Phrase Boundary and Tone Melody as Predictors of Co-Speech Gesture Timing in Igbo

Kathryn Franich
Harvard University

Introduction: Co-speech gestures are timed to occur with metrically-prominent syllables in several languages [1,2,3]. Little research has examined the temporal alignment of co-speech gestures in African tonal languages, where metrical prominence is often hard to identify due to a lack of canonical stress correlates [5]. Existing findings for Niger-Congo languages indicate that co-speech gestures gravitate to stem-initial position [4], which in some languages corresponds with metrical prominence [6]. Lexical tone has not played a large role in gesture timing [4,7]. However, existing research has been limited to languages with relatively short words, barring exploration of factors other than stem position which may influence gesture timing. Here, we look at the timing of co-speech gestures in Igbo, a Niger-Congo language which features longer word length and a (H)igh vs. (L)ow tone system. While stem-initial position does appear to be an attractor of gestures for words in phrase-medial position, word position and tone melody—specifically, melodies involving sequences of H tones—play an important role in gesture alignment in words at the edges of a phrase. Results suggest that metrical organization at both the segmental and tonal tiers serves to constrain gesture alignment.

Methodology: Data consists of conversational speech produced by 4 Igbo speakers recorded in pairs in Northern Nigeria (Fig. 4). 1,300 gestures of the hands were coded manually in video data by a team of coders, with inter-coder reliability established [8]. Apexes of co-speech gestures—defined as peak velocity of manual movement (based on visualized blurring and kinematic measures using MediaPipe [9])—were extracted along with the phones with which they overlapped. We focus here on data involving words of 3-5 syllables (total of ~300 tokens). Most words consisted of at least one prefix, a monosyllabic stem, and one or more suffixes, though about 1/3 of words were polysyllabic stems. Data were coded for word-, intonational phrase-, and stem-position, and aligned tone. For tone melody, visual inspection of data suggested gestures were more likely to align with the initial H of a HH sequence; we therefore coded each syllable to indicate if it was in this position.

Results: Results of a mixed effects logistic regression model revealed that gestures were more likely to occur in stem-initial than non-initial position (β=0.99, z=2.20, p<.05) and more likely to occur in word-final than initial or medial positions (β=1.24, z=3.24, p<.01). However, interactions were observed with phrase position: stem-initial position was more likely to be targeted for a gesture phrase-medially (β=1.04, z=2.16, p<.05), while word-final position was more likely to be targeted in phrase-initial and final positions (β=1.70, z=2.80, p<.01) (Figs. 1&2). While there was no significant effect of tone on gesture occurrence (β=0.27, t=1.35, p>.05), there was a significant effect of tone melody (β= 1.32, z = 2.45, p < .01): gestures were more likely to align to the first H in a HH sequence, if it was unbroken by a downstep (Fig. 3).

Discussion: The gravitation of gestures to word-final position (and odd-numbered syllables, more generally) is consistent with arguments from Clark [10] suggesting this position is metrically strong in Igbo. We propose that word-final position is the default position for metrical prominence—and, hence, gestures—in the language, but that this position is disfavored for gestures phrase-medially due to the prevalence of vowel assimilation effects which obscure the boundary between word-final and word-initial vowels [11]. Even more striking was the effect of tone melody, which reveals a preference for gestures on syllables which initiate a HH sequence. Results are consistent with the presence of *tonal feet* in Igbo [12,13]. We demonstrate how both tonal and segmental tiers can exert metrical prominence effects, and that gestures are most likely to be attracted to syllables in which prominence aligns across these two tiers.
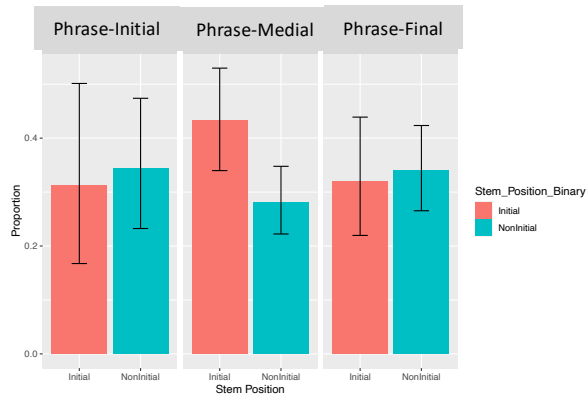
Figure 1: Proportion of gestures aligned across stem-initial (red) vs. non-initial (blue) positions, by phrase position
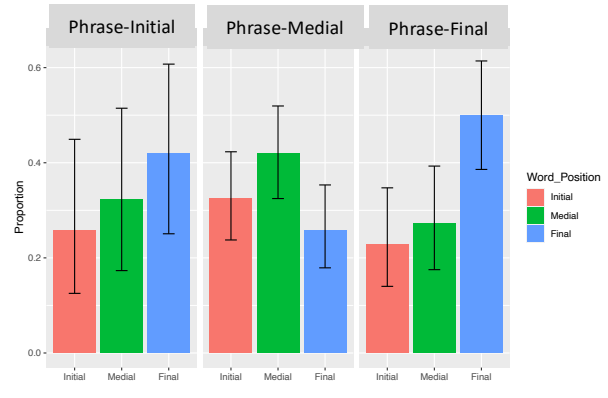

Figure 2: Proportion of gestures aligned across word-initial (red), medial (green), and final (blue) positions, by phrase position
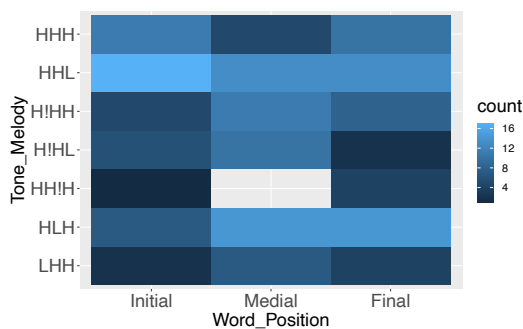

Figure 3: Gesture alignment across word positions, by tone melody (trisyllabic words only). Light blue shading indicates greater counts at a position; transparency indicates a lack of tokens in a given cell. Note the lighter shading in initial position for HHH and HHL vs. H!HH and H!HL, where word-medial position has more counts.

Figure 4: Igbo speakers in conversation task

References

[1] T. Leonard and F. Cummins, "The temporal relation between beat gestures and speech." Language and Cognitive Processes, 26, 10, 2011, pp. 1457-1471. [2] D. Loehr, Gesture and Intonation. Washington, D.C.: Georgetown University dissertation, 2012. [3] N. Esteve-Gibert and P. Prieto, "Prosodic structure shapes the temporal realization of intonation and manual gesture movements." JSLR, 56, 3, 2013, pp. 850-64. [4] Author, 2022 [5] Hyman, L. (2014). Do all languages have word accent? In H. Van der Hulst (Ed.), *Word Stress: Theoretical and Typological Issues* (pp. 56-82). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139600408.004. [6] Author, 2021. https://doi.org/10.1353/lan.2021.0021. [7] H. S. H. Fung and P. P. K. Mok, 'Temporal coordination between focus prosody and pointing gestures in Cantonese', J. Phon., vol. 71, pp. 113–125, 2018, doi: 10.1016/j.wocn.2018.07.006. [8] Kita, S., van Gijn, I., and van der Hulst, H. 2006. Movement phases in signs and co-speech gestures, and their transcription by human coders. *Gesture and Sign Language in Human-Computer Interaction*, 23-35. [9] Lugaresi et al. 2019. MediaPipe: A Framework for Building Perception Pipelines. [10] Clark, M.M. 1990. The tonal system of Igbo. Publications in African Languages and Linguistics. Dordrecht: Foris. Pp. xi+323. [11] Zsiga, E. Features, gestures, and Igbo vowels: An approach to the phonetics-phonology interface. *Language*, 73, 2, 227-274. [12] Leben, W.R. 1996. Tonal feet and the adaptation of English borrowings into Hausa. Studies in African Linguistics, 25, 139-154. [13] Bickmore, L. 2003. The use of feet to account for binary tone spreading. In *Frankfurter Afrikanistische Blatter* vol. 15, R. J. Anyanwu (ed.), Rudiger Koeppe Verlag, Koln.