

DURATIONAL ADJUSTMENT IN CANTONESE FOCUS MARKING OF NATIVE AND NON-NATIVE SPEAKERS

Yike Yang

Hong Kong Shue Yan University
yyang@hksyu.edu

ABSTRACT

The speech production of non-native speakers usually differs from that of native speakers, but very few investigations on the prosody of non-native speech have been conducted. Although duration has been regarded as an important acoustic cue for focus marking in Cantonese, precisely how native Cantonese speakers manipulate duration in different types of syllable structures remains to be explored. This study aims to fill these research gaps by examining how native and non-native speakers of Cantonese mark prosodic focus via durational adjustment. The native speakers demonstrated an asymmetrical effect of focal lengthening in different types of syllable structures in Cantonese, and the word-final nasals were expanded substantially. The non-native speakers revealed different patterns, which was attributed to their failure to fully acquire Cantonese focus marking. Discussions about the amount of focal lengthening in Cantonese and the acquisition and learnability of second language (L2) prosody are provided.

Keywords: second language, speech production, speech prosody, focus, duration

1. INTRODUCTION

1.1. Durational adjustment in focus marking

Focus is a linguistic device that introduces new information and highlights the portion of a sentence that contains this information [1]. Cross-linguistically, there are various linguistic means of focus marking [2], and there are even different ways to manipulate acoustic cues to differentiate the new information from the old information within languages that make use of prosody to mark focus [2]. More specifically, while cues related to fundamental frequency (F0), such as mean F0 and maximum F0, have been found to correlate with focus in Mandarin [3], duration is an important acoustic cue that marks focus in Cantonese [4]. However, although the durational adjustment of focus marking has been examined previously in Mandarin [5] and Yolochochitl Mixtec [6], a thorough investigation of the durational adjustment of focus marking in Cantonese is still lacking.

The effect of focus-induced lengthening is not always symmetrical. In stressed languages such as

English and Swedish, the stressed syllables often exhibit greater lengthening within the focal words than do their unstressed counterparts [7], [8]. Moreover, whether the focal lengthening is symmetrical for the segments within a syllable is a controversial issue [5], [8].

1.2. Non-native speech

Non-native speakers usually differ from native speakers in various aspects (e.g., word choice and cohesion), and there has been reports of a perceived foreign accent in non-native speakers' speech production [9]. Although some domains of linguistic knowledge have been proven to be acquirable at the end state of second language (L2) acquisition (such as syntax [10] and lexicon [11]), it has been argued that the attainment of native-like pronunciation is unlikely for late L2 learners [12]. The foreign accent in L2 ultimate attainment has been well documented, but consensus regarding which segmental and suprasegmental features may contribute to a non-native accent has not been reached. Some attempts have been made to explore the effects of segments on accent, but very few investigations have concerned the suprasegmental level [13].

1.3. This study

As discussed above, although duration has been found to be a salient acoustic cue in Cantonese focus marking, precisely how native speakers of Cantonese manipulate duration in different types of syllable structures remains to be explored. While many studies have focused on the production of L2 segments [14], there are few investigations of non-native speakers' marking of focus prosody [15]. To fill these gaps, this study examines the manipulation of duration in Cantonese focus marking by native and non-native speakers.

Two research questions will be addressed in this paper:

- 1) How do native speakers of Cantonese mark focus via durational adjustment?
- 2) Can non-native speakers of Cantonese acquire the durational adjustment patterns in Cantonese focus marking?

2. METHODS

2.1. Speakers

Forty-three participants were recruited to participate in a production experiment at a speech laboratory of a local university. Twenty-one of the participants were native speakers of Hong Kong Cantonese (10 females and 11 males, aged 20.78 ± 2.56), who were born and raised in Hong Kong; the other 22 were native speakers of Mandarin (19 females and 3 males, aged 30.14 ± 4.30), who were born in Northern China and had been exposed to Cantonese since their arrival in Hong Kong. The Mandarin speakers had arrived in Hong Kong after puberty (average age: 22.73 ± 4.21) and the average length of residence was 7.41 ± 3.11 years. The participants completed a language background questionnaire [16] prior to the recording session, according to which the non-native speakers were all fluent speakers of Cantonese. None of the participants had any history of speaking, hearing or language difficulties.

2.2. Materials and procedures

Two types of syllable structures were considered, namely consonant-vowel combinations (CV) and consonant-vowel-consonant combinations (CVC). Two monosyllabic verbs with the high-level Tone 1 were chosen for each syllable type, and the verbs were then embedded in meaningful Cantonese sentences, as shown in Sentence 1 below. The target syllables are presented in Table 1. As Cantonese is a subject-verb-object (SVO) language, the verb normally occurs in the sentence-medial position.

- (1) *go2 go3 silgeil zaal cyunlbaal*
 that CL driver drive residential_bus
 ‘That driver drives the residential route.’

Table 1: Target syllables.

Initial consonant	Vowel	Final consonant	Meaning
/ts/	/a/	N/A	to drive
/f/	/ei/	N/A	to fly
/l/	/i/	/ŋ/	to carry
/ts/	/ɔ/	/ŋ/	to assemble

Different types of focus were elicited via precursor questions that were asked by the experimenter and the participants were instructed to answer the questions as naturally as possible using the target sentences. The question and answer pairs were presented randomly on a computer screen in E-Prime 2.0 [17] and the dialogues were recorded at the sampling rate of 44,100 Hz in Audacity [18] on

another computer. As the broad focus and the verb focus were of particular relevance to the current study, only the results for these two focus types are reported in this paper.

2.3. Data analysis

The author, who is a trained phonetician, performed the segmentation of the target verbs manually at both the segmental and syllabic levels using Praat [19]. Following the segmentation, the duration of each segment and syllable was extracted automatically via a Praat script. The duration was measured in milliseconds (ms).

The duration values at the segmental and syllabic levels were analysed via linear mixed-effects modelling using the ‘lme4’ package [20] in R [21], [22]; the *focus type* was the fixed effect, and the *speaker* and *repetition* were included as the random effects. The broad focus type was the baseline condition. The native and non-native data were first analysed separately and were then compared. The figures were plotted using the ‘ggplot2’ package in R [23].

3. RESULTS

3.1. Durational adjustment of native speakers

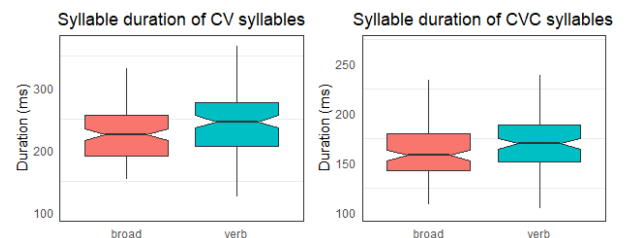


Figure 1: Syllable duration of native speakers.

We first examined how native speakers of Cantonese made durational adjustments for different syllable types at the syllabic level. Linear mixed-effects models suggested that the native speakers lengthened the syllable duration for both syllable types, but the lengthening effect was more robust for CV syllables (28.656 ± 4.908 ms; $\chi^2(1) = 31.918$, $p < .001$) than it was for CVC syllables (14.662 ± 3.115 ms; $\chi^2(1) = 21.229$, $p < .001$). The mean syllable duration of the syllables produced by native speakers is plotted in Figure 1.

Separate models for the segment duration of CV and CVC syllables were then fitted. For the CV syllables, there were main effects of focus ($\chi^2(1) = 7.733$, $p = .005$) and segment ($\chi^2(1) = 569.570$, $p < .001$) as well as a marginal interaction between focus and segment ($\chi^2(1) = 3.746$, $p = .053$) on the duration values. Post-hoc tests revealed a minor

lengthening of the consonant duration (8.94 ± 3.94 , $p = .107$) and a significant lengthening of the vowel duration (19.72 ± 3.94 , $p < .001$) under focus. For the CVC syllables, there were main effects of focus ($\chi^2(1) = 6.526$, $p = .011$) and segment ($\chi^2(2) = 235.520$, $p < .001$) on duration values, but the interaction between focus and segment did not reach significance ($\chi^2(1) = 2.027$, $p = .363$). Post-hoc tests showed no differences for the first consonant ($p = .964$) and the vowel ($p = .614$) in the two focus types, but the second consonant (syllable-final nasal sound) was slightly lengthened under focus (11.87 ± 4.23 ms, $p = .057$). The mean segment duration of the syllables produced by native speakers is plotted in Figure 2.

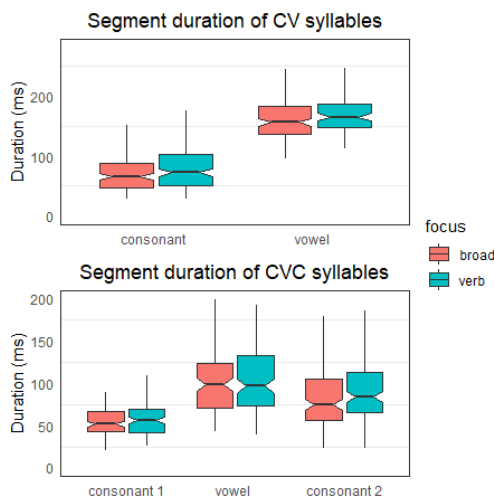


Figure 2: Segment duration of native speakers.

3.2. Durational adjustment of non-native speakers

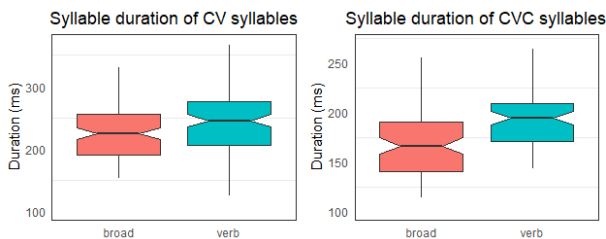


Figure 3: Syllable duration of non-native speakers.

This section reports the non-native speakers' data. At the syllabic level, there was a main effect of focus on syllable duration for the CV syllables ($\chi^2(1) = 44.038$, $p < .001$) and for the CVC syllables ($\chi^2(1) = 41.129$, $p < .001$), lengthening the on-focused syllable by 56.39 ± 7.92 ms and 26.56 ± 3.88 ms, respectively. The mean syllable duration of the syllables produced by non-native speakers is plotted in Figure 3.

Separate models were then fitted for the segment duration of CV and CVC syllables. For the CV syllables, there were main effects of focus ($\chi^2(1) =$

13.287 , $p < .001$) and segment ($\chi^2(1) = 385.17$, $p < .001$) as well as an interaction between focus and segment ($\chi^2(1) = 9.658$, $p = .002$) on the duration values. Post-hoc tests revealed a lengthening of the consonant duration (15.7 ± 5.69 , $p = .032$) and a lengthening of the vowel duration (40.7 ± 5.69 , $p < .001$) under focus. For the CVC syllables, there were main effects of focus ($\chi^2(1) = 14.023$, $p < .001$) and segment ($\chi^2(2) = 303.35$, $p < .001$) as well as an interaction between focus and segment ($\chi^2(1) = 9.843$, $p = .007$). Post-hoc tests showed no differences for the first consonant ($p = .942$) between the two focus types, but the vowel (23.66 ± 4.44 ms, $p < .001$) and the second consonant (12.12 ± 4.44 ms, $p = .071$) were lengthened under focus. The mean segment duration of the syllables produced by non-native speakers is plotted in Figure 4.

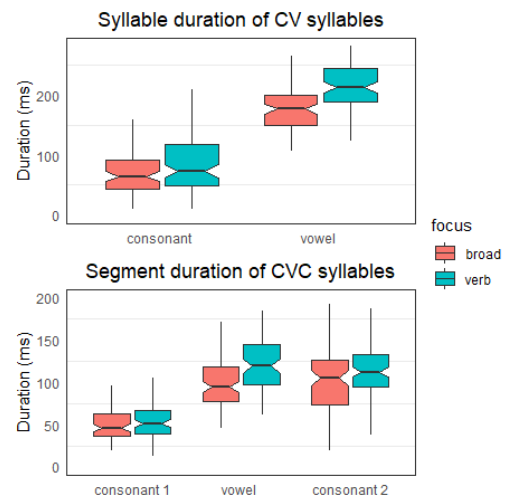


Figure 4: Segment duration of non-native speakers.

3.3. Comparing the two speaker groups

The duration patterns of the CV and CVC syllables produced by the native and non-native speakers are compared in this section.

Table 2: Duration patterns of the CV syllables in milliseconds.

	Focus type	C	V	CV
Native speakers	Broad	71.42	160.59	232.01
	Verb	80.36	180.30	260.66
	Increase	12.52%	12.28%	12.35%
Non-native speakers	Broad	70.46	175.61	246.06
	Verb	86.12	216.33	302.45
	Increase	22.23%	23.19%	22.91%

Table 2 presents the duration patterns for the CV syllables produced by the two speaker groups. The native and non-native speakers lengthened the duration of the initial consonant and the vowel equally, but the non-native speakers lengthened the

duration to a much greater extent than did the native speakers (around 12.5% for native speakers and 23% for non-native speakers).

Table 3: Duration patterns of the CVC syllables in milliseconds.

	Focus type	C1	V	C2	CVC
Native speakers	Broad	56.22	103.95	83.72	243.89
	Verb	59.67	110.62	95.59	265.88
	Increase	6.15%	6.41%	14.17%	9.02%
Non-native speakers	Broad	51.55	99.29	104.35	255.19
	Verb	55.62	122.94	116.47	295.03
	Increase	7.89%	23.83%	11.61%	15.61%

Table 3 shows the duration patterns for the CVC syllables produced by native and non-native speakers. The native speakers lengthened the first consonant and vowel slightly (6.15% and 6.41%), and the major lengthening was on the syllable-final second consonant (14.17%); this suggested that although the Cantonese speakers also marked focus via duration for CVC syllables, the lengthening was asymmetrical. Specifically, the initial consonant and the vowel were lengthened to a lesser extent and the final consonant was lengthened to a greater extent. However, the non-native speakers showed an extremely different pattern from the native speakers, as the duration of the vowel was expanded extensively (23.83%), and the initial and final consonants were expanded to a much lesser extent (7.89% and 11.61%, respectively).

4. DISCUSSION AND CONCLUSION

This study examined the manipulation of duration in Cantonese focus marking by native and non-native speakers using a production experiment. The first research question concerned how native speakers of Cantonese marked focus via durational adjustment. Two types of syllable structures were included in this study, namely CV and CVC syllables. Although there was on-focus lengthening in both syllable types, the durational adjustment for the CV and the CVC syllables differed. For the CV syllables, the consonant and the vowel were equally lengthened, and the contribution of the consonant and vowel to the lengthening of the syllable was similar. For the CVC syllables, the initial consonant and vowel underwent similar degrees of lengthening, and the final consonant experienced the greatest expansion. A question that then arose was why the segments of the CV syllables were expanded equally regardless of whether the vowel was a monophthong or diphthong, while the segments of the CVC syllables varied dramatically in the degree of focal lengthening. One may argue that the obvious lengthening of the final nasal sounds was an additive effect of word-final lengthening and focal lengthening, but if this was the

case, a similar pattern should also have been observed for the CV syllables. Moreover, an earlier study of English suggested that the lengthening effects did not add up [24]. Based on the current data, whether the patterns in the CV syllables or in the CVC syllables were the fundamental ones in Cantonese focus-induced durational adjustment is far from conclusive. Thus, it is necessary to use all the possible syllable structures available in Cantonese to obtain a full picture of Cantonese focal lengthening in future studies. In addition to syllable structure, it is interesting to examine whether there are interactions of prosodic focus and lexical tones in Cantonese focal lengthening, as has been shown in other tonal languages (e.g., [25]).

Note that the amount of focal lengthening in Cantonese was around 10% (12.35% in CV syllables and 9.02% in CVC syllables), which was far less than indicated in the results reported previously in other languages [8]. A possible explanation for this might be that the target words in this study were all sentence-medial verbs. As the lengthening effect may be sensitive to the position of the focal components [5], further studies should include more positions to test whether this was caused by the present test design or by the nature of Cantonese focus lengthening.

With regard to the second research question, we showed that the non-native speakers of Cantonese did not have similar patterns to those of the native speakers. In general, the non-native speakers expanded the syllable duration to a greater extent than did the native speakers. For the CVC syllables, whereas the native speakers mainly lengthened the final consonant and only lengthened the initial consonant and the vowel slightly, the non-native speakers mainly lengthened the vowel, thus suggesting that the non-native speakers' Cantonese was non-native-like in terms of durational adjustment. Despite the non-native speakers all being fluent speakers of Cantonese, they had failed to fully acquire focus prosody in Cantonese. This indicates that suprasegmental features may be problematic for advanced L2 speakers and require more time and effort to acquire. It also remains to be explored whether the non-native-like lengthening patterns would be salient in revealing foreign accent.

In summary, this study examined the durational adjustment in Cantonese focus marking by native and non-native speakers. The native speakers demonstrated an asymmetrical effect of focal lengthening in different types of syllable structures. There were differences between the native and non-native speakers, which were attributed to the failure of the non-native speakers to acquire focus marking in Cantonese.

5. ACKNOWLEDGEMENTS

The author would like to thank all the informants for their participation and the ICPHS committee and reviewers for their constructive suggestions. This work was partially supported by an ASA International Student Grant from the Acoustical Society of America.

6. REFERENCES

- [1] D. Büring, "Focus and intonation," in *The Routledge Companion to the Philosophy of Language*, G. Russell and D. G. Fara, Eds. London: Routledge, 2012, pp. 103–115.
- [2] Y.-C. Lee, "Prosodic focus within and across languages," University of Pennsylvania, 2015.
- [3] Y. Yang and S. Chen, "Revisiting focus production in Mandarin Chinese: Some preliminary findings," in *Proc. Speech Prosody 2020*, 2020, pp. 260–264.
- [4] W. L. Wu, "Cantonese prosody: Sentence-final particles and prosodic focus," Ph.D dissertation, Univ. College London, London, 2013.
- [5] Y. Chen, "Durational adjustment under corrective focus in Standard Chinese," *J. Phon.*, vol. 34, no. 2, pp. 176–201, 2006.
- [6] C. DiCanio, J. Benn, and R. Castillo Garcia, "The phonetics of information structure in Yoloxóchitl Mixtec," *J. Phon.*, vol. 68, pp. 50–68, 2018.
- [7] A. E. Turk and L. White, "Structural influences on accentual lengthening in English," *J. Phon.*, vol. 27, no. 2, pp. 171–206, Apr. 1999.
- [8] M. Heldner and E. Strangert, "Temporal effects of focus in Swedish," *J. Phon.*, vol. 29, no. 3, pp. 329–361, 2001.
- [9] J. E. Flege, G. H. Yeni-Komshian, and S. Liu, "Age Constraints on Second-Language Acquisition," *J. Mem. Lang.*, vol. 41, pp. 78–104, 1999.
- [10] A. Sorace and F. Filiaci, "Anaphora resolution in near-native speakers of Italian," *Second Lang. Res.*, vol. 22, no. 3, pp. 339–368, Jul. 2006.
- [11] K. Saito, "The Role of Age of Acquisition in Late Second Language Oral Proficiency Attainment," *Stud. Second Lang. Acquis.*, vol. 37, no. 4, pp. 713–743, 2015.
- [12] D. Singleton, "The Critical Period Hypothesis: A coat of many colours," *IRAL - Int. Rev. Appl. Linguist. Lang. Teach.*, vol. 43, no. 4, pp. 269–285, 2005.
- [13] K. Idemaru, P. Wei, and L. Gubbins, "Acoustic Sources of Accent in Second Language Japanese Speech," *Lang. Speech*, vol. 62, no. 2, pp. 333–357, Jun. 2019.
- [14] J. E. Flege, I. R. A. MacKay, and D. Meador, "Native Italian speakers' perception and production of English vowels," *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2973–2987, Nov. 1999.
- [15] Y. Yang, "First Language Attrition and Second Language Attainment of Mandarin-speaking Immigrants in Hong Kong: Evidence from Prosodic Focus," Ph.D dissertation, The Hong Kong Polytechnic Univ., Hong Kong, 2022.
- [16] D. Birdsong, L. M. Gertken, and M. Amengual, "Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism," *COERLL, University of Texas at Austin*, 2012. [Online]. Available: <https://sites.la.utexas.edu/bilingual/>.
- [17] W. Schneider, A. Eschman, and A. Zuccolotto, *E-Prime User's Guide*. Pittsburgh: Psychological Software Tools Inc, 2012.
- [18] Audacity Team, "Audacity(R): Free Audio Editor and Recorder." 2019.
- [19] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." 2015.
- [20] 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.
- [21] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [22] RStudio Team, "RStudio: Integrated Development for R." RStudio, Inc., Boston, MA, 2016.
- [23] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer, 2016.
- [24] 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.
- [25] 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.