figure 4: Means (and ± two standard errors) of the LorRange F_0 rise of the Rising tone within different frequency ranges

With regard to the percentage of F_0 rise duration, however, there was no significant effect of high frequency on the Low tone, as shown in Figure 5. The most salient frequency effect on the Rising tone (Fig. 6) is shown in the range of 10-100 where the Rising tone seems to start its F_0 rise earlier.

The results thus raise the question of whether higher frequency words are more resistant to the 3rd tone sandhi and therefore some may have remained Low even in the context of 3rd tone sandhi, which consequently lowered the mean of the LogRange F_0 rise. This possibility is ruled out when we further examined the high frequency words. It was found that most of the tokens within the 1000- bin (983 out of 1218 tokens) is the word *keyi* 'ok/fine'. And the mean LogRange F_0 rise of *keyi* is 0.055. This suggests that the lowered LogRange F_0 rise of higher frequency words is probably due to the fact that the Sandhi Rising of some frequent words are realized with less F_0 rise.
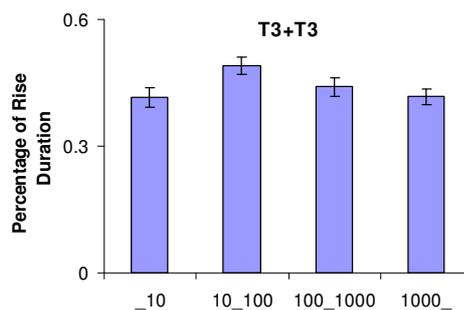
Figure 5: Means (and ± two standard errors) of the percentage of rise of the Low tone within different frequency ranges.
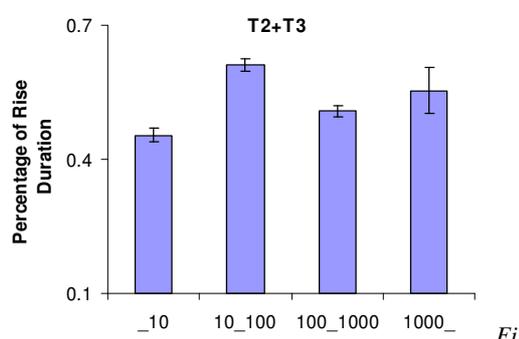
Fi gure 6: Means (and ± two standard errors) of the percentage of rise of the Rising tone within different frequency ranges.

3.3. Constituent Boundary

For tri-syllabic constituents, as shown in Figures 7-8, there was a significant interaction of syllable position and the structure of the constituent on the LogRange of F_0 rise [F (2, 7478) = 6.8, p = .001] as well as on the percentage of F_0 rise duration [F (2, 7478) = 38.5, p < .0001]. We examined the two syllables separately.

For the 1st syllable, there was a significant effect of constituent structure [F (2, 3739) = 3.6, p = .028] on the LogRange F_0 rise. Bonferroni Post-hoc tests showed that the only the 1+2 structure and the 1+1 structure differed significantly. Constituent structure also affected the percentage of F_0 rise duration significantly [F (2, 3739) = 10.6, p < .0001]. Bonferroni Post-hoc tests showed that the 1+2 structure differed significantly from both the 1+1+1 and 2+1 structure.

For the 2nd syllable, there was a significant effect of constituent structure [F (2, 3739) = 15.9, p < .0001] on the LogRange of F_0 rise. Bonferroni Post-hoc tests showed that only the 2+1 structure differed significantly from the other two. Significant effect of constituent structure was also found on the percentage of F_0 rise duration [F (2, 3739) = 38.7, p < .0001]. Bonferroni Post-hoc tests showed that only the 2+1 structure differed significantly from the other two structures.

These results are rather puzzling. There are three possibilities with regard to word boundaries in our data: the Pattern 2+1 has a word boundary between S2 and S3, as shown in Fig. 9a. The pattern 1+2 has a word boundary between S1 and S2, as shown in Fig. 9b. The Pattern 1+1+1

has two word boundaries, as shown in Figure 9c. We may predict that for S1, its acoustic realization should show similar patterns in Fig. 9b and Fig. 9c as in both cases, S1 is followed by a word boundary. For S2, its acoustic realization should show similar patterns in Fig. 9a vs. Fig. 9c, as in both cases, S2 is followed by a word boundary. Furthermore, the pattern in Fig. 9a should be different from that in Fig. 9b. Contrary to our predictions, however, for the first syllable, both the LogRange F_0 rise and the percentage of rise duration show that there were significant difference between the pattern 1+2 and the pattern 1+1+1. As for the second syllable, the pattern 1+2 and 2+1 were different, as expected. But surprisingly, 1+2 showed similar pattern to 1+1+1, again contrary to our expectation. The data thus seem to suggest that while word boundary does matter in structures as Fig. 9a and 9b, where word boundary affected the application of the 3rd tone sandhi, word boundary in Fig. 9c does not have the same effect. Rather, it seems that all three mono-syllabic words were grouped into one tonal domain where both the 1st and the 2nd syllable had sandhi changes, as the 1st syllable in 2+1 and the second one in 1+2 respectively.
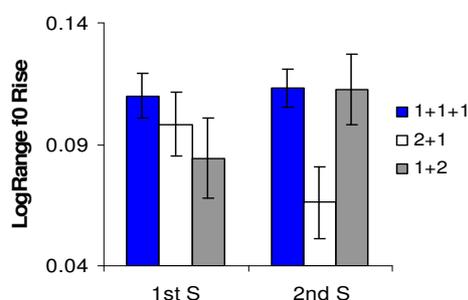
Figure 7: Means (and ± two standard errors) of the log range of F_0 rise in the first and second syllables within tri-syllabic constituents of three different constituent structures.
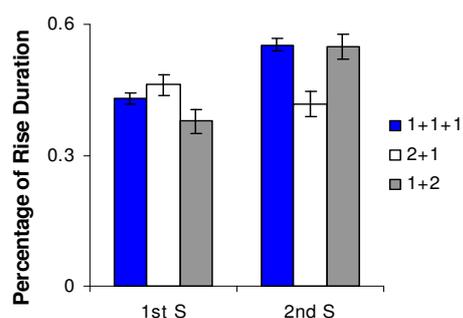
Figure 8: Means (and ± two standard errors) of the percentage of rising time over the first or second syllables within tri-syllabic constituents of three different constituent structures.
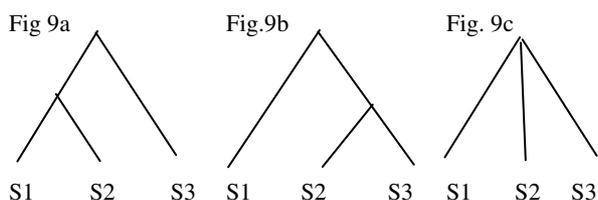
Figure 9: Possible higher-level structural patterns of tri-syllabic constituents.

4. Conclusions

This paper is the first, to our knowledge, that examines the 3rd tone sandhi phenomenon in a large corpus with naturally occurring conversational data. Our results confirm previous reports that there are indeed low-level acoustic differences between an underlying Rising tone and a Sandhi Rising tone both in terms of the magnitude of F_0 rise as well as the rise duration. We also found that given a bi-syllabic word, a 3rd tone sandhi domain, the frequency of a word has an effect on the realization of the Sandhi Rising tone although such an effect was significant only in highly frequent words. Specifically, Sandhi Rising tones in highly frequent words showed less F_0 rise than less frequent words. Such an effect, however, was not observed on the underlying Rising tone. This suggests that the difference between a Rising and a Sandhi Rising tone is manifested at higher level of speech production. Last, we observed interesting yet puzzling data with regard to the role of constituent boundary on the realization of the Low tone. When a tri-syllabic constituent contains a bi-syllabic word, we observed an effect of word boundary. When a tri-syllabic constituent contains three independent words, however, no effect of word boundary was observed. Further studies, with detailed information on the syntactic and prosodic grouping of these constituents, are needed to confirm and to understand this pattern.

5. Acknowledgements

6. References

[1] Chen, M. 2000. *Tone Sandhi*. Cambridge University Press. Cambridge.

[2] Chen, Y. 2003. *The phonetics and phonology of contrastive focus in Standard Chinese*. PhD dissertation. Stony Brook University.

[3] Chen, Y. 2004. Focus and Intonational Phrase Boundary in Standard Chinese. 2004. *International Symposium on Chinese Spoken Language Processing*. Hong Kong.

[4] Kuo, Y., Xu, Y., & Yip. M. (to appear). The Phonetics and phonology of apparent cases of iterative tonal changes in Standard Chinese. Gussenhoven, C., & T. Riad (eds.), Proceedings of Tone and Intonation.

[5] Pellom, B., Sonic. 2001. The University of Colorado Continuous Speech Recognizer, tech report # TR-CSLR-2001-01, Boulder, Colorado.

[6] Peng, Shu-Hui. 2000. Lexical versus 'phonological' representations of Mandarin Sandhi Tones. In M.Broe & J. Pierrehumbert (eds.), *Papers in laboratory phonology 5: acquisition and the lexicon*, pp. 152-167. Cambridge: Cambridge University Press.

[7] Shih, Chilin. 1986. The Prosodic Domain of Tone Sandhi in Chinese. PhD dissertation. University of California at San Diego.

[8] Shih, Chinlin. 1997. Mandarin third tone sandhi and prosodic structure. In J. Wang & N. Smith (eds.), *Studies in Chinese Phonology*, pp. 81-124. Dordrecht: Foris.

[9] Speer, S, R., Shih, C.-L., & Slowiaczek, M.L. 1989. Prosodic structure in language comprehension: Evidence from tone sandhi in Mandarin. *Language and Speech, 32*, 337-354.

[10] Zhang, Z. S. 1988. Tone and Tone Sandhi in Chinese. PhD dissertation. Ohio State University.