

Learning prosodic phrasing in infancy: The role of gestures and visual attention

Victoria Reshetnikova¹, Roy Hessels², Aoju Chen¹

¹*Institute for Language Sciences, Utrecht Universit*, ²*Experimental Psychology, Helmholtz Institute, Utrecht University*

Speakers use prosodic cues (i.e. pitch change, final lengthening, pause) to group words into meaningful units [1-4], and listeners rely on such cues in speech comprehension. This process are known as prosodic phrasing. Interestingly, the use of the prosodic cues to mark prosodic phrase boundaries varies across languages. Developmental research on prosodic phrasing has primarily focused on auditory cues. However, gestures accompanying speech (or co-speech gestures) were found to co-occur with prosodic boundaries [5, 6]. Visual cues could then potentially influence the learning of prosodic phrasing not only through their presence but also through infants' attention to them, as attention to mouth movements and referents of pointing gestures supports lexical development [7-9]. Thus, adopting a foreign language exposure paradigm that allows for controlled input [10], this study investigates (1) how the availability of co-speech gestures influences infants' learning of prosodic phrasing and (2) how infants' gaze behaviour during interactions relates to this learning.

Forty-one Dutch-learning infants (F:18, mean = 7;1, range = 4;1-9;28) participated in twelve weekly interactive exposure sessions with native speaker of German. German differs from Dutch in relative importance of boundary cues with German putting more weight on pitch change and lengthening and Dutch on pause in prosodic boundary perception [3, 4]. Gesture availability was manipulated between participants: free hand use in the high-availability condition and restricted hand movement in the low-availability condition (Fig. 1). Learning outcomes were assessed after the exposure period using electroencephalography (EEG), focusing on the neurophysiological response to boundary processing Closure Positive Shift (CPS) in name-sequences (*Moni und Lilli und Manu*), produced with or without a boundary after the second name (three conditions: no-boundary, pitch cue only, pitch+final lengthening) [11] (Fig. 2). In addition, eye-tracking during a subset of sessions measured infants' gaze to the speaker's hands and face.

It was hypothesised that infants in the high-availability condition would show stronger prosodic phrasing learning, reflected in larger CPS responses, than those in the low-availability condition. We also examined whether infants' gaze to gestures related to learning outcomes.

Linear-mixed-effects modelling on ERP amplitude in the time window from 500 to 800 ms after the onset of the final syllable of the second name provide no evidence for a consistent interaction between 'prosodic boundary type' (no cue, pitch cue, and combined pitch + final lengthening cues) and 'gesture availability' (high vs. low availability) across scalp regions. However, a significant interaction of prosodic boundary type and gesture availability was found in the frontal region. Follow-up comparisons indicated that infants differentiated between pitch-cue and no-cue conditions in the high-availability condition but not in the low-availability condition. Importantly, the observed ERP pattern did not resemble a canonical CPS: instead, waveforms showed a broadly distributed positivity between 500-800 ms across all conditions, suggesting a late positivity related to linguistic interpretation of pitch (e.g., P600-like) rather than boundary-specific processing.

To address question 2, we examined infants' gaze behaviour toward the tutor. Across learning conditions, infants spent more time looking at the tutor's head than the hands, despite similar visit rates to both areas, reflecting longer average visits to the head. Overall visual engagement differed by gesture availability, with infants in the high-availability condition showing higher visit rates and greater total looking time than those in the low-availability condition. Statistical analyses of gaze-ERP relationships in the high-availability condition did not reveal reliable associations between gaze measures and ERP responses.

Together, these findings indicate that increased availability of co-speech gestures may facilitate infants' learning of prosodic phrasing. Although infants did not show a canonical CPS response, differentiation between pitch-cue and no-cue conditions suggests emerging sensitivity to novel prosodic boundary cues. Eye-tracking results further indicate that gesture availability may influence infants' engagement with the speaker during interactions. Future research could examine how gesture-prosodic boundary alignment in the input relates to infants' boundary perception, to further clarify the role of gesture in prosodic learning.

References

- [1] Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, 22(8), 1151-1177.
- [2] Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*, 25(7-9), 905-945.
- [3] Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Råling, R., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2016). How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgements. *Language, Cognition and Neuroscience*, 31(7), 904-920.
- [4] Geutjes, J., Ganga, R., van Niekerk, E., Reshetnikova, V. and Chen, A. (2026). Cue-weighting in processing of prosodic boundaries in Dutch: an ERP study. *Acc Psychonomic Bulletin & Review* 33:53
- [5] Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 56(3), 850-864.
- [6] de La Cruz-Pavía, I., Gervain, J., Vatikiotis-Bateson, E., & Werker, J. F. (2020). Coverbal speech gestures signal phrase boundaries: A production study of Japanese and English infant-and adult-directed speech. *Language Acquisition*, 27(2), 160-186.
- [7] Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of child language*, 42(6), 1173-1190.
- [8] de Villiers Rader, N., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language Sciences*, 34(5), 559-568.
- [9] Chen, A. (submitted). Prosodic development across early life: From biologically motivated biases to language-specific systems.
- [10] Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100(15), 9096-9101.
- [11] Holzgrefe-Lang, J., Wellmann, C., Höhle, B., & Wartenburger, I. (2018). Infants' processing of prosodic cues: electrophysiological evidence for boundary perception beyond pause detection. *Language and speech*, 61(1), 153-169.

Figure 1. High- and low-gesture availability learning conditions.



Figure 2. Prosodic boundary conditions (no cue, pitch cue, combined cue). *Figure taken from [10].*

