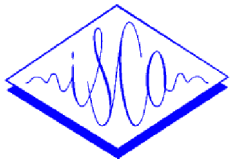


Speech Prosody 2008
Campinas, Brazil
May 6-9, 2008

Revisiting the Phonetics and Phonology of Shanghai Tone Sandhi

Yiya Chen

Center of Lingusitics
Leiden University, Leiden
y.chen.2@umail.leidenuniv.nl

ABSTRACT

This study revisits the phonetics and phonology of Shanghai tone sandhi by examining the f0 contours of non-initial syllables within sandhi domains which start with different lexical tones. Results show significant f0 variation due to the initial lexical tones; the variation, however, diminishes as the number of non-initial syllables increases, resulting in near convergence of f0 values by the end of the 3rd syllable. This suggests the existence of a low tone target for non-initial syllables, the phonetic implementation of which is weak and remarkably comparable to the neutral tone in Standard Chinese [1]. Shanghai Chinese thus suggests the possible existence of weak-strong tonal contrast, like the neutral vs. lexical tonal contrast in Standard Chinese, which manifests at a prosodic level higher than syllable.

10.21437/SpeechProsody.2008-55

1. INTRODUCTION

Shanghai Chinese is a syllabic language with lexical tones. When syllables are combined into words or phrases, lexical tones undergo sandhi changes ([2, 5, 6, 7, 9], among others). The general consensus is that given a sandhi domain (e.g., prosodic word in [6]), the canonical tone of the word-initial syllable determines, to a large extent, the f0 trajectory of the whole tonal domain. There is, however, much less consensus in the phonological representation as well as the phonetic implementation of the sandhi tones.

With regard to the first non-initial syllable, one possibility is that it forms a tone-bearing unit with the initial syllable and both syllables are associated with the lexical tone of the initial syllable [5]. The other possibility is that it is associated with a tone that is spread from one of the level tones decomposed from the tonal contour of the initial lexical tone [2, 6, 7, 9]. With regard to the later syllables within a larger tonal domain, the general consensus is that they are associated with low tone(s). Zee & Maddieson [7] further posit different phonetic tones as well as a number of very complex rules to explain the f0 variation of the low tone(s) over the time course of the domain. Note the only phonetic study on the f0 of poly-syllabic domains in Shanghai Chinese (i.e. [7]) was based on one speaker and the f0 trajectories were obtained by averaging f0 values over long temporal intervals, which made it difficult to examine the precise alignment of the peaks and valleys of the f0 trajectories

This study aimed to extend previous works by investigating in more detail the f0 trajectories of tonal domains. Two patterns were of particular interest. First, the effect of domain-initial lexical tones on the f0 contour of the first non-initial syllable, aiming to shed new light on the existing debate. Second, the f0 realization of the later non-initial syllables, particularly with regard to the nature of the phonetic modification of their low tone(s) as proposed in [7], and how to properly characterize such changes over the time course of the domain.

2. METHODOLOGY

2.1. Test material

The test materials (1) varied in 1) the tone of the initial syllable ***X*** (in bold), where all lexical tones in Shanghai Chinese were included (Tone1: falling; Tone2: high rising; Tone3: low rising; Tone4: short high; and Tone5: short low rising); 2) the number of within-domain non-initial syllables *Y*, ranging from one to three (which all have the lexical Tone3); and 3) the focus condition for which utterances were elicited as answers to two different questions. In the focused condition, the question was on the target noun phrase ($XY_{1\text{-}3}$). In the pre-focus condition, the question was on the phrase that follows the target noun phrase. In total, there were 15 sentences and 30 different renditions of these sentences (5 (X)*3(Y)*2 (focus condition) = 30 renditions). The template sentence started with the tonal domain (underlined), and was followed by *ganghao*, meaning 'just', and then a locative NP.

$\underline{X\ Y_1 Y_2 Y_3}$
(1) **Ga** li mi mei ganghao lala Beijing.
Galimimei just locative Beijing
'Galimimeimao is right in Beijing.'

2.2. Subjects

Six native speakers of Shanghai Chinese, one male and five female, participated in the experiment. All were born in Shanghai, between 1935 and 1950, and

had lived most of their lives mainly in the Xuhui District, the center of the metropolitan city.

2.3. Data collection

The stimuli were presented in randomized order. Each subject produced the test materials in a different order at each repetition. The program presented one question-answer pair at a time. The speakers were asked to utter the pair at their normal speaking rate and as naturally as possible. During the recording, they would be asked to repeat the pair whenever a mistake was made. This included improper intonation, as judged by the author, a native speaker of a closely related Wu dialect. The recordings were conducted in a sound-treated booth at Shanghai Normal University. Two repetitions of the same stimulus in the same focus context were elicited from each speaker.

2.4. F0 analyses

Segment boundaries of the tonal domains (ranging from two to four syllables) were manually labeled within Praat. F0 information was also obtained in Praat, first the smoothed f0 curves with Praat's default setting, which were then manually corrected for f0 tracking errors such as missing of pulse marking or double tagging.

For graphical comparison, 20 f0 points were taken at equal intervals of the syllables (for initial syllables, only the rhyme was included because these syllables differed greatly in their consonant onsets). Each data point was transformed into z-score via Lobanov's (1971) *z*-score metric ($z = (F - \mu)/\sigma$). The normalized f0 value (z) was calculated as the difference between the raw f0 in Hertz (F) and the grand mean f0 of each subject in Hertz (μ), divided by the standard deviation of the overall f0 of the same speaker (σ). Similar procedures can be found in [4, 8].

The f0 curves were also time-normalized. First, an average duration was computed of the syllables in each position in sentences which contained tonal domains that had the same number of syllables. In other words, the average durations of Y_1 in bi-, tri-, and quadric-syllabic tonal domains ($\mathbf{X}Y_1$, $\mathbf{X}Y_1Y_2$, $\mathbf{X}Y_1Y_2Y_3$) were all computed separately. These averaged durations were then used as the time-axis for displaying the f0 contour of each corresponding syllable position in the sentence. Tonal contours of different sentences across speakers could then be compared directly.

3. RESULTS

Results showed a general pattern (which included about 90% of the total data), and some variations. I will first focus on the general pattern (§3.1) and then discuss variations on the theme in §3.2.

3.1. General pattern

Figures 1a-c show the time- and speaker- normalized f0 contours of tonal domains produced with focus. All domains start with one of the five lexical tones, and are followed by a string of non-initial syllables ranging from one to three, all of which have the underlying lexical Tone3 (low rising). These tonal domains are then followed by a falling tone (which is not shown in the figures). Here the size of the tonal domain increases, by adding one more non-initial syllable in each figure (bi-syllabic in Fig. 1a, tri-syllabic in Fig. 1b, and four-syllable in Fig. 1c). Syllable boundaries are indicated with space in the f0 contours.

Three things are to be noted. First, the tone of the initial syllable shows a significant influence on the f0 contour of the domain. This is particularly clear in the first two syllables, where there are five overall rather distinguishable f0 contours.

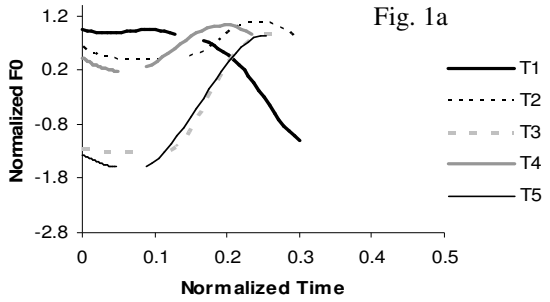
Second, the effect of the preceding lexical tones on the non-initial syllables diminishes over the time course of the domain. Specifically, starting from the second non-initial syllable (Fig. 1b-1c), all f0 contours after the five lexical tones are clearly falling, showing near convergence by the end of the 3rd non-initial syllable (Fig. 1c).

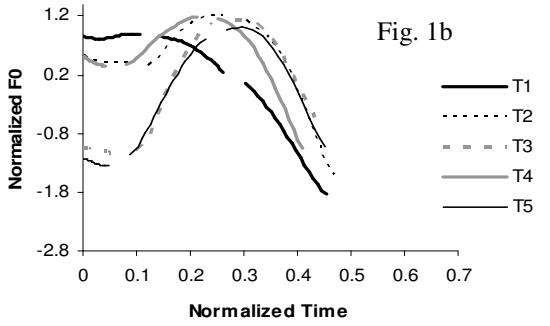
The impressionistic observation was further confirmed by results of ANOVA with the endF0 value of the within-domain non-initial syllables as the dependent variable. Lexical Tone (of the initial syllable), Length (of the tonal domain), and Focus Condition were fixed factors, and Subject and Repetition were treated as random factors. There was a significant effect of Lexical Tone on the endF0 of the non-initial syllables throughout the domain (1st σ: [$F(4, 450) = 72.3$, $p < .0001$]; 2nd σ: [$F(4, 320) = 61.4$, $p < .0001$]; 3rd σ: [$F(4, 201) = 5.6$, $p < .0001$]). Bonferroni post-hoc tests showed that for the 1st non-initial syllable, its endF0 value following the five lexical tones differed significantly from each other across all lengths. For the 3rd non-initial syllables, however, there was only significant difference of Tone1 and Tone5 from Tone2, Tone3, and Tone4. The magnitude of differences was negligible.

A third point to note is that while f0 lowering was salient in the 2nd and 3rd non-initial syllables, in a bi-syllabic tonal domain, f0 lowering, though a much

more subtle pattern, can be observed near the end of the first non-initial syllable (Fig. 1a). When the domain becomes longer, lowering may not start until the 2nd non-initial syllable, particularly in the case of the f0 that follows the initial lexical Tone3 and Tone5 (low rising) (Fig. 1b-c).

We thus further examined the f0 peak alignment, a good indicator of the start of f0 lowering, over the first two syllables across domains of different lengths. Here, we calculated the relative distance of the f0 peak to the syllable edge (i.e. 1st syllable for Tone 1 and 2nd syllable for the other tones) as a percentage of the syllable duration. ANOVA was run on this variable with Lexical Tone (of the initial syllable), Length (of the domain), and Focus Condition as fixed factors. Again, Subject and Repetition were treated as random factors. There was a significant effect of Lexical Tone [$F(4, 450) = 1026$, $p < .0001$], Length of the domain [$F(3, 450) = 73.4$, $p < 0.001$], as well as significant interaction of Tone*Length [$F(12, 450) = 3.9$, $p < .0001$]. This suggests that f0 lowering was dependent on the identity of the initial lexical tone and further regulated by the length of the domain. In particular, we noted that for the bi-syllabic domains (Fig.1a), f0 started falling before the end of the 2nd syllable and pairwise comparisons showed significant differences among all tones (from the right edge of the 1st σ: 11% after Tone1; From the right edge of the 2nd σ: 37% after Tone2; 15% after Tone3; 30% after Tone4; and 10% after Tone5).





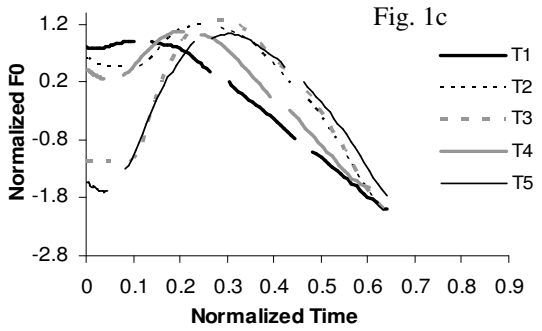


Figure 1a-c: ***F0 contours of tonal domains starting with five lexical tones and followed by falling tone (bi-syllabic in 1a; tri-syllabic in 1b; four-syllable in 1c), produced with focus.***

3.2. Variation

There are three ways in which the data varied, all of them observed in three-syllable or four-syllable tonal domains. One is due to one single speaker's idiosyncratic rise near the end of the last syllable within a tonal domain, regardless of the domain-initial lexical tone. The second concerns initial Tone5 only (short low rising) and was observed over all speakers as free variation. While the most typical pattern for domains with initial Tone5 is shown in Fig. 1a-c, the variant shows a plateau of low f0 which lasts throughout the domain till the last syllable during which f0 rises. A third pattern of variation was produced by the three youngest speakers only over four-syllable domains regardless of the initial lexical tone. As shown in Fig. 2, these speakers sometimes produced a sharp falling f0 over the 3rd syllable of the domain (the left f0 contour), as a contrast to the more general pattern, where there is a gradual lowering of the f0 contour (the right f0 contour in the figure).

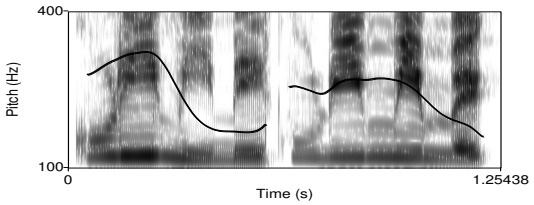


Figure 2: ***Illustration of sharp f0 falling (left) in the 3rd syllable, as compared to the more general pattern of gradual f0 lowering (right).***

4. DISCUSSION AND CONCLUSION

This study re-visited the phonetics and phonology of Shanghai tone sandhi by examining in detail the f0 trajectories of the non-initial syllables within tone sandhi domains which varied in both the domain-

initial lexical tones and the number of non-initial syllables within the tonal domain. Results confirmed, to a large extent, the observations of the existing literature but also revealed patterns of f0 trajectories which are not readily explained by the previous analyses of tone sandhi in Shanghai Chinese. Specifically, we observed a significant effect of the domain-initial lexical tones on the f0 contours of the non-initial syllables. Such influence exhibited two interesting characteristics which require some revisions of the previous proposals. First, the influence lasted, at least statistically, throughout the whole domain. While the most salient effect was manifested over the first non-initial syllable, the effect was persistent even in the 3rd non-initial syllable, though with a negligible magnitude. Second, a clear f0 lowering pattern was observed over larger domains. This has been repeated reported in the literature. The new finding here is that the lowering started well before the end of the first non-initial syllable. This suggests the possibility that a low tone target may have been implemented even over the 1st non-initial syllable, though not very effectively. Clear manifestation of the low tone requires longer domains as shown in the gradual convergence of the f0 lowering through Fig. 1a-c. Such a sluggish realization of non-initial low tone reveals a remarkable similarity with the f0 contour of the neutral tone in Standard Chinese ([1]). In both cases, they show great influence from their preceding lexical tones and both take more than one tone-bearing syllable to manifest their own f0 realization.

If this interpretation is indeed on the right track, then the account of Shanghai tone sandhi can be greatly simplified as the following. Given a tonal domain in Shanghai Chinese, the initial syllable maintains its underlying lexical tone, but the tonal contrasts over a non-initial syllable neutralize to a low tone, comparable to the neutral tone in Standard Chinese. This low tone is weak, compared to the underlying lexical tones which are strong, in that it allows more influence from the preceding lexical tone and it takes longer time to attain its ideal low tone target. This analysis attributes f0 variation over the non-initial syllables to the weak implementation of a low tone due to the neutralization of their underlying lexical tones. The strong-weak contrast is similar to the lexical full/neutral tone contrast in Standard Chinese, but Shanghai Chinese suggests a case of strong-weak contrast at a prosodic level higher than syllable. One question that arises is whether we should attribute the weak-strong difference as an intrinsic specification of a tone target. Or the way a tone is implemented as a weak or strong one should be attributed to the prosodic position of the tone-bearing syllable only. We will leave this topic for future research.

Further studies are also needed to understand the tonal variation (§3.2). With regard to the pattern in Fig. 2, a possibly relevant observation is that the younger generation (below 30) now breaks a long domain into bi-/tri- syllabic tonal domains ([3]). Given a four-syllable domain, they may produce f0 contours similar to that in Fig. 2. This suggests that they may have reinterpreted the weak low tone as a strong one given its domain-initial status. It is likely that my youngest speakers of the older generation (who are above 50) have been influenced by the younger generation and have implemented the weak tone over the 3rd syllable as a strong one although their sandhi domain was maintained. If so, this provides an interesting case where strong vs. weak contrast in tonal implementation can be a source of change in grammars across generations.

5. ACKNOWLEDGEMENT

I thank Carlos Gussenhoven and Bob Ladd for valuable comments on an earlier version of this paper and, of course, all my speakers for their participation. Support by a VENI grant from the Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged.

6. REFERENCES

[1] Chen, Y. & Xu, Y. 2006. Production of Weak Elements in Speech: Evidence from Neutral Tone in Standard Chinese. Phonetica 63: 47-75.

[2] Duanmu, S. 1993. Rime Length, Stress, and Association Domains. *Journal of East Asian Linguistics* 2, 1-44.

[3] Gu, Q. 2004. *The investigation and analysis of the phonetic evolution of the newest variety of Shanghai urban dialect.* MA Thesis, Shanghai Normal University.

[4] Rose, P. 1987. Some considerations in the normalization of the fundamental frequency and linguistic tone. *Speech Communication* 6:4, 343-52.

[5] Sherard, M. 1972. *Shanghai Phonology*. PhD thesis. Cornell University.

[6] Selkirk, E. & Shen, T. 1990. Prosodic domains in Shanghai Chinese. In: Inkelas, S., Zec, D. (eds.), *The phonology-Syntax Connection*. University of Chicago Press: 313-338.

[7] Zee, E., & Maddieson, I. 1980. Tones and Tone sandhi in Shanghai: Phonetic evidence and phonological Analysis. *Glossa*, 14:1, 45-88.

[8] Zhu, X. 2003. *An Experimental Study in Shanghai Tones*. Shanghai: Shanghai Educational Publishing House.

[9] Yip, M. 1980. *The Tonal Phonology of Chinese*. PhD dissertation. MIT