Post-focus pitch register lowering as a phrasal marker

--- An acoustic study of focus and phrasing in Shanghai Chinese

Lei Sun[1,2], *Yiya Chen*[1,2]

1. Leiden University Center for Linguistics;
2. Leiden Institute for Brain and Cognition Leiden University, the Netherlands

l.sun@hum.leidenuniv.nl, yiya.chen@hum.leidenuniv.nl

ABSTRACT

This paper reports an experiment designed to investigate whether post-focus tonal realization is constrained by prosodic structure in Shanghai Chinese (SH). Previous studies have shown that in SH, focus expanded the f0 range of bi-syllabic tone sandhi words while pitch register lowering, instead of pitch range compression, was consistently found on post-focus bi-syllabic tone sandhi words. The present study examined durational adjustments and f0 realizations of bi-syllabic tone sandhi words at different prosodic levels under different focus conditions. The results showed that focus expanded the pitch range of the target syllables at both the levels of Prosodic Word and Major Word, but post-focus pitch register lowering was only found at the level of Major Word. Based on the results, this study concludes that post-focus tonal realization in SH is constrained by prosodic structure and post-focus pitch register lowering is a phrasal marker.

1. INTRODUCTION

It is widely accepted that different information structure notions such as focus can be prosodically encoded across languages. However, the relation between focus and its prosodic expression in languages is still controversial. Different models have been proposed to analyse their relation and early analyses can be divided into two subgroups depending on whether prosodic phrasing is referred to [1]. One view holds that focus is assigned directly to morpho-syntactic elements and then is encoded directly into prosody without any direct reference to prosodic phrasing [2]. In Chinese languages, focus has been long considered to be directly encoded via pitch range manipulation [3, 4]: the pitch range of the focused region was expanded while that of the post-focus region was compressed and the pre-focus f0 modification was kept neutral. However, this has been challenged by recent studies of post-focus tonal realizations in other Chinese languages such as Taiwanese and Cantonese in which post-focus compression was found to be absent [5, 6]. Another view claims that the relation between focus and its prosodic encoding is more indirect and mediated by prosodic phrasing. According to this view, focus requires the highest level of prominence over the focused constituent and as a result the focused constituent becomes the head of a prosodic domain. In this way, focus affects the location of prosodic prominence over a prosodic domain and therefore may force the modification of prosodic phrasing by deleting or inserting a prosodic boundary [7, 8]. Chen [9] found that in SC post-focus lexical tones were realized with a more expanded f0 range than their pre-focus counterparts in some tonal contexts. Therefore, pitch range compression cannot be the primary characteristic of post-focus tonal realization. Instead, this study proposed that different post-focus tonal realizations are examples of the weak implementation of post-focus tonal targets associated with non-prominent constituents and post-focus tonal realization was constrained by prosodic phrasing.

This study investigates whether post-focus tonal realization is constrained by prosodic phrasing in Shanghai Chinese (SH). It has been claimed that in SH, pitch range/register compression (PFC) can be found on the post-focus words, but only at certain prosodic level [10]. Selkirk & Shen [10] analysed it as a post-focus tone deletion rule and proposed two prosodic domains to explain it: Prosodic Word (PW) at which PFC was absent and Major Phrase (MP) where PFC was present. However, [11] reported that the primary phonetic feature of post-focus tonal realization in SH was pitch register lowering instead of post-focus pitch range compression. Therefore, if post-focus tonal realization in SH is indeed constrained by prosodic phrasing as [10] claimed, pitch register lowering should be only found at the phrasal level. The basic idea of this experimental design is to establish different prosodic boundaries with the same pre-boundary and post-boundary words at different prosodic levels and then examine durational adjustment and f0 realization of these words under focus, No-focus, pre-focus/post-focus conditions. If the pitch register of the post-boundary word is found to be lowered at only certain level when focus is on the pre-boundary word, pitch register lowering can be a phrasal marker.

To make sure that the prosodic boundary conditions created are indeed of varying strength, three cues of prosodic boundaries were examined: pre-boundary lengthening, pause and domain-initial strengthening. A big number of studies have shown convincingly that the three durational cues are correlated with prosodic boundary strength [12]. The magnitudes of pre-boundary lengthening effect and pause tend to vary with levels in the prosodic hierarchy: the higher level a pre-boundary segment/a pause is, the longer it is than the same segment/ the pause at lower levels. Phonetic realizations of domain-initial segments also correlate with prosodic boundary strength, but bounded by language-specific phonological constraints [13]. Therefore, the existence of these durational patterns can act as a means of confirming the design of the experiment.

Pitch is also examined to see the distribution of post-focus pitch register lowering at different prosodic levels created. SH has five lexical tones (Tone 1: high falling; Tone 2: high rising; Tone 3: low rising; Tone 4: short high[1]; Tone 5: short low rising). When two syllables are combined into a bi-syllabic noun, verb or adjective, one tone sandhi domain is formed where lexical tonal contrasts over the non-initial syllable are neutralized [14]. The five tones also display interesting co-occurrence patterns with the onset and coda of the tone-bearing syllable [15]. Syllables carrying a high-register tone only allow voiceless onsets while syllables carrying a low-register tone can have only voiced onsets. Long tones occur in open syllables or syllables with a nasal coda while short tones only co-occur with closed syllables with a glottal coda.

2. METHOD

2.1. Test materials and procedure

The design of this experiment followed the method used in [16]. Three morpho-syntactic boundaries were manipulated and bi-syllabic words were used: within a compound (Boundary 1: B1), between a verb and its direct object (B2) and between the subject and the transitive verb (B3). These constructions are commonly recognized to have fundamental syntactic distinctions which are expected to map onto different prosodic domains. At each level, two target syllables S1 and S2 were separately constructed at the pre-boundary position and the post-boundary position. Therefore, at B1, S1and S2 constituted a compound. At B2, S1 was the ending syllable of the bi-syllabic verb and S2 was the beginning syllable of the bi-syllabic object. At B3, S1 was the ending syllable of the bi-syllabic subject while S2 was the beginning syllable of the bi-syllabic verb. Different focus locations were elicited: corrective focus on constituents containing S1 or S2 (CF1/CF2) and on the final part of the carrier sentences (No-Focus: NF).

Three tonal combinations were chosen for S1 and S2 (T1T1, T1T3, T3T1) and four pairs of homophones were constructed for each combination. T1T1 was preceded by T1 while the other two tones were preceded by T3. To measure boundary-related durations, the specific choice of onset consonants for S1 and S2 were made based on the results of a pilot study and semantic meanings of target syllables. Therefore, only unaspirated voiceless and voiced stops and affricates were included as the pilot study showed that the lengthening of fricatives and aspirated segments in SH was very sensitive to prosodic boundaries. Given the co-occurrence patterns of tone and onset mentioned, the S2 onsets for T1T1 were all voiceless affricates; for T1T3 they were all voiced stops; for T3T1, they were three voiceless stops plus one voiceless affricate.

The experiment was carried out on a laptop using E-prime [17]. Eighteen native speakers of SH were recruited as participants and all of them were college students aging from 19 to 24. Participants were asked to read the sentences on the screen which were randomized after they heard the questions played first. To make sure that focus can be elicited naturally, the experiment used a within-subject and between-block design in which every speaker produced every target sentence only once.

2.2. Acoustic and statistical analyses

In total 1941 sentences were recorded and consonants and vowels in S1 and S2 in these sentences were manually labeled in Praat [18], based on waveforms and spectrograms. Durations of S1, pause and the onset consonant in S2 were measured. Measure of the S2 onsets was taken from the point of the stop release to the voice onset of F2 in the vowel. As a result, measure of the pause includes the period of the stop closure in S2. The normalized duration is the mean duration of S1/S2 averaged across speakers.

F0 contours were obtained by taking 20 points (in Hertz) of S1and S2 separately and the raw values in Hz were then transformed into semitones by following formula (1) (ST: semitone; f_0: raw $f0$ value in Hz; f_{ref}: the reference frequency with 100 Hz for females and 50 Hz for males) and were averaged across speakers. Maximum and minimum f0 were also taken over S1 and S2 separately. An index for f0 range in semitone was derived via Formula (2).

(1) $ST = 12 \times \log_2(f_0 / f_{ref})$

(2) Range = 12 x $\log_2$(Maxf0/Minf0)

To examine the effects of focus and boundary on duration and f0 realization, two-way Repeated Measures (RM) ANOVA were conducted in SPSS separately for S1 and S2 with subjects (F1) as a random factor.

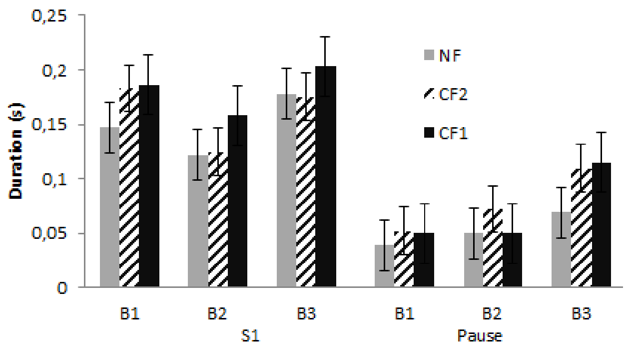
3. RESULTS

3.1. Durational adjustment

Both boundary and focus have main effects on the duration of S1 [Boundary: $F(2, 34) = 79.70$, $p<0.001$; Focus: $F(2, 34) = 91.87$, $p<0.001$]. S1 was produced with the longest duration at B3 and the shortest duration at B2 while S1 had the longest duration under CF1 condition and the shortest under NF condition. The two factors also had a significant interaction [$F(3.18, 54.11) = 25.6$, $p<0.001$]. Post-hoc pairwise comparisons showed that at B1, the durational patterns of CF1 vs. CF2 and CF2 vs. NF differed significantly from those at B2 and B3 ($p<0.001$). The results were illustrated in Fig.1. Three things are of note. First, whether unfocused or on focus, the durations of S1 at different boundaries always differed significantly from each other with the longest duration at B3 and the shortest at B2 (B2<B1<B3). Secondly, at all the three boundaries, S1 was always produced with significantly longer duration under CF1 than under NF ($p<0.001$). Thirdly, when comparing CF1 vs. CF2, it was found that only at B1, the effect of CF2 was as robust as CF1 ($p=0.925$) and significantly lengthened the duration of S1 than NF ($p<0.001$), which showed that the durational adjustment of S1 was also influenced by focus on S2 at B1. The results of pause were also reported in Figure 1. As Fig. 1 showed, under three focus situations, the duration of pause increased gradually with the shortest duration at B1 and the longest at B3 (B1<B2=B3, $p<0.001$).

The duration of S2 onsets were analyzed in the same way within each tone group as [15] reported that focus did not have any significant effect on voiceless unaspirated and voiced stops in the tone sandhi domain-medial position. Three things are worth noting. First, for T1T1, both boundary and focus had main effects on the duration of the S2 onsets and the interaction between them was also significant (Boundary: $F(2,34) = 20.33$; Focus: $F(2, 34) = 41.82$; B*F: $F(4, 68) = 6.16$, $p<0.001$). Under both CF2 and NF, S2 onsets were produced with the shortest duration at B1 and longer duration at B2 and B3 (B1<B2=B3). At each level, the S2 onsets under CF2 were significantly longer than that under NF ($p<0.001$), but at B1, there was no difference between the durations under CF1 and CF2. Secondly, in T1T3 only boundary had a significant main effect on the duration of the S2 onsets ($F(2,34) = 33.06$, $p<0.001$). The onset was produced with the shortest duration at B1 and the longer duration at B2 and B3 (B1<B2=B3). Thirdly, in T3T1, both focus and boundary had significant main effects on the duration of the S2 onsets, but the effect of focus was only near significant (B: $p=0.02$; F: $p=0.42$). The duration at B3 was significantly longer than that at B1 and B2 (B1=B2<B3). Only at B3, the duration of the onsets was longer under CF2 than under NF.

Figure 1: Mean duration and standard error (±2σ) of S1 and Pause as a function of Boundary and Focus.



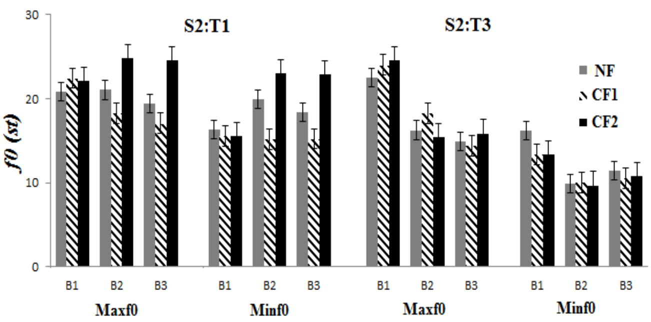
3.2. F0 realization

Boundary and focus had significant main effects on all of the three cues for all the three tone groups except the effect of boundary on Maxf0 of S2 in T1T1 and the effect of focus on Minf0 of S1 in T3T1 (T1T1: Boundary: $F(2, 34) = 1.3$, $p=0.289$; T3T1: Focus: $F(2, 34) =2$, $p=0.15$). The interaction between the two effects was also significant on them in the three groups. Separate ANOVA were further conducted to examine the effect of focus on the three cues of S1 and S2 at different prosodic boundaries. Due to limited space, only the relevant results were reported here.

First, at B1, given a tone sandhi domain consisting of S1 and S2, all of the three tone groups displayed tone sandhi patterns. Furthermore, at B1, there was no significant difference between the effects of CF1 and CF2 on the three cues although both differed significantly from the No-focus condition. To examine post-focus tonal realizations, the three cues for S2 under the post-focus condition (CF1) and NF were compared (T1T1: Maxf0 CF1=NF, p=1, Minf0 CF1<NF, p<0.001; T1T3: Maxf0 CF1>NF, p=0.001, Minf0 CF1>NF, p<0.001; T3T1: Maxf0 CF1>NF, p<0.001; Minf0 CF1>NF, p=0.41). The results showed that for T1T1, only the Minf0 was lowered while for T1T3 and T3T1, the whole pitch registers were raised. Therefore, pitch register lowering was not observed at B1. The results also showed that the domain of f0 adjustment under focus at B1 was always the whole compound whether focus was on S1 or S2, which confirmed the findings in [11].

At B2 and B3, the Maxf0 and the Minf0 under CF1 varied as a function of the tone groups. In T1T1, CF1 was found to lower both the Maxf0 and the Minf0 ($p<0.001$). In T1T3, at B2, CF1 lowered only the Minf0, but the Maxf0 was significantly raised ($F(2,34) = 26.86$, $p<0.001$). At B3, CF1 only raised the Maxf0 significantly ($F(2,34) = 5.53$, $p = 0.008$). In T3T1, at B2, CF1 only lowered the Minf0 and at B3, both the Maxf0 and the Minf0 were significantly lowered. These results contradicted the findings in [11]. As mentioned above, S1 and the preceding tone consisted of a verb at B2 and a subject at B3 while S2 and the following tone constituted an object at B2 and a verb at B3, all of which were tone sandhi domains. T1T1 was preceded by T1 and both T1T3 and T3T1 were preceded by T3. Therefore, at B2 and B3, the tone carried by S1 (the first T1 in T1T1, T1 in T1T3 and T3 in T3T1) was neutralized and the underlying tone of the preceding tone (T1 for T1T1 and T3 for T1T3 and T3T1) was phonetically realized. As a result, the raising Maxf0 in T1T3 and the unchanged Maxf0 in T3T1 was probably related with the tonal coarticulation effect, in particular the delayed alignment of the f0 peak of the preceding tone T3 relative to the edge of the tone-carrying syllable, which is a low rising tone [19]. To examined the effect of CF1 on post-focus tonal realizations of S2, we combined T1T1 and T3T1 as S2 in both of them had the same tone T1, and compared the results with that in T1T3. As Figure 2 illustrated, at B2 and B3, both the Maxf0 and Minf0 of T1 were significantly lowered under CF1 than under NF (B2: Maxf0 $F(2,34) = 151.57$, $p<0.001$; Minf0 $F(1.38, 23.46) = 118.42$, $p<0.001$. B3: Maxf0 $F(1.35, 23.02) = 138.43$, $p<0.001$; Minf0 $F(1.36, 23.19) = 123.77$, $p<0.001$).

Figure 2: Mean F0 maxima and minima and standard error (±2σ) of S2 as a function of Boundary and Focus. When S2 is T1, S1 is T1 and T3. When S2 is T3, S1 is T3.



4. DISCUSSION & CONCLUSION

4.1. Durational adjustment

The results showed that at B1, the pre-boundary duration was unusually lengthened and even longer than B2, which seemed to contradict the hypothesis of varying boundary strength. However, this resulted from the experimental design. At B1, S1 and S2 consist of a compound, which is also a PW. S1 was not only at the manipulated pre-boundary position, but also at the left boundary of a PW while at B2 and B3, S1, the second syllable of a bi-syllabic word (the first syllable is the syllable preceding S1S2), was only at the manipulated pre-boundary position. Therefore, the lengthening at B1 included both the pre-boundary lengthening and the PW boundary-induced lengthening. Ignoring this, the results supported the hypothesis that the boundaries established were of varying strength: B1 was the weakest boundary showing the shortest pause and domain-initial durations, and B3 was the strongest, showing the longest pre-boundary, pause and domain-initial durations. Focus had a lengthening effect on the pre-boundary duration across three prosodic levels, but such effect was only observed on the domain-initial duration of the specific consonant type (unaspirated voiceless affricates).

4.2. F0 realization

The results showed that the f0 realizations of S2 under the post-focus condition (CF1) at B2 and B3 were different from B1. Focus expanded the f0 range and produced a magnified pitch contour of an on-focus tone sandhi word at different prosodic levels, which was also found in [11]. Pitch register lowering was only found at B2 and B3, not at B1. Even though in specific tonal contexts such as T1T3 where the Maxf0 was raised at B2, this was probably caused by the f0 peak delay in the preceding tone T3. This can be confirmed by the results in [11] in which T3 was preceded by five lexical tones. When preceded by a focused T5 which is a short low rising tone, the Maxf0 of T3 was actually raised, not lowered. But when preceded by a focused T1 which is a falling tone, its Maxf0 was lowered. A new experiment will be done with more tonal combinations included.

The results of f0 realizations have provided empirical evidence for the “indirect” view of the relation between focus and its prosodic encoding. In SH, pitch register lowering is absent at the level of PW, but can be observed only in a prosodic phrase (MP or phonological phrase) and therefore is a phrasal marker. In SH, a MP is minimally binary and consists of at least two prosodic words. Focus on the first prosodic word requires the highest level of prominence over a MP, As a result, at the non-prominent position, the pitch register of the post-focus prosodic word is lowered. Focus inserts a MP boundary to the left of the focused constituent in SH. A detailed OT analysis will be further developed to explain the relation between prosody, syntax and focus in SH.

Combing the durational data and the f0 data, this study also shows that the prosodic structure is a mixture of two types of relations across prosodic levels. On one hand, the data of durational adjustment confirms that the prosodic levels established are of varying boundary strength. On the other hand, the data of f0 realizations shows that the prosodic structure consists of qualitatively distinct domains.

4.3. Conclusion

This study investigates whether prosodic encoding of focus in Shanghai Chinese is constrained by the phonological prosodic structure of this language. Prosody is widely known to be employed to convey discourse-level information, but the relation between focus and its prosodic expression is at issue. Taking together our findings in durational adjustment and f0 realizations, this study concludes that in Shanghai Chinese, prosodic encoding of focus is constrained by prosodic phrasing and post-focus pitch register is a phrasal marker. Focus in this language inserts a Major/Phonological phrase boundary to the left of the on-focus constituent.

5. ACKNOWLEDGEMENTS

We thank the support from the Chinese Scholarship Council to the first author as well as the support from the European Research Council (206198) to the second author. We also thank Jos Pacilly at the Phonetics Lab of Leiden University for writing the scripts for the experiment.

6. REFERENCES

[1] Féry, C., Ishihara, S. 2009. How focus and givenness shape prosody. In Zimmermann, M. and Féry, C. (eds), *Information Structure from Different Perspectives*. Oxford University Press, 36-63.

[2] Büring, D. 2006. Focus projection and default prominence. In V. Molnár. and S. Winkler. (eds), *The Architecture of Focus*. Berlin/New York: Mouton De Gruyter, 321- 346.

[3] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*. 27, 55-105.

[4] Jin, S. 1996. An acoustic study of sentence stress in Mandarin Chinese. Ph.D. dissertation. Ohio State University, Columbus.

[5] Xu, Y., Chen, S., Wang, B. 2012. Prosodic focus with and without post-focus compression (PFC): a typological divide within the same language family? *The Linguistic Review*. 29, 131- 147.

[6] Wu, W. L., Xu, Y. 2010. Prosodic focus in Hong Kong Cantonese without post-focus compression. In *Proceedings of Speech Prosody* Chicago.

[7] Selkirk, E. 1986. On derived domains in sentence prosody. Phonology Yearbook 3, 371- 405.

[8] Truckenbrodt, H. 2006. Phrasal stress. In Brown, K. (ed), *The Encyclopaedia of Languages and Linguistics*. Oxford: Elsevier, 9, 572-579.

[9] Chen, Y. 2010. Post-focus F0 compression --- Now you see it, now you don't. *Journal of Phonetics*. 38, 517-525.

[10] Selkirk, E., Shen, T. 1990. Prosodic domains in Shanghai Chinese. In Inkelas, S., Zec., D. (eds), *The Phonology-Syntax Connection*. Chicago & London: University of Chicago Press, 312-332.

[11] Chen. Y. 2009. Prosody and information structure mapping: Evidence from Shanghai Chinese. *Chinese Journal of Phonetics*. 2, 123-133.

[12] Wagner, M., Watson, D. G. 2010. Experimental and theoretical advances in prosody: a review. *Language and Cognitive Processes*. 25, 905-945.

[13] Cho, T., McQueen, J. M. 2005. Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*. 33, 121-157.

[14] Chen. Y. 2008. Revisiting the phonetics and phonology of Shanghai tone sandhi. *Proceedings of Speech Prosody* Brazil, 253-256.

[15] Chen. Y. 2011. How does phonology guide phonetics in segment-f0 interaction? *Journal of Phonetics*. 39, 612-625.

[16] Chen. Y. 2011. To Phrase or Not to Phrase: The Effect of Focus in Standard Chinese. *Proceedings of GLOW* 34.

[17] MacWhinney. B., James. S., Schunn. C., Li. P., Schneider. W. 2001. STEP – A system for teaching experimental psychology using E-Prime. *Behaviour Research Methods, Instruments & Computers*. 33, 287–296.

[18] Boersma, P., Weenink, D. 2004. Praat: doing phonetics by computer. Version 5.3.5.1. http://www.praat.org/

[19] Xu, Y. (2001). F0 peak delay in Mandarin. *Phonetica*. 58, 26-52.

[1] Short tones (T4 &T5) refer to tones carried by syllables that end with a glottal coda.