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

The phonetic realization of contrastive focus in Shanghainese

Jia Tian, Jianjing Kuang

Department of Linguistics, University of Pennsylvania, USA
jiatian@sas.upenn.edu, kuangj@ling.upenn.edu

Abstract

This study investigates the phonetic realization of contrastive focus on disyllabic words in Shanghainese. Speakers produced disyllabic target words in three focus conditions: without contrastive focus (broad focus), contrastive focus on the first syllable, and contrastive focus on the second syllable. Results show that contrastive focus in Shanghainese is optionally realized. Among speakers who realize contrastive focus, two strategies are found, but they both involve changes in phrasing, which in turn lead to adjustments of f0, duration, and intensity. Focus tends to block tone sandhi application, so that the focused syllable does not form tone sandhi domain with either the preceding or following syllables. The effect of focus is much bigger on the second syllable of the disyllabic word.

Index Terms: focus, prosodic structure, tone sandhi, Shanghainese

1. Introduction

Focus highlights a part of the sentence. Cross-linguistically, focus can be realized via different aspects of grammatical structure, including syntactic word order, morphological markers, prosodic resources, or any combination of these, and perhaps in some cases, not at all [1, 2]. For languages that realize focus prosodically, it is very common that focus is cued by suprasegmental features: for example, expanded pitch range, longer duration, and higher amplitude [3, 4, 5]. In addition, focus can also lead to changes in phrasing. For example, in Seoul Korean, focus initiates an accentual phrase (AP), and tends to include the following words in the same AP [6, 7]. Similar patterns have been described for Japanese [8] and Greek [9], among others. Focus can also insert a boundary to its right, for example in Chichewa [1] and Tianjin Mandarin [10].

This study examines the realization of contrastive focus in Shanghainese, a Northern Wu dialect spoken in the city of Shanghai. Shanghainese has five syllable tones: T1 (53, high falling), T2 (34, high rising), T3 (13, low rising), T4 (55, short high), and T5 (12, short low rising). Tones 1, 2, and 4 are considered to have high pitch register while Tones 3 and 5 low register. When syllables are combined into prosodic words in Shanghainese, tones on non-initial syllables are first deleted, and tones of the initial syllables spread over the entire domain [11, 12]. For example, disyllabic words starting with a T1 (53) syllable show a high level tone on the first syllable and a falling tone on the second syllable (55+31) regardless of the underlying tone of the second syllable. It has been proposed [13, 14] that sandhi domain non-initial syllables in Northern Wu dialects are weaker than initial syllables, and that this strong vs weak contrast is similar to the full vs neutral tone contrast in Standard Chinese.

For Shanghainese, Selkirk & Shen [15] observed impressionistically that when a word is focused due to contrast or emphasis, the phonetic realization of its lexical tone employs a wider-than-normal pitch range. The valley of the contour tone is realized lower than normal, and the peak higher. Following a focus, the entire pitch register is lowered, and its range is considerably compressed.

More recent experimental studies showed that focus in Shanghainese is cued by suprasegmental features. Chen [16] showed that compared to non-contrastive topic (which provides old information and serves as the background about which the new information is provided), focused disyllabic words (which provides new information) show significant lengthening in both syllables. Focus also raises the max f0 and lowers the min f0 of the entire word. Chen [17] examined the effect of focus on the f0 of vowels following sandhi domain-medial consonants. She found that focus exaggerates the f0 perturbation effect at the beginning of the vowel. Focus in this study was elicited in the same way as that in [16].

Ling and Liang [18] studied the realization of contrastive focus in Shanghainese. In their study, constrastive focus was put on either the first or the second syllable of a disyllabic word. They found that for disyllabic words that undergo tone sandhi, the duration of both syllables were lengthened regardless of whether focus falls on the first or second syllable. The intensity of the focused syllable is increased while that of the other syllable is decreased. The f0 contours of both syllables are enhanced with higher f0, regardless of focus position. Contrastive focus does not alter tone sandhi domain by inserting or deleting tone sandhi boundaries. However, only results of T2/T3/T4/T5 + T2/T3/T4/T5 words (4 * 4 = 16 combinations in total) were described in detail, and it was noted that results of T1+X and X+T1 words (X represents any of the five tones) were not reported because there was considerable variation in the application of sandhi rules.

In this paper, we investigate the realization of contrastive focus in Shanghainese in more detail, and how consistent the pattern is across speakers. We will examine the duration, intensity, and f0 patterns of each syllable in a focused disyllabic word. We are especially interested in whether contrastive focus affects prosodic phrasing.

2. Method

2.1. Test Materials

Target words in this study were disyllabic names indicated as XY, with its first syllable X bearing Tone 1 (a high falling tone) and its second syllable Y bearing Tone 2 (a high rising tone). The target words were embedded in a set of three carrier sentences:

- Broad focus: This person is XY; that person is PQ.
- Contrastive focus on the second syllable: The person I mentioned is X*Y*, not X*M*. (M shares the same segments with Y, but bears Tone 3, a low rising tone.)
- Contrastive focus on the first syllable: The person I mentioned is *X*Y, not *Z*Y. (Z shares the same tone with X, but

10.21437/SpeechProsody.2020-54

differs from X in segments.)

Each speaker produced 90 sentences (30 sets) in total. Due to time limit, 9 sentences from each speaker were analyzed in the current study.

2.2. Speakers and Recording

Seventeen native speakers of Shanghainese (thirteen females and five males) participated in the study in Shanghai in the summer of 2019. Audio recordings were made in a quite room using a Shure WH30 head-mounted microphone in Audacity at a sampling rate of 44,100 Hz. Stimuli were presented in Chinese characters in OpenSesame, an open-source graphical experiment builder [19]. The three sentences of each set that contain the same target word were presented adjacently. Each sentence was presented once.

2.3. Data analysis

Sentences produced with disfluency or background noise were discarded. Rimes of both syllables of the target words were manually labeled in Praat [20]. Duration, intensity (root mean square energy), and f0 were automatically extracted by VoiceSauce [21]. For each rime, intensity and f0 were calculated at every millisecond, and then re-interpolated into nine equal time intervals. Mean values of the entire rime were also calculated. Tokens with f0 tracking errors (e.g., average f0 of 0 Hz and octave error where the f0 track abruptly changes by an octave due to irregularities in glottal pulse) were removed. All measures were z-score normalized for each speaker based on the given speakers mean value.

Linear mixed-effects (LME) models were used to compare intensity and duration. Growth curve analyses (GCA) [22] using quadratic orthogonal polynomials were used to compare the f0 contours among different focus conditions. The time terms for orthogonal polynomials were uncorrelated, hence their parameter estimates are independent of each other. The intercept term indicates the average height of the curve; the linear term indicates the overall slope of the curve; and the quadratic term indicates the sharpness of the centered peak of the curve. In both models, broad focus was the reference category. Random intercepts and slopes were included when they improved model fit. Analyses were carried out in R [23] using the lme4 package [24].

3. Results

3.1. Overall pattern

Figure 1 gives the f0 contours of both syllables of words produced in three focus conditions. Averaged across all speakers and words, contrastive focus (focus on either the first or second syllable) induces higher pitch on the second syllable of the word (first: Estimate = 0.43, t = 12.74, p < .05; second: Estimate = 0.50, t = 14.49, p < .05). There is a significantly steeper falling on the first syllable when it is focused, though the effect size is small (Estimate = -0.54, t = -6.64, p < .05).

Overall duration results are given in Figure 2. Constrastive focus induces longer duration on the second syllable than broad focus (first: Estimate = 21.45, t = 2.885, p < .05; second: Estimate = 27.00, t = 3.72, p < .05). The lengthening effect in the two focus conditions does not differ from each other (Estimate = 5.55, t = 0.75, p = .45). Duration of the first syllable is not significantly affected by focus (p = .87).

The results for intensity (see Figure 3) also show significant effect of focus on the second syllable but not on the first syllable (p = .46). Constrastive focus leads to higher intensity on the second syllable (first: Estimate = 0.54, t = 2.37, p < .05; second: Estimate = 0.90, t = 4.04, p < .05), and the two focus conditions do not differ significantly from each other (Estimate = 0.36, t = 1.59, p = .11).

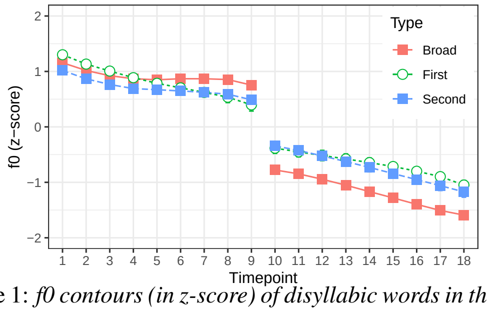

Figure 1: *f0 contours (in z-score) of disyllabic words in three focus conditions, averaged across all speakers and words. Error bars represent standard error of the mean.*

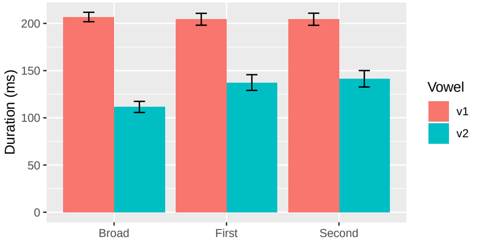

Figure 2: *Duration (in ms) of disyllabic words in three focus conditions, averaged across all speakers and words. Error bars represent standard error of the mean.*

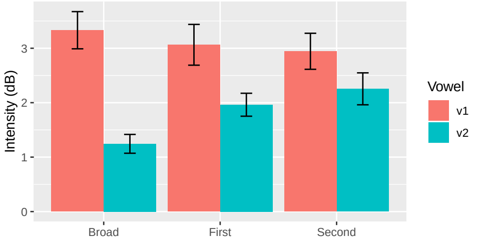

Figure 3: *Intensity (in dB) of disyllabic words in three focus conditions, averaged across all speakers and words. Error bars represent standard error of the mean.*

3.2. Individual variation

The previous analyses were based on results averaged across all tokens produced by all speakers. However, we noticed significant individual variation among our speakers, especially on the second syllable in the contrastive focus conditions. For example, standard deviation of duration of the second syllable in the broad focus condition is 42.48 ms, while those in the two focused conditions are 58.12 ms and 63.35 ms, respectively. In fact, we noticed that speakers consistently apply tone sandhi in the broad focus condition, but show different patterns in focused conditions. In this section, we examine individual variation in detail.

3.2.1. Group 1, “Sandhi” speakers

Six of our speakers do not produce much difference in focus conditions. As shown in Figure 4, target words are consistently produced with Tone 1 sandhi: the first syllable shows a level tone, and the second syllable has a slightly falling contour. Results of GCA models show slightly heightened pitch on the first syllable when it is focused (Estimate = 0.16, t = 5.07, $p < .05$), and slightly heightened pitch on the second syllable when either syllable of the word is focused (first: Estimate = 0.31, t = 12.56, $p < .05$; second: Estimate = 0.16, t = 6.41, $p < .05$). However, as shown in Figure 4, the effect size is very small.

Figure 5 and Figure 6 show the duration and intensity for both syllables of the target words in different focus conditions. Linear mixed-effects models were built for mean duration and intensity for first and second syllables separately to investigate the effect of focus condition on duration and intensity. The effect of focus condition is not significant in any model ($p > .05$).

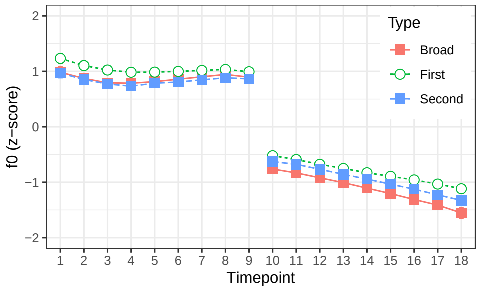

Figure 4: *f0 (in z-score) by focus conditions for Group 1 speakers. Error bars represent standard error of the mean.*

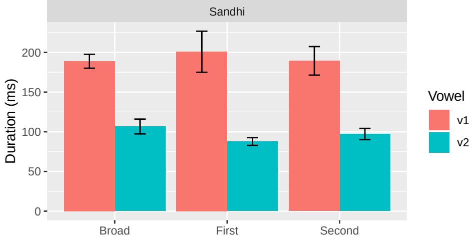

Figure 5: *Duration (in ms) by focus conditions for Group 1 speakers. Error bars represent standard error of the mean.*

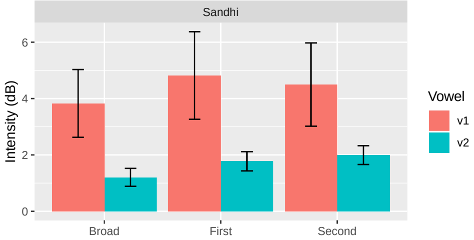

Figure 6: *Intensity (in dB) by focus conditions for Group 1 speakers. Error bars represent standard error of the mean.*

3.2.2. Group 2, "Underlying tone" speakers

Five of our speakers produce both syllables with underlying tones when there is contrastive focus on either syllable of the word. As shown in Figure 7, tokens with broad focus were produced with Tone 1 sandhi, and the f0 pattern is comparable to that in Figure 4. By contrast, tokens with contrastive focus on either the first or second syllable were produced with underlying tones. The first syllable had a falling contour, which is different from the level tone in the broad focus condition. The second syllable shows a slightly rising contour, which is different from the falling tone in the broad focus condition. These observations are confirmed by the GCA analysis. In addition, GCA on the first syllables shows that those in words whose first syllable is focused have the sharpest falling slope. For the second syllables, those in disyllables whose second syllable is focused show a higher f0 than those in disyllables whose first syllable is focused (Estimate = 0.74, t = 11.12, $p < .05$).

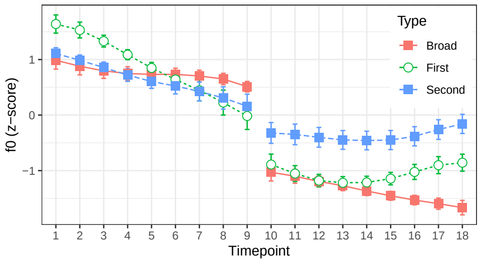

Figure 7: *f0 (in z-score) for each focus condition for Group 2 speakers. Error bars represent standard error of the mean.*

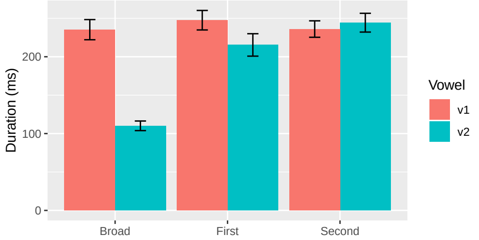

Figure 8: *Duration (in ms) for each focus condition for Group 2 speakers. Error bars represent standard error of the mean.*

Figure 8 and Figure 9 show the duration and intensity for both syllables of the target words in different focus conditions. Linear mixed-effects models show that contrastive focus induces longer duration on the second syllable. Second syllables in disyllables with contrastive focus on the second syllable are marginally longer in duration than those in disyllables with contrastive focus on the first syllable (Estimate = 28.94, t = 1.73, p = .09). Duration of the first syllable is not significantly affected by focus condition ($p > .05$). With regard to intensity, contrastive focus on the first syllable of the disyllabic word induces a marginally higher intensity on the first syllable (Estimate = 4.23, t = 1.79, p = .08). Similarly, contrastive focus on the second syllable of the disyllable induces higher intensity on the second syllable of the disyllabic word (Estimate = 3.35, t = 3.0, $p < .05$).

3.2.3. Group 3, "Falling tone" speakers

Five of our speakers show a third pattern. While they consistently produce tokens in broad focus with canonical tone sandhi, they only produce canonical tone sandhi around half of the time in the focused conditions. In the remaining half of the time, they produce a high falling tone on the second syllable of the disyllabic word, and a high level or high falling tone on the first syllable.

Figure 10 shows the f0 contours of these speakers. Results are based on values averaged across all tokens (therefore, the broad focus condition consists solely of tokens produced with canonical tone sandhi, while the two focused conditions consist of tokens produced with a high falling tone on the second syllable). f0 model shows that contrastive focus induces higher pitch

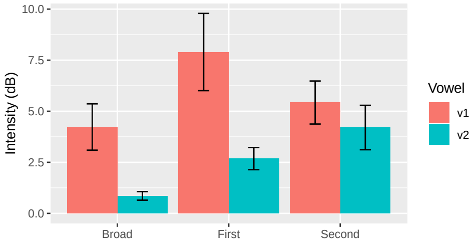

Figure 9: *Intensity (in dB) for each focus condition for Group 2 speakers. Error bars represent standard error of the mean.*

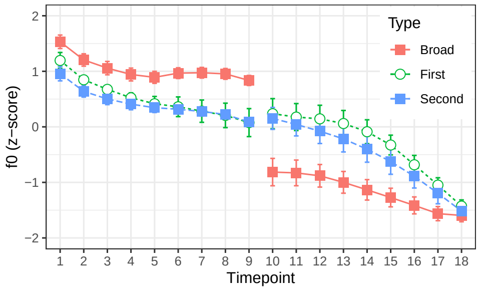

Figure 10: *f0 (in z-score) by focus condition for Group 3 speakers. Error bars represent standard error of the mean.*

and steeper falling on the second syllable (p < .05). Contrastive focus also induces lower pitch on the first syllable (first: Estimate = -0.53, t = -9.14, p < .05; second, Estimate = -0.62, t = -10.72, p < .05). Focus on the first syllable induces a slightly sharper falling on the first syllable (Estimate = -0.47, t = -2.72, p < .05).

Duration and intensity results are given in Figure 11 and Figure 12, respectively. Models show that focus induces marginally longer duration on the second syllable (first: Estimate = 30.66, t = 2.05, p = .053; second: Estimate = 26.82, t = 1.938, p = .067). The effect of focus on the first syllable is not significant (p> .05). The effect of focus on intensity is not significant for either syllable of the disyllable (p> .05).

4. Discussion and Conclusion

This study showed that there is individual variation in the realization of contrastive focus in Shanghainese. For about one third of our speakers (Group 1 speakers), focus does not induce much change in prosody.

In addition to being optionally realized, there is also variation among speakers who mark focus prosodically. For the Group 2 speakers in this study, tone sandhi is blocked when there is contrastive focus on either syllable of the disyllabic word, and both syllables are produced with their underlying tones. The change in sandhi application is signaled by f0 corresponding to underlying tones, lengthened duration in the second syllable, and increased intensity of the focused syllable. In addition to change in phrasing, focus is also cued by an increase in f0: the second syllable shows higher f0 when the second syllable of the word is focused compared with when the first syllable of the word is focused.

For the Group 3 speakers in this study, when there is contrastive focus, the second syllable shows a high falling tone and longer duration. We speculate that these changes indicate a change in phrasing: tone sandhi is partially applied (or blocked) in this case. As a result, the second syllable loses its rising contour and shows a falling pattern, but its high register (high pitch onset) is still retained.

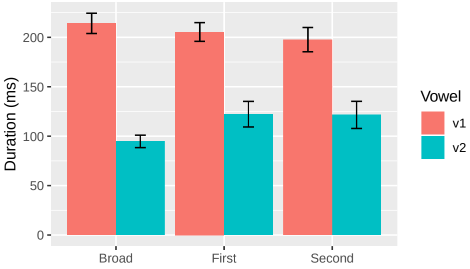

Figure 11: *Duration (in ms) by focus condition for Group 3 speakers. Error bars represent standard error of the mean.*

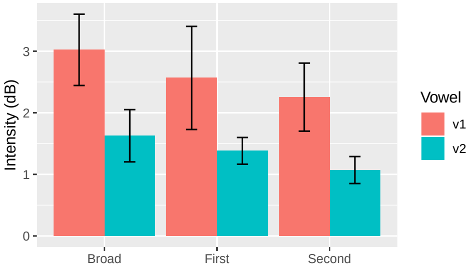

Figure 12: *Intensity (in dB) by focus condition for Group 3 speakers. Error bars represent standard error of the mean.*

Most speakers use either one of the two strategies mentioned above, and only one speaker was found to employ both.

Despite of the variation in the particular strategy used to signal contrastive focus, for both groups, focus tends to block tone sandhi application, so that the focused syllable does not form sandhi domain with either the preceding or following syllables. We also find that for both strategies, the effect of focus mostly appears on the second syllable of the disyllabic word. This phenomenon can be explained by the nature of tone sandhi in Shanghainese. In Shanghainese (and other Wu dialects in general), when syllables form sandhi domain, non-initial syllables become weaker (shorter and narrower in pitch range), while the first syllable is less affected [14, 25, 26]. Therefore, when sandhi is blocked or undone due to contrastive focus, the second syllable becomes strong again, while the first syllable does not change much.

To sum up, this study shows that contrastive focus in Shanghainese is optionally realized. For speakers who realize focus prosodically, focus is cued mainly by changes in phrasing. Focused syllables tend to form a prosodic word by themselves, without forming sandhi domains with adjacent syllables.

We notice that our results differ significantly from results of previous studies on Shanghainese (e.g. [16, 17, 18]). While we report contrastive cues to be primarily signaled by changes in phrasing, previous studies have never reported changes in phrasing (except for Ling and Liang [18], who noted in a footnote that T1+X and X+T1 words show considerable variation in tone sandhi application). We speculate that the different results are due to different tonal combinations being examined and different ways in which focus is defined (e.g., constrastive focus vs. new information, focus on a single syllable vs focus on the entire word). Future work should be done to gain a more comprehensive understanding of the phonetic realization of focus in Shanghainese.

5. References

[1] D. Büring, “Towards a typology of focus realization,” in *Information Structure: Theoretical, Typological, and Experimental Perspectives*, 2009.

[2] J. Kim and S.-A. Jun, “Prosodic Structure and Focus Prosody of South Kyungsang Korean,” *Language research*, vol. 45, no. 1, pp. 43–66, 2009.

[3] W. E. Cooper, S. J. Eady, and P. R. Mueller, “Acoustical aspects of contrastive stress in questionanswer contexts,” *The Journal of the Acoustical Society of America*, vol. 77, no. 6, pp. 2142–2156, 1985.

[4] S. Baumann, M. Grice, and S. Steindamm, “Prosodic marking of focus domains - Categorical or gradient?” in *Speech Prosody 2006*, Dresden, Germany, 2006, pp. 301–304.

[5] M. Swerts, E. Krahmer, and C. Avesani, “Prosodic marking of information status in Dutch and Italian: A comparative analysis,” *Journal of Phonetics*, vol. 30, pp. 629–654, 2002.

[6] S.-A. Jun, “The Phonetics and Phonology of Korean Prosody,” Ph.D. dissertation, Ohio State University, 1993.

[7] S.-A. Jun and H.-J. Lee, “Phonetic and phonological markers of contrastive focus in Korean,” in *Proceedings of the 5th International Conference on Spoken Language Processing*, Sidney, Australia, 1998, pp. 1295–1298.

[8] M. E. Beckman and J. B. Pierrehumbert, “Intonational structure in Japanese and English,” *Phonology Yearbook*, vol. 3, pp. 255–309, 1986.

[9] M. Baltazani and S.-A. Jun, “Focus and topic intonation in Greek,” in *Proceedings of the 14th International Congress of Phonetic Sciences*, 1999, pp. 1305–1308.

[10] Q. Li and Y. Chen, “Neutral Tone Realization in Tianjin Mandarin,” in *Proceedings of the 17th International Congress of Phonetic Sciences*, 2011, pp. 1214–1217.

[11] E. Zee and I. Maddieson, “Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis,” *UCLA Working Papers in Phonetics*, vol. 45, pp. 93–129, 1979.

[12] B. Xu and Z. Tang, *[A Description of the Urban Shanghai Dialect]*. Shanghai: Shanghai Educational Publishing House, 1988.

[13] Y. Chen, “The acoustic realization of vowels of Shanghai Chinese,” *Journal of Phonetics*, vol. 36, pp. 629–648, 2008.

[14] J. Kuang, J. Tian, and Y. Zhou, “The common word prosody in Northern Wu,” in *Proc. TAL2018, Sixth International Symposium on Tonal Aspects of Languages*, Berlin, 2018, pp. 7–11.

[15] E. Selkirk and T. Shen, “Prosodie Domains in Shanghai Chinese,” in *Phonology-Syntax Connection*. Chicago: University of Chicago Press, 1990, pp. 313–337.

[16] Y. Chen, “Prosody and information structure mapping: evidence from Shanghai Chinese,” , vol. 2, pp. 123–131, 2009.

[17] ——, “How does phonology guide phonetics in segment-f0 interaction?” *Journal of Phonetics*, vol. 39, no. 4, pp. 612–625, 2011.

[18] B. Ling and J. Liang, “Focus encoding and prosodic structure in Shanghai Chinese,” *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. EL610–EL616, 2017.

[19] S. Mathôt, D. Schreij, and J. Theeuwes, “OpenSesame: An opensource, graphical experiment builder for the social sciences,” *Behavior Research Methods*, vol. 44, pp. 314–324, 2012.

[20] P. Boersma and D. Weenink, “Praat: doing phonetics by computer,” 2018.

[21] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, “VoiceSauce: A program for voice analysis,” in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 2011, pp. 1846–1849.

[22] D. Mirman, *Growth Curve Analysis and Visualization Using R*. Boca Raton: CRC Press, 2014.

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

[24] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, G. Grothendieck, and P. Green, “lme4: Linear Mixed-Effects Models using ’Eigen’ and S4,” 2016.

[25] Y. Chen, “Revisiting the Phonetics and Phonology of Shanghai Tone Sandhi,” in *Speech Prosody 2008*, Campinas, Brazil, 2008, pp. 253–256.

[26] S. Duanmu, “Metrical and Tonal Phonology of Compounds in Two Chinese Dialects,” *Language*, vol. 71, no. 2, pp. 225–259, 1995.