10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Revisiting focus production in Mandarin Chinese: Some preliminary findings

Yike Yang[1], *Si Chen*[1,2]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong S.A.R., China

[2]The Hong Kong Polytechnic University-Peking University Research Centre on Chinese Linguistics, Hong Kong S.A.R., China

yi-ke.yang@connect.polyu.hk, sarah.chen@polyu.edu.hk

Abstract

Prosodic focus has been well documented in many languages, and various acoustic cues have been identified in focus production. However, the issue of focus domain has not been thoroughly studied. This study investigated the production of prosodic focus in Mandarin declarative sentences, and designed stimuli with complex sentence subjects and with different focus widths. Eleven native speakers of Mandarin participated in the recording experiment. Production data with various focus conditions were elicited with precursor questions and then analysed with linear mixed-effects modelling. Our data revealed focus-induced change of F0, duration and intensity values in pre-focus, on focus and post-focus regions. The results suggest that focus size may not interfere with focus realisation in Mandarin. Concerning the role of F0 range in Mandarin focus marking, we provided conflicting results compared with previous studies. Moreover, it is suggested that focus realisation in non-sentence-final positions and within complex nominal phrases should be considered for a better understanding of focus domain.

Index Terms: speech production, prosody, focus, focus domain, Mandarin Chinese

1. Introduction

Focus is used to separate presupposed and new information [1], and it can be conveyed by various linguistic means (e.g., syntax, morphology, and prosody) [2]. Prosodic focus makes use of acoustic cues to highlight or emphasise part of an utterance [3]. The focused constituent is generally realised with an expanded F0 range, lengthened duration and increased intensity [3], [4], while the post-focus component may be associated with reduced or compressed F0, duration and intensity [5], [6], the phenomenon of which has been further coined as post-focus compression [7]. The pre-focus counterpart, on the other hand, is reported to remain intact [5], [8]. However, the implementation of prosodic focus varies cross-linguistically [8]–[17].

Mandarin Chinese (Mandarin) is a tone language that uses F0 to distinguish lexical items. Several studies have examined the interaction between prosodic focus and lexical tones in Mandarin. Their results suggested that the local shape of F0 contours is maintained for the lexical tones, while focus is conveyed by F0 range and is manipulated for both on-focus and post-focus components [18], [19]. Also, when focus is placed at the end of a Mandarin sentence, this sentence with narrow focus does not acoustically differ from a sentence with broad focus, which in turn makes it difficult to distinguish these two sentences in perception [3], [20]. Note that both studies adopted simple three-word sentence structures, with five syllables [3] and six syllables [20] in one sentence, respectively, and the sentence-final focus was always placed on a disyllabic word. It is yet unknown whether the pattern will be different when focus is placed on a larger unit.

This brings us to the issue of focus domain, which receives some attention in speech perception (e.g., focus projection in [21]) but has rarely been studied in speech production. [22] included stimuli varying in focus breadth and found that VP focus has a lower F0 peak and shorter duration values than narrow object focus in Korean. To the best of our knowledge, no study has worked on issues concerning focus domain in Chinese languages. This study aims to fill this gap by examining the production of prosodic focus in simple declarative subject-verb-object (SVO) sentences in Mandarin. The current study attempts to investigate the acoustic correlates of Mandarin prosodic focus when the focused components differ in size and position.

2. Methods

2.1. Participants

Eleven native speakers of Mandarin (six females, five males; aged: 24.72 ± 4.39) participated in a production experiment at the speech lab of the Hong Kong Polytechnic University. All participants started to speak Mandarin from birth and had spent most of their lives in Mandarin-speaking regions. No participants reported any history of speaking, hearing or language difficulty.

2.2. Stimuli and procedures

Following previous studies [18], [23], we used simple SVO sentences as our stimuli. Because lexical tones are irrelevant to this research, we restricted the syllables to Tone 1, the high level tone in Mandarin. To test the effect of focus size, we used complex nominals (nouns with a determiner and a classifier) as the subject. There were six target sentences with exactly the same syntactic structure, as in (1):

(1) *na zhi maomi lin shubao.*
DET CL cat carry bag
'The cat carries a bag.'

As shown in Table 1, five *wh*-questions were used to elicit various focus types. The focused components in the answers were underlined. For each target sentence, there were three repetitions. In total, we collected 990 trials (11 speakers * 5 focus types* 6 target sentences * 3 repetitions).

 10.21437/SpeechProsody.2020-53

Table 1: *Focus types and question answer pairs.*

Focus types	Precursor questions	Answers
Broad focus	*ni shuo shen me?* 'What did you say?'	*na zhi maomi lin shubao.* 'The cat carries a bag.'
Subject focus	*shei lin shubao?* 'Who carries a bag?'	*na zhi maomi lin shubao.* 'The cat carries a bag.'
Verb focus	*na zhi maomi zenme shubao?* 'What does the cat do to the bag?'	*na zhi maomi lin shubao.* 'The cat carries a bag.'
VP focus	*na zhi maomi zuo shenme?* 'What does the cat do?'	*na zhi maomi lin shubao.* 'The cat carries a bag.'
Object focus	*na zhi maomi lin shenme?* 'What does the cat carry?'	*na zhi maomi lin shubao.* 'The cat carries a bag.'

The stimuli (question and answer pairs written in Chinese characters) were randomly presented on a computer screen in E-Prime 2.0 [24]. During the recording session, the first author, a native speaker of Mandarin, asked the precursor questions to the participants, and the participants were instructed to answer them as naturally as possible. The question and answer pairs were recorded at the sampling rate of 44,100 Hz in Audacity [25] on another computer. Only the answers were processed for further analysis.

This project has been approved by the Human Subjects Ethics Sub-committee (HSESC) of the Hong Kong Polytechnic University (Reference #: HSEARS20190102001). All participants gave their written consent prior to the recording sessions.

2.3. Data processing and analysis

Following the conventions in [26], trained phoneticians manually segmented the sonorants of all the seven syllables in the target sentences in Praat [27]. Twenty time normalised F0 points and intensity points, mean, maximum and minimum F0 values as well as mean duration and intensity were extracted using the ProsodyPro Praat script [28]. The F0 values, measured in Hz, were then converted to semitones (st) individually, with mean F0 of each speaker as reference [29].

The F0 and intensity points were used for visualisation only. The remaining values were analysed with linear mixed-effects modelling using the 'lme4' package [30] in R [31], [32]. Focus type was the fixed effect, where broad focus was used as the baseline for comparison. Speaker, sentence and repetition were included as the random effects. The figures were plotted with the 'ggplot2' package [33].

3. Results

3.1. F0

Figure 1 shows the averaged F0 contours of the six target sentences and provides pairwise comparisons between the broad focus and the other focus conditions. The black curve indicates the contour of the broad focus, while the grey ones each stand for one focus type. The shaded areas represent the standard errors for each curve. The vertical dotted lines mark the boundaries between the subject and verb and between the verb and object. As shown in the figure, the subject and verb foci exhibited an increase of F0 under focus and a sharp decrease of F0 in the post-focus positions. The VP and object foci did not reveal much divergence in the focused regions, but they both showed pre-focus compression of F0 before the focused regions.

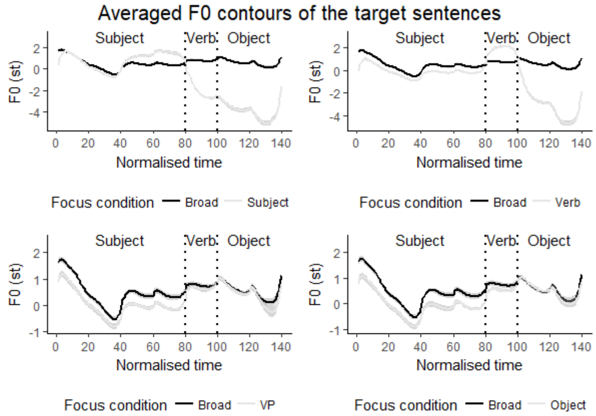

Figure 1: *Averaged F0 contours of the target sentences.*

Next, we fitted linear mixed-effects models with focus as the fixed effect to investigate if, and if so, how focus would influence various F0 values (i.e., mean F0, maximum F0, minimum F0 and F0 range). Models were constructed separately for each syntactic position, namely, subject, verb, and object. Further likelihood ratio tests comparing the models in question with null models suggested that focus affected these F0 values in different syntactic positions.

The mean F0 of the subject, verb and object under different focus conditions is shown in Figure 2. In the subject position, the model suggested that focus affected mean F0 ($\chi^2(4)$ = 152.48, p < .001), raising it by 0.42 ± 0.09 st for the subject focus ($\chi^2(968)$ = 4.49, p < .001), and lowering it by 0.56 ± 0.09 st for the verb focus ($\chi^2(968)$ = -5.00, p < .001), 0.46 ± 0.09 st for the VP focus ($\chi^2(968)$ = -4.96, p < .001), and 0.51 ± 0.09 st for the object focus ($\chi^2(968)$ = -5.48, p < .001). In the sentence medial verb position, the effect of focus was also significant ($\chi^2(4)$ = 937.97, p < .001). More specifically, the mean F0 was raised by 1.02 ± 0.09 st for the verb focus ($t(914)$ = -27.39, p < .001), and it was lowered by 2.57 ± 0.09 st for the subject focus ($t(914)$ = -27.39, p < .001) and 0.24 ± 0.09 st for the object focus ($t(914)$ = -2.73, p < .01). Focus also affected the mean F0 in the object position ($\chi^2(4)$ = 899.16, p < .001), but such effect was limited to the subject and verb foci, lowering the mean F0 by 3.05 ± 0.12 st ($t(800)$ = -25.08, p < .001) and 2.75 ± 0.11 st ($t(789)$ = -25.35, p < .001), respectively. The data suggested that focus increased the mean F0 for the subject and verb foci but not for the VP and object foci. There was evidence for both pre-focus and post-focus lowering of mean F0.

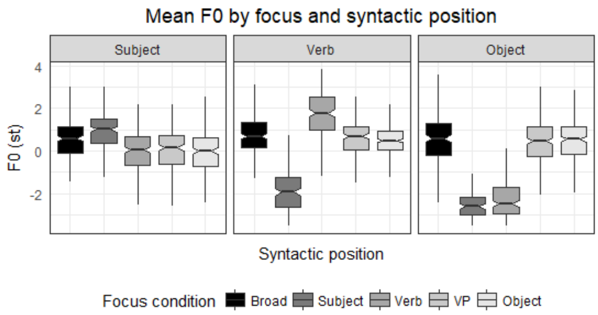

Figure 2: *Mean F0 by focus and syntactic position.*

Figures 3 presents the maximum F0 of different focus types in each syntactic position. In the subject position, focus raised the maximum F0 of subject focus ($t(968) = 4.40$, $p < .001$) and lowered the maximum F0 of the remaining focus conditions ($p < .001$ for all three conditions). In the verb position, focus raised the maximum F0 of the verb focus ($t(914) = 9.25$, $p < .001$) and lowered the maximum F0 of the subject ($t(914) = -14.36$, $p < .001$) and object foci ($t(914) = -2.73$, $p < .01$). In the object position, focus lowered only the maximum F0 of the subject ($t(885) = -19.50$, $p < .001$) and verb foci ($t(884) = -14.26$, $p < .001$).

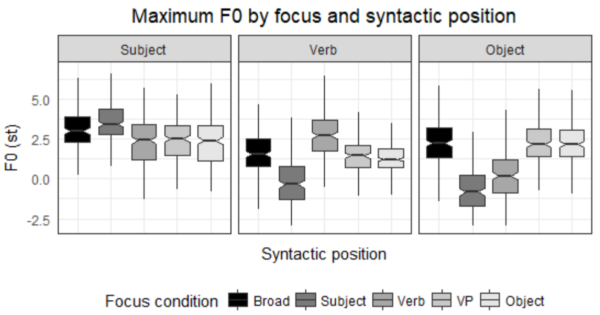

Figure 3: *Maximum F0 by focus and syntactic position.*

Figures 4 shows the minimum F0 of the subject, verb and object under different focus conditions. The only observed raising of the minimum F0 was the verb position for the verb focus (0.86 ± 0.10 st ($t(953) = 8.36$, $p < .001$)). Pre-focus lowering was found for the object focus only (0.39 ± 0.16 st ($t(968) = -2.39$, $p = .02$) in the subject position and 0.24 ± 0.10 st ($t(953) = -2.38$, $p = .02$) in the verb position). Also, post-focus lowering was evident for the subject focus (3.35 ± 0.10 st ($t(953) = -32.54$, $p < .001$) in the verb position and 3.86 ± 0.12 st ($t(851) = -31.89$, $p < .001$) in the object position) and verb focus (3.76 ± 0.12 st ($t(851) = -31.22$, $p < .001$)).

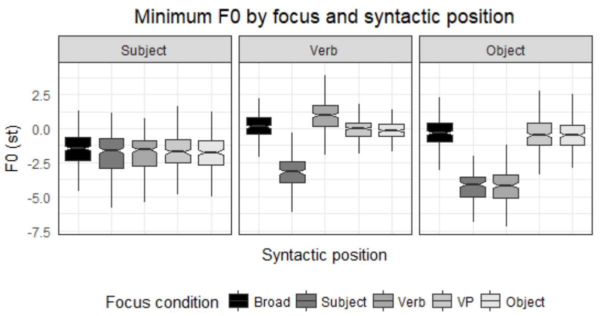

Figure 4: *Minimum F0 by focus and syntactic position.*

The F0 range was measured by the difference between maximum F0 and minimum F0 for each component and the results are plotted in Figure 5. In the subject position, there was an on-focus expansion of F0 range for the subject focus (1.00 ± 0.20 ($t(970) = 5.01$, $p < .001$)) and pre-focus compression of F0 range for other focus types ($p < .05$ for the three types) comparing with the baseline broad focus. In the verb position, there was a post-focus expansion of F0 range for the subject focus (1.23 ± 0.13 ($t(948) = 9.37$, $p < .001$)) and a marginal on-focus expansion of F0 range for the verb focus (0.22 ± 0.13 ($t(948) = 1.71$, $p = .09$)). In the object position, there was a post-focus expansion of F0 range for the subject focus (0.56 ± 0.14 ($t(855) = 3.91$, $p < .001$)) and the verb focus (1.80 ± 0.14 ($t(855) = 12.41$, $p < .001$)), as well as a marginal on-focus expansion of F0 range for the object focus (0.25 ± 0.14 ($t(855) = 1.80$ $p = .07$)). No post-focus compression of F0 range was found in our data.

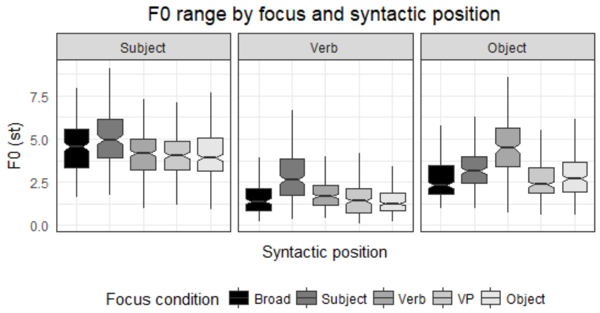

Figure 5: *F0 range by focus and syntactic position.*

3.2. Duration

The syllable duration values of the target sentences are presented in Figure 6. In the sentence initial subject position, focus lengthened one syllable by 5.34 ± 1.09 ms for the subject focus ($t(970) = 4.89$, $p < .001$) and shortened one syllable by 3.58 ± 1.09 ms and 3.35 ± 1.09 ms for the VP focus ($t(970) = -3.28$, $p < .01$) and object focus ($t(970) = -3.07$, $p < .01$), respectively. In the verb position, focus lengthened one syllable by 24.64 ± 1.69 ms for the verb focus ($t(964) = 14.50$, $p < .001$) and shortened one syllable by 5.41 ± 1.69 ms for the subject focus ($t(964) = -3.19$, $p < .01$). In the object position, the only significant effect was the shortening of the subject focus by 6.53 ± 1.66 ms for one syllable ($t(961) = -3.94$, $p < .001$).

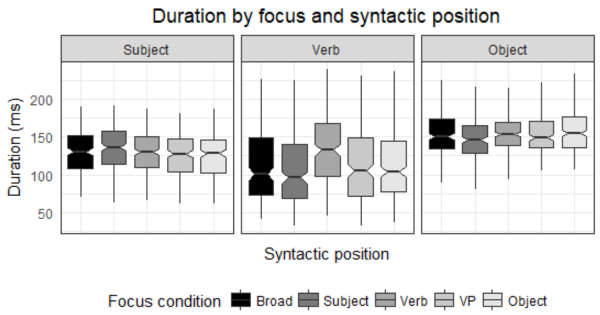

Figure 6: *Duration by focus and syntactic position.*

3.3. Intensity

For intensity, we first plotted the averaged intensity contours of the six target sentences in Figure 7, which also shows the pairwise comparisons between the broad focus and the other focus conditions. The black curve indicates the contour of broad focus, while the grey ones each stand for one focus type. The shaded areas represent the standard errors for each curve. The vertical dotted lines mark the boundaries between the subject and verb and between the verb and object. Similar to F0 contours, the subject and verb foci also showed an increase of intensity under focus and a sharp decrease of intensity in the post-focus regions. Pre-focus compression of intensity can be observed for the verb, VP and object foci.

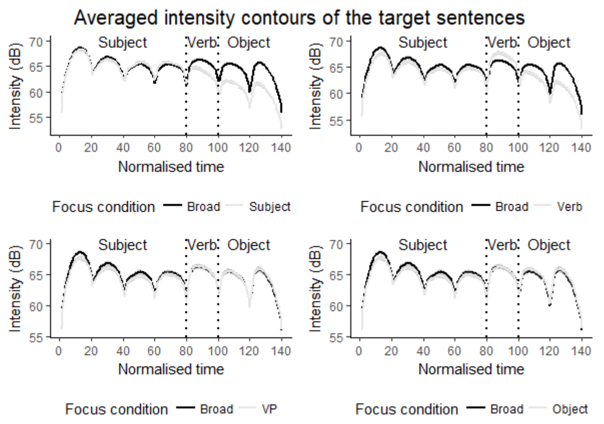

Figure 7: *Averaged intensity contours of the target sentences.*

Figure 8 presents the mean intensity of the subject, verb and object under different focus conditions. In the subject position, there was a compression of intensity for the verb focus (0.78 ± 0.19 dB (t(968) = -4.23, $p < .001$)), the VP focus (0.58 ± 0.19 dB (t(968) = -3.14, $p < .01$)) as well as the object focus (0.58 ± 0.19 dB (t(968) = -3.12, $p < .01$)). In the sentence medial verb position, focus increased the intensity of the verb focus by 1.37 ± 0.24 dB (t(919) = 5.68, $p < .001$) and decreased the intensity of the subject focus by 1.63 ± 0.24 dB (t(919) = -6.82, $p < .001$). In the object position, focus lowered the intensity of the subject focus and verb focus, by 3.59 ± 0.22 dB (t(917) = -16.70, $p < .001$) and 2.38 ± 0.22 dB (t(917) = -11.10, $p < .001$), respectively.

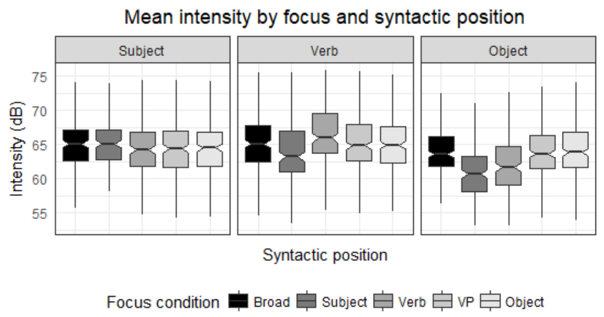

Figure 8: *Mean intensity by focus and syntactic position.*

4. Discussion

This study investigated how F0, duration and intensity values were implemented in the realisation of prosodic focus in Mandarin Chinese. Despite the complex design (determiner + classifier + head noun, which doubles the size of the subject in [18], [34]), the subject focus differed from the broad focus in all the F0 and duration values under consideration, and there was also post-focus compression of all the acoustic cues for the subject focus, suggesting that focus size may not interfere with prosodic realisation, at least not for the sentence-initial focus in our Mandarin data. The VP focus and object focus, however, did not show any difference from each other; nor did they differ from the broad focus in the focused regions, which echoed the findings from previous studies that Mandarin final focus is difficult to be distinguished from the broad focus [3], [20]. Contrary to their findings, our results revealed pre-focus compression of various cues for the VP and object foci, suggesting that the VP and object foci may be identifiable from the broad focus in the perception.

This study aimed to test the effect of focus domain, but the results proved no significant difference regardless of the focus width (VP vs object foci). One possible explanation is that we only included one type of focus condition in our test design, namely, narrow focus as an answer to a *wh*-question. A recent study showed some differences between narrow focus and corrective focus in Mandarin [35]. The VP focus might differ from the object focus if we add more focus types. Moreover, it should be noted that both the VP focus and object focus were placed at the sentence final position in our test, which according to [18], has nothing to suppress as it is the last constituent in the utterance. This seems plausible when we consider focus realisation in Korean as reported in [22]. Korean is a subject-verb-object (SOV) language, and conventionally the object precedes the verb, which makes the verb the post-focus position to be suppressed when the object is under focus. Subsequent studies should consider testing focus domain in sentence positions other than final position in Mandarin.

Although the effect of focus was significant for most of the tested cues for the verb focus, the use of F0 range was not different from broad focus, namely, when the verb was under focus, there was no expansion of F0 range. This was inconsistent with the findings in [34], where focus always induced an expansion of F0 range. Also, while previous studies suggested the post-focus compression of F0 range in Mandarin [18], [34], our results showed an expansion of F0 range in the post-focus positions for both subject and verb foci, which is similar to the pattern of some tones in Chongming Chinese [36]. It can be seen from Figures 3 and 4 that there was a lowering of maximum and minimum F0 in the post-focus regions, but such lowering effect was more robust for the minimum F0, which resulted in the expansion of F0 range. This may have to do with our stimuli, which are syllables of the high tone, allowing for more space to drop. Different tonal combinations should be included to test whether F0 range is a reliable measurement for focus production and perception.

Also, it was surprising that the subject focus did not differ from the broad focus in the mean intensity when we actually observed some differences between them from Figure 7. If we examine the curves in detail, there was a decrease of intensity in the first two syllables (the determiner and the classifier), which may have cancelled out the increase of intensity in the subject head noun. Focus realisation within complex nominal phrases should be addressed and compared with [37].

In sum, our data revealed focus-induced change of F0, duration and intensity values in pre-focus, on focus and post-focus regions, suggesting that focus size may not interfere with focus realisation in Mandarin. Concerning the role of F0 range in Mandarin focus marking, we provided conflicting results compared with previous studies, which is worth further examination. Moreover, focus realisation in non-sentence-final positions and within complex nominal phrases should also be considered for a better understanding of focus domain.

5. Acknowledgements

This study was partially supported by a research grant from Faculty of Humanities, the Hong Kong Polytechnic University (grant #: 1-ZVHJ). The authors thank all the informants for their participation and the reviewers for their comments.

6. References

[1] M. Krifka, "Basic notions of information structure," *Acta Linguist. Hung.*, vol. 55, no. 3–4, pp. 243–276, 2008.

[2] C. Gussenhoven, "Focus, mode and the nucleus," *J. Linguist.*, vol. 19, no. 2, pp. 377–417, 1983.

[3] Y. Xu, S. W. Chen, and B. Wang, "Prosodic focus with and without post-focus compression: A typological divide within the same language family?," *Linguist. Rev.*, vol. 29, no. 1, pp. 131–147, 2012.

[4] W. Cooper, S. Eady, and P. Mueller, "Acoustical aspects of contrastive stress in question-answer contexts," *J. Acoust. Soc. Am.*, vol. 77, no. 6, pp. 2142–2155, 1985.

[5] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *J. Phon.*, vol. 33, no. 2, pp. 159–197, 2005.

[6] S. Ishihara, "Syntax-Phonology Interface of Wh-Constructions in Japanese," in *Proc. 3rd Tokyo Conf. Psycholinguistics*, 2002, pp. 165–189.

[7] Y. Xu, "Post-focus compression: Cross-linguistic distribution and historical origin," in *Proc. ICPhS 2011*, 2011, pp. 152–155.

[8] M. S. Alzaidi, Y. Xu, and A. Xu, "Prosodic encoding of focus in Hijazi Arabic," *Speech Commun.*, vol. 106, pp. 127–149, 2019.

[9] W. L. Wu and Y. Xu, "Prosodic Focus in Hong Kong Cantonese without Post-focus Compression," in *Proc. Speech Prosody 2010*, 2010, pp. 1–4.

[10] A. Lee and Y. Xu, "Conditional realisation of post-focus compression in Japanese," in *Proc. Speech Prosody 2018*, 2018, pp. 216–219.

[11] Y. Yang, S. Chen, and K. Li, "Effects of focus on duration and intensity in Chongming Chinese," in *Proc. ICPhS 2019*, 2019, pp. 3578–3582.

[12] Y. Lee and Y. Xu, "Phonetic realization of contrastive focus in Korean," *Proc. Speech Prosody 2010*, pp. 1–4, 2010.

[13] S. Roessig, D. Mücke, and M. Grice, "The dynamics of intonation: Categorical and continuous variation in an attractor-based model," *PLoS One*, vol. 14, no. 5, pp. 1–36, 2019.

[14] C. DiCanio, J. Benn, and R. Castillo Garcia, "The phonetics of information structure in Yoloxóchitl Mixtec," *J. Phon.*, vol. 68, pp. 50–68, 2018.

[15] H. S. H. Fung and P. P. K. Mok, "Temporal coordination between focus prosody and pointing gestures in Cantonese," *J. Phon.*, vol. 71, pp. 113–125, 2018.

[16] T. Wang, J. Liu, Y. Lee, and Y. Lee, "The interaction between tone and prosodic focus in Mandarin Chinese," *Lang. Linguist.*, vol. 21, no. 2, pp. 331–350, 2020.

[17] B. Wang, Y. Xu, and Q. Ding, "Interactive Prosodic Marking of Focus, Boundary and Newness in Mandarin," *Phonetica*, vol. 75, no. 1, pp. 24–56, 2018.

[18] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours," *J. Phon.*, vol. 27, no. 1, pp. 55–105, 1999.

[19] Y. Chen and C. Gussenhoven, "Emphasis and tonal implementation in Standard Chinese," *J. Phon.*, vol. 36, no. 4, pp. 724–746, 2008.

[20] S. Jin, "An acoustic study of sentence stress in Mandarin Chinese," The Ohio State University, 1996.

[21] J. Bishop, "Focus projection and prenuclear accents: evidence from lexical processing," *Lang. Cogn. Neurosci.*, vol. 32, no. 2, pp. 236–253, 2017.

[22] S.-A. Jun and H.-S. Kim, "VP focus and narrow focus in Korean," in *Proc. ICPhS 2007*, 2007, pp. 1277–1280.

[23] Y. Chen, Y. Xu, and S. Guion-Anderson, "Prosodic Realization of Focus in Bilingual Production of Southern Min and Mandarin," *Phonetica*, vol. 71, no. 4, pp. 249–270, 2014.

[24] W. Schneider, A. Eschman, and A. Zuccolotto, *E-Prime User's Guide*. Pittsburgh: Psychological Software Tools Inc, 2012.

[25] Audacity Team, "Audacity(R): Free Audio Editor and Recorder." 2019.

[26] Y. Zhang, S. L. Nissen, and A. L. Francis, "Acoustic characteristics of English lexical stress produced by native Mandarin speakers," *J. Acoust. Soc. Am.*, vol. 123, no. 6, pp. 4498–4513, 2008.

[27] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." 2015.

[28] Y. Xu, "ProsodyPro - A tool for large-scale systematic prosody analysis," in *Proc. Tools and Resources for Analysis of Speech Prosody*, 2013, pp. 7–10.

[29] F. Nolan, "Intonational equivalence: an experimental evaluation of pitch scales," in *Proceedings of ICPhS 2003*, 2003, pp. 771–774.

[30] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015.

[31] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2018.

[32] RStudio Team, "RStudio: Integrated Development for R." RStudio, Inc., Boston, MA, 2016.

[33] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer, 2016.

[34] A. Yang and A. Chen, "The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children," *First Lang.*, vol. 38, no. 1, pp. 26–46, 2018.

[35] Y. Chen, "Prosodic comparisons of two types of realization of focus in mandarin," in *Proc. Speech Prosody 2018*, 2018, pp. 917–921.

[36] Y. Yang, S. Chen, and K. Li, "Pitch realization of post-focus components in Chongming Chinese," *J. Acoust. Soc. Am.*, vol. 144, no. 3, pp. 1938–1938, 2018.

[37] Y.-Y. Hsu and A. Xu, "Focus Acoustics in Mandarin Nominals," in *Proc. Interspeech 2017*, 2017, pp. 3231–3235.