# Perceptual Weightings of Prosodic Cues in Mandarin Prosodic Focus Processing

Wenxi Fei & Yu-Yin Hsu

*Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University*

Understanding how continuous variations across multiple acoustic dimensions are mapped onto linguistic representations is an important question in the field of language comprehension [1-2]. In Mandarin, prosodic cues like F0, duration, and intensity are used to mark focus, but their specific contributions to interpreting intentions behind prosodic focus remain unclear [3]. Thus, the current study investigates how native Mandarin speakers process prosodic focus, by addressing the following three research questions: 1) How do native listeners weigh prosodic cues when identifying communicative intentions in Mandarin? 2) Which cue plays a more dominant role? 3) How is pragmatic information processed in terms of temporal features?

**Methods.** We recruited 30 native Mandarin participants ($N_{female}$ = 15, $M_{age}$ = 22.17, $SD_{age}$ = 2.90), and adopted a modified Visual World Paradigm with two options [4]. The participants were asked to listen to an audio sentence and then to decide whether it responded to a broad-focus question or a narrow-focus question of the verb (Fig. 1). We used resynthesized Mandarin audio stimuli to manipulate three acoustic cues of focus. Two base stimuli of a five-syllable sentence in Tone 1, e.g., crow eats watermelon, were recorded corresponding to “What happened?” (broad focus) and “What does crow do to watermelon?” (narrow verb focus). The base stimuli were morphed with WORLD [5] to create seven-step continua (Table 1). In total, 210 stimuli were created (3 cues×7 levels×10 sentences). The listeners’ responses and fixations were recorded for analysis.

**Results.** Bayesian logistic mixed-effects models’ results showed that continuum step credibly impacted focus-interpretation responses, particularly in the F0 condition. In Fig. 2, as continuum step increased, “verb focus” responses increased ($B$ = 0.72, 95%$CI$ = [0.64, 0.80]). Which was most visible in the F0 condition. Based on the model estimates, the perceptual cue weightings showed that listeners had a greater reliance on F0 ($B$ = 1.78, 95%$CI$ = [1.61, 1.96]) than intensity ($B$ = 0.19, 95%$CI$ = [0.06, 0.32]) and duration ($B$ = 0.08, 95%$CI$ = [-0.04, 0.20]), consistent with previous findings [6]. Eye-tracking data showed that the listeners’ perceptual divergence for different steps appeared only in the F0 condition (Fig. 3). In the suggested analysis window [3], the significant divergences between step 1 and 7, the most acoustically salient pair, appeared from around 451 ms, indicating the occurrence of processing acoustic information. Between step 2 and 3, the acoustically ambiguous pair and showing the most robust difference in response results, there are significant divergences from about 754 ms to 1158 ms. This pair comparison indicates that the pragmatic information was not involved in the early processing together with acoustic information but occurred independently later.

**Conclusion.** Our findings show that native listeners effectively map the altered acoustic weightings to different communicative demands raised by prosodic focus in Mandarin, not just the mappings to word or sentence meanings [1-2]. The results also show that the focus intention interpretation happens only at a late time window, indicating that listeners require sufficient time to process pragmatic information [3], and the communicative task can play a critical role [7].

**Table 1.** Step manipulation of audio stimuli from 0% to 100%, 50% represents the ambiguous step.

| Condition | Mean F0 of verbs (Hz) | Duration (ms) | Intensity (dB) |
|---|---|---|---|
| **F0** | 283, 292, 302, 313, 324, 336, 349 | 850 (50%) | 80.94 (50%) |
| **Duration** | 313 (50%) | 793, 812, 831, 850, 870, 889, 908 | 80.94 (50%) |
| **Intensity** | 313 (50%) | 850 (50%) | 79.23, 80.02, 80.88, 80.94, 81.03, 81.08, 81.14 |

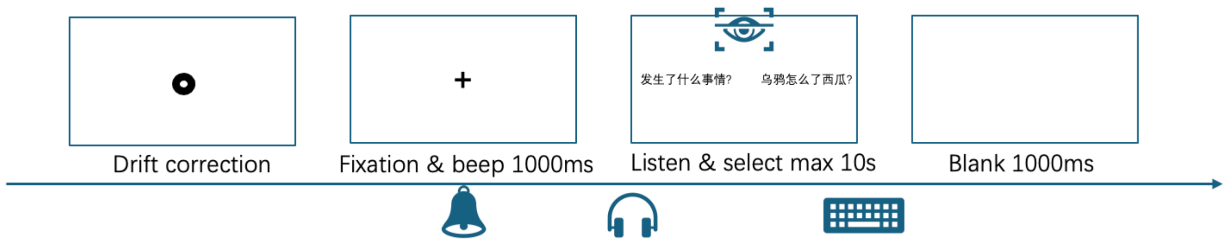


**Fig 1.** Procedure of the experiment.

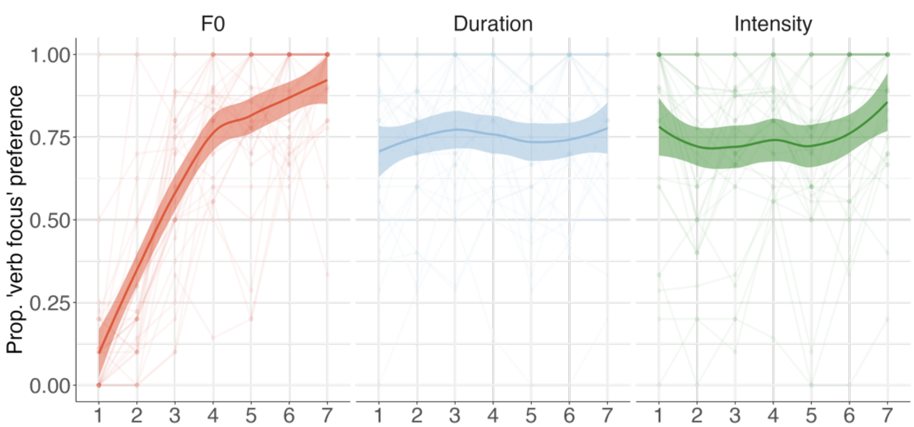


**Fig 2.** Response patterns by condition. Shade areas represent 95% Bayesian confidence intervals. Background lines show individual patterns.

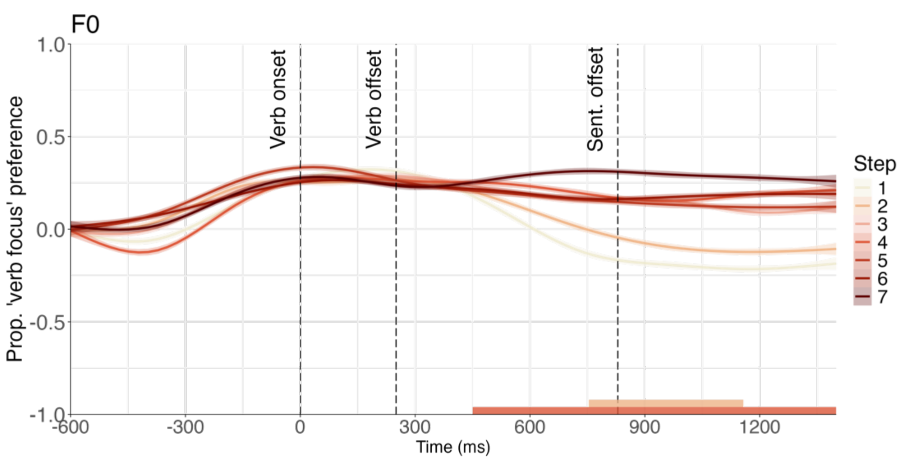


**Fig 3.** Eye movement data of the F0 condition. The colored horizontal lines indicate the significant time window reported by GAMMs between step pairs.


**References.**

[1] Holt & Lotto (2006). *JASA*.

[2] Toscano & McMurray (2010). *Cogn. Sci*.

[3] Gao et al. (2024). *NeuroImage*.

[4] Steffman (2021) *Cogn. Neurosci*.

[5] Morise et al. (2016). *IEICE Trans. Inf.& Syst*.

[6] Zeng et al. (2022). *Brain Lang*.

[7] Pickering & Strijkers (2024). *topiCS*.