INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Organizing syllables into sandhi domains —Evidence from F0 and duration patterns in Shanghai Chinese

Bijun Ling, Jie Liang

Tongji University

lingbijun@me.com, liangjie56@vip.tom.com

Abstract

In this study we investigated grouping-related F0 patterns in Shanghai Chinese by examining the effect of syllable position in a sandhi domain while controlling for tone, number of syllables in a domain, and focus condition. Results showed that F0 alignment had the most consistent grouping-related patterns, and syllable duration was positively related to F0 movement. Focus and word length both increased F0 peak and F0 excursion, but they had opposite influence on F0 slope, which indicated that focus and word length had different mechanisms in affecting F0 implementation, as focus increased articulation strength while word length influenced speaker's pre-planning.

Index Terms: tone sandhi, F0, duration, Shanghai Chinese

1. Introduction

It is generally believed that in a multi-syllabic utterance individual syllables are not evenly arranged, but organized into separate groups, even when there are no pauses involved. Listeners rely on such systematic phonetic variations to locate the boundaries of structural constituents, like words or phrases [1]. One such set of cues, i.e., the systematic variations in duration, has been well studied in intonation languages and several durational mechanisms have been proposed. The most influential proposals are "domain-edge lengthening", which includes initial lengthening [2,3] and final lengthening [4-6]; and "polysyllabic shortening" [7, 8]. In addition, pitch accent is also a influencing factor of duration pattern[9], because polysyllabic shortening appeared greater when words were accented [10] and accentual lengthening was greater in a monosyllable than in a disyllable [11], which indicated the possibility of a link between polysyllabic shortening and accentual lengthening.

Although the speech timing processes have been well studied in intonation languages, the nature of such grouping and how it is phonetically realized are not well understood in tone languages. [12] studied the syllable organization in short words/phrases with 1–3 syllables in Standard Chinese (SC) and proposed that syllable grouping was primarily encoded with durational adjustments, because each syllable needed to retain its tone features. Furthermore, the F0 excursion of rising/ falling tones showed patterns commensurate with syllable duration. [13] studied the accentual lengthening of mono-morphemic four-syllable words in SC and found that the distribution of lengthening was non-uniform: there was a strong tendency of edge effect with the last syllable lengthened the most.

Although Shanghai Chinese (SHC) is also a syllabic language with five lexical tones (Table.1-left), it is in difference to SC, because when syllables are combined into words/phrases in SHC, lexical tones would undergo sandhi changes [14-19]. The general consensus is that given a sandhi domain, the initial syllable is assigned stress and its tone type determines the F0 contour of the whole domain, and the tonal contours of non-initial syllables never surface (Table. 1-right).

Table 1: the value of citation tones and sandhi tones (using Chao's five-level numerical scale, which divides a speaker's pitch range into five scales with 5 indicating the highest and 1 the lowest).

	Duration		
Register	Long [CV]		Short [CVʔ]
High	T1 [HL]	T2 [LH]	T4 [LHq]
Low		T3 [LH]	T5 [LHq]
	Falling	Rising	
	Contour		

Citation Tone	Sandhi Tone T+X	T+X+X
T1 [51]	55+31	55+33+31
T2 [34]	33+44	33+55+31
T3 [14]	22+44	22+55+31
T4 [55]	33+44	33+55+31
T5 [13]	11+13	11+22+13

As we can see, in SHC the sandhi domain represented the boundary of word/phrase and its corresponding tone sandhi rule can help listeners to locate the boundaries. Since syllable grouping is primarily encoded with tone sandhi processes, would there still be the systematic durational variations (i.e. domain-edge lengthening and polysyllabic shortening) in SHC? Furthermore, within the sandhi domain, each syllable is assigned with its specific pitch target. Would the pitch target influence the syllable duration, i.e. the larger its pitch excursion involving the longer duration?

Secondly, there is a hypothesis that utterance length is a determining factor of initial F0 peak, with longer utterances involving a higher initial F0 peak [20-23]. Would this hypothesis also stand in the process of tone sandhi? How would word/phrase length (syllable numbers) influence the F0 implementation of tone sandhi? Within a sandhi domain, would the F0 peak raise proportional to the word/phrase length or remain the same?

Last but not least, focus (i.e. accent) has been reported to raise the F0 peak and extended the syllable duration of on-focus constituents [13, 24, 25]. Since focus and word length both increase the F0 peak and extend the duration, do they have the same mechanism? As the distribution of lengthening was non-uniform in Standard Chinese, would SHC have the same distribution of lengthening?

The present study was an attempt to improve our understanding of syllable organization in Shanghai Chinese by examining syllable duration and F0 implementation within sandhi domains of varying lengths and different focus

http://dx.doi.org/10.21437/Interspeech.2016-631

conditions. In this paper, we try to answer the following questions:

(1) What is the duration pattern within a sandhi domain in SHC? Is it the same as SC?

(2) Within the sandhi domain, how is the F0 contour implemented with the word/phrase length increase?

(3) If both focus and word length raise F0 peak and extend syllable duration, do they have the same mechanism?

2. Method

2.1. Stimuli

The stimuli, as shown in Table 2, consist of words/phrases that naturally form 1, 2 and 3 syllable groups when put into the carrier frames shown in Table 3. The stimulus words/phrases were divided into three groups based on their tonal compositions. The first syllable of each stimulus words/ phrases was a proper noun (a family name) with T1[HL], T3[LH] and T5[LHq] respectively, and the rest syllables of the stimulus words/phrases were made of the same word /ma^1/(mum) or /ma^{1}ma/ (mother). In order to make continuous F0 contours, and reduce its perturbation caused by consonants [2, 27], only syllables with initial sonorant consonants /m/ and /l/ were used.

The carrier sentence is shown in Table 3 [26]. Square brackets indicate boundaries of the sandhi domains. The stimulus (X) was always located at the beginning of the sentence, followed by the locative marker /ləʔləʔ/. Two types of questions were asked to elicit the carrier sentence. One question was about which word was on a certain row (2a) and the other about which row a specific word was on (2b). The same stimulus sentence (3) would then be uttered accordingly, with two prosodic patterns - one with the target word X focused, bolded and underlined (3a), and the other X was not focused (3b), in which case focus was on the row number.

Table 2. List of stimuli and their compositions

Word length	Meaing	T1 [HL]	T3 [LH]	T5 [LHq]
Monosyllable	X [Family name]	猫 /mɔ/ 捞 /lɔ/	毛 /mɔ/ 老 /lɔ/	穆 /moɔ/ 陆 /loɔ/
Disyllable	X+/ma^1/ [Mum X]	猫妈 /mɔ ma / 捞妈 /lɔ ma /	毛妈 /mɔ ma/ 老妈 /lɔ ma/	穆妈 /moɔ ma/ 陆妈 /loɔ ma/
Tri-syllable	X+/ma^{1}ma/ [Mother X]	猫妈妈 /mɔ ma ma/ 捞妈妈 /lɔ ma ma/	毛妈妈 /mɔ ma ma/ 老妈妈 /lɔ ma ma/	穆妈妈 /moɔ ma ma/ 陆妈妈 /loɔ ma ma/

Table 3: List of carrier sentences

(1) Carrier sentence	[X] [ləɔ^{5}ləɔ] [di^3 se haŋ]. X/ locative marker/ the 3rd row X is on the 3rd row.
(2) Questions	(a) [sa^2 gə zɿ] [ləʔ^{5}ləʔ] [di^3 se haŋ]? Which word/ locative marker/ the 3rd row Which word is on the 3rd row? (b) [X] [ləɔ^{5}ləɔ] [di^3 tɕɤ haŋ]? X/ locative marker/ which row On which row is X?
(3) Answers	(a) *[X]* [ləɔ^{5}ləɔ] [di^3 se haŋ]. (X focused) (b) [X] [ləɔ^{5}ləɔ] [di^3 *se* haŋ]. (X non-focused)

2.2. Subjects and recording procedure

8 speakers (4 males and 4 females) between the ages of 25 to 35, born and raised in Shanghai urban areas, participated in the study. They were paid for their participation and none reported any hearing, vision, or reading deficiencies.

The recording was conducted in the sound booth at Tongji University. The leading questions were recorded beforehand by a female speaker (not one of the participants). For each trial, an appropriate leading question was played through headphones to the participant and then the participant read aloud the target sentence. The sentences were presented in PPT with random order. Each participant went through a number of practice trials before the start of the real recording. Every participant read the material three times. So we achieved 18 words *8 speakers *2 prominence conditions *3 times =864 tokens.

2.3. Labeling and measurements

The acoustic analysis was prepared by using a Praat [28] script "ProsodyPro" [29]. The onset and offset of each syllable within the target words/phrases was manually labeled. Furthermore, the consonant and vowel of each syllable was segmented manually by referring to the change of F2 in the spectrogram. The script measured the F0 at 10 equidistant points of the sonorant consonant and 20 points of the vowel respectively. Subsequently, in order to normalize for gender differences, the F0 was converted into a semitone scale (a psycho-acoustic scale equal to perceptual intervals). Formula (1) relates frequency in semitones, F, to frequency in Hz, f:

$$F=12*\log_2(f/50) \quad (1)$$

From the F0 curves of each stimulus word/phrase, we measured the F0 maximum (maxF0) and minimum (minF0) values and the duration between ($D_{excursion}$). Then we computed the F0 excursion, the range between the maxF0 and minF0 of each tone group in semitones; and the F0 slope, linear slope of the F0 rise or fall:

$$\text{F0 excursion}=\text{maxF0}-\text{minF0} \quad (2)$$

$$\text{F0 slope}=\text{F0 excursion}/ D_{excursion} \quad (3)$$

Furthermore, we measured the duration of each syllable ($D_{syllable}$) and computed the average syllable duration of each speaker ($D_{speaker}$), and then the relative duration of each syllable ($D_{relative}$) was calculated:

$$D_{relative} = D_{syllable} / D_{speaker} \quad (4)$$

3. Analyses and results

3.1. General description

Prior to any numerical analysis, mean F0 curves were first examined to identify general patterns of various effects. Fig. 1 displays mean F0 curves showing the effects of prominence condition, tone type, word length and within-group position.

There are several items to be noted. (1) With the word length increasing from one syllable to three syllables, the F0 contour is stretched with the general trend/shape retaining. The F0 trajectory of each tone group is the same as the description in literature, except that the F0 trajectory in disyllabic words of the T3 group, in which the F0 contour reaches its peak in the middle of second syllable and then decline; in literature, it reached its peak around the offset of second syllable. This phenomenon needs further investigation. (2) In multisyllabic words, the turning point of F0 contour (F0 peak or valley) locates at the syllable boundary, which indicates that within sandhi domain, each syllable has its specific pitch target. (3) The F0 peak of each tone group rise with the word length increasing, but there is no obvious word length effect on F0 valley, except in T1 group. (4) Focus raised the F0 peak obviously but did not affect the F0 valley, regardless of the word length or its tone type. Detailed observations will be discussed in the next section together with the results of statistics analyses.

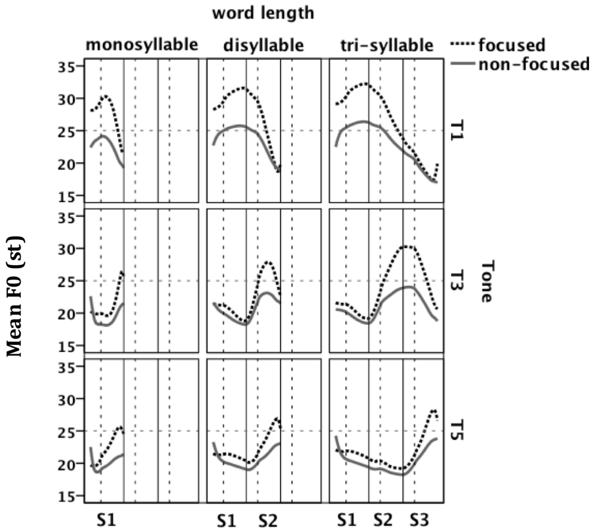

Fig 1: Mean F0 curves under the effects of prominence, tone and word length averaged over three repetitions by 8 subjects.

3.2. F0 measurements analysis

In order to examine the F0 implementation within the sandhi domain, the Multivariate ANOVA was conducted in SPSS (20.0) with maxF0, minF0, F0 excursion and F0 slope as dependent variables, and Tone type (T1, T3, T5), Word length (mono-, di-, tri-syllable) and Prominence condition (focused, non-focused) as fixed factors.

Tone type had significant effect on maxF0 [F(2,17)=29.351, p<0.001], minF0 [F(2,17)=5.714, p=0.003], F0 excursion [F(2,17)=159.722, p<0.001] and F0 slope [F(2,17)=22.043, p<0.001]. Post hoc (LSD) tests showed that the minF0 of T1 group was significantly lower than that of T3 (t=-1.070, p=0.017) and T5 (t=-1.470, p=0.001) groups, and the maxF0 of T1 tone group was significantly higher than that of T3 (t=2.229, p<0.001) and T5 (t=2.659, p<0.001) groups; and there were no significant difference between T3 and T5 tone groups in both minF0 and maxF0. The F0 excursion and F0 slope of each tone group was significantly different from each other, and both of them were in line with the sequence: **T1>T3>T5** (Fig. 2).

Word length had significant effect on maxF0 [F(2,17)=8.273, p<0.001], F0 excursion [F(2,17)=54.474, p<0.001] and F0 slope [F(2,17)=37.338, p<0.001], but had little influence on minF0 [F(2,17)=1.005, p=0.366]. Post hoc (LSD) tests showed that the **maxF0** of monosyllabic words was significantly smaller than that of di- (t=-0.843, p=0.024) and tri-syllabic words (t=-1.511, p<0.001); the maxF0 of disyllabic words was smaller than that of tri-syllabic words, but with no significant difference (t=-0.668, p=0.074). The F0 excursion and F0 slope of mono-, di- and tri-syllabic words were significantly different from each other. **With the word length increasing, the F0 excursion became larger, while the F0 slope became smaller** (Fig. 2)**.** To be noted, there was a significant interaction of "Tone type *Word length" on F0 slope [F(4,17)=28.211, p<0.001]. In T3 group, the F0 slope of disyllabic word was significantly bigger than that of mono- and tri-syllabic words. This phenomenon was caused by the tonal alignment pattern of disyllabic words that the F0 rising only happened at the first half of the second syllable as discussed in 3.1.

Prominence condition had significant effect on maxF0 [F(1,17)=321.138, p<0.001], F0 excursion [F(1,17)=573.508, p<0.001] and F0 slope [F(1,17)=291.913, p<0.001]; but no significant effect was found on minF0. In focused condition, the maxF0, F0 excursion and F0 slope were significantly larger than those in non-focused condition (Fig. 2).

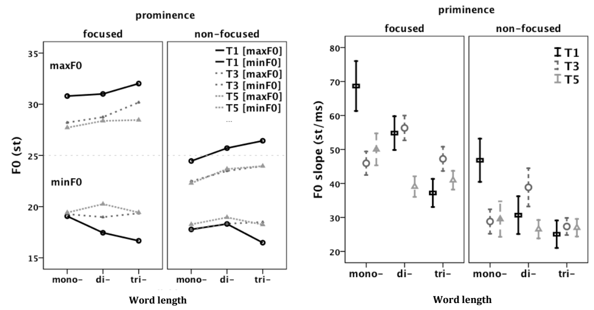

Fig 2: maxF0 and minF0 (left) & F0 slope (right) of each tone group under the effects of prominence condition and word length.

3.3. Duration analysis

A similar ANOVA was conducted with relative duration of syllable as dependent variable. Results showed that the syllable duration was significantly influenced by Tone type [F(2, 17)=103.489, p<0.001], Prominence condition [F(1, 17)=239.177, p<0.001] and Word length [F(2, 17)=110.626, p<0.001]. Post hoc (LSD) Tests showed that in general, syllable duration was significantly different among three tone types: **T3 (1.075)> T1 (1.001) >T5 (0.924)**; and among three word lengths: **mono- (1.148) >di- (1.026) >tri- (0.933)**; in addition, it was significantly lengthened in focused condition: **focused (1.082) > non-focused (0.918)**.

In order to explore the durational pattern within sandhi domains, the data was split by word length and separated ANOVA was run with Syllable position (S1, S2, S3) and Tone type (T1, T3, T5) as fixed factors.

In disyllabic words, the duration of second syllable (S2) was significantly longer than the first syllable (S1) **[S1 (0.846) <S2 (1.207)]** regardless of tone types. As for the effect of tone types, the duration of S1 in T5 group was significantly shorter than that of T1 and T3 groups, but no significant difference between the latter two **[T3 (0.951) /T1 (0.936) >T5 (0.650)]**; the duration of S2 was significantly different among three tone groups **[T5 (1.293) >T3 (1.226) >T1 (1.104)]**.

In tri-syllabic words, the duration of S1, S2 and S3 was significantly different from each other, **S1 (0.777) <S2 (0.960) <S3 (1.062)** regardless of tone types. As for the effect of tone types, the duration of S1 in T5 group was significantly shorter

than that of T1 and T3 groups **[T3 (0.873) /T1 (0.847) >T5 (0.611)]**; the duration of S2 in T3 group was significantly longer than that of T1 and T5 groups **[T3 (0.991) >T1 (0.947)/T5(0.941)]**; the duration of S3 in T5 group was significantly longer than that of T1 and T3 groups **[T5 (1.107) >T3 (1.057) /T1 (1.023)]** (Fig 3).

In order to examine the distribution of accentual lengthening within the sandhi domain, the lengthening ratio of each syllable [$D_{relative}$(focused)/$D_{relative}$(non-focused)] was computed. Then the data was split by word length and tone type, and separated ANOVA was run with Syllable position as fixed factor. The results showed that regardless of tone types, the magnitude of lengthening was **S1>S2** and **S1>S2>S3** in di- and tri- syllabic words respectively, which was not in line with the results of Standard Chinese that the last syllable was lengthened the most [13, 24, 25].

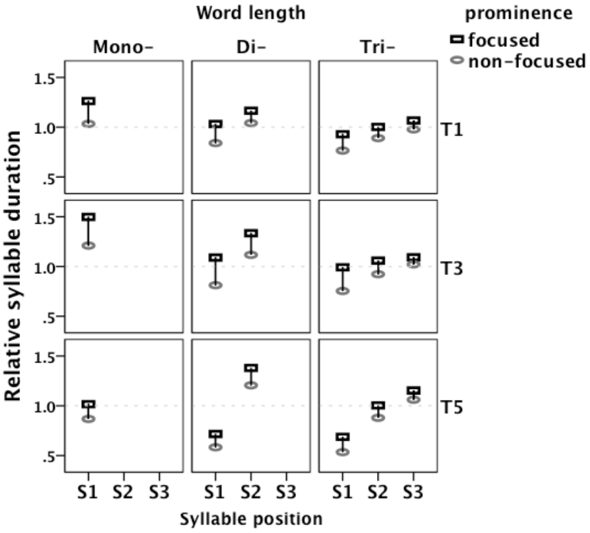

Fig 3: Mean syllable duration under the effects of prominence condition, tone types and word length.

4. Discussion

In summary, when syllables were grouped into sandhi domains in SHC, there was an obvious effect of polysyllabic shortening, i.e. the syllable duration became shorter with the word length increasing. With regard to the duration pattern within sandhi domains, there are two points deserving special interest:

(1) Within the sandhi domain, the duration of the initial syllable was the shortest while the last syllable was the longest, regardless of tone types (Fig. 3). Such duration pattern was in difference to that of Standard Chinese (SC), in which the initial syllable was the second longest. Take tri-syllabic word for example, the duration pattern in SHC was **S3>S2>S1** while that in SC was **S3>S1>S2**. As for the reason why the initial syllable was shortest in SHC, it could be caused by the tonal target of the initial syllable, which was always a level tone. Comparing to contour tones (rising/falling), level tone required less duration to implement [27]. The results indicated that final lengthening might be universal, while initial lengthening was language specific.

(2) There was an obvious compensatory lengthening effect in the last syllable of T5 (checked tone) group. Because the initial syllable of T5 group was a checked tone and significantly shorter than T1 and T3, in both di- and tri-syllabic words the last syllable of T5 group was significantly longer than that in T1 and T3 groups. This indicated that there was a tendency of isochrony to preserve the relative uniformity of the duration of words/phrases.

With regard to the effect of word length on F0 implementation, the F0 peak and F0 excursion increased while the F0 slope decreased with the word length increasing. Crucially, F0 slope represented how fast pitch changed, which reflected the articulation strength [27]. The F0 slope decrease indicated the articulation strength diminished with the word length, which was in accordance with the physiological mechanism. Furthermore, the F0 peak of T1 group increased with the word length, which indicated that speakers incrementally adjusted the height of the initial F0 peak within a planning unit. However, it needs further investigation whether the initial F0 peak correlates directly with the length of the whole utterance or with a smaller planning unit (e.g. word or phrase).

Although focus also raised F0 peak and expanded F0 excursion, focus and word length had different mechanisms in adjusting F0 implementation within sandhi domains, which was reflected by the adjustment of F0 slope. As focus increased the F0 slope while word length decreased it, which meant prominence affected the F0 implementation by increasing articulation strength of focused constituents while word length through duration extension.

Last but not least, when the focus domain was multi-syllabic, the distribution of lengthening is non-uniform [13]. In Shanghai Chinese, the duration of initial syllable increased the most, while in Standard Chinese, the last syllable was lengthened the most. There were two possibilities to explain the difference: (1) In SHC, the initial syllable was assigned with stress as proposed by phonologists [15-17], therefore the initial syllable was the metrically stronger, which should be lengthened more than metrically weaker ones; (2) In this experiment, the information load of initial syllable was larger than other syllables, as the initial syllable offered new information (a family name) while other syllables were the same. Therefore, the syllable with new information got lengthened most. In order to have a better understanding of the distribution of accentual lengthening within sandhi domains, more experiments are required.

5. Conclusions

The goal of the present study is to identify the duration pattern related to syllable grouping in Shanghai Chinese, and to compare the effect of focus and word length on F0 implementation. As for the duration patterns within sandhi domains, in general the syllable duration correlated negatively to the word length. But there are two things to be noted: (1) The last syllable of T5 group showed a significant compensatory lengthening effect in multi-syllabic words, as the initial syllable was shorter caused by glottal coda; (2) Within sandhi domain, the initial syllable was shortest but became lengthened most in focused condition. The mechanism behind such phenomenon requires further investigation. Both focus and word length increased F0 peak and F0 excursion, but they had an opposite influence on F0 slope, which indicated that they had different mechanisms in affecting F0 implementation.

6. References

[1] A. E. Turk, & S. Shattuck-Hufnagel, "Word-boundary-related duration patterns in English", Journal of Phonetics, vol. 28, pp. 397–440, 2000.

[2] M. Cooper, "Laryngeal and oral gestures in English/p, t, k/", In Proceedings of the XIIth international congress of phonetic sciences, Aix-en-Provence, pp. 50–53, 1991

[3] Fougeron, & P. A. Keating, "Articulatory strengthening at edges of prosodic domains", Journal of the Acoustical Society of America, vol. 101, pp. 3728–3740, 1997.

[4] H. Klatt, "Vowel lengthening is syntactically determined in a connected discourse", Journal of Phonetics, vol. 3, pp. 129–140, 1975.

[5] M. E. Beckman, & J. Edwards, "Lengthenings and shortenings and the nature of prosodic constituency", In J. Kingston, & M. E. Beckman (Eds.), Papers in laboratory phonology I: Between the grammar and the physics of speech (pp. 152–178). Cambridge: Cambridge University Press, 1975.

[6] W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, & P. Price, "Segmental durations in the vicinity of prosodic phrase boundaries", Journal of the Acoustical Society of America, vol. 91, pp. 1707–1717, 1992.

[7] I. Lehiste, "The timing of utterances and linguistic boundaries", Journal of the Acoustical Society of America, vol. 51, pp. 2018–2024, 1972.

[8] R. F. Port, "Linguistic timing factors in combination", Journal of the Acoustical Society of America, vol. 69, pp. 262–274, 1981.

[9] A. M. C. Sluijter, & V. J. van Heuven, "Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch", Phonetica, 5vol. 2, pp. 71-89, 1995

[10] E. Turk, & S. Shattuck-Hufnagel, "Word-boundary-related duration patterns in English", Journal of Phonetics, vol. 28, pp. 397–440, 2000.

[11] A. E. Turk, & L. White, "Structural influences on accentual lengthening in English", Journal of Phonetics, vol. 27, pp. 171–206, 1999.

[12] Y. Xu, & M. L. Wang, "Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin", Journal of Phonetics, vol. 37, pp. 502–520, 2009.

[13] Y. Y. Chen, "Durational adjustment under contrastive focus in Standard Chinese", Journal of Phonetics, vol. 34, pp. 176–201, 2006

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

[15] M. Yip, Tone. Cambridge: Cambridge University Press, 2002.

[16] H. Xu, Z. H. Tang, & N.R. Qian, "Tone sandhi in New Shanghainese", Dialects, vol. 3, pp: 145-155, 1981

[17] S. Duanmu, "Rime length, stress, and association domains", Journal of East Asian Linguistics, vol. 2, pp: 1–44, 1993

[18] Y. Y. Chen, "Revisiting the phonetics and phonology of tone sandhi in Shanghai Chinese", Speech Prosody 2008, Campinas, Brazil, 253–256.

[19] N. R. Qian, The history and development of the Shanghai Dialect, Shanghai: Shanghai People's Publishing House, 2003

[20] W. Cooper, & J. Sorenson, Fundamental Frequency in Sentence Production. Heidelberg: Springer, 1981.

[21] M. Liberman, & J. Pierrehumbert, Intonational invariance under changes in pitch range and length. In Language and Sound Structure, M. Aronoff, R. Oehrle, (eds.) Cambridge MA: MIT, pp. 157–233, 1984.

[22] R. Ladd and C. Johnson, "Metrical factors in the scaling of sentence-initial accent peaks", Phonetica, vol. 44, pp. 238–245, 1987

[23] B. Connell, "Tone, Utterance Length and F0 Scaling", International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages Beijing, China, March 28-31, 2004

[24] Y. Y. Chen, "Focus and intonational phrase boundary in Standard Chinese", Proceedings of the international symposium on Chinese spoken language processing, Hongkong, 2004

[25] J. Yuan, Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling (Ph.D. dissertation). Ithaca: Cornell University, 2004

[26] Y. Y. Chen, "The acoustic realization of Shanghai vowels", Journal of Phonetics, vol: 36, pp: 629–748, 2008

[27] Y. Xu and X. Sun, "Maximum speed of pitch change and how it may relate to speech", Journal of the Acoustical Society of America, vol: 111, pp: 1399–1413, 2002.

[28] P. Boersma, Praat, "A system for doing phonetics by computer", Glot International, vol: 5, pp: 341–345, 2001

[29] Y. Xu, http://www.phon.ucl.ac.uk/home/yi/tools.html, 2005-2009