The phonetics and phonology of corrective focus in Xiangshan Wu Chinese: interaction with disyllabic tone sandhi

Yibing Shi, Brechtje Post, Francis Nolan

University of Cambridge
ys538@cam.ac.uk, bmbp2@cam.ac.uk, fjn1@cam.ac.uk

ABSTRACT

Exploring the interaction between tones and prosodic focus has been an important topic in prosodic studies in tonal languages, as f_0 multi-tasks, both cueing prosodic focus and realizing tonal contrasts. This study provides novel data on the phonetic realisations of corrective focus and its interaction with tone sandhi in Xiangshan dialect, an under-documented Chinese dialect spoken in the Northern Wu region in China. Thirty-eight disyllabic lexical compounds and phrases were selected for this study. Corrective focus was elicited on either the whole disyllable or one of the two syllables. Results show that prosodic focus can be realised via gradient acoustic cues: raised intensity and lengthened duration. Meanwhile, focus can also be optionally realised by phonological re-phrasing as evidenced in tone sandhi changes. The cue-trading relations between tone sandhi changes and gradient acoustic cue changes are also investigated.

Keywords: prosodic focus, tone sandhi, Xiangshan Chinese

1. INTRODUCTION

Prosodic focus, which serves to emphasise a part of the utterance, is typically realised through enhancement of variable gradient acoustic cues on the focused items, including expanded pitch range, lengthened duration, and increased intensity [1],[2]. Alternatively, focus can result in phonological re-phrasing, as evidenced by research across different languages, e.g., Korean [3] and Bengali [4]. In tonal languages, the investigation of focus becomes especially compelling as pitch potentially competes to signal prosodic focus, lexical tonal contrasts, as well as tone sandhi patterns, which are categorical and systematic tonal changes in certain environments.

While rather consistent and similar gradient cue changes have been found in several tonal languages, such as duration lengthening in Shanghai Chinese [5] and Standard Chinese [6], the picture of the interaction between focus and tone sandhi remains largely complex and unclear. Take Shanghai Chinese for example: although many studies witnessed no drastic effects of focus on tone sandhi patterns [7], some accounts reported that focus can block tone sandhi, realising both syllables as their underlying tones [8], [9], or it could occasionally produce other specific tone sandhi patterns, e.g., realising the second syllable in a disyllable as a high falling tone [10].

The current study is the first investigation of the realisations of prosodic focus in Xiangshan Wu Chinese, an under-documented Chinese dialect that is mainly spoken in Xiangshan County, Zhejiang Province in a Southeastern part of China. It has six lexical tones, of which four are non-checked tones (HH, HL, LHL, LH) and the remaining two are checked tones that are only associated with syllables ending with a glottal stop (Hq and LHq) [11].

The following research questions are addressed in the study: (1) To what extent is focus realised in Xiangshan dialect using gradient acoustic cues i.e., intensity and duration? (2) How does focus interact with tone sandhi? Does it block, change, or preserve sandhi patterns? (3) If there is a sandhi change under focus, do other acoustic cues still change, and if so, at a larger or a smaller magnitude?

2. METHODOLOGY

2.1. Participants

Eight Xiangshan native speakers were recruited and recorded by the first author (4 female, age range: 47-53, mean age: 50). All of them were born and raised in Xiangshan and speak Xiangshan as their major language in their daily life. No hearing and speaking impairments were reported among them.

2.2. Materials and procedure

Thirty-eight disyllabic tokens were chosen for this study, with the first syllable bearing the tones LHL or HH and the second syllable bearing any of the four non-checked tones, i.e., HH, HL, LHL, LH. Three morphosyntactic structures were covered in the test materials: lexical compounds, Modifier-Head phrases, and Verb-Object phrases.

The test tokens were presented to the speakers in a Microsoft PowerPoint presentation, with each slide showing only one test token. In the first session of the experiment, to elicit the non-focused citation realizations of each test token, the speakers were asked to read the test tokens one by one as naturally as they could.

Corrective focus was elicited on three different domains in the disyllables: the first syllable, the second syllable, or the whole disyllable. For each test token, a sentence containing a disyllable which differs from the target token in the first syllable, the second syllable, or the whole disyllable, was pre-recorded by a native male Xiangshan speaker aged 53. The pre-recorded audio file was played to the speakers while the target token was shown on a laptop screen. In this way, speakers were prompted to correct the speech using the target tokens after listening to the recordings. The scenario is summarised in Table 1, with the symbol 'σ' representing 'syllable' and the bold and italic letters showing the domain of focus. In total, there were 1,216 tokens elicited (8 speakers * 38 tokens * 4 focus conditions).

Table 1: Corrective focus elicitation scenario

Focus	Recording	Desired output
σ_1	These words are CB.	No, they are ***A***B.
σ_2	These words are AD.	No, they are A***B***.
$\sigma_1\sigma_2$	These words are EF.	No, they are ***AB***.

Prior to the actual recording of the focused items, all speakers completed a practice series with three tokens that were not part of the actual experiment, in order to familiarise themselves with the self-managed recording procedure. The participants were recorded in a small quiet room located in Xiangshan, using an H4n Zoom recorder at a sampling rate of 44.1k Hz.

2.3. Data Analysis

All the acoustic data were manually segmented and annotated using Praat [12]. For each test tokens, the rhymes – which include onset glides, vowels, and final nasals – of both syllables were manually labelled. Duration, mean intensity, and mean f_0 values were extracted using the ProsodyPro Praat script [13]. f_0 values at ten equidistant measurement points in each rhyme were also obtained using the same script. The first normalised time point in each rhyme was discarded to eliminate possible f_0 perturbations.

All f_0 values were log-z-score normalised for each speaker based on the given speaker's mean f_0 value. Duration values were also normalised to neutralise the potential final lengthening effect of the final syllables as opposed to the first syllables. To achieve this, actual duration values were divided by mean duration values in the same position (either the first or the second syllable) within each speaker.

The normalised duration ratio (σ_1/σ_2), mean intensity differences ($\sigma_1 - \sigma_2$), and normalised mean f_0 differences ($\sigma_1 - \sigma_2$) between the two syllable rhymes were calculated. The advantage of using the relative values over the absolute ones is to better capture the relative changes between the two syllables, as prominence is argued to represent a relation between elements in an utterance rather than an absolute threshold [14].

The tone contour patterns of each token across different focus conditions were first compared and labelled manually by a phonetically trained native Xiangshan speaker. Tokens with phonetically and perceptually different tone contours were labelled with different tone sandhi categories according to the actual realisations of tone contours.

Linear mixed-effects models were used to compare the duration ratio and mean intensity differences under different focus conditions, using the *Lme4* package [15] in R [16].

3. RESULTS

3.1. Gradient acoustic cues

Linear mixed-effects models were constructed to analyse the relations between gradient acoustic cues (i.e., mean intensity differences and normalised duration ratio) and focus conditions. FOCUSCONDITION (broad, first, second, disyllable) was taken as the fixed effect, and SPEAKER and ITEM were held as the random effects with random intercepts. All the 1,216 tokens were analysed.

The overall duration ratio (σ_1/σ_2) averaged across all the speakers and morphosyntactic structures in four focus conditions are given in Figure 1. The broad focus condition serves as the baseline for observing the durational effect induced by focus. Linear mixed-effects analysis revealed that corrective focus on the first syllable significantly increases the duration ratio (Estimate = 0.06, $t = 2.93$, $p < 0.01$**) and that on the second syllable significantly decreases the duration ratio (Estimate = -0.08, $t = -3.98$, $p < 0.001$***). It indicates that corrective focus lengthens the relative duration of the focused syllable. Corrective focus on the whole disyllable, however, does not affect the duration ratio significantly compared to the broad focus condition (Estimate = -0.03, $t = -1.66$, $p = .10$).

The mean intensity difference ($\sigma_1 - \sigma_2$) patterns for all the tokens and speakers are depicted in Figure 2.

Similarly, the relative intensity of the focused syllable is significantly raised when either the first or the second syllable is focused (first: Estimate = 1.02, t = 3.4, p < 0.001***; second: Estimate = -1.15, t = -3.88, p < 0.001***). When the whole disyllable is focused, however, the intensity difference between the two syllables does not differ significantly from the broad focus condition (Estimate = -0.33, t = -1.1, p = .27).

Figure 1 Duration ratio (σ_1/σ_2) by focus condition

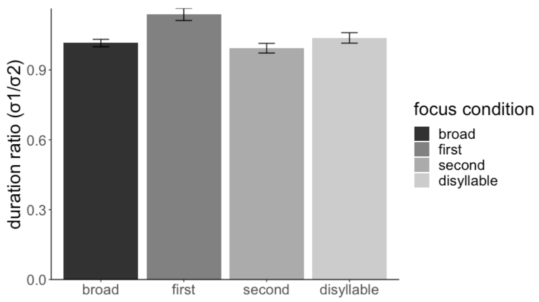
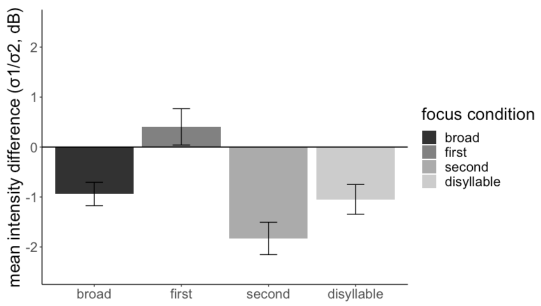


Figure 2 Mean intensity difference (σ_1 - σ_2) by focus condition



3.2. Tone sandhi patterns

Due to the space limit, only the tone sandhi patterns for lexical compounds and Modifier-Head phrases with LHL-σ_2 combinations, as well as their interaction with corrective focus, are reported in this paper. Eight disyllables produced by each speaker were selected, covering all the LHL-σ_2 tone combinations and focus conditions, i.e., 256 tokens in total (8 tokens * 4 focus conditions * 8 speakers).

Figure 3 shows the three sandhi patterns found for those tokens when there is no corrective focus present. Across all speakers and tokens, sandhi1 pattern is the predominant pattern for the LHL-H, LHL-LHL, and LHL-LH tone combinations (78.7% of those combinations), and only appears a few times for LHL-HL. Sandhi2 pattern is found for all except the LHL-HL underlying tone combination, and only 5.7% of all the tokens show this pattern. Sandhi3 pattern is exclusively found in the LHL-HL combination and is the predominant one for this combination (76.5%).

In the current experiment, there are four possible outcomes under focus: broad sandhi on the disyllable (BS, 73%), focus realising syllable 2 sandhi as a rise (σ_2R, 13%), focus realising syllable 2 sandhi as a fall (σ_2F, 9.9%), and the focused syllable carrying its underlying tone (UT, 4.1%). Tokens under the BS type show the sandhi patterns they adopt in the broad focus condition, suggesting that focus has no effect on sandhi patterns for this type. Second syllables of tokens in the σ_2R and σ_2F categories exhibit different sandhi tones when they are under corrective focus, as opposed to when they are under broad focus. The disyllables under the σ_2R type, which show a sandhi2 pattern (L-LML) under broad focus, exhibits a sandhi1 pattern (L-LM) under corrective focus regardless of the focus position. The opposite is observed for σ_2F type, as can be seen in Figure 4.

Figure 3 Tone sandhi patterns (broad sandhi—BS) for LHL-σ_2 disyllables under broad focus

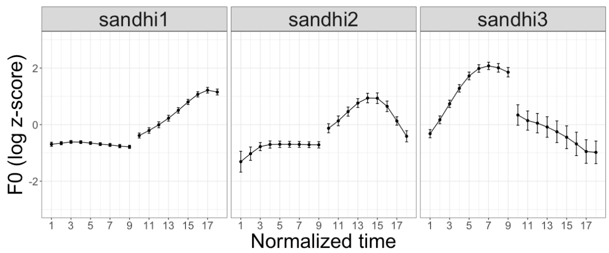
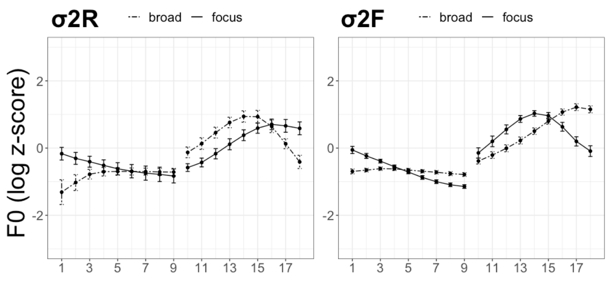


Figure 4 Focus realising σ_2 sandhi as a rise or a fall for LHL-σ_2 disyllables under corrective focus



The fourth type (UT) realises the focused syllable with its underlying tone (Table 2). For example, as shown in the highlighted row in Table 2, a token with underlying LHL-H tones exhibits a sandhi1 pattern (L-LH) under broad focus. However, when the second syllable receives corrective focus, it is realised with its underlying tone H, while the first tone still preserves its broad focus sandhi tone L. This type (with only one exceptional token where both syllables are realised with their underlying tones) is not found when the whole disyllable is focused.

Table 2 Focused syllable carrying its underlying tone under corrective focus for LHL-σ_2 disyllables

Focus	UR tones	Broad sandhi	Focus sandhi
σ_1	LHL-HL	LH-ML	**LHL**-ML
σ_1	LHL-H	L-LH	**LHL**-LH
σ_2	LHL-HL	LH-ML	L-**HL**
σ_2	LHL-H	L-LH	L-**H**

3.3. Tone sandhi changes and gradient acoustic cues

To analyse the cue-trading relations between tone sandhi changes and gradient acoustic cue changes, linear mixed-effects models were constructed with mean intensity difference and normalised duration ratio as the dependent measures, respectively. The fixed factors were FOCUSCONDITION (broad, first, second, disyllable), SANDHITYPE (BS, σ_2R, σ_2R, UT), and their interaction; SPEAKER and ITEM were the random factors with random intercepts. Models with only main effects were compared to the same models with interaction through Likelihood Ratio Tests, where the p-value is derived. Post hoc Tukey HSD tests were conducted when interactions were significant.

The results of mean intensity differences exhibit a significant interaction between FOCUSCONDITION and SANDHITYPE ($\chi2(8) = 21.428$, $p = 0.006$**). Among all the pairwise comparisons, the UT sandhi type displays significantly lower mean intensity differences ($\sigma_1 - \sigma_2$) compared to the other three sandhi types (σ_2R, σ_2F, BS), as shown in the post hoc tests results (σ_2F – UT: Estimate = 6.103, t = 3.233, p = 0.008**; σ_2R – UT: Estimate = 5.359, t = -2.605, p = 0.049*; BS – UT: Estimate = 5.85, t = 3.734, p = 0.002**).

For normalised duration ratio, no statistical significance is found for the main effect of SANDHITYPE ($\chi2(3) = 5.301$, $p = .151$) or the interaction between FOCUSCONDITION and SANDHITYPE ($\chi2(8) = 11.07$, $p = 0.198$).

4. DISCUSSION AND CONCLUSION

The current study has established that corrective focus in Xiangshan Chinese can be realised both phonetically and phonologically. Phonetically speaking, the focused monosyllable within a disyllable domain is generally characterised by a greater relative intensity and lengthened relative duration compared to that under broad focus. This finding aligns with commonly reported prosodic ways of encoding focus in previous research [1], [17], [5], [6]. Corrective focus on the whole disyllable, however, does not induce too much change on the intensity differences and duration ratio between the two syllables.

On the other hand, corrective focus in Xiangshan Chinese can affect phonological tone sandhi patterns. Among disyllabic lexical compounds and Modifier-Head phrases with LHL-initial underlying tone combinations, three divergent tone sandhi change types have been identified under corrective focus: (a) a L-LHL broad sandhi pattern is realised as a L-LH pattern; (b) a L-LH broad sandhi pattern is realised as a L-LHL pattern; (c) the focused syllable is realised as its underlying tone, while the non-focused one maintains its broad sandhi tone. It should be noted that the two sandhi patterns L-LHL and L-LH in types (a) and (b) are not evenly distributed in either broad focus or corrective focus conditions: six out of eight speakers predominantly use the L-LH pattern under broad focus, and switch to the L-LHL pattern under corrective focus; the remaining two speakers show the opposite patterning. It suggests that the two patterns may not be free variants that are randomly produced by the speaker, but are rather associated with the presence or absence of the corrective focus. More data need to be investigated before a reliable account can be made for such sandhi differences.

More interestingly, this study found that focus can potentially change the phonological phrasing, blocking the sandhi for the focused syllable in Xiangshan Chinese. This blocking strategy has been found in other Chinese languages with tone sandhi phenomena, but is usually reported to block sandhi for both syllables [8], [9]. In the present study, Xiangshan Chinese tends, when one syllable is focused, to apply broad sandhi to the underlying disyllable, determining the shape of the unfocused syllable, and then to realise the focused syllable with its underlying tone. As far as the authors know, this has not been found in any other Chinese dialect.

Moreover, this study discovers that the focused second syllable which carries its underlying tone under focus has a significantly higher relative intensity compared to the other three sandhi types. This further confirms an existing correlation between the phonetic and phonological strategies.

To sum up, this study shows that corrective focus in Xiangshan Wu Chinese induces a higher relative intensity and lengthened duration on the focused syllable. Tone sandhi patterns can be realised differently under corrective focus, and in some cases focus potentially blocks the sandhi for the focused syllable while having no effect on the sandhi of the non-focused syllable. There is a potential cue-trading relation between sandhi and gradient acoustic cues.

5. REFERENCES

[1] W. E. Cooper, S. J. Eady, and P. R. Mueller, ‘Acoustical aspects of contrastive stress in question–answer contexts’, *J. Acoust. Soc. Am.*, vol. 77, no. 6, pp. 2142–2156, Jun. 1985, doi: 10.1121/1.392372.

[2] S. Baumann, M. Grice, and S. Steindamm, ‘Prosodic marking of focus domains-categorical or gradient’, in *Proceedings of speech prosody*, Dresden Germany, 2006, pp. 301–304.

[3] S.-A. Jun and H.-J. Lee, ‘Phonetic and phonological markers of contrastive focus in Korean’, in *5th International Conference on Spoken Language Processing (ICSLP 1998)*, ISCA, Nov. 1998, p. paper 1087-0. doi: 10.21437/ICSLP.1998-151.

[4] E. Selkirk, ‘Bengali Intonation Revisited: An Optimality Theoretic Analysis in which FOCUS Stress Prominence Drives FOCUS Phrasing’, in *Topic and Focus*, C. Lee, M. Gordon, and D. Büring, Eds., in Studies in Linguistics and Philosophy, vol. 82. Dordrecht: Springer Netherlands, 2007, pp. 215–244. doi: 10.1007/978-1-4020-4796-1_12.

[5] Y. Chen, ‘Durational adjustment under corrective focus in Standard Chinese’, *J. Phon.*, vol. 34, no. 2, pp. 176–201, Apr. 2006, doi: 10.1016/j.wocn.2005.05.002.

[6] Y. Chen and C. Gussenhoven, ‘Emphasis and tonal implementation in Standard Chinese’, *J. Phon.*, vol. 36, no. 4, pp. 724–746, Oct. 2008, doi: 10.1016/j.wocn.2008.06.003.

[7] B. Ling and J. Liang, ‘The nature of left- and right-dominant sandhi in Shanghai Chinese—Evidence from the effects of speech rate and focus conditions’, *Lingua*, vol. 218, pp. 38–53, Jan. 2019, doi: 10.1016/j.lingua.2018.02.004.

[8] M. Y. Chen, *Tone Sandhi: Patterns across Chinese Dialects*, 1st ed. Cambridge University Press, 2000. doi: 10.1017/CBO9780511486364.

[9] E. Selkirk and T. Shen, ‘Prosodic domains in Shanghai Chinese’, in *The Phonology-Syntax Connection*, Chicago: University of Chicago Press, 1990, pp. 313–337.

[10] J. Tian and J. Kuang, ‘The phonetic realization of contrastive focus in Shanghainese’, in *Speech Prosody 2020*, ISCA, May 2020, pp. 265–269. doi: 10.21437/SpeechProsody.2020-54.

[11] Y. Shi, ‘Tones and Disyllabic Tone Sandhi in Xiangshan Chinese Dialect’, University of Cambridge, 2020.

[12] P. Boersma and D. Weenink, ‘Praat: doing phonetics by computer’. 2018.

[13] Y. Xu, ‘ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis’, in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013, pp. 7–10.

[14] J. Terken and D. Hermes, ‘The Perception of Prosodic Prominence’, in *Prosody: Theory and Experiment*, M. Horne, Ed., in Text, Speech and Language Technology, vol. 14. Dordrecht: Springer Netherlands, 2000, pp. 89–127. doi: 10.1007/978-94-015-9413-4_5.

[15] D. Bates, M. Mächler, B. Bolker, and S. Walker, ‘Fitting Linear Mixed-Effects Models using lme4’. arXiv, Jun. 23, 2014. Accessed: Jan. 03, 2023. [Online]. Available: http://arxiv.org/abs/1406.5823

[16] R Core Team, ‘R: A language and environment for statistical computing’. Vienna, Austria, 2018. [Online]. Available: https://www.r-project.org/

[17] S. Baumann, M. Grice, and S. Steindamm, ‘Prosodic Marking of Focus Domains - Categorical or Gradient?’.