Franich, Kathryn & Keupdjio, Hermann & Nwosu, Vincent. 2024. The timing of speech and gesture in two Niger-Congo languages: Implications for word-level prominence. *Glossa: a journal of general linguistics* 10(1). pp. 1–39. DOI: https://doi.org/10.16995/glossa.17426

Open Library of Humanities

The timing of speech and gesture in two Niger-Congo languages: Implications for word-level prominence

Kathryn Franich, Harvard University, Cambridge, MA, USA, kfranich@fas.harvard.edu

Hermann Keupdjio, Western University, London, ON, Canada, keupsmann2011@gmail.com

Vincent Nwosu, University of Calgary, Calgary, Alberta, Canada, vincent.nwosu@ucalgary.ca

Co-speech gestures are timed to occur with prosodically prominent syllables in several languages. In prior work in Indo-European languages, gestures are found to be attracted to stressed syllables, with gesture apexes preferentially aligning with syllables bearing higher and more dynamic pitch accents. 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, and where a key function of pitch is in distinguishing between words, rather than marking intonational prominence. Here, we examine the alignment of co-speech gestures in two different Niger-Congo languages with very different word structures, Medʉmba (Grassfields Bantu, Cameroon) and Igbo (Igboid, Nigeria). Our findings suggest that the initial position in the *stem* tends to attract gestures in Medʉmba, while the final syllable in the *word* is the default position for gesture alignment in Igbo; phrase position also influences gesture alignment, but in language-specific ways. Though neither language showed strong evidence of elevated prominence of any individual tone value, gesture patterning in Igbo suggests that metrical structure at the level of the *tonal foot* is relevant to the speech-gesture relationship. Our results demonstrate how the speech-gesture relationship can be a window into patterns of word- and phrase-level prosody cross-linguistically. They also show that the relationship between gesture and tone (and the related notion of 'tonal prominence') is mediated by tone's function in a language.

1 Introduction

Co-speech gesture—gestures of the hands, head, face, shoulders, and other parts of the body that accompany speech—have been found to show close temporal alignment with prosodic structure. In particular, gestures in many languages appear to temporally coincide with prosodically prominent events in speech: in several western Indo-European languages, for example, gestures of the hands and head tend to show close temporal alignment with stressed syllables, particularly when they also bear phrase-level pitch accents (Esteve-Gibert et al. 2017; Kendon 1980; McNeill 1992; Rochet-Capellan et al. 2008; Tuite 1993). More specifically, several studies have shown that gestures are timed such that the gesture *apex*—variously defined in terms of the point of peak extension or peak velocity of the articulators (the fingers, for example, in a manual gesture such as that shown in **Figure 1**)—is timed to occur with the pitch accent peak of a stressed syllable (Brentari & Coppola 2013 for Italian; Esteve-Gibert & Prieto 2013 for Catalan; Leonard & Cummins 2011; Loehr 2004; 2012; Pouw & Dixon 2019 for English; de Ruiter & Wilkins 1998 for Dutch). And though most research has focused on the timing of *beat gestures* (which are not, in their most basic form, referential or depictive of anything in the sentence), recent work suggests that most (if not all) gesture types show crucial prosodic timing to speech (Esteve-Gibert & Prieto 2013; Shattuck-Hufnagel & Prieto 2019).

Figure 1: Gesture apex as defined in terms of maximal extension of the fingers in a manual beat gesture.

In the case of English, where pitch accent type is closely linked to information structure and focus interpretation, it has been found that gestures are more likely to occur with stressed syllables bearing pitch accents with higher and more dynamic pitch profiles (Im & Baumann 2020). Indeed, even in languages where the presence of lexical stress is contested—as is the case in French and in Turkish—gestural apexes are still more likely to align with high pitch accents than with other accent or tone categories with lower average f0 (Rohrer et al. 2019; Türk & Calhoun 2023). And though deviations from perfect temporal alignment between pitch accent and gesture do occur (Leonard & Cummins 2011), studies have shown that manipulating the timing of pitch accents relative to co-speech gestures has a negative impact on speech processing (Swerts & Krahmer 2008; Duan et al. 2023; Morett 2023), and that manipulating other factors—such as proximity of a word/gesture to a phrase boundary—dually affects timing of pitch accents and co-speech gestures (Esteve-Gibert & Prieto 2013; Krivokapić et al. 2017). These findings all suggest a functional link in the production of speech and co-speech gesture, and one that may specifically implicate control of gesture and pitch (Pouw et al. 2020).

However, nearly all existing work on the timing of speech and co-speech gesture has been conducted on languages which (1) display clear acoustic evidence of either lexical stress, phrase-level accent, or both; and (2) show a close three-way relationship between pitch patterns, prosodic structure, and information structure (Féry 1993; Ladd 1996; Jun & Fougeron 2000; Ozge & Bozsahin 2010; Prieto 2014). Very little work has investigated the temporal alignment of co-speech gestures in tonal languages, in which the relationship between pitch, prosodic prominence, and information structure is often very different from that observed in non-tonal languages. The present work aims to fill this gap by studying the alignment of speech and co-speech gesture in two tonal Niger-Congo languages, Medʉmba and Igbo. As is the case for most Niger-Congo languages (particularly those situated in the northwestern region of Sub-Saharan Africa), neither of these languages displays typical acoustic correlates of lexical stress, such as increased vowel duration or intensity. Nonetheless, for both languages, there is some evidence to suggest that word-level metrical prominence asymmetries are present (Clark 1990; Franich 2018, 2021). Unlike many of the Indo-European languages examined for gesture alignment so far, tone is primarily used to encode lexical contrasts in both Medʉmba and Igbo. Information-structural relations such as focus and topic tend to be encoded through focus particles and word order, rather than through the use of prominence-lending pitch events (Kouankem & Zimmermann 2013; Osuagwu & Anyanwu 2020; Zimmermann & Kouankem 2024). Therefore, these languages allow us to examine whether the cross-linguistic link between gesture and tone is direct and widespread, or mediated by the relationship between tone/pitch, prominence, and information structure. In the following sections, we provide an overview of each of the languages under investigation.

1.1 Medʉmba

Medʉmba is a Grassfields Bantu language spoken in Cameroon by approximately 200,000 people according to Ethnologue (Eberhard et al. 2024). While demonstrably descended from Proto-Bantu (Voorhoeve 1971), the language looks quite different from Eastern and Southern Bantu languages in having reduced segmental morphology (especially in the use of noun class and concord prefixes), leaving it with a relatively short average word size (around 1.5 syllables). In part as a consequence of this reduced segmental system, the language also boasts an extensive system of tonal morphology. Though the language is analyzed as having only an underlying high vs. low tonal contrast, contour tones can appear at the surface level as a result of several types of processes. Medʉmba utilizes tonal morphemes ('associative markers') for marking possession and other associations, as in various types of compounds (1). In such examples, the tonal associative marker always docks to the right edge of the first member of the construction, either forming a contour on that syllable (if the associative morpheme and initial member's tone differ) or merging with an existing matching tone. Contours can also form as a result of certain types of grammatical 'overwrite' rules, such as in the replacement of a word's underlying tone with a high-low contour as a reflex of A' agreement (2) (Keupdjio 2021).

(1) Tonal contour resulting from association of a tonal morpheme for possessive and compound constructions (vowels with derived contours marked in red)

(a) bàm + H + mɛ́n → bắm ˡmɛ́n
belly + AM + child "the child's belly"

(b) ᵑgʉ́b + L + ⁿʒʉ́ → ᵑgʉ̂bⁿʒʉ́
cut + AM day "morning"

(2) Tonal contour (in red) resulting from morphological tonal overwrite (Keupdjio 2021)

(a) Mɛ́n ˡjə́n = í
Child see = 3SG
'The child has seen him/her.'

(b) Mɛ́n zə̀ à jə̂n ˡmʙʉ́ ˡlɑ
Child COMP 3SG see dog REL
'The child who has seen the dog'

(c) Mɛ́n ˡlàb = í
Child see = 3SG
'The child has hit him/her.'

(d) Mɛ́n zə̀ à lâb ˡmʙʉ́ ˡlɑ́
Child COMP 3SG see dog REL
'The child who has hit the dog'

As mentioned, Medʉmba does not have clear evidence of lexical stress as measured through changes in amplitude or duration, but certain phonological patterns provide evidence for word-level prominence asymmetries in the language. Medʉmba is one of several languages around the

region known as the Macro-Sudan Belt (Güldemann 2007) that display evidence of stem-level prominence asymmetries, such that stem-initial position tends to carry a much broader array of consonantal and vocalic contrasts than non-initial positions in the stem or affixes (Hyman et al. 2019; Idiatov & van de Velde 2016; Lionnet 2017). As can be seen in (3a-b), stem-initial position (underlined and bolded in polysyllabic words) in Medʉmba can realize a range of vowel contrasts, while non-initial positions (stem-final, prefix, and suffix positions) realize only schwa[1]. Consonants are also lenited in non-stem-initial position when they occur intervocalically. The verb *kàɣə́* 'release' is realized without its final vowel and with a final velar stop when the word occurs phrase-medially for *kàk* (3d), but the velar stop is lenited to [ɣ] preceding the final vowel in (3c). Notably, the stem-initial /k/ resists lenition in (3c), though it also occurs intervocalically. Acoustic signals of /k/ realized in different stem positions are shown in (4).

(3) Patterns of stem-initial prominence in Medʉmba

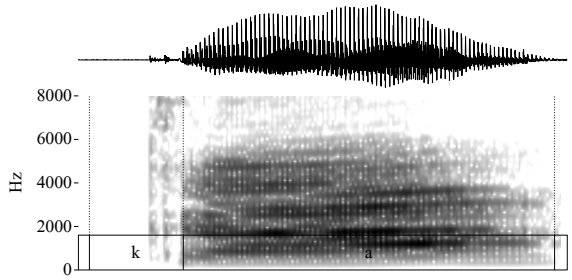
(a) Mɛ́n ˈlɛ̌n ˈnə-**zi̱**nə́
child know INF-walk
'The child knows how to walk'

(b) Mɛ́n ˈkw**ɨ́**'-də́ ˈmbw**ɔ́**ɣə́
child tend-ITER fire
'The child has tended the fire'

(c) Ŋ̂, á ˈlɛ̌n nə-kàɣə́
yes 3SG know release
'Yes, he/she know how to release it.'

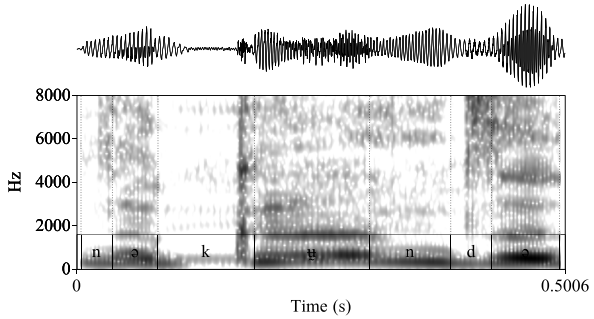
(d) À? kàk ŋgáp
3sg.FUT release hen
'He/she will release the hen.'

(4) Acoustic realizations of /k/ across stem positions

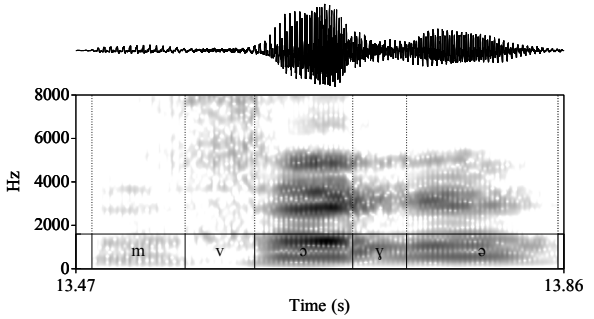


(a) Stem- and word-initially in the word *ká* 'plate,' /k/ is realized as a plosive with a closure and clear release burst

[1] One exception to this generalization is found in words containing medial glottal stops, in which the stem-final syllable will always match in quality with that of the stem-initial syllable (e.g. nə-ʒúʔú 'to listen').



(b) Stem-initially and word-internally in the word *nə̀-kʉ́ʔ-ndə́* 'to knock,' /k/ is realized as a plosive with a closure and clear release burst



(c) Stem-internally, in the word *mvɔ́k* 'bone' (produced *mvɔ́yə́* phrase-finally), /k/ is realized as voiced fricative [ɣ] or approximant [ɰ].

Franich (2018, 2021) demonstrates that stem-initial syllables in Medʉmba show elevated rhythmic prominence in a phrase repetition task. Furthermore, loanwords, which can be longer than native words (up to 3 or 4 syllables), show alternating patterns of consonantal lenition, indicating that 'stem-initial prominence' is actually more fruitfully characterized as metrical prominence. Given the importance of metrical prominence in the patterning of co-speech gestures cross-linguistically, we hypothesize that metrically-prominent stem-initial syllables in Medʉmba will be more likely to attract gestures than non-initial syllables. An open question concerns whether tone will also play a role in determining gesture positioning in the language.

1.2 Igbo

Igbo is a Niger-Congo language of the Volta-Niger subfamily (sometimes referred to as West Benue-Congo or Eastern Kwa) spoken in Nigeria. Like Medʉmba, the language features both high and low tones, though Clark (1990) presents arguments that high tone is underspecified in the language, inserted as a default tone on otherwise toneless syllables. Contour tones can be formed through some morphophonological processes, but are generally rarer in Igbo than in Medʉmba. Unlike Medʉmba, Igbo, like other 'Kwa-type' languages, does not display vocalic and consonantal contrast asymmetries consistent with stem-initial prominence (see Lionnet 2017 for discussion of patterning in Kwa more generally). Igbo also has a relatively more agglutinating

structure than Medʉmba, with a large array of affixes (5–6). As can be seen in (5–6), vowels within a (non-compound) word generally undergo ATR harmony.

(5)	a.	ọ́-ˡsḭ́-álá	3SG-tell-INDIC	‘He/she/they have told’
	b.	ḭ́-ˡzʉ́-rʉ́	2SG-buy-INDIC	‘You have bought’
(6)	a.	ḭ́-bʉ̀	INF-sing	‘To sing’
	b.	ḭ́-ˡsḭ́	INF-say	‘To say’
	c.	í-ˡméchí	INF-close	‘To close’

Nouns, like verbs, tend to begin with vowels that are historically related to a noun class prefix system; however, this system is no longer seen to be productive, and these vowels are now analyzed as initial in the root (7) (Zsiga 1992).

(7)	a.	ákwá	‘cry’
	b.	ènyò	‘mirror’
	c.	òbí	‘ancestral house’
	d.	ùsòlò	‘row’
	e.	ìfúˡlú	‘flower’

Due to the fact that many words in the language are both vowel-initial and vowel-final, vowel hiatus arises with very high frequency in the language. Ihiunu and Kenstowicz (1994) document a process in vowel sequences in which the first vowel assimilates in quality to that of the second, while the tone of the first vowel is maintained (8). Note that an additional process of regressive high tone spreading applies to change the tone of the second vowel in (8a) and (8d) (see Uwaezuoke 2021 for further details).

(8)	a.	ré ìkó	→	rí ˡíkó	‘sell a cup’
	b.	cọ́ ákwà	→	cá ákwà	‘seek a cloth’
	c.	kè ófé	→	kò ófé	‘share out soup’
	d.	rọ̀ ʉ̀tá	→	rʉ̀ ʉ́tá	‘bend a bow’

Zsiga (1993, 1997), taking a quantitative approach, demonstrates that the process of vowel assimilation (or perhaps better termed ‘coalescence’) is not categorical, but gradient. She models the process of assimilation as a process of gestural overlap, drawing on gestural representations from Articulatory Phonology (Browman & Goldstein 1986, 1992). Based on durational data, she proposes that the duration of the first vowel in the VV sequence (the word-final vowel of the initial word) shortens, allowing for the percept of the second vowel to be more robust relative to the first. She notes that, among other things, the process of vowel shortening and gestural overlap does not occur as regularly at phrase boundaries (where domain boundary strengthening serves to increase gestural magnitude cross-linguistically; Byrd et al. 2000, Cho & Keating 2001) as within phrases. The variability of this process of coalescence will become important to our discussion of the results for Igbo co-speech gesture timing presented in Section 4.

Another process that will be relevant to our discussion of gesture timing in Igbo is tonal downstep, whereby the second high in a sequence of two high tones is produced at a relatively lower pitch value than the first. Downstep (ꜜ) is analyzed by Clark (1990) as arising where a sequence of two high tones occurs (9a); downstep does not occur across syllables which are linked to the same high tone (via tone spreading), as in (9b).

(9) a. H H
σ́ ꜜσ́

b. H
σ́ σ́

An example of how this downstep arises is in (10). Here, we see that when the high tone word *ócé* 'chair' takes the high tone 3rd person singular possessive clitic *=yá*, downstep arises (though it does not arise within the root itself, since the noun contains a single multiply-linked high tone).

(10) a. H
ócé
chair
'chair'

b. H H
ócé ꜜyá
chair 3SG.POSS
'his chair'

A challenge to this analysis of Igbo downstep comes from the associative construction, where a noun is prefixed with a floating tone[2]. When that floating tone is high, it will link to the initial vowel of the root, delinking the previously linked high tone (10a,b). This associative tone (labeled AM in examples 11–12) is therefore generally expected to trigger downstep on the second syllable of the root, due to the newly created sequence of two high tones. Clark (1990) points out that this is precisely the outcome for disyllabic words in the associative construction (11), but that this is not the outcome for trisyllabic words in the same construction (12).

(11) a H
eɣu
'goat'

b. H H → H H
AM eɣu eɣu
Surface Form:
éꜜɣú
'of a goat'

[2] While Clark (1990) assumes the associative tone is linked to an empty V slot, we find an analysis with a simple floating tone morpheme to be more parsimonious while still capturing the necessary generalizations.

(12) a. H
akụkwọ
'book'

b. H H → H H
AM
akụkwọ akụkwọ

Surface Form:
ákụ́kwọ́
'of a book'
(c.f. *á^{l}kụ́kwọ́)

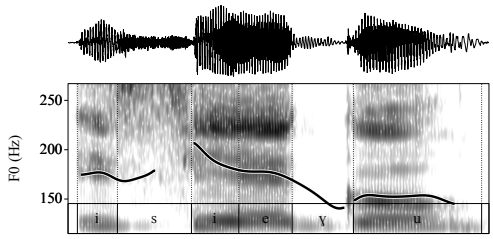
Clark proposes that the lack of downstep on trisyllabic forms such as those in (12) is attributable to metrical prominence asymmetries between initial syllables in di- and trisyllabic words (13). Specifically, she proposes that the language assigns alternating prominences from the right edge of the word, and that initial metrically-prominent syllables actually resist the linking of the associative high tone (and the subsequent downstep that it would trigger).

(13) a. $\overset{*}{\text{é}}{}^{l}\text{ɣú}$

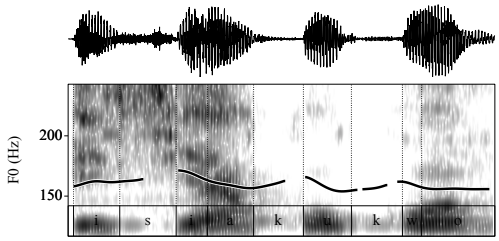
b. $\overset{*}{\text{á}}\text{kwụ́kw}\overset{*}{\text{ọ́}}$

Acoustic representations of the phrases *ísí é^{l}ɣú* 'head of goat' and *ísí ákụ́kwọ́* 'head of book' are given in (14). Note the clear pitch lowering on the last syllable in *é^{l}ɣú*, reflecting the application of downstep, while the word *ákụ́kwọ́* maintains a flat pitch level throughout.

(14) Patterns of downstep in the associative construction in Igbo on disyllabic and trisyllabic words. Think black lines represent pitch traces.



(a) *ísí é^{l}ɣú* ('head of goat') with downstep affecting final syllable



(b) *ísí ákụ́kwọ́* ('head of book') with no downstep and even pitch throughout

Our prediction for gesture alignment in Igbo, based on Clark's analysis, is that word-final syllables (and possibly also word-initial syllables in trisyllabic words) will be the most likely syllables to attract gestures in the language. However, given the prevalence of stem-initial prominence patterns in Niger-Congo, including in languages which neighbor Igbo, such as Ibibio (Akinlabi & Urua 2003; Harris & Urua 2001), we also consider that stem-initial syllables could be gesture-attracting in Igbo. Again, it is an open question whether tone might also influence gesture timing. Given Zsiga's observation that vowel coalescence may be more likely to target word-final vowels in certain prosodic environments, we also examine phrase position as a possible determinant of gesture timing.

2 Method

2.1 Participants and Procedure

This research was granted ethics approval by the Institutional Review Board of Harvard University (Protocol IRB 22-1095). Six Medumba speakers from the village of Bangoulap, Ndé Division of the West Region of Cameroon (3 men, 3 women) and four Igbo speakers from the greater Abuja area in Nigeria (3 men, 1 woman) participated in the study. Three of the Igbo speakers spoke the Ọnịchà dialect, while one spoke the Owerri dialect. The age range of the participants was between 35 and 55 years. Speakers were generally interviewed in pairs (except for two of the Medumba speakers) in a quiet room or outdoor area. They were interviewed about aspects of language and cultural practices, such as traditions around marriage ceremonies and child naming. Medumba interviews were conducted by a team of researchers which included a native Medumba speaker. Interviews with Igbo speakers were conducted by a native Igbo speaker (the third author).

Individual audio tracks were recorded for each participant using a Shure SM10 head-mounted microphone input to a Zoom Q8 video/audio recording device at an audio sampling rate of 48 kHz and a video sampling rate of 30 frames per second. Interviews were conducted for around 25–30 minutes, resulting in between 20,000 and 25,000 phones recorded per language. This yielded a total of 1073 gesture-aligned phones for Igbo and 1302 gesture-aligned phones for Medumba. Though the vast majority of the gestures obtained in the data (73%) were categorized as beat gestures, we opt to include all gesture types in the present analysis in order to maximize our sample size.

2.2 Processing of Audio and Video Data

Audio data were transcribed by native speakers of the two languages and force-aligned using the FAVE aligner (Rosenfelder et al. 2022). Alignments were subsequently checked by the first and second authors for accuracy. Gestures were coded by a team of trained annotators in ELAN software (ELAN 2024). Coding was done according to an adapted set of criteria based on the MIT Speech Communications Group Gesture Studies Coding Manual (MIT Speech Communications Group, 2020) with the file's audio muted in order to avoid speech-related bias in coding decisions. Coders were asked to observe which hand the speaker tended to gesture with most, and to treat that hand as the dominant hand for the sake of coding. Coders were tasked with marking off intervals for the various intentional movements of the gesture, including the *preparation* phase (if present), the *stroke*, any *holds* (pre- or post-stroke, if present), and the *recovery* (if present) (**Table 1**). All of these intervals were marked on a single tier. On a separate tier, coders marked off an interval within the stroke phase which corresponded with the gesture *apex*. In prior work,

the definition of the apex has varied, with some authors using a spatial definition (usually the point of maximum extension of articulators such as the fingers in a hand gesture) (Loehr 2012), while others have relied on a definition based on peak or minimum velocity of movement of an articulator (Pouw & Dixon 2019; Pouw 2020; Trujillo et al. 2018). We opt to use the timing of peak velocity as our measure of the gesture apex for various reasons. First, Pouw & Dixon (2019) demonstrate that, at least for English, it is the peak velocity of the gesture that most closely approximates the pitch peak in a pitch accented syllable. Second, a drawback of using the point of maximum extension in the present case is that some types of gestures—such as those with a circular path of motion—do not have a clear point of maximum extension. All gestures do, however, have a point of maximum velocity. In the present work, we therefore rely on the visualized point of peak velocity as a measure of apex timing. This point was identified by coders as the start of the frame in which the target hand moved the most (typically the frame that involved the blurriest visualization of the hand). In prior work, we have demonstrated that this visualized point of peak velocity closely approximates the true moment of peak velocity as measured computationally (Dych et al. 2023).

Gesture Landmark	Kinematic Definition	Required phase?
Preparation onset	Start of movement of the hand into position, prior to stroke onset	No
Stroke onset	Start of intentional movement of the hand, regardless of direction	Yes
Stroke apex	Timing of peak velocity of manual movement, as observed from relative distance moved/blurriness from one frame to the next	Yes
Stroke offset	Endpoint of intentional movement of the hand	Yes
Recovery onset	Start of less intentional movement of the hand towards a rest position	No

Table 1: Gesture landmarks with kinematic definitions.

2.3 Establishing Coding Reliability

As described by Shattuck-Hufnagel and Ren (2018), a key part of gesture coding involves the identification of *intentional* movements—those which constitute part of the communicative act, and which can be differentiated from unintentional 'fidgets,' which lack communicative intention. These movements are described by Kendon (1980) as constituting part of *the phrase of*

gesticulation or *G-Phrase*. At present, specific kinematic correlates of intentional movements have yet to be defined, but our experience leads us to believe that increased speed and magnitude of movement are both likely to be predictors of perceived intentionality of a movement. Given the lack of a kinematic measure of intentionality, we relied on coders' perceptions of intentional movements, having coders work in pairs in order to establish reliability of their results. Coders were highly consistent in their judgments of gestures vs. fidgets, agreeing approximately 95% of the time on whether a movement should be coded as a gesture. Where two coders disagreed on the status of a given gesture, a third member of the research team was consulted for a tie-breaking judgment.

Coders were also tasked with establishing reliability of the timing of their gesture apexes. To do so, coders worked in pairs, with the goal of having their respective apex times in a given participant's file within one video frame of each other. After establishing reliability on a series of video samples, coders in a given pair worked on a participant's file together, continuing to check that their gestures aligned temporally throughout the file they co-coded, discussing any disagreements that arose between them and bringing in a third research team member for tie-breaking judgments if a consensus could not be reached.

2.4 Data Post-Processing

Timestamps for gesture and speech data were extracted from ELAN .eaf files and Praat .TextGrid files, respectively, and merged using a Python script. In order to understand the overall timing of gestures relative to the words they occurred with, temporal lag between gesture apex and the corresponding phone onset was calculated for all gestures. Each syllable of a gesture-aligned word was then coded for position in word, position in stem, the tone with which the syllable was realized, tone melody of the word as a whole, and the position of the syllable within an intonational phrase. Given that cues to intonational phrasing in these languages are still not fully defined, for the purposes of the present analysis, we use the edges of 'breath groups' (Lieberman 1967) as an approximation of intonational phrase boundaries (Pierrehumbert 1980). As is common in African tonal languages, both languages apply a pattern of pitch reset at the beginnings of units consistent with intonational phrases (Downing & Rialland 2017; Kügler 2017; Kula & Hamann 2017). Pitch reset was present at the boundaries of a large proportion of the breath groups marked in our datasets, giving us confidence in our approach to marking prosodic phrases. Future work will seek to describe acoustic and articulatory cues of intonational phrases in Medʉmba and Igbo in greater detail.

2.5 Hypotheses to be Tested

Based on evidence presented in §1 for greater rhythmic prominence of stem-initial syllables in Medʉmba, we hypothesize that these syllables will be more likely to attract a co-speech

gesture than other positions within the stem. We furthermore hypothesize that stem position will be a better predictor of gesture presence (i.e., whether a given syllable is aligned with a gesture apex or not) than word position. In Igbo, conversely, we predict that word-final position will be more likely to attract a gesture than other positions in the word, and that, given a lack of positional effects linked to stem position in the language, word position will be a better predictor of gesture presence than stem position. For both languages, we incorporate exploratory hypotheses regarding the influence of phrase position, lexical tone, and tonal melody.

2.6 Statistical Analysis

For each language, two sets of linear mixed effects models were fit to the data. The first set of models included fixed effects for Word Position (up to four levels, depending on number of syllables) Stem Position (coded as a binary variable with levels 'Initial' and 'NonInitial'), Phrase Position (three levels: Initial, Medial, Final), and Aligned Tone (the value of the tone or contour carried by the syllable to which a gesture was aligned). For Igbo, less than 2% of gestures aligned to words containing contour-toned syllables. We therefore limited our Igbo analyses to examine words containing high and/or low tone syllables. For Medʉmba, tone was coded as a four-level factor including values of High, Low, Falling, and Rising. Two-way interactions between a) Word Position and Phrase Position and b) Word Position and Aligned Tone were included for Igbo; for Medʉmba, due to our theoretical predictions, we included interactions between a) Stem Position and Phrase Position and b) Stem Position and Aligned Tone. The second set of models were identical to the first, except that the variable Aligned Tone was swapped out for Tone Melody. Due to the sparsity of data for some melodies, we opted to trim from the dataset those melodies which accounted for less than 2% of the data. Following Barr et al. (2013), by-subject random slopes for all main effects were included in each model. Variance Inflation Factors were checked for each model to ensure that collinearity between fixed effects was reasonably low; the maximum GVIF value across models was 5 (and most values were within the 1.5–2.5 range). Significance of fixed effects and interactions for all models were assessed using Wald's tests. Where more than two levels of a variable were being compared, Tukey-adjusted *p*-values are reported.

Because only a small proportion of gesture-containing words in Medʉmba were longer than two syllables (around 1%), we examined gesture patterning across polysyllabic words within a single model. We also opted to separate out compound words from non-compounds in Medʉmba, given the prevalence of compounding in the language and the potential confounds they could introduce in our analysis. For Igbo, since longer words were more common, we conducted separate analyses for each word size (disyllabic, trisyllabic, and quadrisyllabic).

3 Results

3.1 Medʉmba

Looking at overall alignment of gesture apexes within syllables in Medʉmba (across 1302 unique words), we see that apexes tended to align within syllable onsets, but close (within 25 ms) to the onset of the vowel (**Figure 2**). This would suggest, interestingly, that gesture apexes are not aligning to pitch peaks, as we would expect apexes to occur relatively later on average (within the vowel), if this were the case. This finding is also in line with findings from Franich & Keupdjio 2022, in which it was found that pitch peak timing was not a significant predictor of gesture apex timing. While future work will need to examine this timing pattern more in depth, our results provide preliminary evidence that gestures are timed in language-specific ways to speech events. In Medʉmba, the relevant point of alignment with the speech signal could be the articulatory onset of the vowel (which tends to occur earlier than the acoustic signal would indicate; Browman & Goldstein 1988), or perhaps the *perceptual center*, which tends to coincide with listeners' perceived 'moment of occurrence' of the syllable, and usually aligns somewhere around the acoustic vowel transition (Morton et al. 1976; Scott 1993).

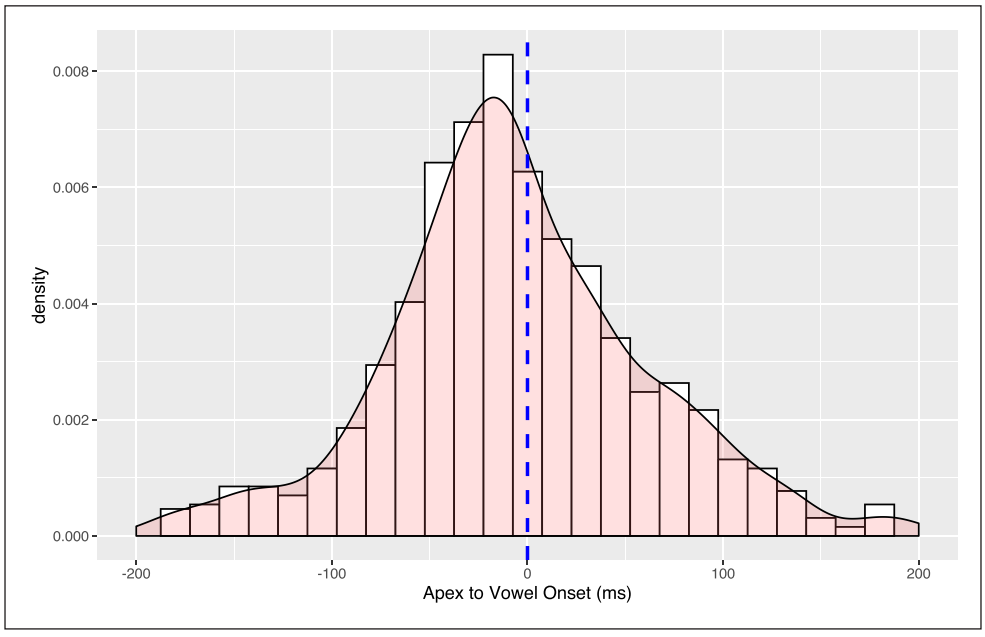


Figure 2: Overall alignment of gesture apexes within the syllable, Medʉmba. Blue dotted line indicates vowel onset.

3.1.1 Effects of Word and Stem Position

Examining how the linguistic factors of interest interact with gesture timing, for non-compound polysyllabic words (for which we had 382 unique words in our corpus), we found that word

position did not reach significance in predicting gesture timing in Medʉmba (Word-initial vs. Word-final: $\beta = 0.54$, $z = 1.245$, $p = .44$; **Figure 3**), but that stem position did: stem-initial position was significantly more likely to attract a co-speech gesture than non-initial (stem-final, prefix, and suffix) positions ($\beta = 1.46$, $z = 3.125$, $p < .01$; **Figure 4**). Thus, despite overlap in stem and word position in many words, variance in the data was better explained by stem position than word position.

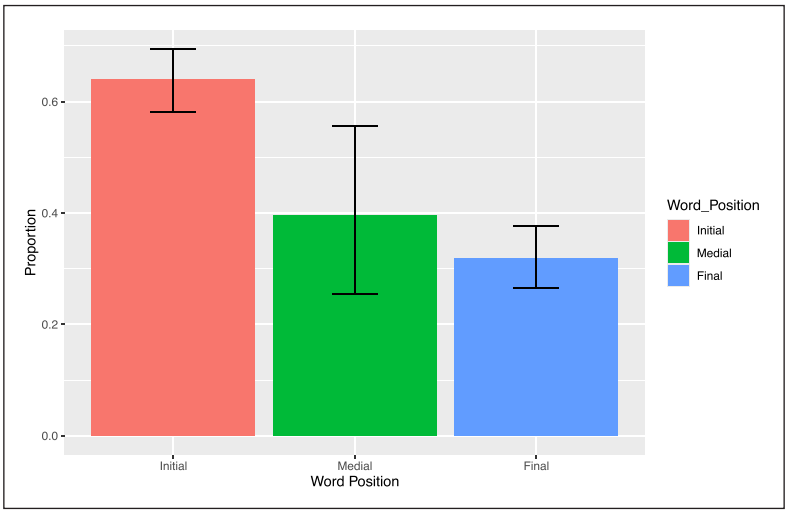


Figure 3: Gesture occurrence by word position in Medʉmba

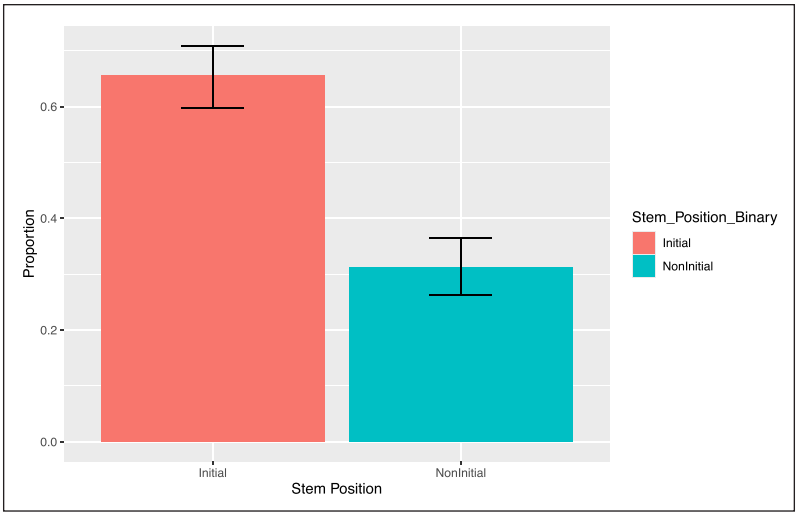


Figure 4: Gesture occurrence by stem position in Medʉmba.

The importance of stem position over word position can be seen more clearly if we examine just the subset of prefixed words (e.g. *nà-zìnɔ́*, INF-walk, or 'to walk') in which we see that medial position becomes a strong attractor of gesture presence (given that the stem-initial syllable is typically

word-medial in infinitives) (**Figure 5**). As demonstrated in **Figure 6**, stem-initial position remains the stronger gesture-attracting position even in these words where stem-initial position is not word-initial.

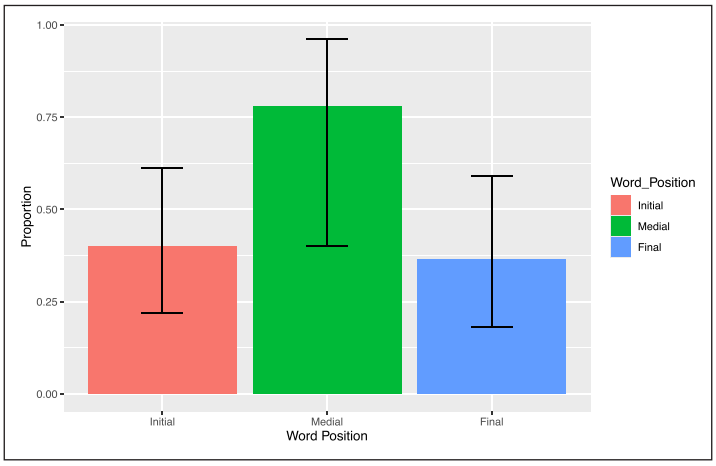


Figure 5: Gesture occurrence by word position in Medumba, prefixed words.

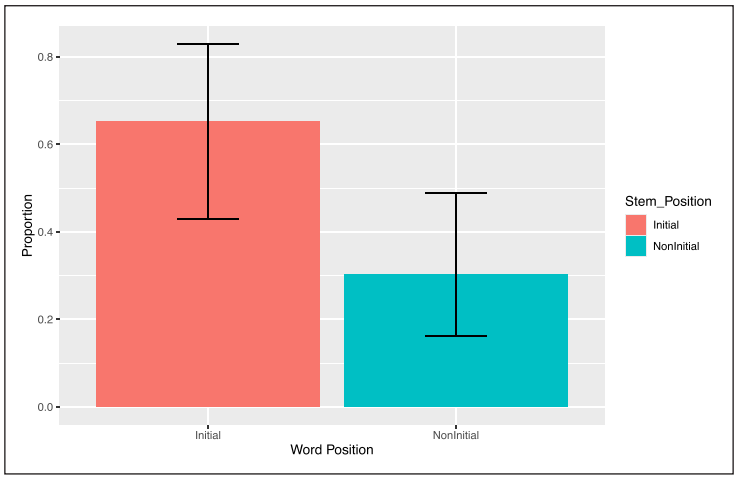


Figure 6: Gesture occurrence by stem position in Medumba, prefixed words.

Examining gesture patterning in compound words (for which there were 86 unique words in our corpus), we find a significant effect of word position on gesture alignment, such that word-initial position was significantly more likely to attract a gesture than final position ($\beta = 1.60$, $z = 3.141$, $p = .01$; **Figure 7**). However, word-medial position did not differ significantly from word-final position in terms of gesture attraction ($\beta = –4.59$, $z = –1.613$, $p = .24$). If we examine word position across words of different lengths (disyllabic vs. trisyllabic), we see that word-medial position is in fact a slightly stronger attractor of gestures for trisyllabic compounds, though patterns are more variable for these words (**Figure 8**).

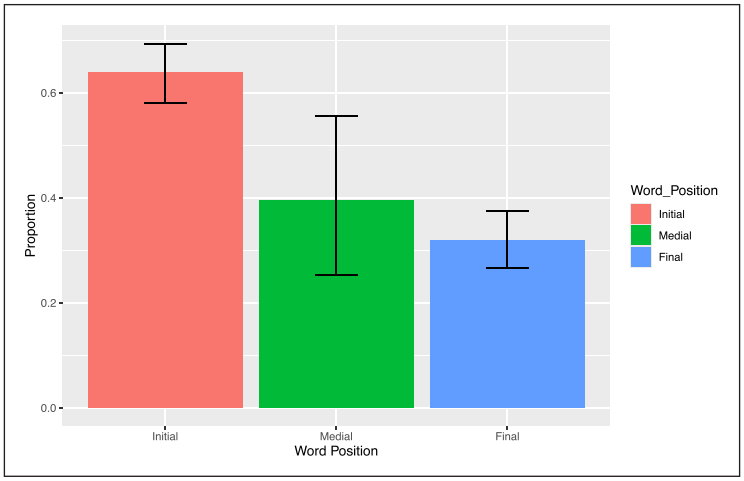


Figure 7: Gesture occurrence by word position in Medumba, compounds.

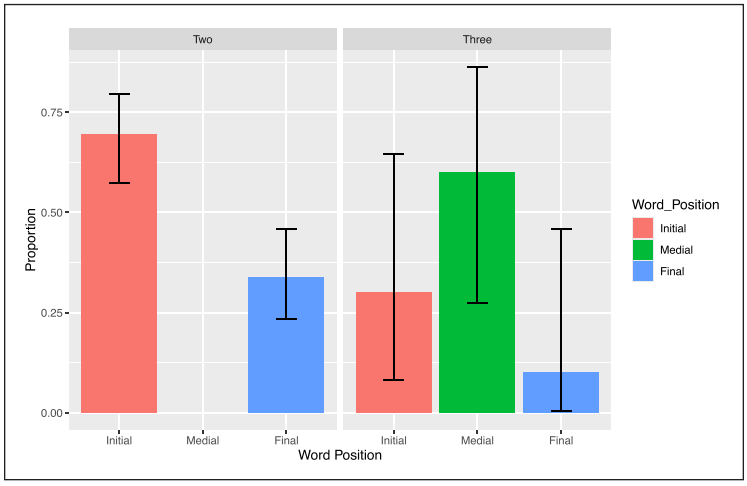


Figure 8: Gesture occurrence by word position and length in Medumba, compounds.

This difference in patterning between disyllabic and trisyllabic compounds suggests that additional layers of structure—prosodic or syntactic—are at play in driving gesture timing. Medumba compounds can be both left-headed (e.g. *ʤʰǔ?-ⁿʧə́?*, place-one, or 'together' and *mèn-zʷí*, person-female, or 'woman') and right-headed (e.g. *tèt-tsʷə́?*, middle-night, or 'middle of the night' and *nû-ntsú?*, ingest-container, or 'cup'), with right-headed compounds more common in the current corpus. We also note that in many of our trisyllabic compounds, the first two syllables were stem-initial, while the third was not. Examples of these structures include *ⁿtǔ?-kə́mə́* (fetch-peace) meaning 'mediator' and the encliticized form *ʃàᵐ-bʰə́=ə̂m* (pass-front=1sg.Poss.cl1) meaning 'my elder.' One possibility is that disyllabic compounds are incorporated, by default, into a single left-headed prosodic constituent (e.g. a prosodic word), while trisyllabic words are parsed into a more complex structure. The relative prosodic weakness of stem-final

syllables and enclitics may force them to prosodify with a preceding (word-medial) syllable, forming a minimal prosodic constituent to which the word-initial syllable then attaches. Under this account, the minimal prosodic constituent would be the one to which the gesture is attracted.

3.1.2 Effects of Phrase Position and Tone

There was also a significant interaction between stem position and phrase position in predicting gesture occurrence in non-compound words in Medʉmba ($\beta = 1.69, z = 3.031, p < .01$; **Figure 9**). Specifically, stem-initial position was seen to be a stronger attractor of gestures than non-initial positions in phrase-medial and phrase-final positions, but this effect diminished to some degree in phrase-initial position. This difference appears to be driven by a tendency for gestures to be realized relatively later when following a phrase boundary. This finding is in line with results from Esteve-Giberta & Prieto (2013) who showed that pointing gesture apexes were timed relatively later relative to an associated stressed syllable when the stressed syllable occurred immediately after an intonational phrase boundary. We also see a higher proportion of gestures realized on the stem-initial syllable in phrase-final position, suggesting that gestures may have occurred relatively earlier in anticipation of an upcoming phrase boundary. This would also be in line with findings from Esteve-Giberta & Prieto for Catalan, although the difference in gesture patterns between phrase-medial and phrase-final positions did not reach significance in Medʉmba ($\beta = 0.59, z = 1.431, p = .15$).

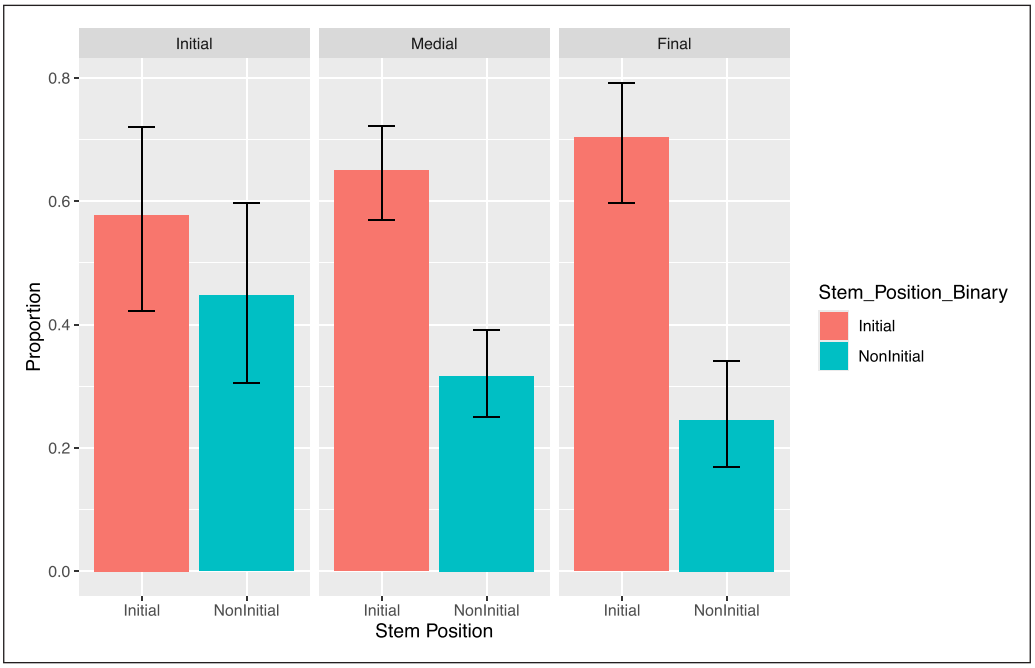


Figure 9: Gesture occurrence by stem position and phrase position, Medʉmba.

Results from the second model revealed that gesture position was not influenced in Medʉmba by either aligned tone (examined in the first model) (βs < 0.700; zs < 1.100; ps > 0.27) or tone melody (examined in the second model) (βs < 0.700; zs < 1.00; ps > 0.36). It is therefore not the case that syllables bearing relatively higher or more dynamic pitch profiles attract gestures at greater rates in Medʉmba.

3.2 Igbo

As can be seen in **Figure 10**, looking across 1020 unique words, gesture apexes were generally aligned very close to the vowel onset in Igbo. This is somewhat in contrast with our alignment results for Medʉmba, where gesture apexes were found to align earlier within the syllable, usually within the syllable onset. We note that the high proportion of vowel-initial words in Igbo, paired with the relatively less complex inventory of onset clusters, may be responsible for this difference. Either way, these results are also consistent with the idea that the gesture apexes are aligning to a landmark such as the vowel onset or perceptual center in Igbo.

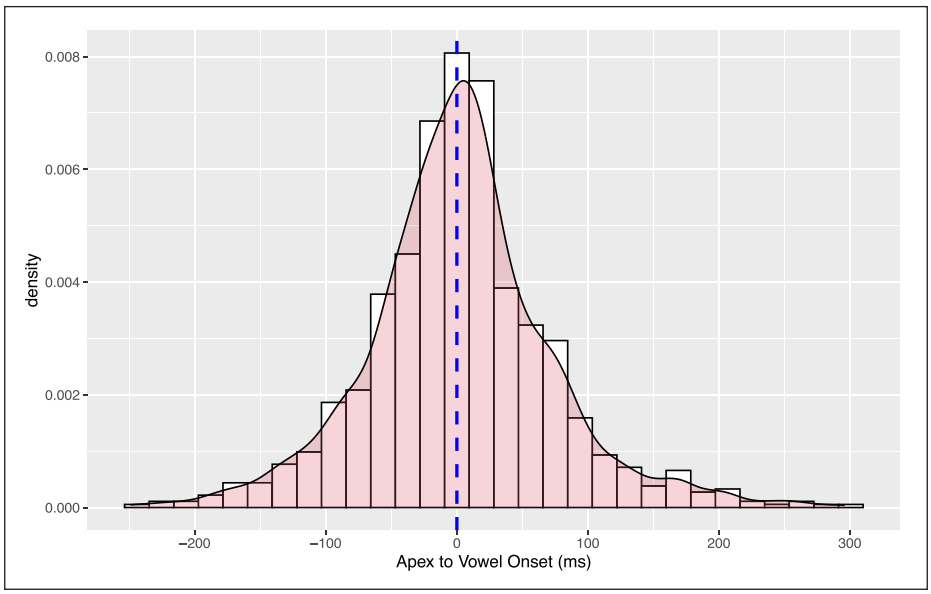


Figure 10: Overall alignment of gesture apexes within the syllable, Igbo. Blue dotted line indicates vowel onset.

3.2.1 Disyllabic Words

Turning now to our model results, we first examine disyllabic words, of which there were 406 in total in our sample. As predicted by Clark's metrical analysis of Igbo, final syllables in disyllabic words were more likely to be targeted for a gesture in Igbo than initial syllables ($\beta = 1.27$, $z =$

3.411, $p < .001$; **Figure 11**). Though the effect size was not as large, stem position also turned out to be a significant predictor of gesture location, with stem-initial syllables more likely to attract a gesture than non-initial syllables ($\beta = 0.61$, $z = 2.508$, $p < .05$; **Figure 12**). Though there were no significant effects of tone on gesture alignment, we did find that low tones were marginally more likely to attract a gesture than high tones ($\beta = 0.40$, $z = 1.684$, $p = .09$), and a significant interaction between tone and word position indicated that word-final syllables were especially likely to attract a gesture if they bore a low tone ($\beta = 1.03$, $z. = 3.124$, $p < .01$; **Figure 13**). Finally, there was a marginal interaction observed between word position and phrase position: though word-final position was overall the most likely position to attract a gesture, this effect became somewhat weaker in phrase-medial position ($\beta = 0.57$, $z = 1.611$, $p = .10$; **Figure 14**).

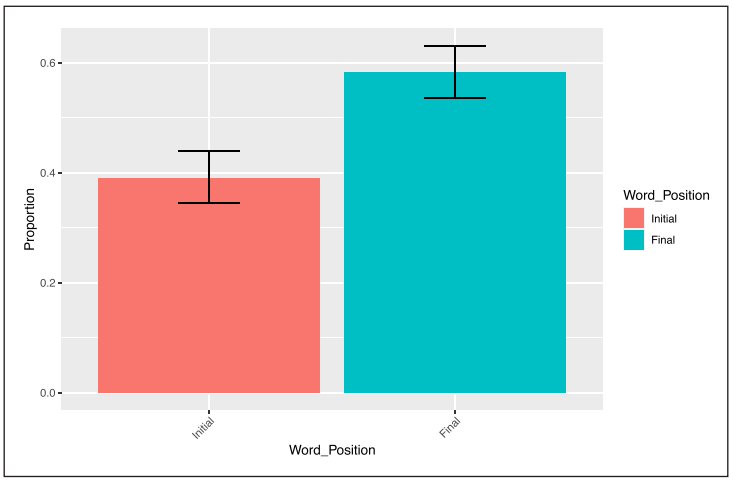


Figure 11: Gesture occurrence by word position in Igbo.

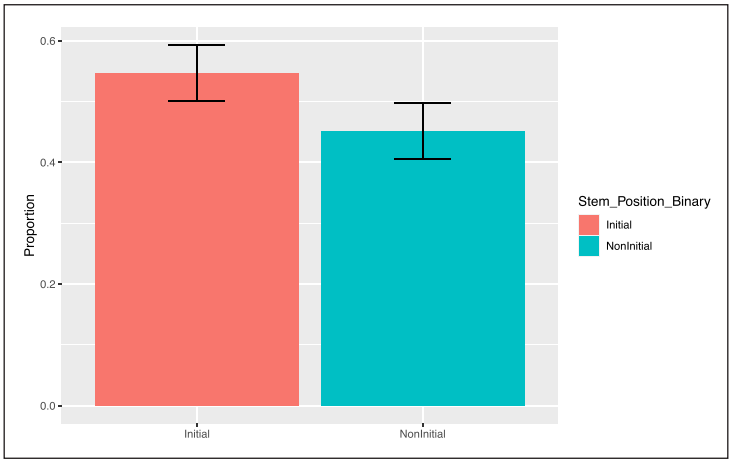


Figure 12: Gesture occurrence by stem position in Igbo.

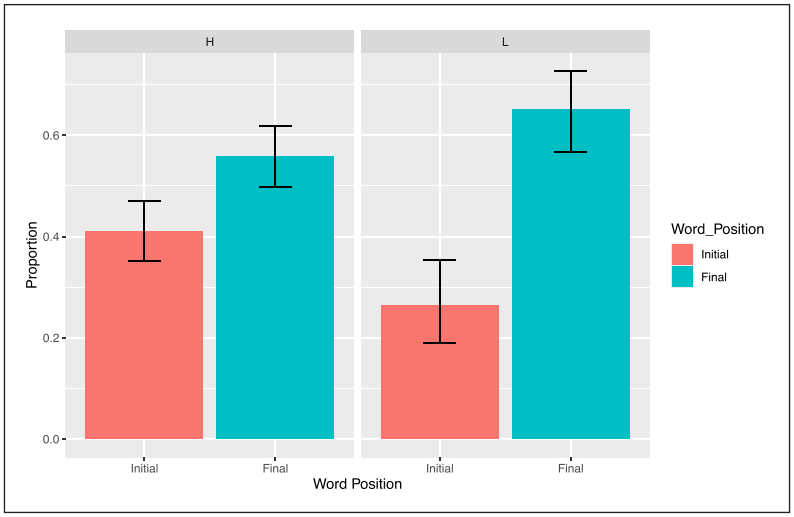


Figure 13: Gesture occurrence by word and aligned tone, disyllabic words, Igbo.

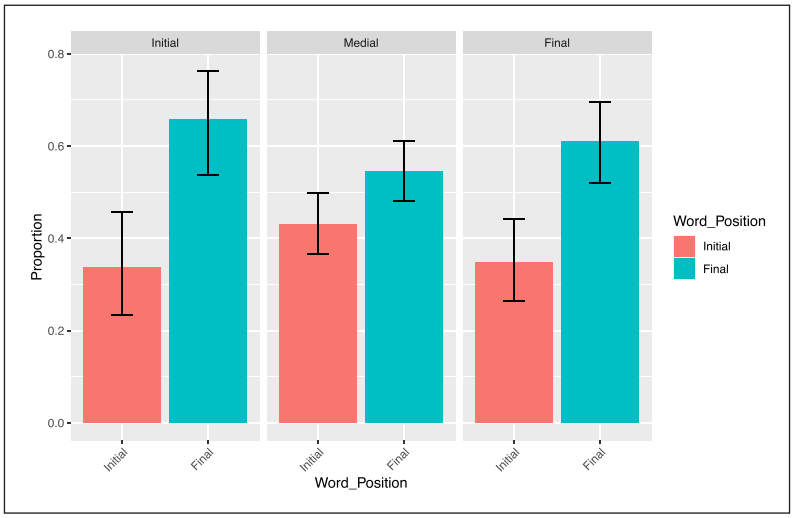


Figure 14: Gesture occurrence by word and phrase position, disyllabic words, Igbo.

Turning now for the results of the second model, we find, furthermore, that there is a significant interaction between Tone Melody and Word Position in conditioning gesture timing. In **Figure 15**, where the lightness of the blue shading indicates the number of tokens per cell, we see that words with HL and LH melodies overwhelmingly have gestures aligned in word-final position. Words with a HH melody, on the other hand, were significantly more likely to have a gesture aligned in *word-initial* position, on the other hand ($\beta = 2.24, z = 5.114, p < .001$; **Figure 15**). Words with LL and H!H melodies both had numerically more gestures aligned to word-final position, though patterns for these words did not differ significantly from those of the HH melody

(for LL: $\beta = 1.02$, $z = 1.459$, $p = .14$; for H!H: $\beta = .57$, $z = 1.215$, $p = .22$). We note that there were fewer words receiving a gesture with these melodies overall in the corpus.

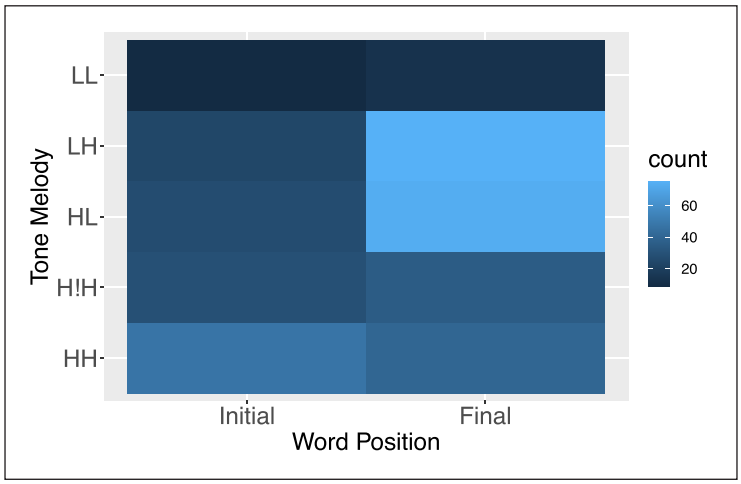


Figure 15: Frequency of gestures across word positions and tone melodies, disyllabic words. Lighter cells correspond to higher counts.

Taken together, these results suggest that tone, in and of itself, is not a strong determinant of gesture position, but that *tone melody* can play a role. The greater frequency of alignment of gestures to word-initial position for words bearing a HH melody, in particular, could suggest that words in which tones are multiply linked—as is the case for most HH sequences which lacks a downstep in Igbo, per Clark's analysis—behave differently from those in which each syllable is individually linked to a tone, as is the case in LH, HL, and H!H melodies, and possibly also some LL melodies, given the prevalence of rightward high tone (but not low tone) spreading in the language. Per Clark's analysis (and following from Goldsmith's 1976 Association Convention), tones are linked from left to right, meaning that spreading is targeting an otherwise toneless syllable at the right edge. We will explore further in the following sections the possibility that the underlying tonelessness of syllables in cases of multiply linked high tones may play a larger role in determining gesture timing.

3.2.2 Trisyllabic Words

For trisyllabic words (195 total words), we find once again a significant effect of Word Position, with word-final position most likely to be targeted for a gesture than initial position ($\beta = 1.95$, $z = 3.855$, $p < .01$; **Figure 16**). Word-medial position attracted a number of gestures intermediate between word-initial and word-final position, and did not differ significantly from word-final position in terms of its likelihood to attract a gesture. A significant effect of Stem

Position was also found, with stem-initial syllables more likely to attract a gesture than non-initial syllables ($\beta = 0.58$, $z = 2.298$, $p < .05$; **Figure 17**). There was no significant effect of Aligned Tone for trisyllabic words. Similar to disyllabic words, we found that, despite an overall pattern of word-final position being the strongest attractor of gestures, patterns varied by phrase position: specifically, phrase-medially, it was actually word-medial syllables, rather than word-final syllables, that were more likely to attract the gesture ($\beta = 1.84$, $z = 3.669$, $p < .001$; **Figure 18**).

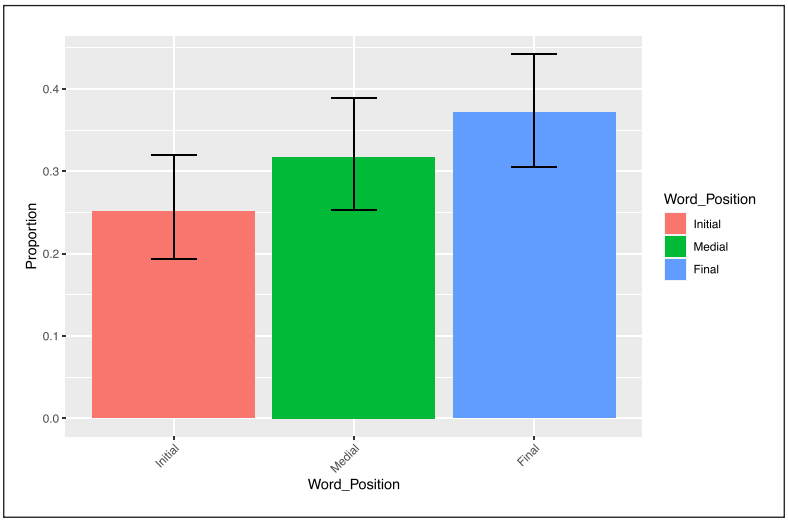


Figure 16: Gesture occurrence by word position, trisyllabic words, Igbo.

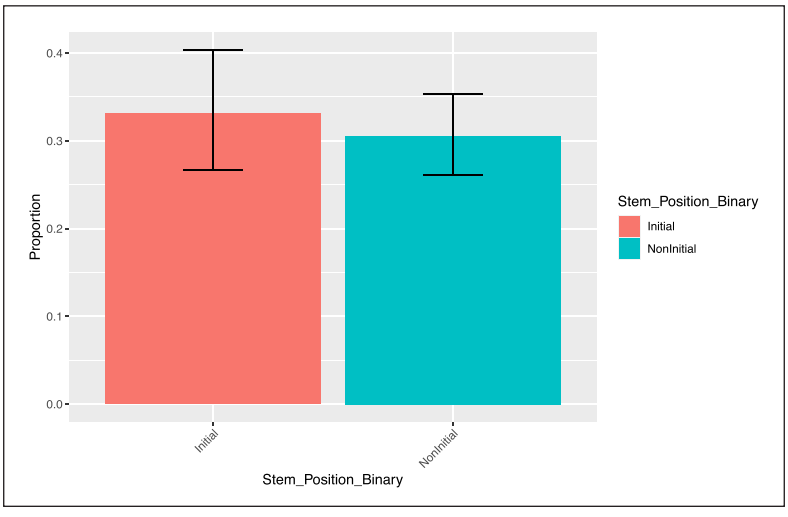


Figure 17: Gesture occurrence by stem position, trisyllabic words, Igbo.

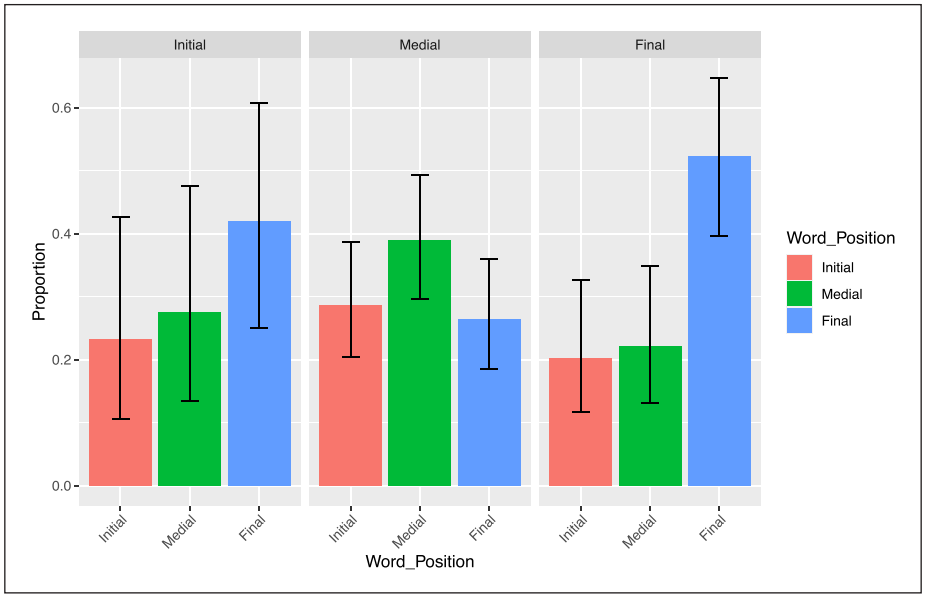


Figure 18: Gesture occurrence by word position and phrase position, trisyllabic words, Igbo.

Our second model once again revealed a significant interaction between tone melody and word position. Specifically, alignment of gestures differed significantly between words bearing a HHH melody and those bearing a H!HH melody or a H!HL melody: compared with word-final syllables, word-medial syllables in H!HL and H!HH words were significantly more likely to attract a gesture (HHH vs. H!HL: $\beta = 4.03, z = 3.342, p < .001$; HHH vs. H!HH: $\beta = 2.21, z = 2.105, p < .05$; Fig. 19). A post-hoc Fisher's Exact test showed that HHH and HHL did not differ significantly in their gesture alignment patterns overall within the word ($p = 0.53$). Similarly, H!HH and H!HL melodies did not differ significantly in their gesture alignment within the word ($p = 0.60$). For ease of interpretation, we have combined these sets of melodies together (indicated by HHX and H!HX) in **Figure 19**. Strikingly, we see that the HHX words have a high likelihood of gestures aligning to both the word-initial syllable and the word-final syllable, but less likelihood of alignment with word-medial syllables. For the H!HX words, the pattern is the opposite: the medial syllable is far more likely to attract a gesture than either the word-initial or word-final syllable. These patterns are in line with findings for the disyllabic words, in that sequences of high tones unbroken by a downstep—where there is presumably a multiply-linked high tone present—behave differently from sequences of unlike tones, or sequences of high tones broken up by a downstep. Furthermore, as with the disyllabic words, word-initial syllables which initiate a HHX sequence are more likely to attract a gesture than word-initial syllables which do not initiate a HHX sequence. We discuss implications of this finding in Section 5.

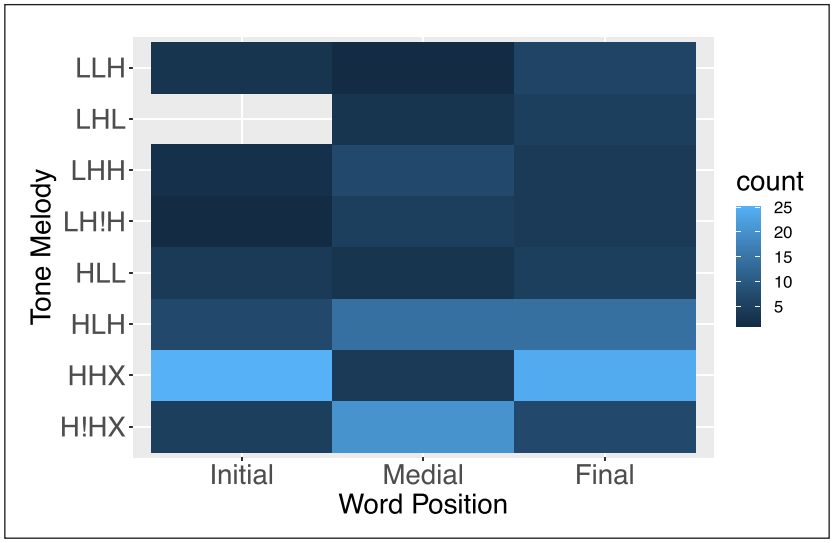


Figure 19: Frequency of gestures across word positions and tone melodies, trisyllabic words, Igbo. Lighter cells correspond to higher counts. Gray/transparent cells represent combinations with no instances.

3.2.3 Quadrisyllabic Words

Quadrisyllabic words were relatively fewer in our Igbo sample, with only 62 quadrisyllabic words containing only a single gesture. We note, too, that this sample was highly heterogenous in terms of morphological makeup (e.g. words with multiple prefixes are not differentiated from those with multiple suffixes or compounds with multiple stems) and tonal patterns. Sixty percent of the words in this sample were morphologically-complex, in comparison with 49% and 47% for trisyllabic and disyllabic words, respectively. Within this sample, patterns by word position trended in a similar direction as with disyllabic and trisyllabic words (with word-final position most likely to attract a gesture), but this effect did not reach significance after correcting for multiple comparisons ($\beta = 3.82$, $z = 1.98$, $p = .23$; **Figure 20**). No significant effect was observed for either stem position or aligned tone. There was also no significant interaction between word and phrase position for quadrisyllabic words.

The second model revealed a significant interaction between Word Position and Tone Melody. Specifically, words with a tone melody of HH!HL showed a significantly different pattern from those with a melody of H!HHH. As can be seen in **Figure 21** (which displays only those tone melodies accounting for more than 2% of the data), gestures were more likely to occur in penultimate position for words with a HH!HL melody than for those with a H!HHH melody, where there were no observations of penultimate gestures. This is in line with our observation for disyllabic and trisyllabic words that downstep is implicated in the groupings of high tones

for the purposes of gesture alignment. No other significant differences were observed for this interaction.

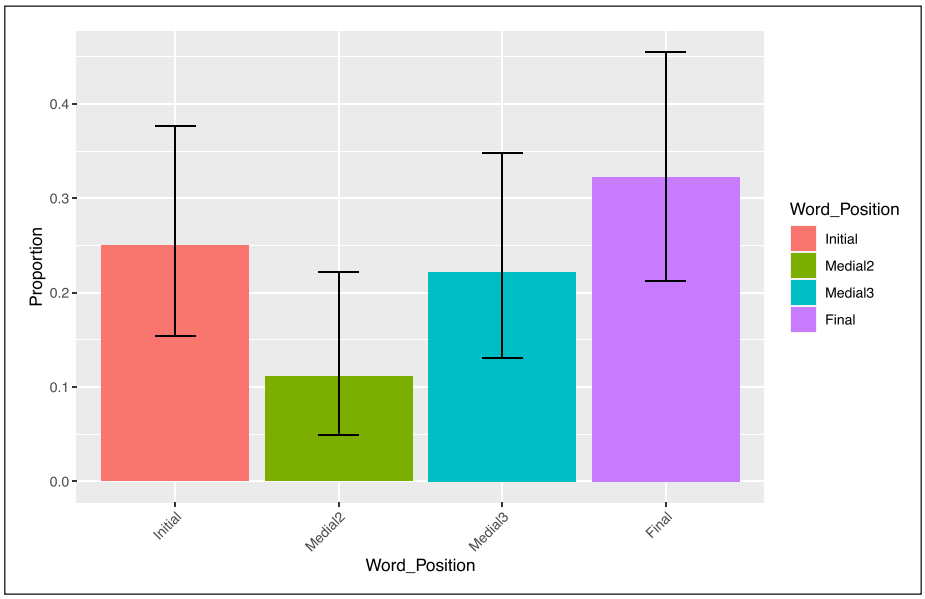


Figure 20: Gesture occurrence by word position, quadrisyllabic words, Igbo.

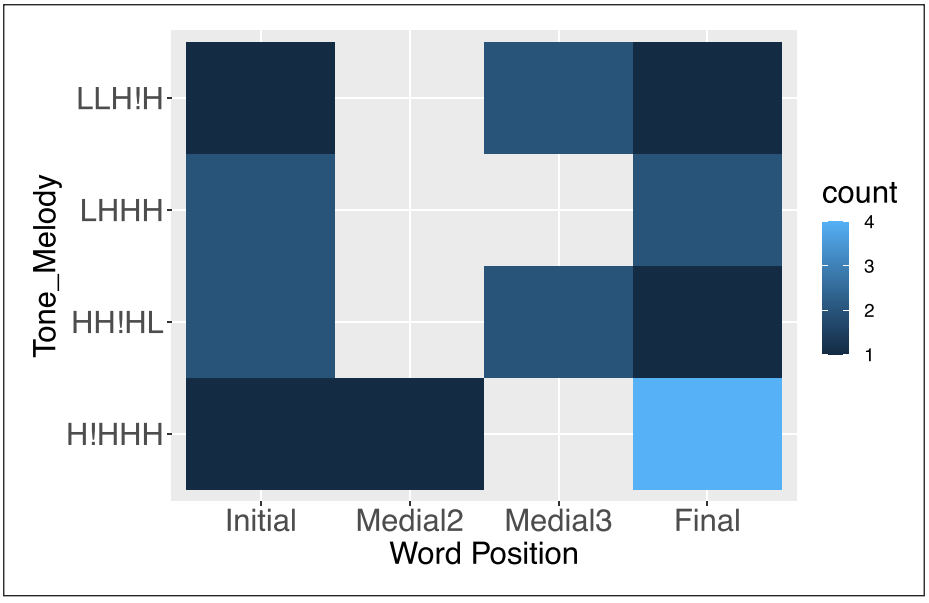


Figure 21: Frequency of gestures across word positions and tone melodies, trisyllabic words, Igbo. Lighter cells correspond to higher counts. Gray/transparent cells represent combinations with no instances.

4 Discussion

Our findings reveal a probabilistic relationship between gesture timing and speech in Medʉmba and Igbo which is mediated by both word-level prominence and prosodic phrasing. In this way, the speech-gesture relationship bears similarities to that found in many stress-based languages (Kendon 1980; McNeill 1992; Tuite 1993; Rochet-Capellan et al. 2008; Esteve-Gibert et al. 2017). However, the individual patterns of gesture timing across the two languages reveal striking differences in the behavior of word-level prominence. In Medʉmba, stem-initial position is the strongest attractor of gestures, consistent with previous work which has argued for metrical prominence of stem-initial syllables (Franich 2018; 2021). In Igbo, on the other hand, word-final position is the strongest attractor of gestures. This finding is in line with work by Clark (1990) arguing for the right-to-left assignment of metrical prominences at the word level in Igbo. Aside from the difference in position of gestures across the two languages, a key observation is that gesture provides a unified source of evidence for word-level prominence which is otherwise cued through quite different means across the two languages (contrasts among vowels and consonants in Medʉmba; patterns of downstep resistance in Igbo).

Though stem-initial position in Medʉmba and word-final position in Igbo both appear to be the sources of 'default' prominence in the two languages, there are clearly many cases where these sources were not sufficient to explain gesture timing. Another intriguing finding in our results is that prosodic phrasing influenced timing of gestures in both languages. Once more, however, the pattern was language-specific. Medʉmba, like Catalan, showed evidence that gestures were 'repelled' from phrase boundaries, being realized later, in particular, when a word followed a phrase boundary. The source of this repulsion, particularly where it occurs after an initial boundary, is unclear, though it could relate to the presence of a boundary tone at the left edge of the phrase which may contribute to the patterns of pitch reset observed in the language. In Igbo, on the other hand, gestures are more likely to drift away from the default word-final position phrase-medially, where they instead gravitate to word-medial (and perhaps stem-initial) positions. Here, the difference in timing appears rather to have to do with the relative strength and duration of segments at prosodic boundaries, where, for example, the process of vowel coalescence (modeled as gestural overlap) described in Section 1.2 is less likely to occur, or is likely to be weaker. Phrase-medially, where word-final vowels may assimilate almost completely with a following word-initial vowel, the line between the ending of one word and the beginning of the next becomes blurred. From a rhythmic perspective, we note that patterns of textsetting in Igbo (Franich & Nwosu, in prep) suggest that coalescence often results in the complete loss of a timing slot between the two vowels, even though phonetic evidence would suggest that remnants of both vowels remain present in production. It may therefore be the case that the acoustic features attributed to word-final vowels in our data belie a similar loss of a timing unit, in which case apparent vowel sequences are better modeled as a single word-initial vowel in

surface representations. Under these circumstances, some syllables treated as non-final in our data might be better treated as word-final from a prosodic standpoint.

Even if speakers are not completely deleting word-final timing slots, a prosodically weakened word-final syllable may still lead speakers to align gestures to other positions within the word. Though it seems somewhat counterintuitive to suggest that metrically-prominent word-final syllables should be subject to this sort of prosodic weakening, there are, in fact, several languages in which stressed vowel deletion or weakening occurs and leads to stress shift (Lightner 1972; Al-Mozainy et al. 1985; Halle & Vergnaud 1987). We note that word-final vowels also tend to be less *informative* relative to word-initial vowels by several metrics: they occur relatively later within the word, and also tend to harmonize in quality with vowels occurring earlier in the word. Greater predictability/less informativity is known to be a contributing factor in patterns of segment reduction (Cohen-Priva 2015; Turnbull 2023). Future work examining the phonetic, phonological, and informational patterning of vowel sequences at phrase-internal word boundaries will help to clarify the mechanisms behind this interaction.

The overall weak effect of lexical tone on predicting gesture timing in both languages is another finding worth discussing. Whereas there have been suggestions in the literature that relatively higher and more dynamic pitch profiles bear inherently greater perceptual prominence (Baumann & Röhr 2015; Cole et al. 2019), our results suggest that tonal prominence is rather more language-specific in nature. Indeed, Igbo speakers showed relatively more gestures on low-toned syllables, suggesting that, if anything, it is the low tone that associates with greater prominence in the language. This is perhaps not surprising given Clark's (1990) proposal that low tone is the only lexically marked tone in Igbo, whereas high tone enters later in the derivation of words and phrases. These findings also underscore the fact that pitch 'prominence' is mediated by many factors, most notably the functional role that pitch plays in a language (e.g. in signaling lexical tonal distinctions vs. phrase-level intonational distinctions) and how it is implicated in the marking of elements of information structure, which also demonstrably influence perceived prominence (Cole et al. 2019).

The weak influence of tone on gesture timing is also interesting in light of our results on the timing of gesture apexes within syllables. In our results for both languages, a very high proportion of words (nearly half) displayed alignment of gesture apexes to consonants, rather than vowels—in these cases, alignment to a pitch peak or minimum seems unlikely. Instead, we suggest that gesture apexes in these languages might more fruitfully be seen to align with the articulatory onset of the vowel, or the perceptual center of the syllable, which can often occur within the acoustic onset of the syllable (Morton et al. 1976; Scott 1991). It might also be the case that an articulatory study would reveal a relationship between the onset of vowel articulation (which often occurs simultaneously with constriction for the preceding consonant; Browman & Goldstein 1988), but future work will need to explore this possibility.

Though tone did not prove to be a major predictor of gesture alignment, tone *melody* turned out to be an important predictor of gesture timing in Igbo. Most notably, gestures were more likely to occur in positions other than word-finally where a word began with a HH sequence unbroken by a downstep, regardless of the length of the word. The same pattern was not observed for H!H sequences (where two highs are broken up by a downstep); in these cases, a high proportion of gestures on words displayed alignment of a gesture to the high tone *following* the downstep (e.g. to word-medial position for H!HH and H!HL words). What kind of a constraint could be responsible for this pattern of findings, particularly where gestures are shown to gravitate away from the apparent default word-final position? Given that the distinction between a HH sequence and a H!H sequence (or a HHH vs. a H!HH sequence) has to do with the presence of additional underlying high tones where a downstep is present, one interpretation—assuming, per Clark's analysis and the Association Convention from Autosegmental Phonology (Goldsmith 1976)—is that gestures tend to avoid *toneless* syllables in Igbo. In (15), we see the high tone linked first to the initial syllable of the word *ócé* 'chair,' and then spread rightward. Under this account, it is the tonelessness of the final syllable in (15) that makes it a poor candidate for receiving a gesture.

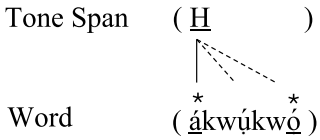
(15)

However, this account does not explain why it is that both word-initial *and* final syllables in HHH sequences are preferentially aligned to gestures over word-medial syllables. This would suggest that tonelessness in itself is not the determinant of whether a syllable can host a gesture (since it is assumed that both the word-final and word-medial syllables in a trisyllabic HHH word are toneless). Instead, it seems that there is another level of prosodic organization taking place within a HHH word, such that both the initial and final syllables are treated with relative prominence. This could simply stem from the default word-final prominence proposed by Clark. However, note that the initial syllables of words with H!HH (and H!HL) melodies do not attract a gesture as frequently—instead, it is the medial syllable that is more likely to attract a gesture. Here, it seems that the downstep initiates a new sequence in which the initial high toned syllable is treated as prominent.

Our results seem to suggest a general preference for alternating syllable prominence within the word, whether prominence stems from the default metrical structure proposed by Clark, or from tonal factors such as being initial in a HH(H) sequence. In the latter case, prominence is assigned to the first high tone in the sequence (which we might refer to as the 'head' of the spreading domain or tone span, following e.g. Cole & Kisseberth 1994 and Key & Bickmore 2014), and then occurs on alternating syllables moving rightward. The idea that tone can exert

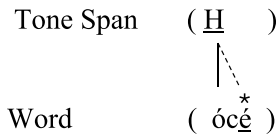
or conform to metrical influence is not new: in many languages, patterns of e.g. binary tone spreading (among other types of patterns) have been explained with reference to the 'tonal foot' (Sietsema 1989; Leben 1997; Zen 1999; Bickmore 2003). Co-speech gesture patterns in Igbo could similarly be driven by such patterns. If this is the case, there appear to be two levels of metrical organization at play in Igbo: one which assigns prominence from the right edge of the *word* and the other which assigns prominence from the left edge of a *tone span* (the first syllable hosting a multiply linked high tone, or the first syllable hosting a high tone after a downstep). In some cases, as in HHH and HHL words, these two prominence assignment patterns are not in conflict, since they both predict that the initial and final syllables of the word will be most likely to be targeted for gestures (16).

(16)

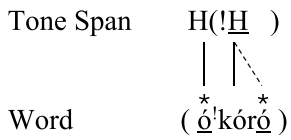


In contrast with the situation in (15), disyllabic words with HH melodies and trisyllabic words with H!HH or H!HL melodies present a conflict between the predicted prominence patterns at the tonal level and at the word level. Here, prominence at the tonal level is expected at the beginning of the high tone span (17), or at the beginning of the downstepped tone (18), while the word-level prominence is expected on the final syllables and alternating syllables to the left.

(17)



(18)



In situations like (17) and (18), it appears that the tone span wins out when prominence is being assigned, since we're more likely to see initial syllables attracting gestures in HH words, and medial syllables attracting gestures in H!HH and H!HL words.[3]

[3] A reviewer points out that, at a high level, it seems as though gesture is highlighting *transitions* between underlying tones, which could be another way in which dynamic changes in pitch are implicated in the speech-gesture relationship. We note that this cannot be the sole explanation for our results, as gestures are still highly likely to occur on final syllables in words with e.g. LL and HHH melodies, where word-initial position would be expected to be preferred based on tone dynamics alone. We do agree that this is an important observation, though, and one that aligns with the possibility (also suggested elsewhere in this paper) that a syllable's relative *informativity* may play a role in determining gesture timing (see e.g. Cohen-Priva 2015; Aylett & Turk 2004).

Our results have implications not just for the typology of speech-gesture relations cross-linguistically, but also for theories of motor control and the functional link between speech and co-speech gesture. It has been proposed, for example, that manual gestures (and perhaps head gestures, as well) are dynamically linked with vocalizations such that movement in one domain can bring about coupling with movements in the other domain (Pouw & Fuchs 2022). Specifically, the authors propose that acceleration of the limbs can lead to impulses which cascade to the respiratory-vocal system, creating an intrinsic physical link between limb movements and the laryngeal movements responsible for pitch changes, possibly explaining the tight relationship between gesture apexes and pitch peaks in many languages. What our results suggest is that the relationship between gesture and pitch in language is not universal, and that the relationship is moderated by the functional role that pitch plays in a language. Thus, should the biomechanical link between gesture and its effects on the respiratory system be demonstrated to be universal, theories of speech motor control must be able to account for differences across languages in terms of whether this link is exploited in speech communication and grammar, or, as in the present case, suppressed in linguistic contexts. We refer the reader to Pouw & Fuchs 2022 for additional discussion on this point. An interesting question concerns whether the association between gesture apex and pitch movements is more likely to arise in languages in which pitch plays a key role in cuing information structure, as opposed to lexical tonal distinctions. Indeed, recent data on the speech-gesture relationship in Mandarin, another lexical tone language, lend support to this view (Rohrer et al. 2024).

Related to this last point, there is a clear need to explore in more depth the role of information structure in conditioning gesture occurrence in both Medʉmba and Igbo. We hypothesize that information structure likely plays an important role in constraining gesture occurrence in tonal languages, just as in non-tonal languages: since there are clearly many more words in our corpus than there are words containing a gesture, word-level prominence is only one piece of the larger puzzle for explaining gesture patterning. Preliminary evidence in our data suggests that, as in non-tonal languages, gestures are more likely to occur on words which constitute new information in the discourse (Im & Baumann 2020). Future work will need to explore whether other types of focus are also associated with the occurrence with gestures, and which types. We have also not explored in depth how different morphoprosodic effects might arise in determining gesture timing. For example, Franich (2021) finds that pronominal enclitics bear greater rhythmic prominence than stem-final syllables in Medʉmba (similar to the effects of secondary stress in English). The present work has treated metrical prominence in a binary fashion, without explicitly investigating how varying levels of prominence might affect gesture timing. Thus, there may be aspects of prosodic structure influencing gesture timing which we have not yet uncovered, and which will be key for future work to explore.

Finally, we want to highlight important next steps in the cross-linguistic study of speech and co-speech gesture. In the present work, we have relied on a notion of gesture 'apex' which is the visualized moment of peak velocity within the gesture stroke. As mentioned in Sections 1 and 2, this is just one of many kinematic landmarks that have been explored with respect to crucial coupling relations between gesture and speech. Preliminary comparisons to other kinematic landmarks such as timing of minimum velocity or peak acceleration do not reveal major differences in the patterning of the phenomena described here for our data. Nonetheless, robust comparisons of gesture timing across different languages necessitates a better understanding of the specific kinematic landmarks relevant for speech-gesture coupling.

Future work should therefore focus on comparing a broader array of kinematic landmarks and their relation to speech events across languages with different prosodic profiles in an effort to further specify these gesture-speech relationships.

5 Conclusion

To conclude, we have shown that the timing of co-speech gestures in both Medʉmba and Igbo is constrained by word-level prominence, as is the case for many non-tonal languages. In this way, gesture attraction serves as a unifying behavior of prosodically prominent syllables across languages which have otherwise very dissimilar acoustic correlates of prominence. Our results also reveal interactions between word-level and phrase-level prosody in determining where gestures will occur. Most strikingly, word-final vowels in Igbo are less likely to attract a gesture when these words occur phrase-medially, suggesting that word-final vowels—while generally the most prominent in the word—lose prominence in contexts where they are subject to extreme levels coalescence with a following vowel. Though no individual tone value was found to attract gestures more than any other, effects of tone *melody* on gesture alignment in Igbo suggest competing pressures between default word-final metrical prominence and prominence at the level of the tonal foot. The patterns we describe also hint at possible informational effects on gesture placement, as gestures appears to avoid syllables with redundant tonal information. Further work will be necessary to assess the degree to which informational effects operate on gesture timing, and how these effects may interact with prosody in shaping gesture timing.

Abbreviations

!:	Downstep
3:	Third Person
AM:	Associative Marker
COMP:	Complementizer
FUT:	Future Marker
H:	High Tone
IND:	Indicative Mood Marker
INF:	Infinitive Marker
ITER:	Iterative Marker
L:	Low Tone
POSS:	Possessive Marker
SG:	singular
REL:	Relative Clause Marker

Data availability

Data and code for analyses presented in this paper can be accessed at the following link: https://osf.io/qmrxc/?view_only=89bbb973ede04ebfa223f9ce2245c46e.

Ethics and consent

This research was granted ethics approval by the Institutional Review Board of Harvard University (Protocol IRB 22-1095).

Funding information

This work was supported by National Science Foundation Linguistics Program Grant No. BCS-2018003 (PI: Kathryn Franich). The National Science Foundation does not necessarily endorse the ideas and claims in this research.

Acknowledgements

We would like to thank the speakers of Medumba and Igbo for their time in participating in this study. We would also like to thank Dr. Ange Bergson Lendja and Fridah Gam for their help with interviewing and coordinating data collection. We thank Eliana Spradling for help with data coding and management of our coding team, and the many research assistants who contributed to the coding of the corpus for this study, including Crystal Akalu, Kylie Boggs, Clarissa Briasco-Stewart, Walter Dych, Lindsay Hawtof, Luc De Nardi, Victoria Ochlan, Anna

Schumeyer, Vitor Lacerda Siqueira, Nicole Taylor, and Juliette Winnard. Thank you to Karee Garvin for assistance with coding coordination and scripts for data wrangling. Finally, we thank members of the Harvard University PhonLab and audiences at the 2023 Annual Conference on African Linguistics, the 2024 Annual Conference on Phonology, the 2024 Symposium Series on Multimodal Communication, Princeton University, UMass Amherst, and Rutgers University for helpful feedback and discussion. All mistakes are our own.

Competing Interests

The authors have no competing interests to declare.

Authors' Contributions

KF contributed to conceptualization of the study, data collection, data preparation, data analysis, and writing. HK contributed to conceptualization of the study and data preparation. VN contributed to conceptualization of the study and data preparation.

References

Akinlabi, Akin & Urua, Eno E. 2003. Foot structure in the Ibibio verb. *Journal of African Languages and Linguistics* 24(2). 119–160. DOI: https://doi.org/10.1515/jall.2003.006

Al-Mozainy, Hamza & Bley-Vroman, Robert & McCarthy, John. 1985. Stress shift and metrical structure. *Linguistic Inquiry* 16. 135–144.

Aylett, Matthew & Turk, Alice. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56. DOI: https://doi.org/10.1177/00238309040470010201

Barr, Dale J. & Levy, Roger & Scheepers, Christoph & Tily, Harry J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Baumann, Stefan & Röhr, Christine T. 2015. The perceptual prominence of pitch accent types in German. In Wolters, M. & Livingstone, J. & Beattie, B. (eds.), *Proceedings of the 18th International Congress of Phonetics Sience (ICPhS)*, Paper number 2981, pp. 1–5. Glasgow, UK: University of Glasgow. ISBN 978-0-85261-941-4.

Bickmore, Lee. 2003. The use of feet to account for binary tone spreading. In *Frankfurter Afrikanistische Blatter* vol. 15, Anyanwu, R. J. (ed.). Rudiger Koeppe Verlag, Koln.

Brentari, Diane & Coppola, Marie. 2013. What sign language creation teaches us about language: Sign Language Creation. *Wiley Interdisciplinary Reviews: Cognitive Science* 4(2). 201–211. DOI: https://doi.org/10.1002/wcs.1212

Browman, Catherine & Goldstein, Louis. 1986. Towards an articulatory phonology. *Phonology* 3. 219–252. DOI: https://doi.org/10.1017/S0952675700000658

Browman, Catherine & Goldstein, Louis. 1988. Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45. 140–155. DOI: https://doi.org/10.1159/000261823

Browman, Catherine & Goldstein, Louis. 1992. Articulatory Phonology: An overview. *Phonetica* 49. 55–180. DOI: https://doi.org/10.1159/000261913

Byrd, Dani & Kaun, Abigail & Narayanan, Shri & Saltzman, Elliot. 2000. Phrasal signatures in articulation. In *Acquisition and the Lexicon: Papers in Laboratory Phonology*, Broe, M. & Pierrehumbert, J. (eds.), pp. 70–88. Cambridge, U.K.: Cambridge University Press.

Cho, Taehong & Keating, Patricia. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29. 155–190. DOI: https://doi.org/10.1006/jpho.2001.0131

Clark, Mary. 1990. *The Tonal System of Igbo*. Walter de Gruyter. DOI: https://doi.org/10.1515/9783110869095

Cohen Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Laboratory Phonology* 6(2). 243–78. DOI: https://doi.org/10.1515/lp-2015-0008

Cole, Jennifer & Hualde, Jose Ignacio & Smith, Caroline L. & Eager, Christopher & Mahrt, Timothy & Naoleão de Souza, Ricardo. 2019. Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics* 75. 113–147. DOI: https://doi.org/10.1016/j.wocn.2019.05.002

Cole, Jennifer & Kisseberth, Charles W. 1994. An optimal domains theory of harmony. *Studies in the Linguistic Sciences* 24. 1–13.

de Ruiter, Jan-Peter & Wilkins, David P. 1998. The synchronization of Gesture and Speech in Dutch and Arrernte (an Australian Aboriginal language): A Cross-cultural comparison. In Santi, S. (ed.), *Oralité et Gestualité* (pp. 603–607). Paris: L'Harmattan.

Downing, Laura J. & Rialland, Annie. 2017. *Intonation in African Tone Languages*. Berlin, Boston: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110503524

Duan, Xu & Zhang, Jie & Liang, Yuan & Huang, Yingying & Yan, Hao. 2023. The effect of speech–gesture asynchrony on the neural coupling of interlocutors in interpreter-mediated communication, *Social Cognitive and Affective Neuroscience* 18(1). nsad027. DOI: https://doi.org/10.1093/scan/nsad027

Dych, Walter & Garvin, Karee & Franich, Kathryn. 2023. Comparing manual vs. semi-automated methods for the coding of co-speech gestures. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, 4170–4174.

Eberhard, David M. & Simons, Gary F. & Fennig, Charles D. (eds.) 2024. Ethnologue: Languages of the World. Twenty-seventh edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

ELAN (Version 6.9) [Computer software]. 2024. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Esteve-Gibert, Nuria & Borràs-Comes, Joan & Asor, Eli & Swerts, Marc & Prieto, Pilar. 2017. The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America* 141(6). 4727–4739. DOI: https://doi.org/10.1121/1.4986649

Esteve-Gibert, Núria & Prieto, Pilar. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research* 56(3). 850–64. DOI: https://doi.org/10.1044/1092-4388(2012/12-0049)

Féry, Caroline. 1993. *German intonational patterns*. Tübingen: Niemeyer. DOI: https://doi.org/10.1515/9783111677606

Franich, Kathryn. 2018. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* 67. 21–33. DOI: https://doi.org/10.1016/j.wocn.2017.11.001

Franich, Kathryn. 2021. Metrical prominence asymmetries in Medʉmba, a Grassfields Bantu language. *Language* 97(2). 365–402. Project MUSE. DOI: https://doi.org/10.1353/lan.2021.0021

Franich, Kathryn & Nwosu, Vincent. in prep. Metrical organization in Igbo: Evidence from children's songs.

Goldsmith, John. 1976. *Autosegmental Phonology*. MIT dissertation.

Güldemann, Tom. 2007. The Macro-Sudan belt: towards identifying a linguistic area in northern sub-Saharan Africa. In Heine, B. & Nurse, D. (eds.), *A Linguistic Geography of Africa*. Cambridge Approaches to Language Contact. Cambridge University Press, 151–185. DOI: https://doi.org/10.1017/CBO9780511486272.006

Halle, Morris & Vergnaud, Jean-Roger. 1987. An essay on stress. (Current Studies in Linguistics 15). (Cambridge, Mass.: MIT Press. Pp. xi + 300. *Phonology* 7. 171–188. DOI: https://doi.org/10.1017/S0952675700001160

Harris, John & Urua, Eno-Abasi. 2001. Lenition degrades information: consonant allophony in Ibibio. *Speech, Hearing and Language: Work in progress* 13. 72–105.

Hyman, Larry M. & Rolle, Nicholas & Sande, Hannah & Clem, Emily & Jenks, Peter & Lionnet, Florian & Merrill, John & Baier, Nico. 2019. Niger-Congo linguistic features and typology. In E. Wolff, E. (ed.), *The Cambridge Handbook of African Linguistics and A History of African Linguistics*. Cambridge University Press. DOI: https://doi.org/10.1017/9781108283991.009

Idiatov, Dmitry & Van de Velde, Mark. 2016. Stem-initial accent and c-emphasis prosody in North-Western Bantu. *Presentation at the 6th International Conference on Bantu Languages (BANTU6)*. University of Helskinki, June.

Ihiunu, Peter and Kenstowicz, Michael. 1994. Two notes on Igbo vowels. M.S., MIT.

Im, Suyeon & Baumann, Stefan. 2020. Probabilistic relation between co-speech gestures, pitch accents and information status. In *Proceedings of the Linguistic Society of America* 5(1). 685–697. DOI: https://doi.org/10.3765/plsa.v5i1.4755

Jun, Sun-Ah & Fougeron, Cecile. 2000. A phonological model of French Intonation. In Botinis, A. (ed.), *Intonation. Text, Speech and Language Technology* 15. Springer, Dordrecht. DOI: https://doi.org/10.1007/978-94-011-4317-2_10

Kendon, Adam. 1980. Gesticulation and Speech: Two Aspects of the Process of Utterance. In Key, M. R. (ed.), *The Relationship of Verbal and Nonverbal Communication*. Berlin, Boston: De Gruyter Mouton, 207–228. DOI: https://doi.org/10.1515/9783110813098.207

Keupdjio, Hermann. 2021. *The syntax of A'-dependences in Bamileke Medumba*. University of British Columbia dissertation.

Key, Michael & Bickmore, Lee. 2014. Headed tone spans: Binarity and minimal overlap. *Southern African Linguistics and Applied Language Studies* 32(1). 35–53. DOI: https://doi.org/10.2989/16073614.2014.925218

Kouankem, Constantine & Zimmermann, Malte. 2013. Contrastive FOCUS and verb doubling in Mədúmbɑ. Presentation at *Universiteit van Amsterdam*. ATW-Lezing, 20th September.

Krivokapić, Jelena & Tiede, Mark K. & Tyrone, Martha E. 2017. A Kinematic Study of Prosodic Structure in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. *Laboratory Phonology* 8(1). 3. DOI: https://doi.org/10.5334/labphon.75

Kügler, Frank. 2017. Tone and intonation in Akan. In *Intonation in African Tone Languages*, L. Downing, J. & Rialland, A. (eds.). Berlin, Boston: De Gruyter Mouton, 89–130. DOI: https://doi.org/10.1515/9783110503524-004

Kula, Nancy & Hamman, Silke. 2017. Intonation in Bemba. In *Intonation in African Tone Languages*, Downing, L. J. & Rialland, A. (eds.). Berlin, Boston: De Gruyter Mouton, 321–364. DOI: https://doi.org/10.1515/9783110503524-004

Ladd, D. Robert. 1996. *Intonational phonology*. (Cambridge Studies in Linguistics 79.) Cambridge: Cambridge University Press. Pp. xv + 334.

Leben, William R. 1997. Tonal feet and the adaptation of English borrowings into Hausa. *Studies in African Linguistics* 25. 139–154. DOI: https://doi.org/10.32473/sal.v25i2.107400

Leonard, Thomas & Cummins, Fred. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes* 26(10). 1457–1471. DOI: https://doi.org/10.1080/01690965.2010.500218

Lieberman, Philip. 1967. *Intonation, Perception, and Language*. Cambridge: MIT Press.

Lightner, Theodore M. 1972. Problems in the theory of phonology: Russian phonology and Turkish phonology. *Linguistic Research*.

Lionnet, Florian. 2017. Stem-initial prominence in West and Central Africa: Niger-Congo, areal, or both? Presentation at *48th Annual Conference on African Linguistics (ACAL 48)*. Indiana University, Bloomington 30 March–2 April.

Loehr, Daniel P. 2004. Gesture and intonation. Washington, DC: *Georgetown University dissertation*.

Loehr, Daniel P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3(1). DOI: https://doi.org/10.1515/lp-2012-0006

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.

MIT Speech Communications Group. *MIT Speech Communications Group Gesture Studies Coding Manual*. Retrieved June 2020 from http://web.mit.edu/pelire/www/manual/

Morett, Laura M. 2023. Observing gesture at learning enhances subsequent phonological and semantic processing of L2 words: An N400 study. *Brain and Language* 246. 1–12. DOI: https://doi.org/10.1016/j.bandl.2023.105327

Morton, John & Marcus, Steve & Frankish, Clive. 1976. Perceptual centers (p-centers). *Psychological Review* 83. 405–408. DOI: https://doi.org/10.1037//0033-295X.83.5.405

Osuagwu, Eunice & Anyanwu, Ogbonna. 2020. Aspects of syntactic focus constructions in Igbo. *California Linguistic Notes* 42. 1–22.

Özge, Umut & Bozsahin, Cem. 2010. Intonation in the grammar of Turkish. *Lingua* 120(1). 132–75. DOI: https://doi.org/10.1016/j.lingua.2009.05.001

Pierrehumbert, Janet. 1980. The phonology and phonetics of English intonation. PhD Dissertation, MIT.

Pouw, Wim & Dixon, James A. 2019. Quantifying gesture- speech synchrony. In Grimminger, A. (ed.), *Proceedings of the 6th Gesture and Speech in Interaction – GESPIN 6* (pp. 75–80). Paderborn: Universitaetsbibliothek Paderborn. DOI: https://doi.org/10.17619/UNIPB/1-815

Pouw, Wim & Fuchs, Susanne. 2022. Origins of vocal-entangled gesture. *Neuroscience and Biobehavioral Reviews* 141. Article 104836. DOI: https://doi.org/10.1016/j.neubiorev.2022.104836

Pouw, Wim & Harrison, Steven J. & Esteve-Gibert, Nuria & Dixon, James A. 2020. Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America* 148(3). 1231–1247. DOI: https://doi.org/10.1121/10.0001730

Prieto, Pilar. 2014. The intonational phonology of Catalan. In *Prosodic Typology II*, Jun, Sun-Ah (ed.), 1st ed., 43–80. Oxford University Press, Oxford. DOI: https://doi.org/10.1093/acprof:oso/9780199567300.003.0003

Rochet-Capellan, Amélie & Laboissière, Rafael & Galván, Arturo & Schwartz, Jean-Luc. 2008. The speech focus position effect on jaw–finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research* 51(6). 1507–1521. DOI: https://doi.org/10.1044/1092-4388(2008/07-0173)

Rohrer, Patrick L. & Prieto, Pilar & Delais-Roussarie, Elisabeth. 2019. Beat gestures and prosodic domain marking in French. In Calhoun, S. & Escudero, P. & Tabain, M. & Warren, P. (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1500–1504). Australasian Speech Science and Technology Association Inc.

Rohrer, Patrick Louis & Hong, Yitian & Bosker, Hans Rutger. 2024. Gestures time to vowel onset and change the acoustics of the word in Mandarin. In *Proceedings of Speech Prosody*. Leiden, pp. 866–870. DOI: https://doi.org/10.21437/SpeechProsody.2024-175

Rosenfelder, Ingrid & Fruehwald, Josef & Evanini, Keelan & Seyfarth, Scott & Gorman, Kyle & Prichard, Hilary & Yuan, Jiahong. 2022. *Fave: Speaker Fix*. Zenodo. DOI: https://doi.org/10.5281/ZENODO.22281

Scott, Sophie. 1993. P-centres in speech: An acoustic analysis. University College London. Doctoral dissertation.

Shattuck-Hufnagel, Stefanie & Ren, Ada. 2018. The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech. *Frontiers in Psychology* 9. 1514. DOI: https://doi.org/10.3389/fpsyg.2018.01514

Shattuck-Hufnagel, Stefanie I. & Prieto, Pilar. 2019. Dimensionalizing co-speech gestures. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences.* Melbourne, Australia 2019 (pp. 1490–1494).

Sietsema, Brian. 1989. *Metrical dependencies in tone assignment*. Ph.D. Dissertation, MIT.

Swerts, Marc & Krahmer, Emiel. 2008. Facial expression and prosodic prominence: effects of modality and facial area. *Journal of Phonetics* 36(2). 219–38. DOI: https://doi.org/10.1016/j.wocn.2007.05.001

Trujillo, James P. & Vaitonyte, Julija & Simanova, Irina & and Özyürek, Asli. 2018. Toward the markerless and automatic analysis of kinematic features: A toolkit for gesture and movement research. *Behavior Research Methods*. DOI: https://doi.org/10.3758/s13428-018-1086-8

Tuite, Kevin. 1993. The production of gesture. *Semiotica* 93(1/2). 83–106. DOI: https://doi.org/10.1515/semi.1993.93.1-2.83

Türk, Olcay & Calhoun, Sasha. 2023. Multimodal cues to intonational categories: Gesture apex coordination with tonal events. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 14(1). 1–50. DOI: https://doi.org/10.16995/labphon.6432

Turnbull, Rory. 2023. The effect of usage predictability on phonetic and phonological variation. In Díaz-Campos, M. & Balasch, S. (eds.), *The Handbook of Usage-Based Linguistics.* DOI: https://doi.org/10.1002/9781119839859.ch8

Uwaezuoke, Aghaegbuna Haroldson. 2021. Tone assimilation in Igbo: A phonological description. *Mgbakoigba, Journal of African Studies* 8(2). 47–53.

Voorhoeve, Jan. 1971. Tonology of the Bamileke Noun. *Journal of African Languages* 10. 44–53.

Zec, Draga. 1999. Footed tones and tonal feet: rhythmic constituency in a pitch-accent language. *Phonology* 16. 225–264. DOI: https://doi.org/10.1017/S0952675799003759

Zimmermann, Malte & Kouankem, Constantine. 2024. Focus fronting in a language with in situ marking: The case of Mə̀dʉ́mbà." *Languages* 9(4). 117. DOI: https://doi.org/10.3390/languages9040117

Zsiga, Elizabeth. 1992. A mismatch between morphological and prosodic domains: Evidence from two Igbo rules. *Phonology* 9. 101–35. DOI: https://doi.org/10.1017/S0952675700001512

Zsiga, Elizabeth. 1993. Features, gestures, and the temporal aspects of phonological organization. New Haven, CT: Yale University Dissertation.

Zsiga, Elizabeth. 1997. Features, gestures, and Igbo vowels: An approach to the phonetics-phonology interface. *Language* 73(2). 227–274. DOI: https://doi.org/10.2307/416019