# Perceptual weighting of prosodic cues to focus by Hong Kong Cantonese listeners

Chris K. C. Lee
*Linguistics Department, Boston University*

Research on prosodic focus marking has shown that for languages that mark focus prosodically, constituents typically vary in some combination of duration, f0, and intensity according to their focus status [1]. Native listeners can reliably detect the location of focus when these prosodic cues are available (see e.g., [2]–[5]), but the relative contribution of each acoustic cue to listeners' perception of prosodic focus remains poorly understood beyond stress-based prominence languages (e.g. [6]). The current study investigates the relative contributions of duration, f0, and intensity to native listeners' perception of prosodic focus in a lexical tone language without phrase-level stress-like prominence, Hong Kong Cantonese (HKC) [7]. HKC encodes focus mainly with durational lengthening and intensity raising and (only consistently in the two rising tones) f0 range expansion in the on-focus region [4].

We conducted a focus location identification experiment with 29 native HKC listeners (23 female and six male, aged 18-42), and resynthesized stimuli were used. To create the resynthesized stimulus set, two versions of the sentence *Zoeng$^{2}$ Wing$^{4}$ jiu$^{3}$ heoi$^{3}$ Fo$^{2}$ Taan$^{3}$* "Zoeng Wing has to go to Fo Tan" were recorded and served as the base stimuli. In one version, the speaker answered, "Who has to go to Fo Tan?" (*early focus*), and in the other, he answered, "Where does Zoeng Wing have to go?" (*late focus*). Table 1 presents the acoustic characteristics of the two base stimuli. The two base stimuli were analyzed and then morphed with Tandem-STRAIGHT [8] such that each resynthesized recording in the stimulus set varies independently along three six-step continua (f0, duration, and intensity, see Table 2). Altogether 216 (unique) resynthesized recordings were created. Participants listened to them and decided whether the speaker responded to an early focus or a late focus question.

Figure 1 shows the percent late focus responses as a function of the three acoustic continua at the group level. All three acoustic dimensions bias participants' responses in the expected way: a longer constituent duration, higher f0 level, and higher intensity of the utterance-final word lead participants to identify the utterance as late focus (duration: $\beta$=.889, $z$=9.25; f0: $\beta$=.214, $z$=4.23; intensity: $\beta$=.211, $z$=6.11). Based on these binomial mixed-effects model estimates (which quantify the change in the likelihood of 'late' focus response per unit change in the continuum step), the perceptual cue weighting for prosodic focus of HKC listeners appears to be **duration > intensity = f0.**

Figure 1 also presents the individual response curves to the three acoustic continua, which reveals substantial individual variation in cue weighting. To investigate the range of individual differences, each participant's responses were fitted to separate binomial fixed-effects models with the three acoustic continua (in steps) as predictors. The model estimates were used as the perceptual weight of the prosodic cues. Four patterns were observed (see Figure 2): (A, $N$=8) all three acoustic continua are significant predictors, with a perceptual cue weighting **duration > f0 > intensity**; (B, $N$=10) all three acoustic continua are significant predictors, with a perceptual cue weighting **duration > intensity > f0**; (C, $N$=7) **duration** is the *only* significant predictor; and (D; $N$=3) **f0** is the most heavily weighted cue.

Taken together, the vast majority of the participants align well with the group-level trend that duration is the primary perceptual cue for prosodic focus, but participants vary in which cues, f0 or intensity, are used as the secondary cue. Our findings also parallel previous studies that find duration as the main cue for prosodic focus in production [4]. However, even in a tonal context (i.e., rising tone) where the f0 range is consistently expanded under focus in production [4], listeners' responses do not seem affected strongly by f0. In the future, participants should be assessed in both production and perception to gain a better understanding of the link between the two modalities in prosodic focus marking.

**Table 1.** Acoustic characteristics of the two base stimuli *Zoeng*$^{2}$ *Wing*$^{4}$ *jiu*$^{3}$ *heoi*$^{3}$ *Fo*$^{2}$ *Taan*$^{3}$. (TP = turning point)

| | Zoeng2 Wing4 (tones: high rising – low falling) | | | Fo2 Taan3 (tones: high rising – mid level) | | |
|---|---|---|---|---|---|---|
| | f0 at TPs (Hz) | Duration (ms) | Intensity (dB) | f0 at TPs (Hz) | Duration (ms) | Intensity (dB) |
| Early focus | 103/198/66 | 778 | 75.11 | 98/113/93 | 451 | 69.95 |
| Late focus | 108/150/94 | 405 | 69.15 | 98/148/108 | 969 | 74.07 |

**Table 2.** Acoustic measures of the three acoustic continua. (Note that the values of Steps 1 and 6 match the acoustic characteristics of the base stimuli. The resynthesized stimuli were checked by three native HKC speakers to ensure that resynthesis did not change the identity of the lexical tones.)

| | Continuum steps | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Zoeng$^{2}$ | duration (ms) | 778 | 675 | 591 | 517 | 458 | 405 |
| Wing$^{4}$ | f0 at TP (Hz) | 103/198/66 | 104/187/74 | 105/175/81 | 106/166/88 | 107/157/93 | 108/150/94 |
| | Mean intensity (dB) | 75.11 | 74 | 72.79 | 71.57 | 70.37 | 69.15 |
| Fo$^{2}$ Taan$^{3}$ | duration (ms) | 451 | 524 | 609 | 701 | 822 | 969 |
| | f0 at TP (Hz) | 98/113/93 | 98/118/96 | 98/125/99 | 98/133/102 | 98/140/105 | 98/148/108 |
| | Mean intensity (dB) | 69.95 | 70.81 | 71.55 | 72.4 | 73.19 | 74.07 |

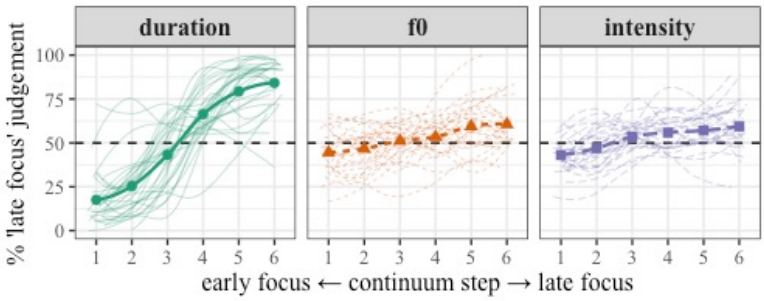


**Figure 1**. Percent 'late focus' response as a function of the three acoustic continua.

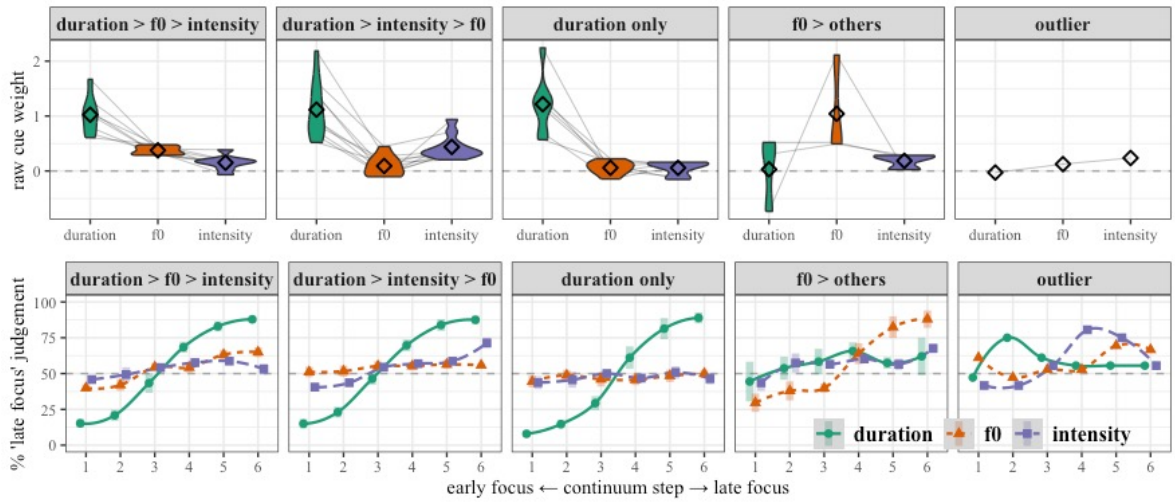


**Figure 2**. Individual differences in the relative use of prosodic cues to focus in Cantonese.

**[1]** F. **Kügler** & S. **Calhoun** (2020). "Prosodic encoding of information structure: a typological perspective," in Gussenhoven, C. & A. Chen (ed.) *The Oxford Handbook of Language Prosody*. **[2]** M. **Breen**, E. **Fedorenko**, M. **Wagner** & E. **Gibson** (2010). "Acoustic correlates of information structure," *Lang. Cogn. Process*. **[3]** T. B. **Roettger**, T. **Mahrt** & J. **Cole**. (2019). "Mapping prosody onto meaning – the case of information structure in American English," *Lang. Cogn. Neurosci.* **[4]** W. **Wu** & Y. **Xu**, (2010). "Prosodic focus in Hong Kong Cantonese without post-focus compression," *Proc. Speech Prosody 2010*. **[5]** Y. **Xu**, S.-W. **Chen** & B. **Wang** (2012). "Prosodic focus with and without post-focus compression: A typological divide within the same language family?" *Linguist. Rev.* **[6]** K. **Jasmin**, F. **Dick**, L. L. **Holt** & A. **Tierney** (2020). "Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities," *J. Exp. Psychol. Gen.* **[7]** P. **Wong**, M. **Chan**, and M. **Beckman**, (2005). "An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese," in S.-A. Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. **[8]** H. **Kawahara**, et al. (2008). "Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation," in *2008 IEEE ICASSP*.