

## How Fast Do Prosodic Cues Update? Evidence from Short-Term Cue Reweighting in Mandarin Prosodic Focus

Wenxi Fei and Yu-Yin Hsu

*Department of Language Science and Technology, The Hong Kong Polytechnic University*

wen-xi.fe@connect.polyu.hk, yu-yin.hsu@polyu.edu.hk

Understanding prosodic prominence requires listeners to integrate multiple acoustic cues, and recent work shows that this cue-weighting strategy can shift rapidly when listeners encounter new input distributions. Evidence from short-term exposure paradigms demonstrates that even brief exposure can change how suprasegmental cues are combined during perceptual categorization tasks [1]. From a broader computational perspective, such adaptation may arise at different stages of processing, from changes to signal-based perceptual expectations to adjustments during later meaning-based interpretations [2]. Distinguishing these processing stages is theoretically critical, because they represent when adaptation affects online comprehension. However, it remains unclear which of these levels is primarily affected during prosodic adaptation. The present study directly explores the processing stages by examining how short-term exposure modulates the time course of cue use in Mandarin prosodic focus, combining response data with eye-fixation patterns during spoken language comprehension.

Forty native Mandarin listeners ( $N_{female} = 20$ ,  $M_{age} = 23.88$ ,  $SD_{age} = 3.00$ ) participated in a short-term exposure study using a 2AFC eye-tracking paradigm. The participants listened to an audio sentence and then decided whether it answered a broad-focus question (“What happened?”) or a narrow-verb-focus question (e.g., “What did crow do to watermelon?”) (Figure 1). Two base stimuli of a five-syllable Mandarin sentence in Tone 1, e.g., “crow ate watermelon”, were recorded corresponding to broad focus and narrow verb focus. Stimuli were then resynthesized, varying along a seven-step duration continuum. The experiment consisted of a Baseline block, two exposure blocks, and two post-exposure blocks. The Canonical exposure reflected Mandarin listeners’ typical cue weighting reported for prosodic focus, higher F0 accompanied by longer verb duration; while the Accented exposure paired duration with an inverse F0 pattern (Figure 2).

The response results aligned well with work on short-term perceptual cue reweighting at the phrasal level [1]. The Bayesian logistic mixed-effects model showed that continuum step credibly impacted focus-interpretation responses, but the magnitude of these effects varied across conditions (Figure 3, top-left). In the baseline, “verb-focus” responses increased with continuum step ( $B = 0.55$ ,  $95\%CI = [0.29, 0.80]$ ), though the difference between responses to step 1 and step 7 was small ( $\Delta_{step} = 13.3\%$ ), consistent with duration’s role as a weak cue for Mandarin focus [3]. After Canonical exposure, the step effect became stronger ( $\Delta_{step} = 17.1\%$ ;  $B = 1.08$ ,  $95\%CI = [0.57, 1.61]$ ), indicating that listeners reinforced the conventional cue-focus mapping. Following Accented exposure, the step had a null effect but showed a slight reversal trend ( $\Delta_{step} = 1.9\%$ ;  $B = 0.31$ ,  $95\%CI = [-0.13, 0.74]$ ), suggesting partial and incomplete adaptation to the acoustic-specific distribution.

The eye-tracking data revealed the processing stages across three conditions (Figure 3), comparing the endpoints of the duration continuum (step 1 vs. 7) over a 0-1600ms window. Only the baseline condition showed an early but small divergence between step 1 and 7, emerging shortly after verb onset, suggesting that listeners’ predictive use of duration is minimal in the absence of primary cue. After exposure, early fixations between step 1 and 7 remained overlapping in both Canonical and Accented conditions, indicating that learning does not immediately recalibrate early perceptual expectations. In contrast, later portions of the time course (about 600-1600ms) showed some exposure-driven effects. The Canonical condition produced a robust divergence over time while the Accented condition showed a less stable one: the differences between step 1 and 7 were greatest at around 1100ms and then the two trajectories gradually moved back toward each other by approximately 1600ms. Canonical exposure strengthened listeners’ existing expectations, leading to stable late divergence, whereas Accented exposure created only a temporary shift that faded as listeners reverted to their long-term knowledge.

In conclusion, these results show that short-term exposure mainly modulates the late integration of prosodic duration cues during discourse-level comprehension, short-term adaptation seems to be located primarily at later, meaning-based interpretive stages, while early signal-based processing remains relatively stable without obvious categorical changes. This dynamic combination of stability and flexibility thus aligns with the stage-differentiated adaptation framework [2] and supports the view that high-level mapping is effortful relative to low-level acoustic encoding.

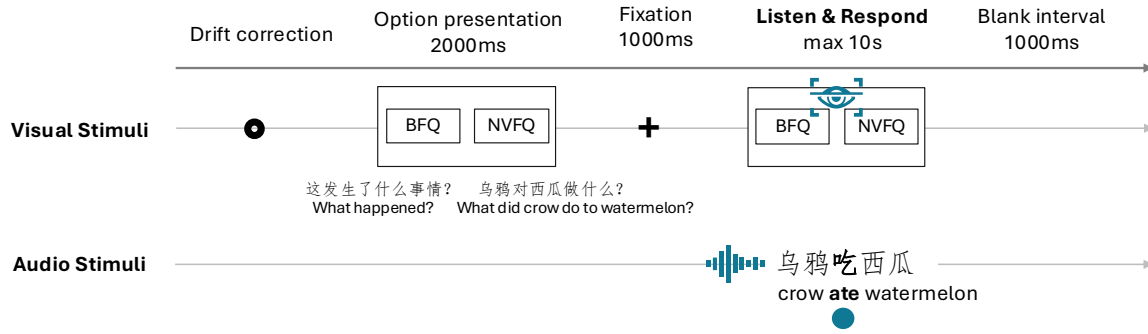


Figure 1: Procedure of each trial. BFQ means the option of broad-focus question and NVFQ means the option of narrow-verb-focus question. The presented sentences are examples from the full set of materials.

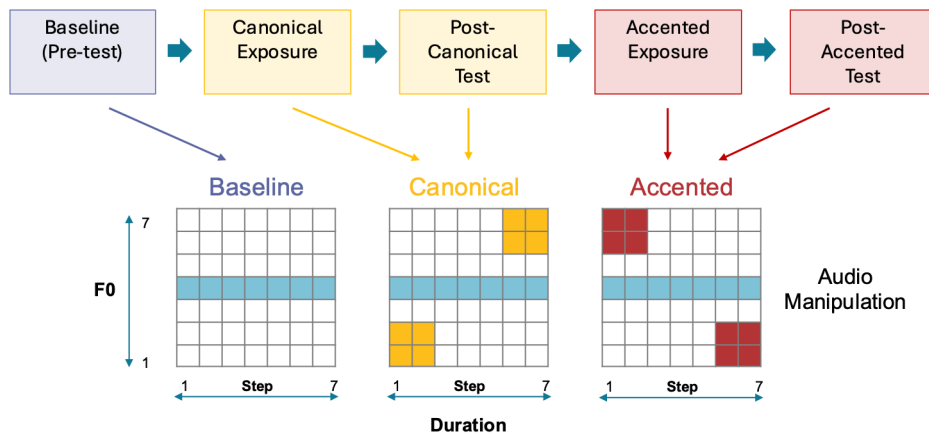


Figure 2: Block procedure and stimuli manipulation for Baseline, Canonical, and Accented conditions. The cyan squares represent the manipulation of pre- and post-test trials, while the yellow and red squares represent the exposure ones.

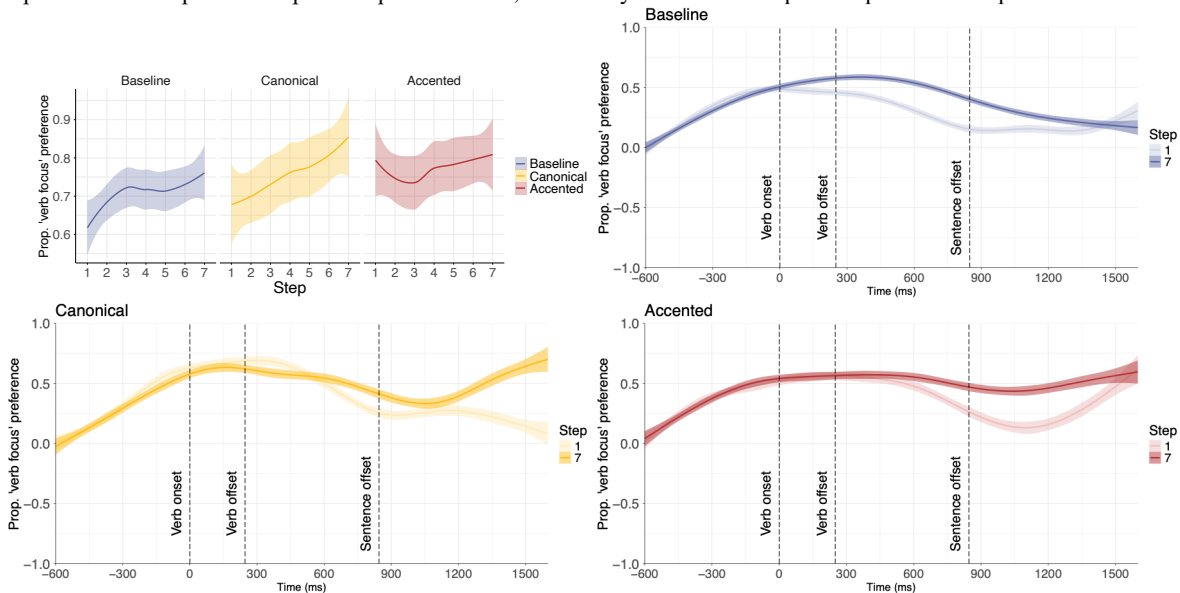


Figure 3: Response curves across the seven-step duration continuum (top-left) and eye-tracking data across the three conditions. Shaded areas represent 95% Bayesian credible intervals. “Verb offset” and “Sentence offset” indicate the mean offsets of verbs and sentences, respectively.

## References:

- [1] Jasmin, K., Tierney, A., Obasih, C., & Holt, L. (2023). Short-term perceptual reweighting in suprasegmental categorization. *Psychonomic Bulletin & Review*, 30(1), 373-382.
- [2] Xie, X., Jaeger, T. F., & Kurumada, C. (2023). What we do (not) know about the mechanisms underlying adaptive speech perception: A computational framework and review. *Cortex*, 166, 377-424.
- [3] Fei, W., & Hsu, Y. Y. (2025). Mapping Acoustic Cues to Pragmatic Functions: Perceptual Cue Weighting of Prosodic Focus in Mandarin. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47).