Focus Acoustics and Prosodic Organization in Hong Kong Cantonese and Taiwan Mandarin

Yu-Yin Hsu[1], Anqi Xu[2]

[1]Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong
[2]Department of Speech, Hearing and Phonetic Sciences, University College London, UK
yu-yin.hsu@polyu.edu.hk, a.xu.17@ucl.ac.uk

ABSTRACT

The acoustic realization of focus can be influenced by the position of a focalized word in a larger constituent and by constraints on prosodic organization of an utterance. Here, we report four production studies that explore the potential effects of local prosodic organization on the realization of focus in Hong Kong Cantonese (HKC) and Taiwan Mandarin (TwM). The materials consisted of sentences in which a syntactic subject noun phrase (consisted of monosyllabic numeral, classifier, and noun) expressing either corrective or wh focus. The span of the focus constituent within such an NP was controlled using short conversations indicating either (i) the numeral only focus, (ii) the noun only focus, or (iii) the whole noun phrase focus.

Our results showed that the acoustic realization of focus in HKC and TwM extends beyond general acoustic highlighting of focus constituents, i.e., the acoustic realization of focus in HKC and TwM are influenced by constraints of prosodic organization.

Keywords: focus, prosodic organization, Hong Kong Cantonese, Taiwan Mandarin, morphosyntactic unit

1. INTRODUCTION

Cross-linguistically, the prosodic realization of focus can depend on a variety of factors, including local constraints on prosodic organization or the position of a word within a larger focus constituent. Focus concerns the way that the part of a sentence which introduces alternatives is related to the discourse context in distinct ways [17]. Prior studies on focus prosody in Chinese have emphasized the relative prominence of a simple noun's phonetic acoustics in a sentence (e.g., [3], [5], [15], [23]). While focus constituents larger than words are well-attested in many languages, little work has been done on the focus prosodic realization of larger linguistic units in Chinese languages, especially more complex phrasal domains in varieties of Chinese (cf. [4]).

In this study, we not only extend the search space from purely prominence-based marking to other phonological features such as phrase boundaries; we also consider how such focus marking may be simultaneously conditioned by syntax (cf. [14], [20], [22]). We conceptualize focus in terms of focused constituents, which may consist of one or more lexical words in a more complex syntactically and semantically related morpho-syntactic structure (i.e., a noun phrase of numeral-classifier-noun sequence, henceforth NP). Previously, with this type of structure, Beijing Mandarin's focus prosody has been reported to show on-focus rise of intensity and F0, and lengthening, together with clear post-focus compression of F0 and intensity in both Tone 1 and Tone 4, contrasting with the same NPs indicating old information (see [12] for HKC). Particularly, it was reported that the non-focused classifier may pattern with the monosyllabic numeral-focus inside of such NPs in Mandarin varieties ([10]-[11], [13]).

Therefore, we adopt this morphosyntactic structure to study two varieties of Chinese: Hong Kong Cantonese (HKC) and Taiwan Mandarin (TwM) in two types of focus: wh-narrow focus (elicited by a wh-question), and corrective-focus (elicited by a statement containing corrective information), and to seek to explore how focus marking is situated in the overall organization of the prosodic system in HKC and TwM through four production experiments. We examined whether the focus acoustic realization involved simple highlighting of words, or whether there was evidence for systematic marking of the edges of the focus constituent, or possible influences of syntactic effects that would lead to asymmetries in how words were prosodically realized inside and outside of a focus.

2. METHODS

2.1. Stimuli

The target items were three-syllable nominals (NP). All syllables bore high-level tone in HKC and TwM.

Table 1: *Examples of target NPs.*

	Targets /IPA/	Gloss
HKC	一/jɐt/ 間/kan/ 屋/ŋɔk /	"one house"
TwM	三/san/ 枝/tʂɻ/ 花/hua/	"three flowers"

8 versions of such target NPs for HKC and for TwM were directly adopted from the items in [10] and [12],

respectively. To avoid potential utterance-initial boundary effects, target NPs were preceded by adverbial phrases (two-syllable adverbials in HKC, and three-syllable adverbials in TwM). Target NPs were then followed by a two-syllable verb phrase and a one-syllable sentence final particle. None of the target sentences had other potentially focus sensitive words (e.g., adverbs equivalent to *no* and *only*).

The span of the focus constituent in the target NP was controlled using short contexts consisting of a wh-question (experiments 1a and 2a) or a corrective statement (experiments 1b and 2b). The wh-element of questions targeted either (i) the entire NP (ANP), (ii) the numeral only (ANUM), or (iii) the noun only (ANOUN). The corrective statement targeted either (i) the entire NP (CNP), (ii) the numeral only (CNUM), or (iii) the noun only (CNOUN). We added one extra condition in experiment 2b as a baseline, in which the target NP was part of the background of a sentence, expressing old information (ODNP), that does not have the same level of acoustic strength focus NPs have ([10], [11]). These contexts were pre-recorded by a female native speaker of HKC and of TwM, respectively, and presented auditorily. Example paradigms are given in Table 2. In total there were 104 target items (8 versions × 6 focus conditions × 2 languages + 8 versions of ODNP in experiment 2b).

Table 2: *Example contexts and target sentences. Focus constituents indicated by underlining.*

Focus	Context	Target sentence
ANP	A: On the balcony, what withered away?	B: On the balcony, three flowers withered away
ANUM	A: On the balcony, how many flowers withered away?	B: On the balcony, three flowers withered away.
ANOUN	A: On the balcony, what of three units withered away?	B: On the balcony, three flowers withered away.

Focus	Context	Target sentence
CNP	A: Yesterday, three bridges collapsed.	B: Yesterday, one house collapsed.
CNUM	A: Yesterday, two houses collapsed.	B: Yesterday, one house collapsed.
CNOUN	A: Yesterday, one bridge collapsed.	B: Yesterday, one house collapsed.
ODNP	A: Yesterday, what happened to one house?	B: Yesterday, one house collapsed.

2.2. Participants

17 female native Cantonese speakers from Hong Kong (mean age ± SD: 22.94 ± 1.25 years) and 12 female native Mandarin speakers born and raised in Taiwan (mean age ± SD: 20.5 ± 1.16 years), who were university students, joined our study. None reported any history of hearing problems. Each participant was paid HK$60 after the experiment.

2.3. Procedure

The experiments were conducted in Hong Kong in a sound-attenuated speech lab with a calibrated Telefunken M-80 dynamic microphone and a Focusrite Scarlett 2i2 sound interface. Participants first signed an informed consent form and filled out a language background form. They were seated in front of a computer screen and wore headphones. Stimuli were presented one at a time (self-paced) on the screen. The order was pseudo-randomized, such that no target item occurred immediately adjacent to itself. Participants were asked to first listen to the context question, and then read the target sentence aloud as casually and naturally as possible; no instructions were given regarding focus or emphasis. Participants produced each sentence twice; additional repetitions were allowed in cases of mispronunciation or hesitation. Productions were recorded in .wav format at a sampling rate of 44.1 kHz with 16-bit quantization. There were three practice trials before the main trials. Each session lasted about 30 minutes.

2.4. Measurements

The target items were segmented by Praat [2]. Syllable boundaries were determined by using both visual 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 ProsodyPro 5.7.6 [24] for duration, mean intensity and fundamental frequency (F0). F0 was time-normalized across tokens by dividing each syllable to 10 equal intervals and calculating the trimmed F0 values. For display purpose, the F0 values in Hz were z-score transformed for each speaker before plotting.

Linear Mixed-Effects models were conducted on the duration and mean intensity using the lme4 package [1]. The initial model included random intercepts of item and speaker and by-speaker random slope for 'focus'. By-item random slope was not incorporated because in the cross-subject design and item was nested in the focus condition [16]. After the random structures were maximized, 'focus' was added as a fixed effect. The significance of the main effect was evaluated by likelihood ratio test. The post-hoc Tukey test were done by the emmeans [21] in R.

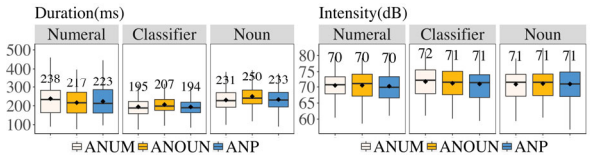
3. RESULTS

3.1. Experiment 1a: wh-focus in HKC

Focus had significant main effect on the duration of all syllables in the target NPs (Numeral: $\chi2 = 8.652$, df = 2, $p = 0.013$; Classifier: $\chi2 = 14.331$, df = 2, $p <$

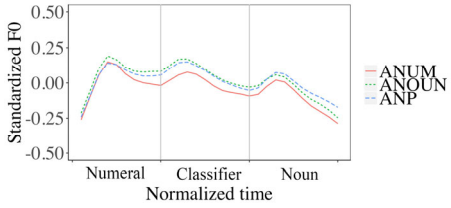
0.001; Noun: $\chi2 = 8.864$, df = 2, $p = 0.012$). The post-hoc tests showed that the whole NP focus had longer duration than the noun focus in the numeral syllable ($p = 0.014$) but shorter in the classifier syllable ($p < .001$). Both the classifier ($p = .008$) and the noun ($p = .010$) showed lengthening effects in the noun focus condition while comparing with the numeral focus condition. Noun in noun focus was longer than the noun under NP focus ($p = 0.054$). Intensity did not seem to be affected by focus types.

Figure 1: Boxplots of duration and intensity by syllable and wh-focus in HKC noun phrases. The dots and the numbers indicate the means.



Though the analyses of F0 did not reach statistical significance, Fig. 2 suggested that while the basic tonal contours were maintained across focus conditions, the numeral or noun in the focus constituent, either because it was the sole focus or within the focused NP, had a higher overall F0. By contrast, F0 on the classifier appeared higher for both conditions of NP-focus and Noun-focus. In other words, F0 of the classifier appeared to pattern with the F0 level of focused noun.

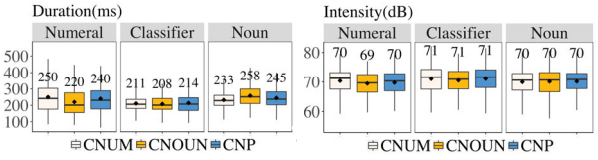
Figure 2: Time-normalized F0 (z-score transformed) by wh-focus of HKC NPs.



3.2. Experiment 1b: corrective-focus in HKC

Focus had significant main effect on the duration of all the syllables in the noun phrase (Numeral: $\chi2 = 11.398$, df = 2, $p = 0.003$; Classifier: $\chi2 = 8.819$, df = 2, $p = 0.012$; Noun: $\chi2 = 12.498$, df = 2, $p = 0.002$).

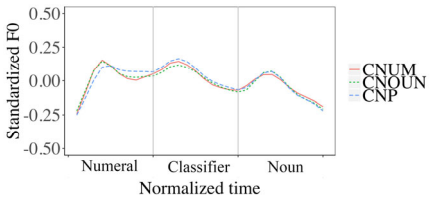
Figure 3: Boxplots of duration and intensity by syllable and corrective-focus in HKC noun phrases. The dots and the numbers indicate the means.



corrected (CNUM), the duration of the numeral syllable was longer than that in corrective noun condition (CNOUN, $p = .003$). The duration of the noun syllable was longer in CNOUN condition ($p = .003$) and was longer in the corrective NP condition (CNP, $p = .005$) than when it was under the CNOUN condition. The duration of classifier patterned with the noun under CNP condition, which was longer than when the classifier was in the CNOUN ($p = .009$).

Similar to wh-focus study, the analyses of F0 did not reach statistical significance, and yet, plots in Fig. 4 suggest that corrective-noun focus triggered higher F0 across all three syllables, and showed much higher F0 level in the noun syllable while comparing with focus conditions of CNP and CNUM.

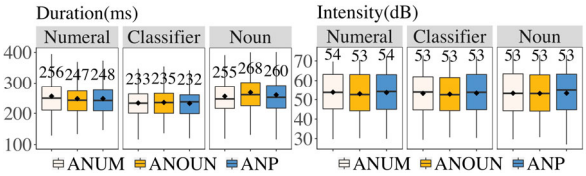
Figure 4: Time-normalized F0 (z-score transformed) by corrective-focus of HKC NPs.



3.3. Experiment 2a: wh-focus in TwM

In this set, focus did not seem to affect the duration. The post-hoc tests showed that the duration of noun syllable under the noun focus condition exhibited marginally lengthening effects while comparing with numeral focus condition ($p < 0.069$).

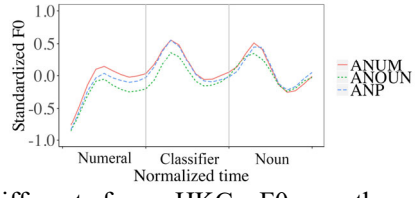
Figure 5: Boxplots of duration and intensity by syllable and wh-focus in TwM noun phrases. The dots and the numbers indicate the means.



Focus showed significant main effects on the intensity of the numeral syllable ($\chi2 = 9.5941$, df = 2, $p = 0.008$), and its post-hoc tests showed that ANUM ($p = 0.024$) and NP focus ($p < 0.001$) exhibited higher intensity than the numeral under ANOUN condition.

Similar to HKC, Fig. 6 suggested that numeral and noun had a higher overall F0 in ANUM and ANOUN.

Figure 6: Time-normalized F0 (z-score transformed) by wh-focus of TwM NPs



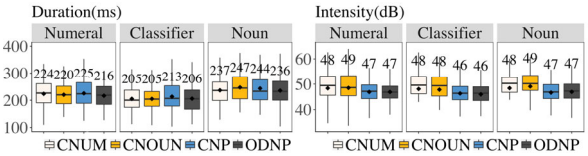
Yet, different from HKC, F0 on the classifier appeared higher for both numeral-focus and NP-focus, suggesting that F0 of classifier appeared to pattern with the F0 level of numeral in ANUM focus.

3.4. Experiment 2b: corrective-focus in TwM

Focus significantly influenced the duration of the numeral ($\chi2 = 11.707$, df = 3, $p = 0.008$), and showed marginal effects on the noun ($\chi2 = 7.143$, df = 3, $p = 0.067$). Post-hoc tests only showed significant lengthening effects of the numeral under corrective numeral (CNUM, $p < .001$) and corrective NP (CNP, $p = .018$), comparing with the numeral in ODNP.

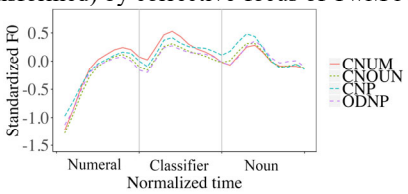
Focus also affected the intensity of the classifier ($\chi2 = 9.453$, df = 3, $p = 0.024$) and the noun ($\chi2 = 16.195$, df = 3, $p = 0.001$). Post-hoc tests showed that intensity of the classifier was higher in CNUM ($p = .008$) and corrective noun (CNOUN, $p = .042$) when compared with ODNP. The intensity of the noun was higher under the condition of CNOUN than under conditions of CNUM ($p = .046$) and ODNP ($p = .002$). Noteworthy is that intensity of noun under CNOUN was much higher than that under CNP ($p < .001$).

Figure 7: Boxplots of duration and intensity by syllable and corrective-focus in TwM noun phrases. The dots and the numbers indicate the means.



Similar to wh-focus in TwM, Fig. 8 showed that numeral and noun in the focus constituent had a higher overall F0. By contrast, F0 on the classifier appeared higher for NP-focus (CNP) and Numeral-focus (CNUM), i.e., F0 of the classifier appeared to pattern with the F0 level of the numeral.

Figure 8: Time-normalized F0 (z-score transformed) by corrective-focus of TwM NPs.



4. DISCUSSION AND CONCLUSIONS

Our study explored whether focus realization in HKC and TwM extended beyond general acoustic highlighting of focus constituents. The results show support for this idea in some respects. First, we found that both the numeral and noun were longer in HKC when inside the focus constituent as compared to outside of it. Concerning TwM, focus condition marginally influenced the duration but showed stronger effects on intensity on the numeral and the noun. In TwM wh-focus, numeral showed higher intensity in ANUM and ANP and higher intensity of noun in ANOUN; in corrective focus, higher intensity of numeral was found in CNUM and CNP, and higher intensity of the noun under CNOUN and CNP conditions. This could be explained if focus-related acoustic prominence concerns primarily the edges of a constituent. Consider that in many languages, edge marking is one of the primary prosodic exponents of focus, and this can occur at either or both the left or right edges (French: [7][8]; Japanese: [9][18]; Basque: [9]).

Second, durational effects of focus were stronger for the noun than for the numeral, and in some cases the noun region showed stronger effects under noun-focus than under NP-focus (e.g., HKC noun was longer in ANOUN than in ANP; TwM noun showed higher intensity in CNOUN than in CNP). This would be surprising under a pure acoustic prominence approach, since such cases involved the same type of focus but showed different strength of effects due to whether the noun was the focus alone or was a part of a larger focused unit. Yet, syntactically, noun tends to carry more information weight and lies at a stronger syntactic juncture (see also [13]); this may be one of the reasons why focusing the noun alone brought more highlight, and it would in turn suggest that prosodic organization interact with focus marking.

Third, the classifier did not pattern with respect to focus in the same way between HKC and TwM. In HKC, the duration of the classifier was lengthened in wh-noun focus, and in corrective-NP, i.e., conditions in which the noun was focus related. In TwM the classifier showed higher intensity under the numeral-focus and the noun-focus conditions. Similarly, the F0 patterns showed that classifier patterned with noun-foci in HKC but patterned with numeral-foci in TwM. TwM's "classifier with numeral" patterns would suggest that a level of prosodic constituency may be influencing how precisely focus marking can target the actual focus constituent, especially if we consider the general assumption that Mandarin prosodic word is minimally disyllabic. Although the differences of classifier patterns between TwM and HKC might be surprising, considering that classifiers in Cantonese have been shown to be syntactically more independent than Mandarin classifiers [6], this would suggest that a different prosodic structural organization interact with focus marking in HKC.

In sum, these results suggest that complex internal organization at different structural levels may interact with the prosody system of focus marking. Considering the mechanism and functions through the interaction of prosodic alignment, structural organization, and focus marking, we expect future studies of Chinese languages and varieties of tone languages to reveal more details about the acoustic representation of information structure.

5. ACKNOWLEDGEMENTS

This research was made possible through support by the general research fund (4-ZZJQ) supported by the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University.

We would like to thank the three ICPhS 2019 anonymous reviewers for their insightful comments, and we would also like to thank James S. German for his suggestions at the earlier stage of this project. We thank Ka Wai Chan, Tsz Shan Lo, Xia Wang and Ka Keung Leon Lee for their technical support. Mistakes remaining are exclusively our own.

6. REFERENCES

[1] Bates, Douglas, Martin Maechler, Ben Bolker , and Steven Walker. 2015. "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software* 67 (1): 1-48. Accessed 11 10, 2017. https://cran.r-project.org/web/packages/lme4/lme4.pdf.

[2] Boersma, Paul, and David Weenink. n.d. *Praat: doing phonetics by computer.* Accessed 11 10, 2017. http://www.fon.hum.uva.nl/praat/.

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

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

[5] Chen, Y. & Gussenhoven, C. 2008. "Emphasis and tonal implementation in Standard Chinese." *Journal of Phonetics* 36 (4): 724-746.

[6] Cheng, Lisa Lai-Shen, and Rint Sybesma. 1998. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4): 509-542.

[7] Di Cristo, A. 1998. "Intonation in French." In *Intonation systems: A survey of twenty languages*, 195-218. Cambridge: Cambridge University Press.

[8] German, J. S. & D'Imperio, M. 2016. The status of the initial rise as a marker of focus in French. *Language & Speech,* 59(2), 165-195.

[9] Gussenhoven, C. 2004. *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

[10] Hsu, Yu-Yin. 2018. "Prosody and corrective focus within the nominal domain of Mandarin Chinese." *Proceedings of the 43th Annual Meeting of the Berkeley Linguistics Society.* 2017 Berkeley.

[11] Hsu, Yu-Yin, and Anqi Xu. 2017. "Focus Acoustics in Mandarin Nominals." *Interspeech.* Stockholm, Sweden. 3231-3235.

[12] Hsu, Yu-Yin, A. Xu and H. Ngai. 2018. "Focus Prosody in Cantonese and Teochew Noun Phrases." Proceedings of the 9th International Conference on Speech Prosody, pp 961-965. DOI: 10.21437/SpeechProsody.2018-194.

[13] Hsu, Yu-Yin, and James S. German. 2018 "Prosodic organization and focus realization in Taiwan Mandarin." The 32nd Pacific Asia Conference on Language, Information and computation.

[14] Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. CambridgeMA: MIT PRess.

[15] Jin, S. 1996. *An acoustic study of sentence stress in Mandarin Chinese.* Columbus, OH: Ohio State University.

[16] Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. Annual Review of Psychology.

[17] Krifka, M. 2007. "Basic notions of information structure." In *Interdisciplinary Studies of Information Structure 6*. Potsdam.

[18] Pierrehumbert, J., & Beckman, M. 1988. *Japanese tone structure*. Linguistic Inquiry Monographs, (15), 1–282.

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

[20] Rooth, Mats. 1992. "A theory of focus interpretation." *Natural Language Semantics* 1:75-116.

[21] Russell Lenth, Henrik Singmann, Jonathon Love, Paul Buerkner, Maxime Herve (2018) Estimated Marginal Means, aka Least-Squares Means https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

[22] Selkirk, Elisabeth. 1996. "Sentence prosody: Intonation, stress and phrasing." In *The handbook of phonological theory*, ed. John A. Goldsmith, 550-569. Cambridge MA and Oxford UK: Blackwell.

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

[24] Xu, Y. "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis," in *TRASP 2013*, Aix-en-Provence, France, 2013.