Focus Prosody in Cantonese and Teochew Noun Phrases

Yu-Yin Hsu[1], *Anqi Xu*[2], *Hang Ngai*[1]

[1]Hong Kong Polytechnic University, Hong Kong
[2] Speech, Hearing and Phonetic Sciences, University College London, UK
yyhsu@polyu.edu.hk, a.xu.17@ucl.ac.uk, winniengaihang@gmail.com

Abstract

We report production data on the prosodic realization of two types of foci (constituent wh-answers, and constituent correction) of two Chinese languages that are very different from Mandarin: Hong Kong Cantonese and Teochew (a variety of Southern Min dialect spoken in Jieyang, Guangdong China). The results indicated that unlike what was reported about focus prosody at the sentential level in Cantonese, on-focus lengthening was observed with wh focus data but nothing about on-focus intensity. F0 cues were not obvious but some tendency of post-focal compression was found in F0 velocity. The Teochew data instead showed no significant acoustic distinction across different types of information structure.

Index Terms: Focus, Cantonese, Teochew, complex nominal

1. Introduction

The prosodic marking of focus involves interactions between many levels of linguistic representation [1] [2]. One of the important issues in acoustic studies has been on how focus prosody is realized and how post-focus materials reflect the information packaging acoustically. Chinese languages, in particular, play an important role because many of them exhibit different phonetic inventories and different tone systems. Comparing the realization of focus prosody among these languages will shed new light on our understanding of prosody and its interaction with information packaging.

Much work has been done for the information focus in the Standard variety of Mandarin by observing bare nouns or monosyllabic constituents in a sentence. A prosodic organization should also be sensitive to metrical organizations, and thus examining focus prosody in different syntactic environments is needed. With respect to the realization of focus prosody in a new morpho-syntactic environment [3] [4], a complex nominal containing a sequence of numeral-classifier-noun was adopted. This environment expresses closely related morpho-syntactic structure therein, naturally provides a phonetically controlled context, and functions as one argument in a sentence. It has been reported that acoustic characteristics of focus in Mandarin observed at the sentential level were also available in this morpho-syntactic environment. It showed that Mandarin speakers use a multidimensional strategy to distinguish information foci from contrastive foci, and from old information (cf. [5]); when the numeral region in a complex NP was focused, the post-focal effects (a clear reduction of intensity and F0 lowering) were observed in the noun region of that complex NP.

Different varieties of Mandarin have been reported to differ in whether post-focal compression of pitch range is available ([6]; Beijing vs. Taiwan Mandarin [7]). Besides, some data showed that the F0 range of different tones in Mandarin is not necessarily compressed post-focally, that a preceding high tone focus may impact on the F0 lowering of its following tonal segment, and that a preceding low tone focus may impact on F0 raising in the post-focal region [8]. In some other Chinese languages, the post-focus compression of F0 value was reported to be absent in Cantonese and only greater duration and intensity were found on the focused segments [9]. For Southern Min varieties, it is reported that Taiwanese does not show post-focus compression [10], but in other Southern Min varieties (Melaka and Penang [11]), the post-focal compression was available under specific segmental contexts.

Thus far, different results have been reported for different varieties of Chinese language concerning on-focus marking, and the post-focal effects. In this study, we used the morpho-syntactic environment (numeral-classifier-noun) to study one variety of Cantonese (Hong Kong Cantonese) and Southern Min (Teochew, spoken in Jieyang, East Guangdong province in China). Both languages are involved with extensive social contact with Mandarin. If language contact influences the emergence of post-focal compression [12], we expect Teochew data show post-focal effects, since Mandarin is the official language in that province and some varieties of Southern Min also show post-focal compression; we also expect similar tendency with Cantonese due to the language contact. In addition, if Cantonese and Mandarin employ different strategies on the prosodic marking of focus, we expect no post-focal compression, and only on-focus lengthening and higher intensity in Cantonese as that is what was reported at the sentential level for Cantonese [9]. We are also curious whether effects of the preceding tonal contexts on the post-focus regions [8] may be found in Chinese languages other than Mandarin. In this study, all target items were in high-level tone; therefore it will be expected that in the condition of narrow foci (i.e., corrective numeral), F0 lowering in the post-focal region (i.e., the noun region) in both languages should be observed.

2. Method

2.1. Stimuli

The target items in both experiments were three-syllable complex nominal (NP) containing a monosyllabic numeral, followed by a monosyllabic measure word, and a monosyllabic noun. Every syllable in the target item bears the same high-level tone in Cantonese and in Teochew, respectively.

10.21437/SpeechProsody.2018-194

Table 1: *Examples of target items*

Tone 1			
Cantonese	一間屋	jat1 gaan1 nguk1	“one house”
Teochew	三张车	sam1 ziang1 cia1	“three cars”

Such complex NPs in Table 1 were then embedded in sentences as the subject illustrating one of the following three different information structures: the answer to a wh-NP (ANP), the correction of a numeral (CNUM), and when the whole NP is part of the background expressing the old information (ODNP). The stimuli consist of 36 target sentences in total (12 items ×3 information structures). Stimuli were checked to make sure no potential focus/prominence bearing words (such as *only*, *exactly*, etc.) other than the target items were included in a sentence. Stimuli were randomized so that no identical target item was immediately adjacent while being presented.

Table 2: *Leading questions and target sentences of three types of information structures*

Information	**Leading questions**	**Target sentences**
Answering the whole NP (ANP)	“What can host a family?”	“One house can host a family.”
Correcting the numeral (CNUM)	“Two houses were robbed yesterday!”	“No, one house was robbed yesterday.”
NP as a part of the old information (ODNP)	“What did you just say about one house?”	“One house was sold yesterday.”

2.2. Participants

Six native speakers of Hong Kong Cantonese (3 female, 3 male; aged between 20 and 28) and six native speakers of Jieyang Teochew from East Guangdong (3 female, 3 male; aged between 20 and 40) participated in our study. None of them reported any history of hearing problems. The ethics approval for data collection and the basic geographic information were obtained before each participant started the experiment. Each participant was paid HK$60 compensation after the experiment.

2.3. Procedure

Each participant first filled out a questionnaire of language background and signed an informed consent form. During each experimental session, all of the stimuli were presented on a computer screen in a sound-attenuated room. Participants listened to the leading questions through a headphone and read the target sentences on the screen. They were instructed to respond to pre-recorded utterances as casual and natural as possible; no instruction was given to emphasize token. After a given trial, the next was presented 2 seconds later. Participants only repeated the sentence once unless they mispronounced the words or paused in the middle of utterances. Recordings were made in WAV format at a sampling rate of 44.1 kHz and a 16-bit quantization. Every participant had three practice trails before the experiment session and a 5-minute break after 18 trials. Each experiment lasted about 35 minutes.

2.4. Measurements

The target items were segmented using Praat [13]. Syllable boundaries were determined by using both visual (the waveform and spectrogram) and auditory information. The vocal pulses were manually checked and corrected when there were pitch halving or doubling and creaky voice. The acoustic measurements were generated by a custom-written script ProsodyPro 5.7.6 [14] for duration, mean intensity, maximum F0, minimum F0 and normalized F0 across speakers. The normalization of F0 was realized by dividing each syllable into 10 intervals equal in time and calculating the trimmed F0 values. The F0 value was converted from Hz to semitone scale, relative to 1 Hz by the formula: 12 ln (x / 1) / ln 2. To observe the F0 realization during tone production, the speed of fundamental frequency shift, namely, F0 velocity (in semitone/s) was also measured.

The Linear Mixed-Effects models were conducted on the duration and mean intensity using the lme4 package [15] in R [16]. We started with the simplest model, which included only the random intercepts (‘speaker’ and ‘item’). By-speaker, by-item, and speaker-by-item random slopes for information structure were then introduced if it achieved convergence and judged to be superior to less fully specified model in likelihood ratio tests. ‘Information structure’ was included as potential fixed effects. The post-hoc Tukey’s comparisons were conducted by the *multicomp* package [17] in R.

To compare the F0 curves of noun phrases with different information structure, we applied Smoothing Spline Analysis of Variance (SSANOVA models) to the time normalized F0 and F0 velocity using the gss package [18] in R [16]. For each tonal condition, SSANOVA model included information structure and normalized time and their interaction as predictors of the dependent variable, i.e., F0/ F0 velocity ~ information structure*normalized time. In all SSANOVA figures, F0 means are displayed by lines and 95% confidence intervals for F0 contours of different foci are displayed by transparent ribbons. Where the ribbons do not overlap, the difference between their represented categories can be considered significant (α =0.05) at that time point.

3. Results

3.1. Cantonese results

Figure 1 and 2 display the duration and intensity of numeral, classifier, and noun in the noun phrase. With regard to the numeral region, the mixed models did not reveal a significant main effect of information structure on the duration ($\chi2$ = 4.580, df = 2, *p* = .101) or the intensity ($\chi2$ = .046, df = 2, *p* = .977). With regard to the classifier region, neither duration ($\chi2$ = 1.938, df = 2, *p* = .380) nor intensity ($\chi2$ = 1.068, df = 2, *p* = .586) seemed to be influenced by the information structure.

However, the information structure showed a significant main effect on the duration of the noun region ($\chi2$ = 8.635, df = 2, *p* = .013) but not on the intensity ($\chi2$ = 5.045, df = 2, *p* = .080). A post-hoc test verified that the duration of the noun region in NP answers (ANP) was significantly longer than that of the noun region in the old information (p = .008) and in the correction of numeral (p = .017).

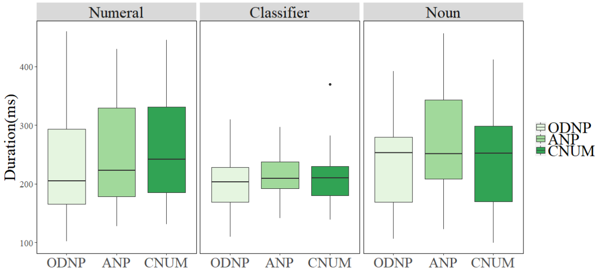

Figure 1: *Boxplot of the duration (ms) of every syllable in Cantonese NPs with three types of information structure*

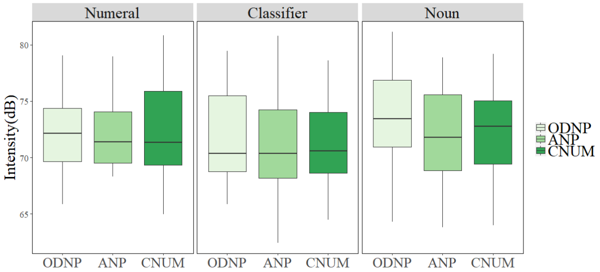

Figure 2: *Boxplot of the intensity (dB) of every syllable in Cantonese NPs with three types of information structure*

The linear mixed models indicated a significant main effect of information structure on maximum F0 (χ2 = 8.223, df = 2, *p* = .016) and a marginal effect on the minimum F0 (χ2 = 6.111, df = 2, *p* = .047) of the numeral region. As shown in Figure 3, the F0 curve of NPs with numeral focus was marginally lower than the one of old information, but the F0 curves were largely overlapped between NP answers and old information.

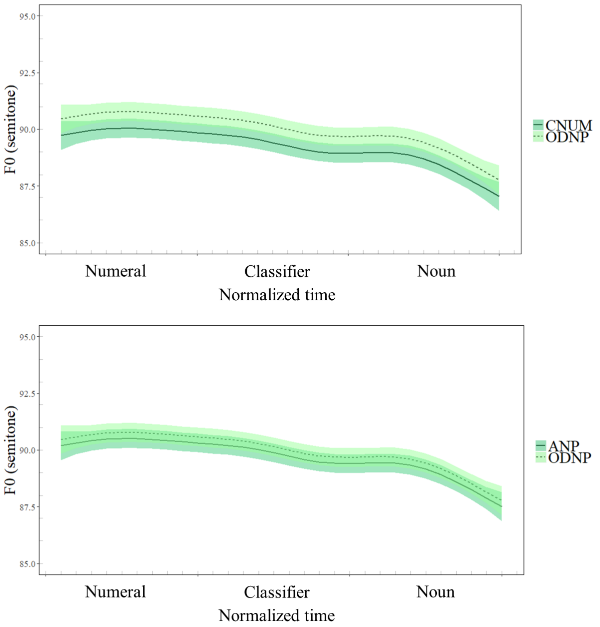

Figure 3: *Time normalized F0 contours (semitone) of Cantonese NPs with three types of information structure.*

Similarly, in Figure 4, the F0 realization of information structure indicated by F0 velocity contours across time showed that the F0 velocity of numeral corrective focus was slightly higher than ODNP at the beginning of numeral region and was significantly lower at the post-focal, noun region. No difference between NP answers and the old information was found.

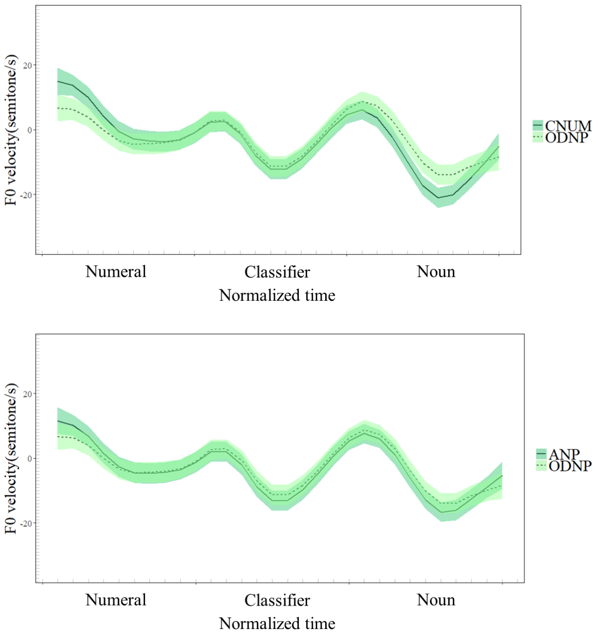

Figure 4: *Time normalized F0 velocity contours (semitone/s) of Cantonese NPs with three types of information structure.*

Considering that the prosodic phrasing may exert an impact on acoustics, we also conducted statistical analysis on the mean duration, intensity, maximum F0, and minimum F0. The main effect of information structure on F0 was confirmed in the mean maximum F0 (χ2 = 15.417, df = 2, *p* < .001) and minimum F0 (χ2 = 9.501, df = 2, *p* = .009) of focused prosodic words. Focused prosodic words in corrective numeral condition had lower maximum F0 (*p* < .001) and minimum F0 (*p* = .004) than the ones with old information. With regard to the whole nominal, the information structure had a significant effect on the duration (χ2 = 6.892, df = 2, *p* = .032). The duration of NP answers was lengthened than that of old information (*p* = .025). The information structure also affected the maximum F0 (χ2 = 13.558, df = 2, *p* = .001) and minimum F0 (χ2 = 9.628, df = 2, *p* = .008). The F0 values of old information were higher than CNUM focus (maximum F0: *p* < .001; minimum F0: *p* = .005).

3.2. Teochew results

The duration and intensity of Teochew noun phrases with different types of foci are displayed in Figure 5 and 6. The mixed models revealed a non-significant main effect of information structure on the duration (χ2 = 1.1647, df = 2, *p* = .559) and the intensity (χ2 = 5.370, df = 2, *p* = .068) of the numeral region. In the classifier region, duration (χ2 = 2.701, df = 2, *p* = .259) and intensity (χ2 = .447, df = 2, *p* = .800) did not seem to be affected by information structure. Likewise, the main effect of information structure was not significant on the duration (χ2 = .422, df = 2, *p* = .810) and the intensity (χ2 = .075, df = 2, *p* = .963) of the noun region.

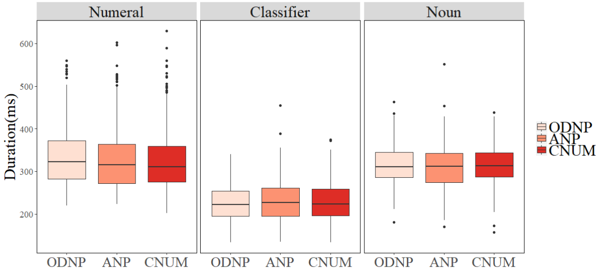

Figure 5: *Boxplot of the duration (ms) of every syllable in Teochew NPs with three types of information structure*

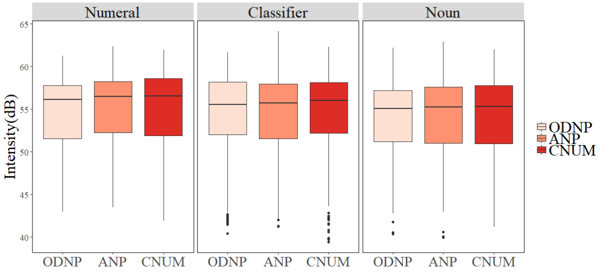

Figure 6: *Boxplot of the intensity (dB) of every syllable in Teochew noun phrases with three types of information structure*

As shown in Figure 7 and 8, unlike Cantonese data, the F0 curves and F0 velocity curves of Teochew noun phrases did not diverge across information structure. Linear mixed effects models on maximum and minimum F0 showed similar results.

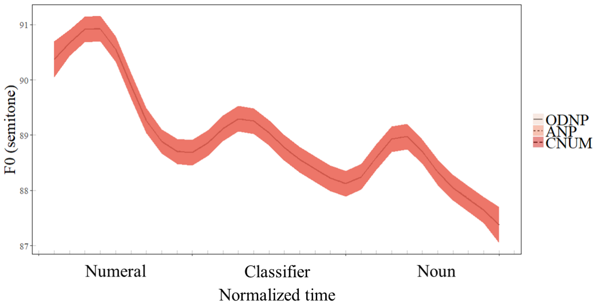

Figure 7: *Time normalized F0 contours (semitone) of Teochew NPs with three types of information structure.*

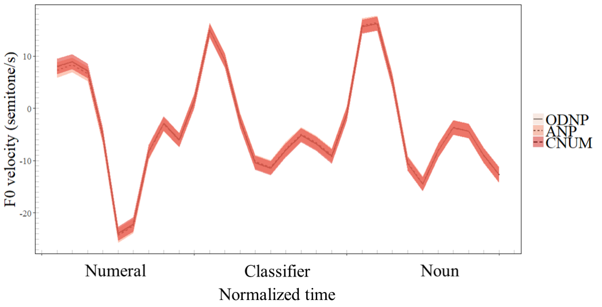

Figure 8: *Time normalized F0 velocity contours (semitone/s) of Teochew NPs with three types of information structure.*

We also conducted statistical analysis on the mean measurements of syllables in different prosodic phrasing pattern, but the main effect of information structure was not significant in all the dimensions we examined.

4. Discussion and Conclusions

Our study partially replicated results in the previous studies and provided observations from a new angle of focus prosody. Cantonese wh focus was reported to be realized by increasing the duration and intensity of on-focus constituents without remarkable change of F0 [9], and we found significant on-focus lengthening of the noun in the NP answer condition, compared with the old information. However, the increase in intensity was not observed. In our study, this may be because segments with high-level tone are disadvantageous in terms of further strengthening in intensity given that they are intrinsically higher in intensity and target items were located at the sentence-initial position. Considering prosodic phrases, we found some on-focus lowering of F0 of corrective foci, which is very different from the on-focus patterns reported with Mandarin data in the previous studies. However, given the negation word preceding target regions in our items, it requires further verification to understand this part of facts properly.

Our results cast doubts on some previous claims. First, despite extensive language contact with Mandarin (cf. [12]), our data of wh information focus in both Hong Kong Cantonese and Jieyang Teochew did not show post-focus compression. Interestingly, Cantonese corrective focus data show that the F0 velocity is modestly higher at focused region and lower at the post-focal noun region than that of the old information. In other words, Cantonese seems to exhibit a certain level of post-focal compression but in a more restricted context. Since the current study only included a small number of participants, it would be interesting to examine this again with more speakers and see if a similar tendency could reach statistical significance. To our surprise, we did not find a significant on-focus rise in the corrective focus condition. We think this may be due to the use of negation item m4 hai6 'no' preceding the target items, which had attracted most of the initial acoustic prominence in this condition, resulting in no acoustic difference in its following segments.

Also worth noting is that while the results about Cantonese corrective focus seemed to be in line with the effect that high tones impact on the end F0 lowering leading to the post-focal domain [8], we did not observe a similar effect in the Teochew study. In Teochew, all acoustic dimensions that we tested were not statistically indistinguishable across different types of information structure. Since variations in phonetic cues were not adopted to encode focal information, whether Teochew speakers resort to other types of methods needs further investigation.

5. Acknowledgments

We would like to express our gratitude to James Sneed German for his insightful suggestions of the earlier stage of this study. We would also like to thank Sui Li, Ka Keung Lee, and Kaly Sze Long Cheung for their technical support. Mistakes remaining are exclusively our own.

6. References

[1] C. Gussenhoven, "Focus, mode, and nucleus.," *Journal of Linguistics.*, pp. 377 - 417, 1983.

[2] Pierrehumbert, J. & J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in *Intentions in communication*, MIT Press, 1990, pp. 271-311.

[3] Y.-y. Hsu, "Prosody and corrective focus within the nominal domain of Mandarin Chinese," in *Proceedings of the 43th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, 2018.

[4] Y.-y. Hsu and A. Xu, "Focus Acoustics in Mandarin Nominals," in *Interspeech*, Stockholm, Sweden, 2017.

[5] Ouyang I. & E. Kaiser, "Prosody and information structure in a tone language: an investigation of Mandarin Chinese," *Language, Cognition and Neuroscience,* vol. 30, pp. 57-72, 2015.

[6] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours," *Journal of Phonetics,* vol. 27, pp. 55-105, 1999.

[7] Y. Chen, Y. Xu and S. Guion-Anderson, "Prosodic realization of focus in bilingual production of Southern Min and Mandarin," *Phonetica,* vol. 71, pp. 249-270, 2014.

[8] Y. Chen, "Post-focus suppression: Now you see it, now you don't," *Journal of Phonetics,* vol. 38, pp. 517-525, 2010.

[9] W. L. Wu and Y. Xu, "Prosodic Focus in Hong Kong Cantonese without Post-focus Compression," in *Speech Prosody*, Chicago, USA, 2010.

[10] Chen, S.-W., Wang, B., & Xu, Y., "Closely related languages, different ways of realizing focus," in *Proceedings of Interspeech*, Brighton, UK., 2009.

[11] T. Huang and F.-f. Hsieh, "Post-focus compression: All or nothing?," *The Journal of the Acoustical Society of America,* vol. 140, p. 3224, 2016.

[12] Y. Xu, S.-W. Chen and B. Wang, "Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?," *The Linguistic Review,* vol. 29, no. 1, pp. 131-147, 2012.

[13] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," [Online]. Available: http://www.fon.hum.uva.nl/praat/. [Accessed 10 11 2017].

[14] Y. Xu, "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP2013)*, Aix-en-Provence,, 2013.

[15] D. Bates, M. Maechler, B. Bolker and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[16] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria (2014) (Version 3.1.0).

[17] T. Hothorn, F. Bretz, P. Westfall, R. M. Heiberger, A. Schuetzenmeister and S. Scheibe, "Simultaneous Inference in General Parametric Models," [Online]. Available: https://cran.r-project.org/web/packages/multcomp/multcomp.pdf. [Accessed 10 11 2017].

[18] C. Gu, "General Smoothing Splines," [Online]. Available: https://cran.r-project.org/web/packages/gss/gss.pdf. [Accessed 21 11 2017].