THE MICROPROSODICS OF TONE SANDHI IN SHANGHAI DISYLLABIC COMPOUNDS

Xiaonong Sean Zhu

Linguistics Department, Faculties, Australian National University

ABSTRACT

This paper examines the F0 variations during tone sandhi due to various prosodic factors such as phonation type, length, stress and pitch height. It will be shown that the F0 height and shape of the second syllable (S2) in disyllabic words are determined by the interaction of four conditions: the intervocalic consonant (C2) voicing, the S2 Truncation, the F0 height of S1, and stress assignment.

1. INTRODUCTION

Lots of work has been done during the last two decades on general F0/pitch in Wu, which is regarded as the most complicated in the world's tonal languages, but nothing on the details of what factors actually condition realisation of F0 even in Shanghai, the lingua franca of the Wu dialects whose tone sandhi is probably the least complex within the Wu group. Using disyllabic words without C2, Zhu 1998 demonstrates smoothly flowing F0 contours for Shanghai tonal rightward spreading in both F0 and Contour element terms. These can be taken to reflect the underlying tension of the vocal cords without influence from supraglottal effects and represent a baseline in terms of which the perturbations observed on items with C2 can be understood.

Shanghai has five citation tones (T1 [51/541] /hl, Upper, Long/, T2 [34] /lh, Upper, Long/, T3 [14/113] /lh, Lower, Long/, T4 [44], /lh, Upper, Short/, and T5 [14/113], /lh, Lower, Short/), three of which have two acoustic allotones each, and five ditone classes (T1+X [55+31], T2+X [33+44], T3+X [22+44], T4+X [33+44], and T5+X [11+14]), defined by the S1 tone. While only T1 is falling or Lo-targeted, the remaining four are (underlyingly) rising or Hi-targeted. The first three ditone classes are Long tone domains, and the remaining two Short tone domains. The first four classes have an [s w] stress pattern, while the last one (T5+X) [w s] (Zhu 1998). There are altogether 25 (=5*5) possible ditone combinations. T1 and T2 are neutralised on S2 (marked as T0), so there are eventually 20 combinations. Each of the five ditone classes has four F0 combinations. The first class (T1+X), with short and sharp falling F0 on S2, are out of question since there is little F0 perturbation on S2 (Fig. 5, Zhu 1998). It is the fairly large second syllable F0 variations of the remaining 16 combinations of the 4 classes with a high offset target that we will deal with in this paper.

This paper reports and discusses the results from an experiment with Shanghai speakers reading fifteen monosyllables and twenty-five disyllabic compounds with C2 under different tonal combinations. It will be demonstrated that the offsets of all the sixteen S2 curves under investigation are congested to a very narrow F0 range which is effectively the same as that of the Hi-target citation tones. So it can be assumed that the offset targets of the S1 input tones spread to S2 during sandhi. The current various F0 shapes of S2 are caused by the factors that bend or raise the onset F0 of S2 and thus shape the whole S2 contours. I will then show that the factors involved in shaping S2 curves are C2 voicing, S2 Truncation, F0 height of S1, and stress. The interaction of these factors results in five degrees of F0 height over the early part of S2, which, in turn, produces five categories in terms of the F0 shapes of S2.

2. PROCEDURE AND RESULTS

Used in the experiment were 15 monosyllables, with three for each tone, and 25 disyllabic compounds for all 25 ditone combinations. As long as I could, I chose syllables with explosive consonants and kept a balance between high and low vowels, which are expected to intrinsically influence F0 height. The testing syllables were written in Chinese characters, mixed with dummy ones, on several pieces of A4-size paper. All the speakers carefully read the word list six times with a short break after each time.

In the recording of monosyllables, five female and six male speakers were used. In the recording of disyllables, three males and a female were used. The monosyllables and disyllable are recorded in separate sessions. All recordings were conducted in a phonetics lab at ANU with professional equipment. Acoustic information was obtained by analogue instrumentation. The measurement procedures used to extract F0 were basically that described in Rose (1990). Tokens were sampled every one third and 20 percent of the duration for Short and Long tones respectively. For the Long tones an additional point at 5% and 10% of the duration was also measured for citation and for that in ditones, respectively, except that after T1 which was sampled every one fourth of the duration.

The F0 data in Hz collected from the experiments were converted to Logarithmic Z-score (LZ) normalised values for each speakers before averaged across speakers. Figure 1 plots the LZ values for both citation and ditones, exclusive of T1(+X), against equalised duration. The ditone data are arranged according to 'same S1 input tones + different S2 input tones' in each panel. The same four input tones on S1 in each panel have been averaged due to their phonetic similarity and phonological non-distinctiveness. '2[]', therefore, is the average of four T2 curves on S1: before T0, T3, T4 and T5. Figure 2 plots the same ditones in another way: different S1 input tones + same S2 input tones. Each panel is divided into two parts by the second '0 ms' line-up, which designates the F0 onset of S2 contours. Note that we are dealing with left-dominant sandhi tones in which the input tone on S2 loses its original contour and receives tonal features from S1. So even the same input tone may have different F0 shapes after different tones. That is why the same S2 input tone is worth comparing since different S1 tones may give rise to different F0 curves for the same input tone on

S2. Otherwise comparing the 'same S2 tone' would sound superfluous or nonsense.

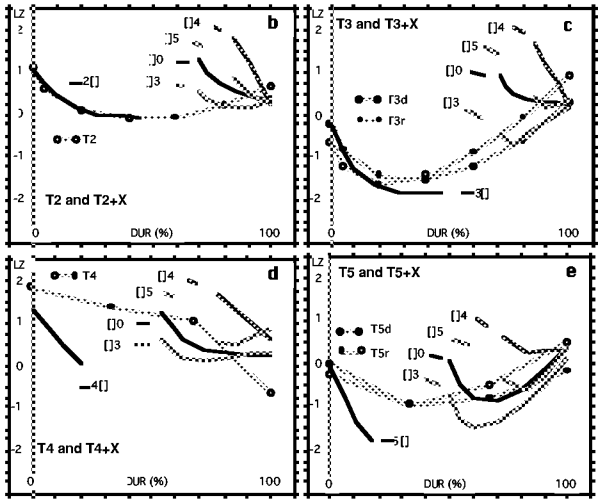

Figure 1: LZ normalised ditones and corresponding citation tones with Hi-target plotted against equalised duration arranged in 'same S1 tone + different S2 tones'.

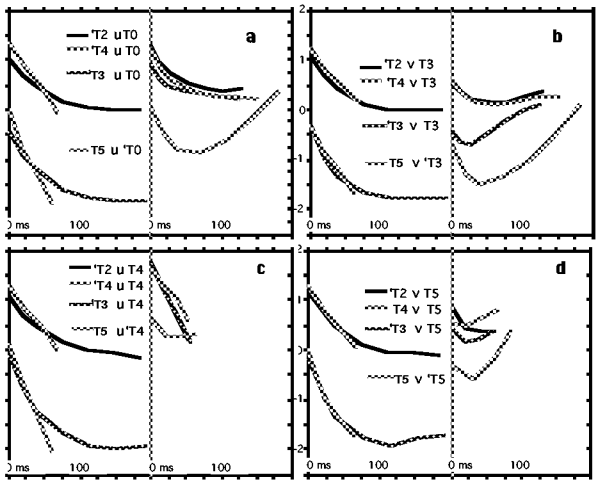

Figure 2: Second-syllable F0 contours of T0 (a), T3 (b), T4 (c), and T5 (d) after Hi-target tones arranged in 'different S1 tones + same S2 tone'. '' indicates stress. Italic and boldface respectively stand for Upper and Lower Register; u and v for unvoiced and voiced C2.

3. DISCUSSION

Zhu 1998 shows clearly that the smoothly flowing ditone curves without C2 have effectively the same onset and offset targets (with a little F0 declination) as that of citation tones. In the case of disyllables with C2, the pictures are more complicated. Yet we still see the same onset and offset targets for T2/3/5(+X) in Figure 1 and for T1+X in Figure 5 in Zhu 1998. T4 here looks exceptional; I will return to this later.

Figure 1 shows that the four S2 curves in each panel differ from each other to a fairly large extent. When they are grouped according to 'the same S2 input tones' in figure 2, some similarities show up, but great differences still exist.

There are, however, two regularities presented in the figures. First we can see a striking between-panel similarity that all the S2 curves offset along a very narrow range of 0.68 standard deviation (between 0.13 and 0.81 SD) with an average of 0.33 SD. There are two offsets, which deviate the average a little farther. One is T5 after T4 (Fig. 2d) which offsets at 0.81 SD. The other is T4, again, after T4 (2c), which offsets at 0.58 SD. If they are viewed as exceptions and excluded, the offset range will be reduced to a 'could-not-be narrower' degree of the order of a quarter SD (between 0.13 and 0.40 SD). We can assume with confidence that all the four S2 input tones (T0 to T5) have the same offset target and that this is the result of rightward spreading, i.e. the S2 receives the spreaded offset target from S1.

Second, although the F0 configurations in each panel of figure 1 differ, there is a constant rank order in terms of the F0 height over the early part of S2. That is, on S2, T4 is always the highest in each panel, T3 always the lowest, and T0 and T5 are intermediate. Since in each panel, all the four S2 contours follow the same input S1 tone, e.g. all the four S1 contours in 1b are T2, this constant rank order cannot be conditioned by factors outside of S2. Two such factors are readily at hand: S2 Truncation and C2 voicing. The F0 over the early part of S2 curves is higher for a Short than for a Long tone, and also higher with a voiceless C2 than with a voiced.

Comparing the S2 curves after T2 (1b) and that after T4 (1d) are very similar, we find they are similar. Since the S1 contours in the two panels differ only in Truncation and it does not affect the F0 configuration of S2, the S1 Truncation does not matter in conditioning the F0 shape of S2. Comparing those after T2 (1b) and after T3 (1c), we find they are also similar except that T3 after T3 (1c) differs markedly from the same T3 after T2 in (1b). Since both S1 curves in the two panels differ only in Register, we can assume it is the Lower Register of S1 that makes T3 contours different after T3 than after T2. Comparing those in 1d and 1e, we find huge differences emerging. The four S2 curves after T4 (1d) lie higher, in a narrower range, with shorter duration. Apart from the S1 Register, those in 1d being Upper while those in 1e Lower, a more important difference in the two panels is the stress patterns: [s w] in 1d while [w s] in 1e. The S2 curves in panels 1c and 1e also differ. This time however the S1 Register is the same Lower in both 1c and 1e. The differences are S1 is Long in 1c but Short in 1e; stress pattern is [s w] in 1c but [w s] in 1e. Combining these observations, we can further assume stress may be a fourth condition on the F0 shape of S2.

Figure 2 shows that the S2 curves in the [s w] stress (T2 to T4+X, i.e. the top three S2 contours in each of the four panels) are different from those in the [w s] pattern (T5+X, i.e. the bottom contour in each panel). The strong S2 curves after T5 are either fully rising or level while the weak S2 curves are basically level in 2a, 2b, and 2d, and falling in 2c. There is only one weak S2 curve (the T3 after T3 in 2b) which looks a little like the strong T3 after T5 in the same panel. Even so, compared with the strong T3 after T5, the weak T3 is much shorter (123 : 193 ms), inflects much higher (-0.68 : -1.46 SD), and does not rise as much (0.81 : 1.60 SD). Such shape differences can only be accounted for in stress patterns.

Next to note are the weak S2 contours in 2a and 2b. Not only do these contours have the same offset target but also similar F0 shape, if we put aside the T3 after T3 for a while. The slightly different shapes over the early part of the three T0 curves (2a) and the top two T3 curves (2b) are accounted for by the onset perturbation. The voiceless C2 before T0 curves raises the onset and makes the main bodies of these contours slightly falling. The voiced C2 before the top two T3 curves pushes down the onset and makes the main body slightly dipping. The onset of T3 after T3, because of the lower F0 of S1, is pulled down further and makes the whole contour rising. So the F0 of the two Long input tones 0 and 3 after strong T2 to T4, although very similar, are subject to C2 voicing (which induces minor onset differences) and F0 height of the Lower S1.

The S2 curves of T4 and T5 in the lower two panels are shorter than those in the top panels due to [+truncated]. The falling shape of T4 in 2c is attributable to 1) the voiceless C2 which raises the onset, 2) the final glottal stop which induces a dropping tail (Zhu 1995, §7) and 3) the Short Truncation which 'swallows' the main body which is supposed to be (underlyingly) rising. It does show a rising curve when the syllable has a zero-*initial* (Zhu 1998). The F0 shape of T5 after T2 in 2d, like T0 on S2 in 2a, is accounted for by the voiceless C2 that raises the onset and makes the main body slightly falling. The F0 shape of T5 after T3, like T3 on S2 in 2b, is caused by the voiced C2, which depresses the onset and makes the main body slightly rising.

In sum we can assume that the weak S2 curves after T2 to T4 have the same contour target and the current various shapes in Fig.2 are accounted for mainly by S2 Truncation and C2 voicing, and reinforced by the F0 height of S1 in cases of T3+T3/T5. Also we can assume that all the sixteen S2 curves have the same offset target, the current shape difference between weak (after T2 to T4) and strong (after T5) S2 contours are attributable to stress.

It is shown in Zhu (1995, §9.4) that all the Hi-target citation tones 2 to 5 have the same tonetic as well as tonological offset target above the overall mean of 0 SD. Now we see the S2 contours after these tones have the same offset above the mean 0 SD. So we can assume the offset target of the Hi-target citation T2 to T5 is realised on the S2 contours during ditone sandhi. In other words, the offset target of the citation tones rightward spreads to the offset of S2 tones.

Below I will further explore how the above-identified four factors (S2 Truncation, C2 voicing, S1 Register and stress) interact to result in ten F0 shapes from the original sixteen S2 curves, and more interestingly, five categories of equal onset height. Since the offsets of these contours are the same, the variation is in the F0 height over the early part of duration and thus the overall F0 shape.

S2 Truncation and C2 voicing. Let us go back to Fig.1 to do some within-panel comparisons. In general, the early part of T4 on S2 is the highest in each panel and T3 the lowest over at least the first half of the duration. The remaining two, namely T0 and T5, lie at effectively the same F0 height and have approximately the same shape in each of panels. Thus we need two rules (R1 and R2 in table 1) to account for these differences in F0 height.

Lower-Register S1 before a voiced C2. We have shown that S1 Register does not influence the F0 height of S2 if the C2 is voiceless. In other words, Lower S1, when before a voiceless C2, has the same effect as Upper S1 has on S2 F0. This can be seen in Fig.2a. The three T0 curves on weak S2 are virtually the same after Lower T3 as after Upper T2 and T4 since C2 is voiceless. So are the three weak T4 curves (2c). On the other hand, S1 Register influences S2 F0 when C2 is voiced. This is shown in 2b and 2d. The second-syllable T3 and T5 are lower after Lower T3 than after Upper T2 and T4. So we need the third rule in table 1.

Stress. We have also shown that stress pattern influences the F0 height of S2. This is clear in all panels of Figure 2. The lowest curve in each panel is the one after T5. In 2b, the above-mentioned three conditions are the same for T3+T3 and T5+T3: Long S2, voiced C2, and Lower S1. However, T3 is still lower after T5 than after T3; this is because the stress assignment is different for the combinations. So is the second syllable T5 in 2d. Therefore we need the fourth rule.

R1	A Long contour is lower than a Short one by one degree.
R2	A contour is lower with a voiced C2 than with a voiceless C2 by one degree.
R3	A contour with a voiced C2 is lower after a Lower S1 than after an Upper S1 by one degree.
R4	A contour is lower on a strong syllable than on a weak one by one degree.

Table 1: Four realisation rules for S2 input tones.

With these rules we can account for the F0 variations over the early part of all sixteen Hi-target S2 curves in Figure 2. First, some contours can be grouped since no rules are applicable. That is, the three T0 curves after T2 to T4 (2a) can be grouped together. Also the three T4 curves after T2 to T4 (2c), the two T3 curves after T2 and T4 (2b), and the two T5 curves after T2 and T4 (2d). After the grouping, we have ten F0 shapes, i.e. 1) the average T4 after T2 to T4, 2) the average T0 after T2 to T4, 3) the average T5 after T2 and T4, 4) T4 after T5, 5) the average T3 after T2 and T4, 6) T5 after T3, 7) T0 after T5, 8) T5 after T5, 9) T3 after T3, and 10) T3 after T5.

When the realisation rules are applied, some of the ten shapes will have the same F0 height over the early part of S2 since some conditions will offset each other. So we will virtually have five categories of F0 height of S2 contours in table 2. Category 1 is higher than category 2 by one degree, than category 3 by two degrees, and so on so forth.

Category 1:	1) ['234 u] T4		
Category 2:	2) ['234 u] T0	3) ['24 v] T5	4) [5 u] 'T4
Category 3:	5) ['24 v] T3	6) ['3 v] T5	
Category 4:	7) [5 u] 'T0	8) [5 v] 'T5	9) ['3 v] T3
Category 5:	10) [5 v] 'T3		

Table 2: The sixteen S2 curves are grouped into ten shapes, and further into five categories in terms of the F0 height over the early part of S2 curves.

The two categories at the extremities have one curve each. In category 1 is the average T4 curve on a weak syllable after T2 to T4 (shape 1). This should be the highest over the early part among all ten shapes, because there is no rule that

lowers this weak, Short T4 with a voiceless C2. In category 5 is the strong T3 after a weak T5 and voiced C2 (shape 10). This should be the lowest among the ten, because all the four rules are applicable to it, therefore it should be four degrees lower than the average T4 in category 1.

In category 2 are three shapes which have equal F0 height over the early part and are one degree lower than category 1. Let us examine them one by one. First, according to R1 (Long lower than Short), the average Long T0 after T2 to T4 (shape 2) should be one degree lower than the average Short T4 (shape 1) since other conditions for both shapes are the same. Second, according to R2 (voiced C2 lower than voiceless C2) the average T5 (shape 3) with a voiced C2 should also be one degree lower than the T4 in category 1. So T5 (shape 3) should equal T0 (shape 2). Third, according to R4, the strong T4 after T5 (shape 4) should be one degree lower than the weak T4 in category 1. Compared with shape 3 in the same category, the strong (-1) T4 with a voiceless C2 (+1) should equal the weak T5 with a voiced C2 because the C2 voicing and stress would offset each other (R2 and R4). Compared with shape 2, the Short (+1) and strong (-1) T4 should be as high as the weak T0, since the Truncation and stress would offset each other (R1 and R4).

Category 3 contains two shapes. The average T3 after T2 and T4 (shape 5) should be one degree lower the average T0 after T2 to T4 of shape 2 due to R2 (Voiced C2 lower than voiceless). It should be two degrees lower than the T4 in category 1 due to R1 (Longer lower than Short) and R2 (Voiced C2 lower than voiceless). T5 after T3 (shape 6) should be as high as T3 of shape 5 in the same category due to R3 (Lower S1 lower than Upper S1) and R1 (Short higher than Long).

Category 4 includes three shape. T5 after T5 (shape 8) should be one degree lower than T5 after T3 (shape 6) in category 3 due to rule 4 (strong lower than weak). T0 after T5 (shape 7) should be as high as T5 after T5 (shape 8) in the same category as R1 (Long lower than Short) and R2 (unvoiced higher than voiced) would offset each other. T3 after T3 (shape 9) should also be one degree lower than T5 after T3 (shape 6) due to R1. It should be one degree higher than T3 after T5 (shape 10) in category 5 due to R4 (weak higher than strong).

The above inferences regarding the five categories of F0 height over the early part of the ten shapes are confirmed in Figure 3. The duration of these F0 shapes is equalised in the figure. Long S2 curves are represented by different kinds of lines, and Short curves by solid lines with different symbols (circles, diamonds, etc). The ten numbers in boldface overwritten on the F0 contours indicate the ten shapes.

Figure 3 shows the five categories expectably lie in five separate ranges. That means the four rules work. The average T4 after T2 to T4 (shape 1) in category 1 is the highest whose early part is above 1 SD, while T3 after T5 (shape 10) in category 5 the lowest whose early part is below -1 SD (ignoring the onset). The three S2 shapes 7, 8 and 9 of category 4 are the next lowest whose early part is between -0.25 and -1 SD. The two shapes 5 and 6 in category 3 (whose early part is between -0.25 and 0.25 SD) are higher than those in category 4 but lower than the three shapes 2, 3 and 4 in category 2 (whose early part is between 0.25 and 1 SD), which, in turn, are lower than T4 of shape 1. There are no exceptions. The only small problem is the T4 after T5 (shape 4) in category 2. It would be closer to expectations if the F0 height at the 33% point of curve 4 was a little higher.

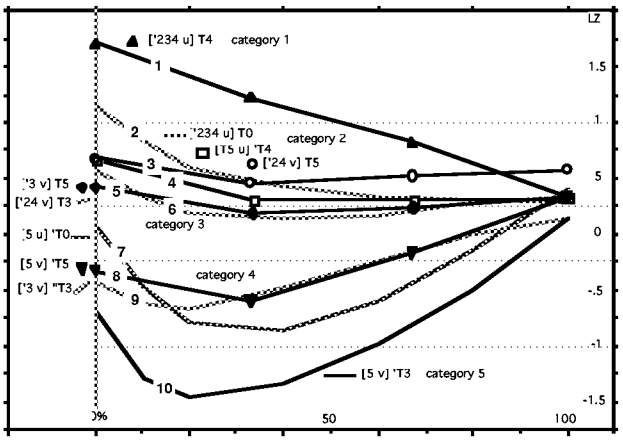

Figure 3: Ten S2 shapes, averaged from sixteen curves, of the five categories, grouped with respect to the four rules.

4. SUMMARY

The overwhelming rightward spreading in Shanghai, as shown in Zhu 1998, may be concealed by various prosodics during phonetic realisation. This paper has shown that the onset and offset targets of Shanghai citation tones spread to the corresponding ditones. Enormous F0 variation was found over the early part of the sixteen Hi-target S2 contours. The apparent complexity of this variation was found to be very regular after the proposal of four rules involving the interaction of four conditions: S2 Truncation, C2 voicing, S1 Register, and stress. The sixteen S2 contours can be combined into ten shapes according to these four conditions. Applying the four rules, the ten shapes further fall into five categories in terms of F0 height over the early part of S2 curves.

5. REFERENCES

1. Rose, P. "Acoustics and phonology of complex tone sandhi," *Phonetica* 47, 1-35, 1990.

2. Zhu, X. *Shanghai Tonetics*. PhD thesis, Australian National University, Canberra, 1995.

3. Zhu, X.S. "What spreads, and how? Tonal rightward spreading on Shanghai disyllabic compounds". Paper for The Fifth International Conference on Spoken Language Processing, 30 Nov - 4 Dec, Sydney, 1998.