

The temporal organization of Seoul Korean prosody in multimodal gestures

Yoonjeong Lee¹ and Jelena Krivokapić¹

¹Department of Linguistics, University of Michigan (USA)

Introduction: In day-to-day conversations, speakers utilize co-speech gestures, including non-referential beat-like movements such as vertical eyebrow, head, and manual gestures, to mark prosodic structure. Co-speech gestures have been found to align with pitch-accented syllables [1,2], prompting the question of how they are coordinated in languages lacking lexical stress and pitch accents. This study examines the coordination of manual and head beat gestures with speech gestures in Seoul Korean, a language that marks prominence through phrasing by introducing a phrase boundary before a prominent word [3]. Given the significant influence of prosodic structure on the production and temporal organization of both speech and co-speech gesture observed in other languages, we predict inter-articulator coordination via prosodic boundaries in Seoul Korean. Specifically, we expect phrase-edge speech gestures to closely align with the beat gesture landmarks, guided by the role of phrasing in this language.

Methods: Eight Seoul Korean speakers (5F, 3M) participated in the experiment, reading a children's story. Speakers differed in their predominant mode of gesturing. Manual beats from six speakers and head beats from two speakers were analyzed. The movements of the lips, tongue, and (uncorrected) head were point-tracked using electromagnetic articulography, while vertical hand motions from video recordings were auto-tracked using a region-of-interest geometrical centroid tracking method [4]. The co-speech gestures were visually checked for perceptibility and non-referentiality. The multimodal signals were time-aligned for precise identification of cross-modal temporal coordination [6]. Lags for manual beat gestures were calculated: (1) between the beat gesture onset (BGONS) and a set of phrase-initial landmarks, and (2) between the beat gesture target (BGMAX) and a set of phrase-final landmarks (Table 1). Due to their small size, head beat gestures had coordination landmarks limited to nearby segmental gestures: BGONS-to-GONS, BGONS-to-GMAX, BGONS-to-GOFF, BGMAX-to-GONS, BGMAX-to-GMAX, and BGMAX-to-GOFF. Using linear mixed-effects models, inter-articulator coordination targets were identified as the most stable and shortest intervals between temporal landmarks, i.e., lag values with the smallest variance and closest to 0.

Results and Discussion: Results indicate systematic inter-articulator coordination patterns. Manual beat gestures, functioning as global phrase markers, unfold with the entire prosodic phrase and are found to be synchronous with both phrase-edge segment and tone gestures. The temporal distance between the four phrase-initial landmarks and the beat gesture onset is similar, with the boundary tone gesture onset and constriction gesture target identified as the most stable landmarks (Fig. 1A). We observe a departure from languages that employ pitch accents, where the typical inter-articulator coordination target is the beat gesture max to F0 max. Notably, the onset and target of the highest pitch gesture emerge as the farthest landmarks, while the rest of the phrase-final landmarks maintain equal proximity to the beat gesture target (Fig. 1B). The phrase-final constriction gesture and boundary tone gesture targets exhibit the highest stability, indicating coordination at the phrasal level, potentially involving laryngeal gestures. Head beats, functioning as local phrase-edge markers, signify the initiation or conclusion of a prosodic phrase. The head gesture is precisely coordinated with the phrase-edge syllable, aligning the preparatory component with the onset consonant and the downward component with the vowel. The cross-modal coordination consistently adheres to onset-to-onset and max-to-max patterns (Fig. 2). The close relationship between prosodic structure and the recruited modalities highlights the use of phrasing in Seoul Korean to mark both prominence and grouping. These findings suggest coordination driven by prosodic structure, albeit manifested uniquely to reflect the specific prosodic structure of the language.

Table 1. Coordination landmarks.

Phrase-initial landmarks for BGONS	Phrase-final landmarks for BGMAX
GONS: constriction gesture onset	GONS: constriction gesture onset
GMAX: constriction gesture target	GMAX: constriction gesture target
BTONS: boundary tone gesture onset	GOFF: constriction gesture offset (release gesture target)
BTMAX: boundary tone gesture target	BTONS: boundary tone gesture onset
	BTMAX: boundary tone gesture target
	F ₀ ONS: onset of the highest F ₀
	F ₀ MAX: highest F ₀

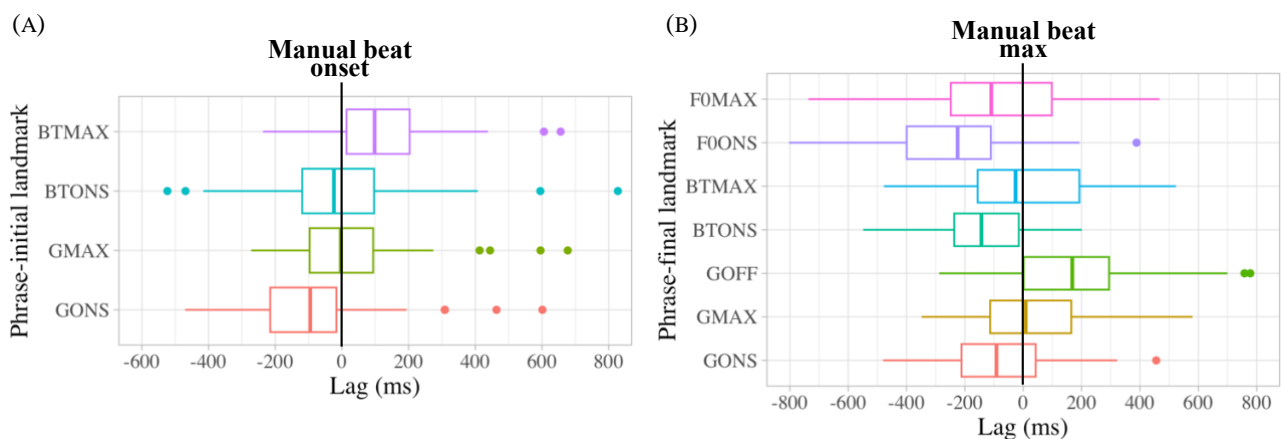


Fig. 1. Coordination at the onset (A) and target (B) of the manual beat gesture.

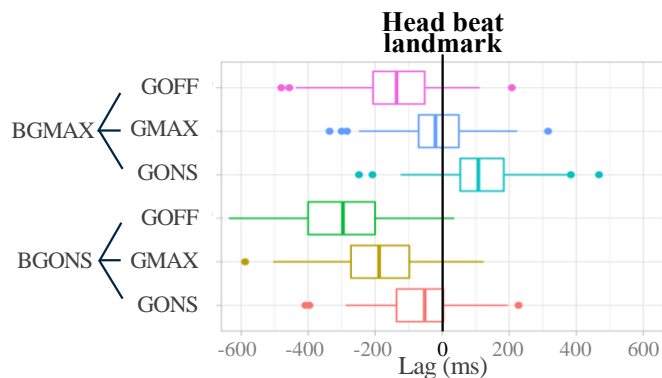


Fig. 2. Coordination at the onset (bottom three) and target (top three) of the head beat gesture.

References.

- [1] Krahmer, E., Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *JML*, 57(3), 396–414.
- [2] Shattuck-Hufnagel, S., Ren, A. 2018. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, 9, 1514.
- [3] Jun, S.-A. 1993. *The Phonetics and Phonology of Korean Prosody*. Ph.D. dissertation, The Ohio State University.
- [4] Oh, M., Lee, Y. 2018. ACT: An Automatic Centroid Tracking tool for analyzing vocal tract actions in MRI speech production data. *JASA*, 144(4), EL290–EL296.
- [5] Krivokapić, J., Tiede, M., Tyrone, M., Goldenberg, D. 2016. Speech and manual gesture coordination in a pointing task. *Proc. of Speech Prosody*, 1240–1244.