Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Does prosody influence segments differently in Cantonese and Mandarin? A case study of the open vowel /a/

Yike Yang[1,2], *Si Chen*[2,3]

[1]Department of Chinese Language and Literature, Hong Kong Shue Yan University, Hong Kong S.A.R.

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

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

yyang@hksyu.edu, sarah.chen@polyu.edu.hk

Abstract

The interaction between segment and prosody has been receiving increasing attention. While speakers of European languages are found to hyper-articulate their speech to maintain the distinction between the focused and unfocused portions, little is known about focus effects on vowels in Chinese languages. This study investigated the potential interaction between prosodic focus and vowels and tested whether the effects of focus function differently in Cantonese and Mandarin, two closely related Chinese languages. In a focus production experiment, the target vowels were analysed on the duration, formants and distances. The results showed that prosodic focus influenced the open vowel /a/ differently in Cantonese and Mandarin. Although focus increased the vowel duration in both languages, the on-focus vowels were lengthened to a greater extent in Cantonese. The effect of focus was minimal on the vowel formants, especially in Cantonese. For the Euclidean distances between the vowels under broad focus and those under the remaining focus types, no difference was found, but Cantonese and Mandarin diverged in the directions in which each focus type moved away from broad focus. These results suggest that, while speakers of both languages hyper-articulate on-focus vowels, there are more differences than similarities between the two languages.

Index Terms: speech production, speech prosody, focus, vowel, Chinese

1. Introduction

While earlier research on speech production had investigated segments and prosody separately, some recent studies have begun to examine the interaction between the two [1]–[3]. Speech rate, for example, has been found to consistently influence vowel space in Catalan and Spanish, and vowels in normal speech have been reported to be more peripheral than vowels in fast speech in both languages [3].

Prosodic focus, also referred to as *stress* in some studies, highlights a portion of an utterance using prosodic means. The highlighted portion always receives prominence in the utterance, with focus-induced changes in F0 and intensity [4], [5]. However, there is still no consensus as to whether and how prosodic focus influences the production of vowels. In American English, for example, Fourakis found that focus has a marginal effect on vowel formants [6], but Erickson suggested that focused high vowels have lower first formant (F1) values and higher second formant (F2) values than unfocused ones [7]. In a prominence rating task of American English, Mo et al. demonstrated a positive correlation between F1 values and perceived vowel prominence as well as a negative correlation between F2 values and perceived vowel prominence for nearly all of the vowels tested [8].

It is predicted that speakers need to hyper-articulate their speech in order to maintain the distinction between the focused and unfocused portions [9]. According to this hypothesis, focused vowels will be less central than their unfocused counterparts, resulting in a larger vowel space. This hypothesis has been partially supported by data from European languages [3], [10], but little is known about the effects of focus on vowels in Chinese languages [11].

To fill this gap in the research, the present study aimed to investigate the potential interaction between prosodic focus and vowels and to test whether the effect of focus functions differently in Cantonese and Mandarin, two closely related Chinese languages. Specifically, three research questions are to be examined in this study: 1) Is there a difference between Cantonese and Mandarin in how prosodic focus influences *vowel duration*? 2) Is there a difference between Cantonese and Mandarin in how prosodic focus influences *vowel formants*? and 3) Is there a difference between Cantonese and Mandarin in how prosodic focus influences *vowel distances*?

2. Methods

2.1. Participants

Forty-two participants were invited to attend a recording session at the Hong Kong Polytechnic University. Half of the participants are native speakers of Hong Kong Cantonese (10 females, 11 males; aged: 20.78 ± 2.56), who were born and raised up in Hong Kong; the second half are native speakers of Mandarin (11 females, 10 males; aged: 25.23 ± 3.72), who were born in Northern China and reported having spent most of their lives in Mandarin-speaking regions. The Mandarin speakers had little, if any, exposure to Cantonese, and none of them spoke Cantonese. No participants had any history of speaking, hearing or language difficulties.

2.2. Materials and procedures

The data presented in this paper were collected in a previous study [12], which only explored prosodic focus. In the present

10.21437/SpeechProsody.2022-137

study, the data were reanalysed and the interactions between prosody and segment were tested. The open vowel /a/ was chosen as a target, as it is a common vowel shared by Cantonese and Mandarin. Next, an actual action verb, which is a monosyllabic word with a combination of an alveolar initial consonant and this vowel, was selected in each language (/tsa/ in Cantonese and /ta/ in Mandarin). Although Cantonese and Mandarin are tone languages, the syllables were restricted to the high level Tone 1 only in order to eliminate the potential effects of lexical tones on vowel production. The verbs were then embedded in meaningful simple SVO (subject-verb-object) sentences, which conform to the canonical word order in Chinese languages.

Table 1: *Sample dialogues in Cantonese.*

Focus types	Precursor questions	Target sentences
Broad focus	*nei5 waa6 mat1je5* you say what 'What did you say?'	*go2 go3 si1gei1 zaa1 cyun1baa1* that CL driver drive residential_bus 'The driver drives the residential bus.'
Subject focus	*bin1go3 zaa1 cyun1baa1* who drive residential_bus 'Who drives the residential bus?'	*go2 go3 si1gei1 zaa1 cyun1baa1* that CL driver drive residential_bus 'The driver drives the residential bus.'
Verb focus	*go2 go3 si1gei1 mat1je5 cyun1baa1* that CL driver what residential_bus 'What does the driver do to the residential bus?'	*go2 go3 si1gei1 zaa1 cyun1baa1* that CL driver drive residential_bus 'The driver drives the residential bus.'
Object focus	*go2 go3 si1gei1 zaa1 mat1je5* that CL driver drive what 'What (type of) bus does the driver drive?'	*go2 go3 si1gei1 zaa1 cyun1baa1* that CL driver drive residential_bus 'The driver drives the residential bus.'

To examine the effects of prosodic focus on Cantonese and Mandarin vowels, four focus types were manipulated, with focus placed on the whole sentence (broad focus), on the subject position (subject focus), on the verb position (verb focus) and on the object position (object focus). The broad focus condition was always the baseline condition for comparisons with the remaining three focus conditions (subject, verb and object, each corresponding to post-, on- and pre-focus). The different focus types were elicited with precursor questions asked by the experimenter, as illustrated in Table 1 (the focused constituents have been underlined). The question and answer pairs were randomly presented on a computer screen in E-Prime 2.0 [13]. The participants were briefed about the question and answer pairs prior to the recording and were instructed to answer the questions as naturally as possible. The dialogues were recorded at a sampling rate of 44,100 Hz in Audacity [14] on another computer.

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

2.3. Data processing and analysis

The segmentation of the vowel /a/ was manually performed by trained phoneticians in Praat [15]. Subsequently, duration was measured for each vowel token, and values of the first, second and third formants (F1, F2 and F3) were extracted over the middle 50ms of each vowel interval with the Burg algorithm [16] using a Praat script. To capture the distances between the baseline condition (broad focus) and the remaining three focus conditions, the Euclidean distances were calculated using Equation 1 [17]:

$$d(b,x) = \sqrt{(F1_b - F1_x)^2 + (F2_b - F2_x)^2} \quad (1)$$

where d is the distance between two points in a two-dimensional Euclidian vowel space defined by F2 on the x-axis and F1 on the y-axis, and b and x stand for the baseline vowel under broad focus condition and the target vowel under one of the remaining three focus conditions, respectively.

The duration and formant values were first analysed with linear mixed-effects modelling using the 'lme4' package [18] in R [19], [20], wherein focus type, language and gender were the fixed effects, and speaker and repetition were included as the random effects. The broad focus type was always the baseline condition. Post-hoc tests were employed to test further effects when a main effect was found in the linear mixed-effects models. The figures were plotted using the 'ggplot2' package in R [21].

3. Results

3.1. Effect of focus on vowel duration

Linear mixed-effects models were first fitted to examine the effect of focus on vowel duration. There were main effects of focus ($\chi^2[3] = 69.544$, $p < .001$) and language ($\chi^2[1] = 25.424$, $p < .001$) but no effect of gender ($\chi^2[1] = .574$, $p = .449$), suggesting that focus influences vowel duration in both languages regardless of gender. There was also an interaction between focus and language ($\chi^2[3] = 9.303$, $p < .001$), which suggests that prosodic focus functions differently when influencing duration in Cantonese and Mandarin. The mean duration values of the vowels under different focus conditions have been plotted in Figure 1, with different colours representing different focus types.

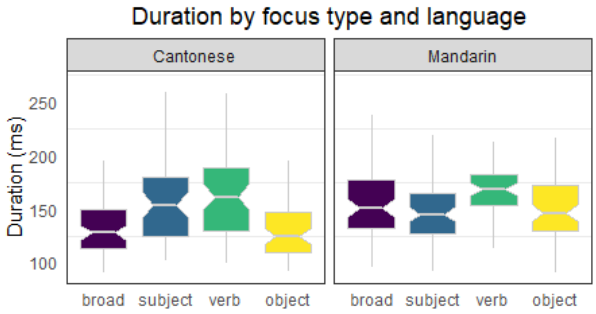

Figure 1: *Duration by focus type and language.*

Tukey's post-hoc tests were used to compare the effects of focus on duration in each language and examine if there were differences of mean duration under the various focus types. In Cantonese, the vowel of the verb was lengthened by 19.374 ± 4.557 ms and 39.780 ± 5.956 ms under subject focus ($p = .002$) and verb focus ($p < .001$), respectively, but there was no significant difference between broad focus and object focus ($p = .477$). In Mandarin, the vowel of the verb was lengthened by 24.894 ± 2.384 ms under verb focus ($p < .001$), but the focus effect was not significant under subject focus ($p = .568$) or object focus ($p = .976$).

3.2. Effect of focus on vowel formants

Linear mixed-effects models showed main effects of language ($\chi^2[1] = 47.292$, $p < .001$) and gender ($\chi^2[1] = 149.320$, $p < .001$) as well as a marginal effect of focus ($\chi^2[1] = 7.467$, $p = .058$) on F1 values. For F2 values, there was only a main effect of gender ($\chi^2[1] = 203.920$, $p < .001$). The main effects of language ($\chi^2[1] = 12.529$, $p < .001$) and gender ($\chi^2[1] = 45.726$, $p < .001$) reached significance for F3 values. The gender effect on the formants may have been a result of female speakers having consistently higher formant values than the male speakers. Therefore, in the following analysis, the data were not divided into gender groups. The formant values of the vowels produced by the Cantonese and Mandarin speakers have been plotted in Figure 2.

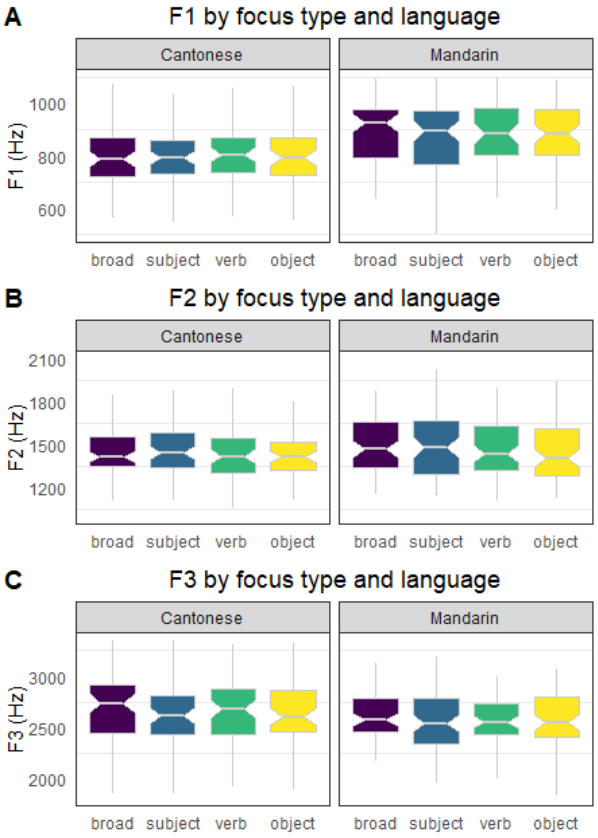

Figure 2: *Formant values by focus type and language.*

Next, further statistical tests were employed for the F1, F2 and F3 values of each focus pair for each language. For F1, post-hoc tests showed no difference between subject focus (p = .534), verb focus (p = .972) or object focus (p = .445) compared with broad focus in Cantonese. In Mandarin, verb focus raised the F1 values by 31.566 ± 10.670 Hz (p = .044), but neither subject focus (p = .326) nor object focus (p = .937) made changes to the F1 values.

For F2, there was no focus effect on the formant values under subject focus (p = .863), verb focus (p = .980) or object focus (p = .308) in Cantonese. In Mandarin, no focus effect was found for subject focus (p = .585) or verb focus (p = .118), but there was a marginal effect of focus under object focus (p = .053), which lowered the F2 by 31.421 ± 31.374 Hz.

For F3, there was a lowering of the formant values by 71.199 ± 30.520 Hz under subject focus (p = .023) but not under verb focus (p = .335) or object focus (p = .165) in Cantonese. An effect of subject focus was found in Mandarin (p = .026), lowering the formant values by 53.391 ± 23.343 Hz. However, no difference was found for the F3 of verb focus (p = .151) or object focus (p = .257) compared with broad focus in Mandarin.

3.3. Effect of focus on vowel distances

The Euclidean distances between the vowels under broad focus and the vowels under other focus types were calculated using Equation 1. The distances for each vowel pair, between broad focus and other focus types, were then submitted to linear mixed-effects models to test if there were any main effects of language, focus or gender. No effect of language ($\chi^2[1] = .022$, $p = .882$), focus ($\chi^2[2] = .430$, $p = .430$) or gender ($\chi^2[1] = 1.718$, $p = .190$) reached significance, suggesting that the vowels under different focus types had similar distances with the vowels under broad focus, regardless of language or gender. The Euclidean distances between broad focus and the other focus types have been plotted in Figure 3.

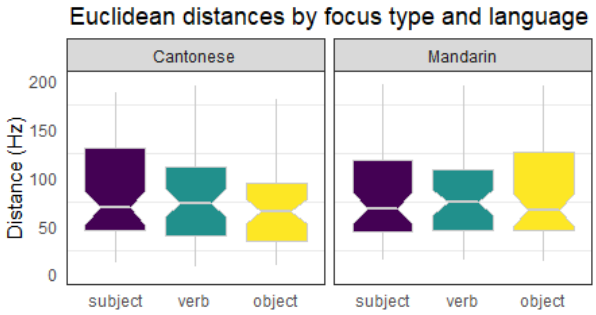

Figure 3: *Euclidean distances by focus type and language (with broad focus as the baseline).*

Again, post-hoc tests were used to investigate if there were any differences between the distances of each focus type with broad focus as the baseline. In Cantonese, the differences in distance between subject focus and verb focus (p = .091) and between subject focus and object focus (p = .098) reached marginal significance. No differences were found for the differences of distance between verb focus and object focus (p = .628). In Mandarin, there were no differences in the distances between subject focus and verb focus (p = .512), subject focus and object focus (p = .735) or verb focus and object focus (p = .854).

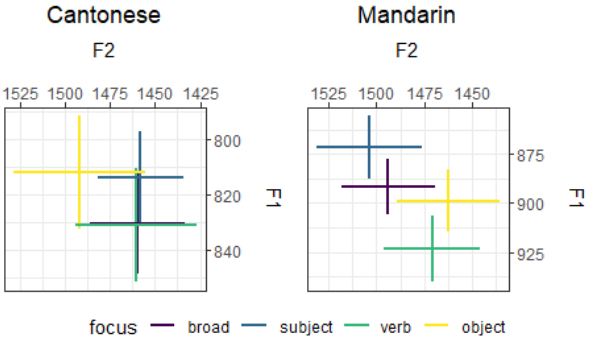

Figure 4: *Vowels under different focus types.*

Lasty, the Euclidean directions between each focus type and broad focus were compared. To clearly present the vowel

space, the vowels under different focus conditions have been plotted with double error bars in Figure 4. It is evident from the figure that although the Euclidean distances were similar for each focus type compared with broad focus, the actual Euclidean directions were very different. In Cantonese, the pre- and post-focus vowels (under object focus and subject focus) were higher than broad focus and the on-focus vowels (under verb focus) overlapped with the vowels under broad focus. The pre-focus vowels were also more frontal than the baseline vowels. In Mandarin, however, the three focus types each had a separate direction away from broad focus, with pre- and on-focus vowels moving lower and more backward and the post-focus vowels becoming higher and more frontal than the baseline vowels.

4. Discussion and Conclusion

This study aimed to examine whether and how prosodic focus influences the open vowel /a/ differently in Cantonese and Mandarin, concentrating on the focus effects on vowel duration, formants and distances.

With regard to the first research question, the data presented evidence for focus-induced durational changes. Moreover, the two-way interaction of focus and language further suggests that native speakers of Cantonese and Mandarin manipulate vowel duration differently according to focus conditions. Whereas the Cantonese speakers lengthened both on- and post-focus vowels and also shortened pre-focus vowels, the Mandarin speakers lengthened on-focus vowels only and left the pre- and post-focus vowels unchanged. The lengthening of on-focus vowels in both languages found in this study is consistent with the results of previous studies in that vowels receiving prominence were longer than the baseline vowels (e.g., [22]–[24]). Even when both languages expanded the vowel duration under focus, the extent to which the vowels were expanded was different, with Cantonese expanding the vowels to a larger extent. This points to the possibility that Cantonese makes more use of duration than Mandarin when marking prosodic focus, which is supported by the manipulation of pre- and post-focus duration in Cantonese but not in Mandarin. Previous studies on focus production also supported this claim. Duration is a crucial acoustic cue in Cantonese focus marking [25] while F0 is used more extensively than duration in Mandarin focus marking [26]. A plausible explanation is that Cantonese has a much more complex tonal system (six lexical tones plus three checked tones) than Mandarin (four lexical tones). Consequently, Mandarin speakers are allowed to manipulate F0 in focus marking, while Cantonese speakers can only resort to duration because they must make the least possible change to F0 to maintain the tonal contrasts. Note that there was also post-focus lengthening of duration in Cantonese. This has been interpreted as a carryover effect from the lengthening of the preceding constituent (the subject position under subject focus) [27], but the nature of contextual effects on syllable and word duration remains to be explored.

The second research question was concerned with the effect of prosodic focus on the F1, F2 and F3 values of the open vowel /a/. If the hypothesis regarding hyper-articulated vowels is true, then the focused vowels should have been lower and more frontal in terms of the vowel height and backness. Namely, the F1 and F2 values should have tended to be higher for the focused vowels. The vowels in this study were treated as front vowels because of the influence from the initial alveolar consonants, namely, assimilation of place of articulation. In the data, there was an increase of on-focus F1 values in Mandarin but not in Cantonese. Also, there was a post-focus decrease of F1 values in Cantonese and Mandarin. For F2 values, no effect of focus was found. This contrast suggests differences in prosody-vowel interactions in Cantonese and Mandarin: the Mandarin low vowel is lower when it receives the most prominence and it is more central when it receives the least prominence; Cantonese does not seem to change the F1 or F2 depending on prominence. Although both Cantonese and Mandarin are syllable-timed languages, Cantonese lacks lexical stress [28], whereas Mandarin has lexical stress and unstressed syllables [29]. Consequently, Mandarin speakers are accustomed to producing reduced unstressed syllables and are able to manipulate vowel formants according to different conditions. Because Cantonese does not have unstressed syllables, Cantonese speakers are expected to be more conservative when producing the segments, as is demonstrated in our data. Also, the absence of a focus effect on F2 values may be explained by the fact that the vowel /a/ in both languages is preceded by alveolar consonants, which brings about some consonant-to-vowel co-articulation in terms of place of articulation. As a result, there may be no space left for the vowels to move forward. Future studies should include different consonant and vowel pairs to better test these focus effects. For F3 values, there was only post-focus lowering, suggesting that the vowels are more rounded in the post-focus position. Rounded vowels should be used to gain a better understanding of the relationship between prosodic focus and vowel roundedness.

To answer the third research question, Euclidean distances between broad focus and the remaining focus types were calculated, and there were no differences between the focus pairs. This is not surprising, given that focus effects on F1 and F2 were minimal in Section 3.2. Further plotting of the vowels under different focus conditions indicates focus-induced different directions of movement from the baseline broad focus condition, although the differences of the Euclidean distances were not significant. As a case study, only the low vowel /a/ was included in this investigation, and the surrounding environment was not carefully controlled, which may explain the non-significant results obtained in the analyses. To provide a fuller picture of vowel movement caused by prosodic focus, studies with more vowels and various surrounding contexts are needed.

To conclude, this study showed that prosodic focus influences the open vowel /a/ differently in Cantonese and Mandarin. Focus increased the vowel duration in both languages, but the on-focus vowels were lengthened to a greater extent in Cantonese. The effect of focus was minimal on the vowel formants and vowel distances, especially in Cantonese. These results suggest that, although speakers of both languages hyper-articulate on-focus vowels, there are more differences than similarities between the two languages in terms of prosody-segment interaction, as has been shown in other areas of speech production (e.g., divergence in F0 patterns in Cantonese and Mandarin statements [30]).

5. Acknowledgements

This research was partially supported by an ASA International Student Grant awarded to the first author. The authors would like to thank all the informants for their participation and the three anonymous reviewers for their constructive suggestions.

6. References

[1] S. Graetzer, J. Fletcher, and J. Hajek, "Prosodic effects on vowel spectra in three Australian languages," in *Proc. Speech Prosody 2014*, 2014, pp. 718–722.

[2] E. Wojtkowiak, "Prosody-segment Interactions in the Acoustics of Polish Front Vowels," *Stud. Polish Linguist.*, vol. 15, no. 3, pp. 151–175, 2020.

[3] M. Nadeu, "Stress- and speech rate-induced vowel quality variation in Catalan and Spanish," *J. Phon.*, vol. 46, pp. 1–22, Sep. 2014.

[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.-C. Lee, "Prosodic focus within and across languages," University of Pennsylvania, 2015.

[6] M. Fourakis, "Tempo, stress, and vowel reduction in American English," *J. Acoust. Soc. Am.*, vol. 90, no. 4, pp. 1816–1827, Oct. 1991.

[7] D. Erickson, "Articulation of extreme formant patterns for emphasized vowels," *Phonetica*, vol. 59, no. 2–3, pp. 134–149, 2002.

[8] Y. Mo, J. Cole, and M. Hasegawa-Johnson, "Prosodic effects on vowel production: Evidence from formant structure," in *Proc. Interspeech 2009*, 2009, pp. 2535–2538.

[9] K. J. de Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 491–504, Jan. 1995.

[10] B. Lindblom, A. Agwuele, H. M. Sussman, and E. E. Cortes, "The effect of emphatic stress on consonant vowel coarticulation," *J. Acoust. Soc. Am.*, vol. 121, no. 6, p. 3802, 2007.

[11] Y. Yang, B. Li, and S. Chen, "Effects of focus and tone on vowel space in Chongming Chinese," *J. Acoust. Soc. Am.*, vol. 145, no. 3, pp. 1929–1929, Mar. 2019.

[12] Y. Yang, "First Language Attrition and Second Language Attainment of Mandarin-speaking Immigrants in Hong Kong: Evidence from Prosodic Focus," Ph.D. dissertation, Dept. Chin. & Bilingual Stud., The Hong Kong Polytech. Univ., Hong Kong, 2022.

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

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

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

[16] D. G. Childers, *Modern spectrum analysis*. New York: John Wiley & Sons, 1978.

[17] R. Wright and P. Souza, "Comparing identification of standardized and regionally valid vowels.," *J. Speech, Lang. Hear. Res.*, vol. 55, no. 1, pp. 182–93, 2012.

[18] 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.

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

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

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

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

[23] 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.

[24] Y. Chen, "Durational adjustment under corrective focus in Standard Chinese," *J. Phon.*, vol. 34, no. 2, pp. 176–201, 2006.

[25] W. L. Wu, "Cantonese prosody: Sentence-final particles and prosodic focus," University College London, 2013.

[26] Y. Yang and S. Chen, "Revisiting focus production in Mandarin Chinese: Some preliminary findings," in *Proc. Speech Prosody 2020*, 2020, pp. 260–264.

[27] J. Wang, "An acoustic study of the interaction between stressed and unstressed syllables in spoken Mandarin," in *Proc. ICSLP '96*, 1996, vol. 3, pp. 1616–1619.

[28] R. S. Bauer and P. K. Benedict, *Modern Cantonese Phonology*. Berlin: Walter de Gruyter, 1997.

[29] Y. R. Chao, *A Grammar of Spoken Chinese*. Berkeley: University of California Press, 1968.

[30] Y. Yang, S. Chen, and X. Chen, "F0 Patterns in Mandarin Statements of Mandarin and Cantonese Speakers," in *Proc. Interspeech 2020*, 2020, pp. 4163–4167.