

How Does Prosody Adapt? Linking Perceptual Learning to Pragmatic Decisions in Eye-Tracking

Wenxi Fei and Yu-Yin Hsu

Department of Language Science and technology, The Hong Kong Polytechnic University

Understanding short-term adaptation in cue-meaning mappings requires not only tracking changes in overt responses but also characterizing how perceptual sensitivity to acoustic differences becomes incorporated into interpretive decisions over time. Most research on adaptive speech perception has focused on perceptual or lexical-level mappings, leaving the cognitive mechanisms that link perceptual learning to higher-level interpretation underspecified [1]. This gap is particularly salient for prosody, where acoustic cues are integrated to support discourse-level meanings rather than categorical lexical differences. Prosodic focus therefore provides a critical test case for examining how adaptation shapes post-perceptual decision processes that map acoustic evidence onto pragmatic interpretations. Thus, this study examines the temporal alignment between fixation-based and response-based sensitivity to acoustic differences in Mandarin prosodic focus perception.

Mandarin listeners ($N = 40$, $N_{female} = 20$, $M_{age} = 23.88$, $SD_{age} = 3.00$) completed a 2AFC eye-tracking task. In each trial, participants listened to an audio sentence and decided whether it answered a broad-focus question or a narrow-verb-focus question with their eye movements recorded. Prosodic focus was cued by duration, a secondary cue in Mandarin relative to F0 with target sentences containing a one-syllable verb manipulated along a seven-step duration continuum (e.g., *crow eats watermelon*). Participants first completed a **Baseline** phase, followed by exposure to a **Canonical** condition and an **Accented** condition (Fig. 1). The Canonical condition preserved the natural mapping (higher F0 with longer duration), whereas the Accented condition created a systematic cue conflict.

Bayesian logistic mixed-effects models revealed clear exposure-dependent changes in cue use (Fig. 2A). Canonical exposure strengthened the step7-step1 contrast in focus interpretation ($\Delta step = 17.1\%$; $B = 1.08$, $95\%CI = [0.57, 1.61]$) relative to Baseline ($\Delta step = 13.3\%$; $B = 0.55$, $95\%CI = [0.29, 0.80]$), whereas the Accented condition showed a reduced and non-credible effect ($\Delta step = 1.9\%$; $B = 0.31$, $95\%CI = [-0.13, 0.74]$), indicating weakened reliance on duration. Eye-tracking analyses further revealed condition-specific differences in the time course of cue use across the 0-1500 ms window (Fig. 2B).

To link these response-level effects to online processing, we computed a contrast-based consistency measure (step7-step1) that quantifies whether stronger duration effects in fixation patterns are linked to stronger duration effects in overt responses within the same time window. As shown in Fig. 3, fixation-response consistency differed between baseline and post-exposure conditions. In Baseline, credible consistency emerged relatively early at about 600-1500 ms, whereas consistency after exposure was delayed or restricted to late windows just before responses. This pattern suggests that short-term learning affects the timing with which acoustic evidence contributes to interpretation. Within the exposure conditions, fixation-response consistency differed by cue structure. Consistency emerged earlier in Accented (after about 900 ms) than in Canonical condition (after about 1200 ms). This pattern is compatible with cue-reweighting accounts, in which adaptation modulates the functional contribution of cues without requiring restructuring [3].

To summarize, these results indicate that short-term prosodic adaptation reshapes when acoustic evidence is integrated into higher-level interpretation. Exposure delayed the alignment between perceptual sensitivity and interpretive decisions, with cue structure influencing the timing at which acoustic information becomes decisively informative. This study further highlights late-stage integration as a key mechanism linking perceptual learning to pragmatic-related interpretation in real-time speech comprehension but leaves open whether this shift reflects changes in cue weighting, decision strategy, or processing demands.

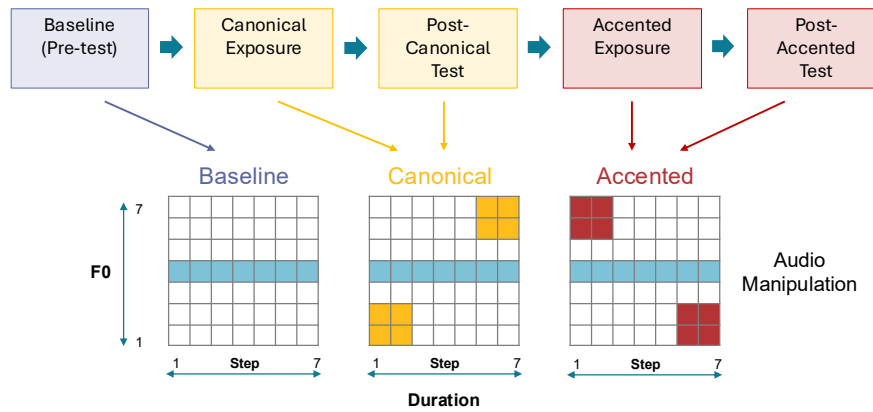


Fig. 1. Procedure and stimulus mappings by condition. Yellow and red squares indicate exposure stimuli and blue ones indicate test stimuli.

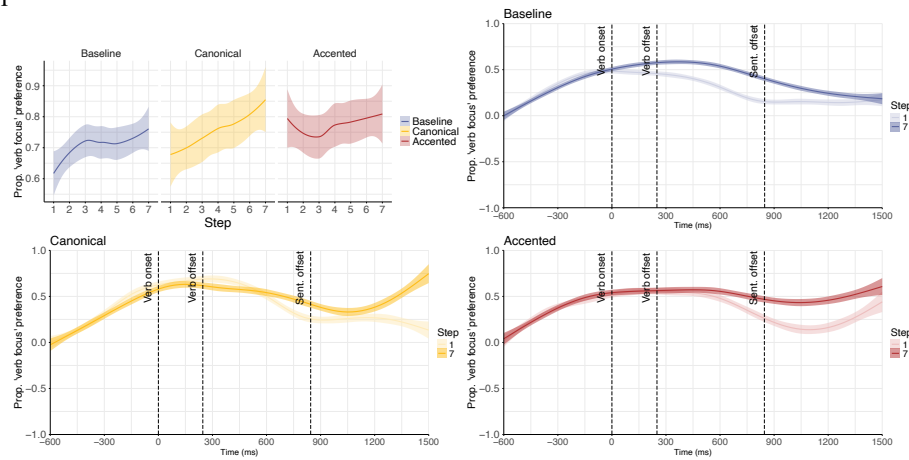


Fig. 2. A: Proportion of “verb focus” responses by participant. **B:** Eye-tracking results showing processing differences across conditions, comparing step1 vs. step7 over the 0-1500ms window.

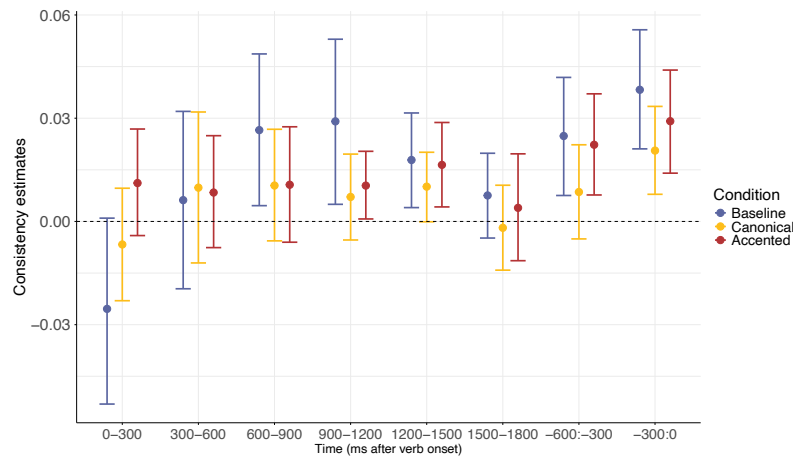


Fig. 3. Fixation-response consistency over time. Negative time windows (i.e., -600:-300, -300:0 ms) indicate intervals before responses, with the last 0 marking the decision time point.

References

- [1] 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.
- [2] Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, 34(3), 434-464.
- [3] Llanos, F., Wu, Y. C., Abel, T. J., & Holt, L. L. (2025). Accented speech modulates multiple event-related potential components across multiple levels of language processing. *Communications Psychology*, 3(1), 186.