Language and Cognition (2025), 17, e92, 1–22
doi:10.1017/langcog.2025.10043

ARTICLE

Gesture, prosodic prominence and stresslessness in Indonesian

Alessa Farinella[1], Constantijn Kaland[2] and Daniel Kaufman[3]

[1]Department of Linguistics, University of Massachusetts Amherst, Amherst, Massachusetts, USA; [2]Linguistik I - Phonetik und Phonologie, Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany and [3]CUNY Queens College and the Endangered Language Alliance, Queens, New York, USA
Corresponding author: Alessa Farinella; Email: afarinella@umass.edu

(Received 17 November 2022; Revised 08 November 2023; Accepted 29 October 2025)

Abstract
Previous studies on a variety of languages have demonstrated that manual gesture is temporally aligned with prosodic prominence. However, the majority of these studies have been conducted on languages with word-level stress. In this paper, we investigate the alignment of manual beat gestures to speech in local varieties of Standard Indonesian, a language whose word prosodic system has been the subject of conflicting claims. We focus on the varieties of Indonesian spoken in the eastern part of the archipelago and Java. Our findings reveal that there is a strong tendency to align gesture to penultimate syllables in the eastern variety and a tendency to align gesture to final syllables in the Javanese variety. Additionally, while the eastern patterns appear to be word based, the Javanese pattern shows evidence of being phrase based. Surprisingly, the penultimate syllable emerges as a gestural anchor in the eastern variety even for two of the three speakers who showed little to no regular prosodic prominence on this syllable. This suggests that gestural alignment may serve to uncover prosodic anchors even when they are not employed by the phonology proper.

Keywords: gesture; Indonesian; Javanese; prosody; stresslessness

1. Introduction

Gesture and speech have been shown to be closely linked as part of a system of multimodal communication (Kendon, 1980; McNeill, 1992, *a.o.*). This is increasingly well-supported by a growing number of studies (Jannedy & Mendoza-Denton, 2005; Loehr, 2004, 2012; Shattuck-Hufnagel & Ren, 2018, *a.o.*). However, the vast majority of studies on the temporal alignment of gesture and speech have been conducted on European languages sharing certain basic prosodic properties. Considerably less is known about the coordination of gesture and speech in non-European languages. As gesture has been shown to coincide with prosodic prominence, in particular pitch accents and stressed syllables, it is unclear what, if

any, prosodic anchors gesture might have in languages that appear to lack pitch accents and word-level prominence.

In the present study, we investigate gesture alignment in local varieties of Indonesian, a language whose prosodic system has been the subject of conflicting claims in the literature and is now often analyzed as stressless (Athanasopoulou et al., 2021 and references therein). Other studies have shown that regional varieties of the national language may differ in their prosody in ways that reflect local substrate languages (Goedemans & van Zanten, 2007) and references therein, with some varieties offering more evidence for word-level stress and others none at all. In this study, we ask how speakers of a language purported to lack word-level stress and phrasal pitch accents coordinate gesture and speech. Given the claims of regional prosodic variation, we also ask whether there are systematic differences in gesture to speech alignment in different regional varieties of Indonesian and what those differences might signify for the prosody of these varieties. We conceive of prosody as a phonological property of words or phrases that is often conveyed with acoustic correlates that may vary from language to language (see, e.g., Bolinger, 1958).

We focus our attention on Indonesian spoken in the eastern part of Indonesia, which Himmelmann (2023) identify as an area with languages with predominately penultimate stress patterns, and the variety of Indonesian spoken in Java, which has been argued to be truly stressless (Himmelmann, 2023; Goedemans & van Zanten, 2007; Himmelmann & Kaufman, 2020; Kaufman & Himmelmann 2024). This work aims to provide insights into co-speech gesture in a language whose prosodic type is underrepresented in the gesture literature and, conversely, uses findings from gesture alignment to corroborate analyses of the prosody of regional varieties of Indonesian. In this way, we contribute a novel approach of using gesture as a window into prosody. Building on the hypothesis that gestures are anchored to prosodically prominent syllables across languages, we also investigate the acoustic properties of gesture-aligned syllables to determine whether they show evidence of being prosodically prominent in the sense of being distinguishable from other syllables in their surrounding environment (see Ladd & Arvaniti, 2023 for discussion of the notion of 'prominence').

The paper is organized as follows: The remainder of this section provides an overview of the current landscape of word-level prosody in Indonesian and regional varieties of Malay. It is followed by a review of previous work demonstrating the relationship between gesture and prosody in other languages. Section 2 outlines the present study and includes details about the speakers, the data and the methodology. Section 3 presents the results of the analysis of gesture alignment, as well as the acoustic analysis of gesture-aligned syllables. Differences between the varieties are highlighted, and the implications of these differences for a future analysis of a difference in prosody among local varieties of Indonesian are discussed. This is expanded upon further in Discussion in Section 4. Finally, Section 5 concludes.

1.1. Word-level prosody in Indonesian

The status of word-level prosody in Indonesian is currently a topic of debate and has engendered contradictory claims in the literature. Early, impressionistic studies claimed that stress regularly occurs on the penultimate syllable, but shifts to the final syllable when the penult contains schwa (Cohn, 1989; Halim, 1981; Stoel, 2007), or that stress regularly occurs on the final syllable (Alisjahbana, 1964; Samsuri, 1971). However,

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

recent work rejects the claim that Indonesian displays any evidence of regular word-level stress, claiming instead that the language is stressless (Athanasopoulou et al., 2021; Goedemans & van Zanten, 2014; Odé & van Heuven, 1994).

At the same time, it has been demonstrated that the prosody of Indonesian varieties is influenced strongly by local languages, which are generally either maintained alongside Indonesian at present or were maintained until relatively recently (Goedemans & van Zanten, 2007; van Zanten et al., 2010). With over 700 languages, Indonesia is one of the most linguistically diverse countries in the world, and it is rare that speakers of the Malay-based national language, *Bahasa Indonesia*, are monolingual in that language. Many Indonesians speak a regionally inflected variety of the national language, in addition to a more divergent Malay dialect that had already served as a lingua franca before the colonial period, possibly in addition to an Austronesian language that is only distantly related to Malay/Indonesian or a non-Austronesian language that has no relation to Malay/Indonesian at all.[1] To take a concrete example, a middle aged person from rural north Sulawesi will typically speak the national language in a way that could be broadly identified as eastern in addition to the local variety of Malay (Manado Malay), which, in this case, is different enough from Indonesian to qualify as a separate language. Additionally, they may speak the local Minahasan language (e.g., Tondano, Tontemboan), which, despite belonging to the same Austronesian family, shares very little with Malay phonologically or grammatically. This example does not even consider older generations who may speak the colonial language, Dutch, or members of the Chinese community who may speak a Chinese language, nor Javanese internal migrants who may continue speaking Javanese in some form. The existence of these varied local influences on Indonesian raises the question of whether we can speak of a uniform Indonesian prosodic system. Instead, it is likely that regional varieties of Indonesian differ in their prosody in ways that are only now being systematically explored.

Studies on local dialects of Malay provide some evidence for regional prosodic variation. Many eastern varieties have been claimed to have penultimate stress, for example, Kupang Malay (Steinhauer, 1983), Manado Malay (Stoel, 2007) and Papuan Malay (Kaland, 2019, 2020, 2021; Kenstowicz, 2021; Kluge, 2017). Ambonese Malay (van Minde, 1997), among other eastern varieties (Paauw, 2008), is described as having phonemic stress which developed from a general penultimate stress pattern that alternated with final stress when the penultimate syllable contained a schwa. This pattern remained even after the merger of *ə and *a, yielding unpredictable final stress on words whose penultimate vowel descends historically from schwa. While several studies describe penultimate stress in the Malay varieties of eastern Indonesia, Maskikit-Essed and Gussenhoven (2016) claim that Ambonese Malay lacks word stress entirely, similar to claims that have been made for Betawi Malay (van Heuven et al., 2008), a western variety spoken in Jakarta. Thus, while certain patterns have come into focus over the last two decades, much work remains to fill out the broader empirical picture. In an attempt to synthesize certain generalizations across Austronesian languages of Island Southeast Asia, Himmelmann (2023) propose an areal

[1]The older local lingua francas are generally referred to as Malay (e.g., Manado Malay, Kupang Malay and Papua Malay), while the more recent locally inflected varieties of the national language are referred to here as varieties of Indonesian. Note that the term Malay is also used to refer to autochthonous varieties of the language in the west of the country, for example, Riau Malay, which were not necessarily spread as lingua francas as in other parts of the country.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

prosodic typology based on available descriptions and instrumental studies. They identify four potential prosodic regions, to be discussed in more detail in Section 2.1.

While the studies discussed above potentially point to differences in regional varieties of Malay, differences in methodology could also play a role in the reported prosodic differences, in particular, for varieties spoken in close geographic proximity. For example, the recent findings by Maskikit-Essed and Gussenhoven (2016) for Ambonese Malay, mentioned above, are perhaps surprising given that Papuan Malay, also spoken in eastern Indonesia and otherwise quite similar to Ambonese Malay, was shown by Kaland (2019) to exhibit regular penultimate stress. A notable difference between these two studies is that Kaland (2019) relied on spontaneous narratives (19 speakers), while the data in the study by Maskikit-Essed and Gussenhoven (2016) on Ambonese Malay consisted of speech recorded in a lab setting (4 speakers). This type of scripted, read speech may not be the best suited to investigating prosody (Kaland, 2019; Roettger & Gordon, 2017). Furthermore, Kaland (2019) investigated 18 acoustic measures as potential candidates of word stress cues, whereas Maskikit-Essed and Gussenhoven (2016) took four measures into account. This is worth noting as, in these varieties, it is possible that stress as a phonological category may manifest different acoustic properties as compared to stress in more familiar (and more widely studied) languages (Himmelmann, 2023). In light of the unresolved questions regarding the word-level prosody of eastern Malay/Indonesian varieties, it is worthwhile to explore this issue further with a range of methodologies, including through the lens of gesture.

In sum, previous studies suggest that word-level prosody varies across in Malay and Indonesian varieties. However, conflicting conclusions for the same or similar varieties indicate a need for corroborating evidence from alternative sources. In the present study, we contribute to this goal through the investigation of co-speech gesture in Indonesian, focusing on spontaneous, naturalistic speech. The next section motivates our decision to look to co-speech gesture through a brief review of previous literature on the relationship between gesture and prosody in other languages.

1.2. Gesture and prosody

A growing body of work has demonstrated the close temporal alignment between gesture and prosodic prominence. Studies on gesture and prosody are often limited to non-referential gestures, sometimes referred to as 'beat' gestures (McNeill, 1992). These types of gestures typically involve quick, rhythmic motions, and they are closely related to speech in terms of timing, but not meaning. By contrast, referential gestures include deictic (pointing) gestures, metaphoric gestures and iconic gestures, which are often timed with respect to the semantic content of an utterance (McNeill, 1992).

However, not all authors divide gestures according to whether they are referential or non-referential. Furthermore, while referential gestures may co-occur with the content they refer to, they may also be timed with respect to prosody (Prieto et al., 2018). For instance, Esteve-Gibert and Prieto (2013) find that deictic gestures, a type of referential gesture, are temporally aligned with pitch accents in Catalan.

The majority of studies on the temporal alignment of gesture and prosody have focused on European languages, and previous studies on English, specifically, have

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

found a strong tendency for gesture to align with pitch accents. For instance, Loehr (2004, 2012) observes a strong tendency for apices (the 'endpoint' or 'goal') of gestures of various types, including beat and deictic gestures, to align to pitch accents. Shattuck-Hufnagel and Ren (2018), using more naturalistic data in the form of a public lecture, report that approximately 83% of non-referential 'beat' gestures were aligned to a pitch accent. This figure was as high as 90% in Yasinnik et al. (2004), which also limited their study to beat gestures, and 95% in Jannedy and Mendoza-Denton (2005), which included referential gestures. Both of these latter studies were also based on video recordings of natural speech.

While it is clear that there may be some variability across speakers, and perfect alignment between gesture and pitch accents is not to be expected, studies of European languages have consistently shown a very strong tendency for speakers to align gestures to pitch accents in languages with diverse stress and intonational patterns (e.g., English, Catalan [Esteve-Gibert & Prieto, 2013], Italian [Esposito et al., 2007]).

Despite the growing consensus with regard to languages showing regular word stress, we know far less about gesture coordination in apparently stressless languages. Recent studies on French have begun to fill this gap. French has been classified as an 'edge-prominence' language (Jun, 2005), with pitch accents demarcating phrase boundaries rather than aligning to stressed syllables (Jun & Fougeron, 2002; Post, 2000). Rohrer et al. (2019) report a similar tendency to align beat gestures to pitch accents in French despite the lack of word stress. However, this alignment was less consistent than in languages with stress: 73% of beat gestures coincided with pitch accents in their study, suggesting that in languages without word-level stress, gesture alignment may be more variable than in languages with stress.

A recent study by Franich and Keupdjio (2022) on Medumba (Grassfields Bantu) is one of the few studies to examine the timing of alignment of gesture, both referential and non-referential, to speech in a tonal language. In Medumba, prominence is located on stem initial syllables and is independent of tone. While the authors found no correlation between gesture and high or rising tones, they suggest that the stem initial syllable may be a target for gesture alignment (though they did not systematically explore this in their study). The fact that gesture is not simply attracted to high f0 or rises suggests that it is anchored to linguistic prominence, which in Medumba, unlike in a language with postlexical pitch accents, is not tied to f0.

The present study builds on Kaufman and Farinella (2022), which examined the alignment of beat gesture to speech in two varieties of Indonesian, finding a strong tendency toward word-final and phrase-final syllables in a Sumatran (western) variety of Indonesian and toward penultimate syllables in all conditions in an eastern variety from Ambon. It also expands on work by Farinella et al. (2023), which focused on gesture alignment and the acoustic correlates of gesture-aligned syllables in Ambonese Indonesian. Their results suggest that there might be real differences in gesture alignment across varieties of Indonesian, and these differences might reflect prosodic differences as evidenced by f0 patterns, duration and vowel quality.

2. Present study

2.1. Data and speaker background

The data for this study come from publicly available video recordings of religious figures preaching in or being interviewed in Indonesian. A number of factors favor

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

this type of data. First, there are a large number of these videos, which allows for working with a relatively large amount of easily accessible, high-quality data. Second, these videos are excellent examples of spontaneous (or in some cases semi-spontaneous) speech, making the data more naturalistic than elicited or laboratory speech. Finally, the speaking style is animated, and the videos contain a large number of gestures. The video clips ranged in time from around 20 minutes to almost 2 hours, but not all parts of the videos were usable for gesture analysis. For each video, we selected portions in which the speaker's hands and upper body are clearly visible, which contained frequent gestures. More information about the clips for each of the speakers is provided below. Stills of each of the speakers are shown in Figure 1.

Four videos of four different speakers were analyzed. The videos themselves as well as background information about the speakers is publicly available and found online, as they are all public religious figures.

Three of the videos feature speakers from the eastern part of Indonesia. Two of the speakers are from Ambon, an island in eastern Indonesia. The first, Pastor Elifas Tomix Maspaitella, who we will refer to as Maspaitella, was born in Rutong, on the island of Ambon in 1974 and raised on the island. The video used in this study is a sermon studio recorded for the Gereja Protestan Maluku (Protestant Church of Maluku). In the video, he is preaching in front of a lectern and giving a semi-scripted sermon lasting just over 23 minutes. A total of 7 minutes of in which the speaker is animated and his hands clearly visible was selected for gesture annotation. Nadia Manuputty-Jambormias, who we will refer to as Manuputty, was born around 1978 and also raised in Ambon. She is also giving a semi-scripted sermon in front of a lectern for the Gereja Protestan Maluku. Her sermon is just under 17 minutes in length, of which 11 minutes were annotated for gesture.

The third eastern speaker, Pastor Mell Atock, who we will refer to as Atock, was born in 1981 in Kupang, on the island of Timor in eastern Indonesia. He runs a popular YouTube channel for which he records and uploads sermons. In the sermon

Figure 1. Images from the videos of the four speakers used in this study. Top left: Maspaitella, top right: Manuputty, bottom left: Atock, bottom right: Bhante.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

selected for this study, he is seated in front of a microphone, and his sermon does not appear to be scripted. The video lasts slightly over an hour, of which 14 and a half minutes were annotated for gesture.

The final speaker is Bhikkhu Bhante Uttamo, henceforth Bhante, a Buddhist bhikkhu, or monk, born in 1960 in Yogyakarta, on the island of Java. In the video, Bhante is being interviewed for an Indonesian television program and is seated in front of a microphone. The interview lasts nearly 2 hours; however, not all of this was usable for gesture annotation, as the parts in which the interviewer was speaking and the camera was trained on the interviewer had to be excluded. Out of the remaining portion of the recording, nearly 15 minutes were annotated for gesture.

As the genre of this video differs from the preaching style of the other three videos, it is possible that this resulted in differences in the number or type of gestures. However, we expect that the strategy of gesture-speech alignment will remain constant, even if Bhante's speaking style is perhaps not as emphatic as that of the other three. How the difference in speaking style affects prosody and gesture (both independently and their temporal synchrony) in Indonesian is an open question that will not be addressed here.

In addition to the regional differences of interest in this study, the speakers vary in their gender, age and likely a number of other factors, including level of education. Individual differences in gesture alignment as well as the influence of social factors on gesture alignment is an area that requires a great deal of further work, and we set these additional factors aside for the present study while acknowledging that they may play a role.

In the videos used for this study, all four speakers are speaking in regionally accented Standard Indonesian rather than regional varieties of Malay (the local lingua francas which preceded the spread of the standardized national language). However, impressionistically, the speech of Maspaitella, Manuputty, and Atock is clearly recognizable as 'eastern Indonesian', and the speech of Bhante can be described as 'Javanese Indonesian' (known colloquially as [məd^{h}o̤k], meaning 'to speak with a Javanese accent').

Table 1 displays speaker names and background information. The geographic location of each of these varieties is shown in Figure 2 in relation to the tentative boundaries of the areal prototypes posited by Himmelmann (2023). The prototypes make generalizations based on regional patterns observed in past work. And while the status of word-level prosody in varieties of Indonesian and regional languages is still far from settled, as discussed at length in Section 1.1, their classification intends to make sense of the varied and conflicting claims. The properties of the two relevant prototypes are as follows:

- **Eastern prototype:** No phonemic length/stress distinctions. Predictable penultimate prominence at the phrase level in unmarked declarative statements,

Table 1. Speakers and their backgrounds

Speaker	City	Region
a. Maspaitella	Ambon	Eastern
b. Manuputty	Ambon	Eastern
c. Atock	Kupang	Eastern
d. Bhante	Yogyakarta	Java

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

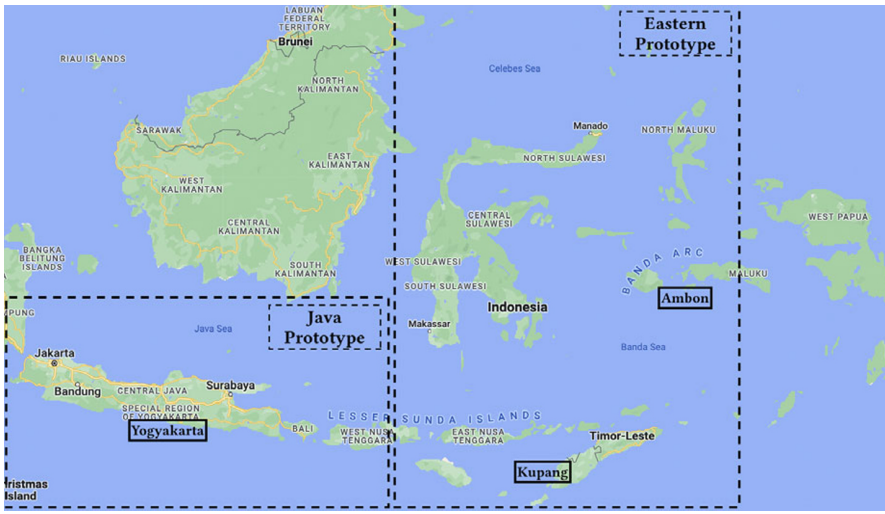

Figure 2. Location of Indonesian varieties.

commonly shifting to final position if the penult contains schwa. Word-level penultimate prominence is attested to highly different degrees, ranging from barely noticeable to very regular and conspicuous prominences on the penultimate syllable of every phonological word. Suffixes and possibly certain enclitics are included in the stress window.

- **Java prototype:** No length distinctions and no other type of word-level prominence. Prominence in pitch, duration and intensity is inherited from higher prosodic levels (prosodic phrase and intonational phrase). Effects of suffixes and enclitics on prominence are highly variable and often difficult to discern.

On this basis, we might expect that speakers of the Javanese variety of Indonesian will show an inconsistent alignment of gesture to any particular syllable of the word or may tend to align gesture to phrase edges (Kaufman & Farinella, 2022). Speakers of the variety of Indonesian spoken in the eastern part of the archipelago, on the other hand, may show more consistent alignment of gesture to penultimate syllables in the word.

2.2. Methodology

For each of the four videos, gesture was annotated without audio in ELAN (Wittenburg et al., 2006) by an experienced coder. We limited our investigation to manual (hand) gestures for the present study. However, we do not limit our sample to non-referential gestures, as it has been shown that even referential gestures are timed with respect to prosody (Prieto et al., 2018). We, therefore, included all gestures that were (i) accompanied by speech, (ii) clearly intentional, which meant excluding gestures that appeared to be unplanned and unfocused (following Shattuck-Hufnagel & Ren, 2018), and (iii) produced with a clear target or endpoint, which meant excluding 'large' gestures that lacked a clear goal. Research on the way in which these types of more continuous gestures are timed with respect to prosody is lacking

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

Figure 3. Stills from the recording of Atock demonstrating the type of gesture excluded in this study, taken from the recording of Atock.

(Wagner et al., 2014). An example of a gesture that was not included in the present study is shown in Figure 3. The images depict stills from the video of Atock making a gesture lasting about 2 seconds. In the images, the speaker is making a wide, circling motion with his hands without a clear target. For gestures such as these, it was not possible to tell where they were aligned in the absence of a target that was a beat-like moment in time, and all gestures of this type were excluded from analysis.

For all gestures that fit the criteria above, the stroke and the apex of the gesture were annotated. Kendon (1972, 1980) defines the stroke as the 'action' part of the gesture, containing the movement interval. We annotate the stroke from the onset, or beginning of movement, up to and including the apex. The apex is the target or point of maximal extension (Loehr, 2004). In our results, we focus on the apex, as it is the apex that appears to show the closest alignment with prosodic prominence (Esteve-Gibert & Prieto, 2013; Leonard & Cummins, 2011). Additionally, our aim is to identify patterns in gesture alignment at the syllable level. Depending on the size of the gesture, a stroke may span several syllables, while the apex is a single point in time and is, therefore, more easily associated with a single syllable. Gestures that involved multiple successive movements were excluded, as it has been shown that speakers tend to time them to finish before a stressed syllable (McClave, 1991).

Figure 4 shows a sample ELAN annotation, demonstrating the type of gesture used in this study and the way it was coded. The stills are taken from the video of Bhante, the speaker from Java, and each consecutive one is two frames from the one preceding (except for the penultimate and final images, which are a single frame apart). Under each video is the corresponding ELAN annotation, with the red line showing the time point. In the first (labeled '1' in the upper left corner), the speaker is just beginning the stroke of the movement. In the second and third images, he continues the pointing gesture, and the stroke spans this entire movement as well. In the fourth image (bottom left corner), the image is quite articulator (in this case, the speaker's hand), but in the subsequent image (labeled '5'), the speaker reaches the apex of the gesture. During coding, we found that in the moments before the apex, the image was blurred (as the speaker was moving), but at the apex, the articulator became clear again, as movement reached its target and was stopped for an instant. The apex is annotated on a separate tier. Because ELAN is only able to mark intervals and not points, the left boundary of the apex interval was always taken as the time point of the apex; the right boundary of the apex was not used. The last image, labeled '6' in the bottom right corner, shows the retraction, which was not annotated as part of the stroke.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

Figure 4. Stills from the recording of Bhante showing an example of the type of gesture used in this study and the annotation in ELAN.

Speech was transcribed in Praat (Boersma & Weenink, 2022) by a native speaker. A coder then segmented the transcription by syllable following the accepted syllable structure of Indonesian (Adisasmito, 1993). Syllable counts for the entire transcribed portion were made, and syllables were annotated as ultimate, penultimate, antepenultimate, or pre-antepenultimate in the word. Monosyllables were excluded from the analysis, as they are uninformative with respect to word prosody. A script was run to match the gestures to the syllables they occurred in, in order to obtain counts of the number of gestures aligned with each syllable position. We excluded any gesture that fell in between two syllables or was too close to the boundary between two syllables to reliably determine which syllable it fell in. Counts were also made of all syllables that did not have a coinciding gesture.

Phrase boundaries were marked impressionistically in Praat in a separate tier as either present or absent for each of the words in the data set using common cross-linguistic correlates to boundaries (pauses, final lengthening, pitch reset). To check this coding, an average of the word duration divided by the number of phonemes in the word was compared for words coded as phrase-final and those coded as non-final. Phrase-final words were found to have greater duration than non-final words. We excluded words whose status was unclear as to phrase-final versus non-final. A script was run to identify, for each gesture, whether it occurred in a word that was phrase-final or phrase non-final (regardless of whether the syllable itself was final or not).

In addition to examining gesture to syllable alignment, acoustic measures were taken to determine whether gesture-aligned syllables showed greater acoustic prominence than non-gesture-aligned syllables. The two variables measured were duration and f0. Raw duration values for each syllable were divided by the number of phonemes in the syllable as a way of accounting for different syllable structures, as in Kaland (2019). This is a coarse-grained method intended to get around the fact that certain syllable structures will be longer simply due to the fact that they contain more segments, which may not reflect any underlying prominence patterns. Note, however, that duration per phoneme does not entirely capture the plausibly high durational variation in consonantal material. 'Duration' in what follows will refer to this transformed value.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

The average f0 in the voiced portion of each syllable was measured, and the raw values were converted to semitones. This was done because stress might affect all voiced material (also consonants) in the syllable. Taking the voiced portion ensures that those effects are taken into account. In addition to the f0 level, we determined whether the voiced subinterval contained a perceptible rise, fall or a rise-fall. This was determined by calculating, for each syllable, the slope of the f0 difference within the interval before and after the f0 maxima within the syllable and comparing that to a perceptual threshold for the given time of the interval (Kaland & Baumann, 2020). However, the perceptual threshold was based on work on English (following Hart et al., 1990, p. 32). It is, therefore, conceivable that listeners of other languages are attuned to more fine-grained f0 excursions, which this method filters out. Even so, this was seen as a necessary initial step in order to factor out extremely small f0 excursions which may not be picked up on by listeners. It should be noted that, for the acoustic measures taken here, we analyze all results descriptively by speaker. Due to the relatively small amount of data and the variability between speakers, we do not conduct statistical tests and instead take a careful look at emerging patterns, which will have to be followed up with more data in order to make any claims of significance.

The videos used in this study, as well as the annotated textgrids and csv files containing the acoustic measurements taken via Praat scripts, are publicly available on OSF, accessible at this link.

3. Results

3.1. Alignment

The results of this study reveal clear patterns in the alignment of gesture to speech, as well as differences in alignment strategies across the four speakers. As one of the goals of this study was to determine whether there were differences in gesture to speech alignment between varieties of Indonesian, results will be discussed separately by speaker. We also report the results of chi-squared tests performed in R (R Core Team, 2021) individually by speaker. We interpret the statistical results with caution, however, given the limited amount of data.

Table 2 shows the number of syllables annotated in the data for each of the speakers, broken down by whether they were aligned with a gesture or not. The second column ('No gesture') shows the number of syllables without gesture, the third column ('Penult/Ultima') shows the number of penultimate and ultimate syllables with gesture, and the final column ('Pre-penult') shows the number of syllables earlier than the penultimate (pre-antepenultimate and antepenultimate) with gesture. As can be seen, syllables earlier than penultimate position in the word are a minority of gesture-aligned syllables for all speakers and were not included in the subsequent analysis. The decision to exclude these syllables was also motivated by

Table 2. Total number of syllables by gesture alignment for each speaker

Speaker	No gesture	Penult/Ultima	Pre-penult
a. Manuputty	312 (59.8%)	191 (36.6%)	18 (3.5%)
b. Maspaitella	277 (71.6%)	94 (24.3%)	16 (4.1%)
c. Atock	292 (67.4%)	120 (27.7%)	21 (4.8%)
d. Bhante	190 (58.3%)	123 (37.7%)	13 (4%)

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

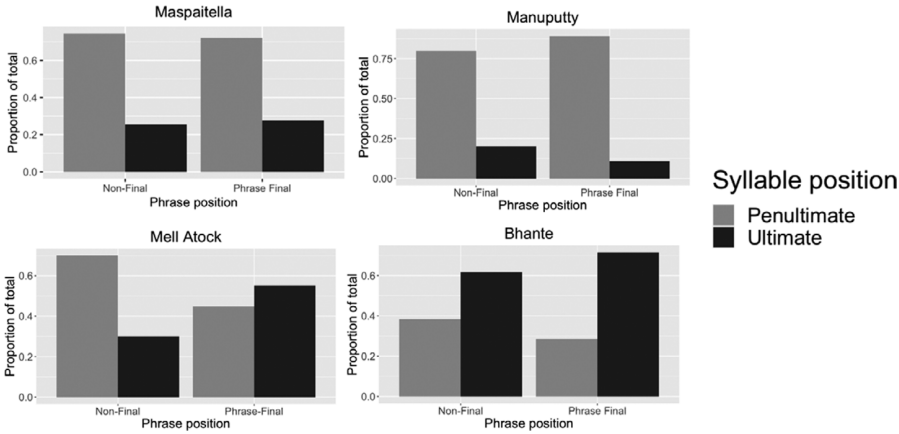

Figure 5. Gesture to syllable alignment by phrase position for all speakers.

the fact that claims of stress in Indonesian have been made for the penultimate and the final syllable, so there are clear predictions for these positions. Note that from this rough-grained view of the alignment patterns, the speaker of the Javanese variety (Bhante) cannot be distinguished from the three speakers of the eastern varieties.

Focusing on patterns of alignment for the gesture-aligned syllables, Figure 5 plots the proportion of gestures aligned to the penultimate and ultimate syllable in the word by phrase position (final or non-final) for each of the speakers individually. For the two eastern speakers whose results are shown in the top row, it is clear that there is a strong tendency to align gesture to the penultimate syllable in the word. For Maspaitella (top left panel), in phrase non-final words, 75% of gesture-aligned syllables were penultimate in the word, and this finding was significant (χ^2 = 91.49, df = 1, p < .001). In phrase-final words, 72% were penultimate in the word, a finding which was also significant (χ^2 = 8.93, df = 1, p = 0.003).

An even stronger tendency for penultimate alignment is observed for Manuputty's speech, shown in the top right panel. Almost 80% of syllables aligned to gesture were penultimate in the word when they were in phrase non-final position, and this difference between penultimate and ultimate word alignment was significant (χ^2 = 24, df = 1, p < .001). In phrase-final words, almost 90% of gestures fell on syllables that were penultimate in the word, a finding that was likewise significant (χ^2 = 41.522, df = 1, p < .001). For both speakers, gesture most frequently coincided with word-penultimate syllables regardless of phrase position, and the difference between alignment to penultimate vs. ultimate syllables in the word was significant at an α of .01. The fact that the pattern of gesture alignment is unaffected by phrase position suggests that gesture targets a position in the word, not the phrase, for these two speakers.

The results of the two speakers presented in the bottom row in Figure 5 display different alignment patterns. Atock (bottom left panel), the speaker from Kupang in eastern Indonesia, also shows a tendency for gesture to coincide with the penultimate syllable in the word in phrase non-final words. This tendency appears weaker than for the other two eastern speakers, with 70% of gestures aligned to the penultimate syllable and 30% aligned to the ultimate syllables in the word, yet still significant (χ^2 = 84.137, df = 1, p < .001). However, for words in phrase-final position, this tendency disappears. Gesture does not display a strong preference for either the penultimate or

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

ultimate syllable in the word, though there is a slightly greater tendency for ultimate alignment (55%) than penultimate alignment (45%). The difference between alignment with the penultimate and ultimate syllable in the word in phrase-final position is not significant for this speaker ($\chi^2 = 0.69145$, $df = 1$, $p = 0.406$).

Finally, Bhante, the speaker from Java, displays a markedly different pattern of gesture alignment from the other three speakers, as can be seen in the panel on the bottom right in Figure 5. In phrase non-final words, there is a weak tendency to align gesture to the ultimate syllable in the word: 61% of gestures on phrase non-final words aligned with the ultimate syllable in the word and 39% with the penultimate syllable in the word. There was, however, no significant difference between alignment with penultimate or ultimate syllables in the word in phrase non-final position for this speaker ($\chi^2 = 1.773$, $df = 1$, $p = 0.183$). For Bhante, unlike the other three speakers, there appears to be no preference for alignment with a particular syllable in the word in phrase non-final position. The tendency to align gesture to the ultimate syllable in the word is stronger in phrase-final words, with 71% of gestures aligned to ultimate syllables and 29% to penultimate syllables, which was significant ($\chi^2 = 12.081$, $df = 1$, $p < .001$).

3.1.1. Interpreting alignment results

Based on the findings from gesture alignment, there is an across the board preference to align gesture to the penultimate syllable in the word for the two eastern speakers from Ambon, a preference to align gesture to the ultimate syllable in the word (especially in phrase-final position) for the speaker from Java and an intermediate pattern for the speaker from Kupang, which is considered part of the eastern region although it is closer to Java than Ambon is.

If the close temporal alignment between manual gesture and prosodic prominence for other languages can be generalized, these results are consistent with penultimate syllables carrying word-level prominence in the eastern variety and ultimate syllables carrying prominence in the Java variety. In phrase non-final position for this speaker of the Javanese variety, there is no significant relationship between the position in the word and gesture alignment, which suggests lack of a syllable-level anchor. However, in phrase-final position, gestures are significantly more likely to occur on final syllables, which suggests that alignment for this speaker may be phrase based not word based.

There are at least two ways of accounting for the alignment pattern in the speech of Atock (Kupang), who shows a significant preference to align gesture to the penultimate syllable in the word in phrase non-final position, as with the other two eastern speakers, but no significant preference for aligning gesture to the penultimate or ultimate syllable in the word in phrase-final position, diverging from the other eastern speakers. It may be the case that penultimate position is prominent at the word level but that this speaker also employs gesture to mark phrase edges. It may also be that word-level prosodic prominence follows a penultimate pattern but phrase-level prominence follows an ultimate pattern and that the latter tends to override the former. This would imply a hybrid of two of the geographical prototypes posited by Himmelmann (2023), the eastern penultimate pattern and a pattern that allows ultimate prominence (either the highly mobile type found in the Java area or the more regular final prominence found further west).

In fact, Halim (1981, pp. 59–62), who describes Indonesian as having a basic penultimate stress patterns, argues that this pattern is subverted in favor of an iambic

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

one in certain phrase-final contexts. Stoel (2005, pp. 120–121) similarly discusses a calling contour and 'emphatic accent shift' which displaces the regular penultimate stress found in Manado Malay to the final syllable. Although we cannot explore this possibility fully here, opposing patterns of word-based and phrase-based (or pragmatically marked) prominence could account for some of the variation observed here.

3.2. Duration

To further test the relation between gesture alignment and prosodic prominence, we compared the duration of gesture-aligned syllables to syllables that were not aligned to gesture for all speakers. The results are displayed individually by speaker in Figure 6. Duration values plotted on the y-axis represent the total duration of the syllable in seconds divided by the number of phonemes in that syllable.

For the three eastern speakers (top two panels and bottom left panel in Figure 6), both penultimate and ultimate syllables aligned to gesture are longer than those that are unaligned in phrase non-final position. This difference is larger for syllables that are penultimate than syllables that are ultimate in the word. There is also some variability between speakers, with Manuputty (top right panel) showing a larger duration difference than the other two eastern speakers, Maspaitella and Atock. In particular, the differences are quite small for Atock, the speaker from Kupang. While his median duration values are higher for gesture-aligned syllables than for unaligned syllables in phrase non-final position, there is overlap in the interquartile range (represented by the boxes), which indicates that this difference may not be very robust.

For the phrase-final syllables, the picture is slightly less clear. For Maspaitella, syllables aligned to gesture are longer only for ultimate syllables; syllables that are penultimate in the word show no difference on the basis of whether they are aligned with a gesture or not. For the other speaker from Ambon, Manuputty, in phrase-final

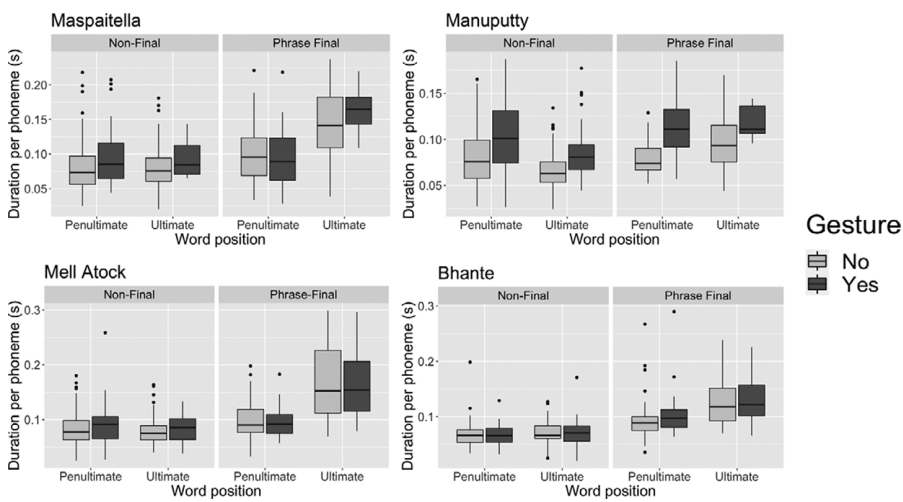

Figure 6. Duration in seconds by word position for phrase non-final words (left) and phrase-final words (right) for syllables with and without gesture by speaker. Note that y-axis scale differs by speaker.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

position, gesture-aligned syllables are longer than other syllables when they are both penultimate and ultimate in the word. For the speaker from Kupang, Atock, there is no difference between syllables aligned to gesture and those that are not. In phrase-final position, syllables that are ultimate in the word are longest for all speakers, with the interesting exception that in Manuputty's speech, penultimate and ultimate gesture-aligned syllables in phrase-final position are of similar duration. It is likely that the effects of phrase-final lengthening on phrase-final words obscure any word-level prominence patterns, a point to which we return in Discussion.

Turning to the speaker from Java, whose results are displayed in the bottom right panel in Figure 6, we see that syllables aligned to gesture and those that are not aligned to gesture are roughly equal in their duration when they occur in words that are phrase non-final. Words in phrase-final position also fail to display durational differences when gesture aligned (though penultimate syllables aligned to gesture in phrase-final position appear slightly longer than those that are not).

3.3. F0

In addition to duration, the second acoustic measurement taken was f0. It has been shown for a number of languages that gesture targets pitch accents (see references in 1.2). However, there is no widely accepted analysis of the intonational phonology of Indonesian at present. Therefore, in order to evaluate the coordination of gesture and f0, measurements of central tendency of f0 and f0 movement were taken. The f0 values are given in semitones (ST) to account for gender differences and to obtain a measure that resembles pitch perception more closely than the Hertz scale.

Figure 7 shows f0 values in semitones (ST) for gesture-aligned and non-gesture-aligned syllables phrase non-finally and phrase-finally. For the first speaker from Ambon, Maspaitella, shown in the top left panel of Figure 7, in phrase-final position, both penultimate and ultimate syllables in the word have higher f0 when aligned to a

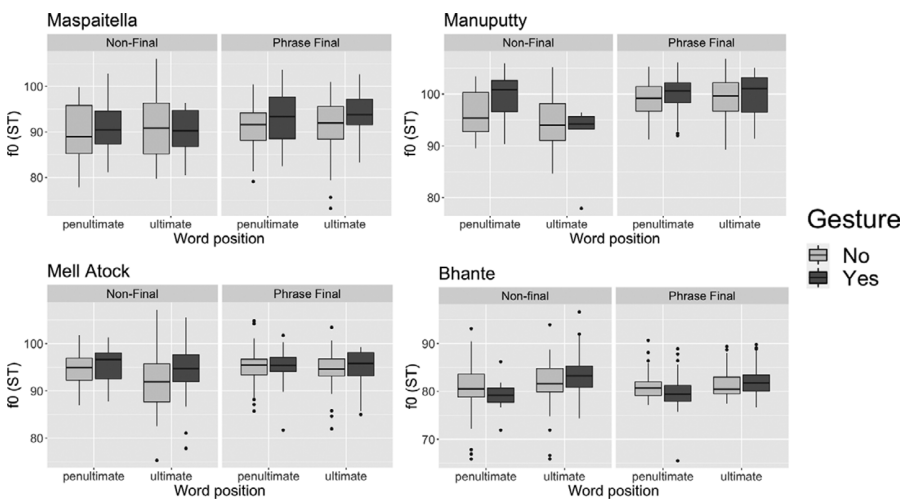

Figure 7. F0 by word position for syllables with and without gesture phrase non-finally (leftmost panels) and phrase-finally (rightmost panels) by speaker.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

gesture, but this is not true in phrase non-final position, where gesture alignment shows no correlation with f0.

The other speaker from Ambon, Manuputty (shown in the top right panel of Figure 7), also has slightly higher median values for syllables aligned to gesture in phrase-final position, but this difference is quite small. However, in phrase non-final position, penultimate syllables in the word have higher f0 when aligned to gesture than when they are not.

F0 values for the final eastern speaker, Atock, are shown in the bottom left panel of Figure 7. Phrase non-finally, f0 is slightly higher for penultimate and ultimate syllables in the word that are aligned to gesture in comparison to those that are not. Phrase-finally, the differences in f0 between aligned and unaligned syllables appear negligible.

Finally, for the speaker from Java, Bhante, shown in the bottom right panel of Figure 7, a completely different pattern emerges. Word-penultimate syllables aligned to gesture in fact have *lower* f0 values than those that are not aligned in both phrase-final and phrase non-final words. Word-ultimate syllables aligned to gesture in phrase non-final and phrase-final words have slightly higher median f0 values.

To sum up the f0 results presented so far, with regard to f0 value, we note that the difference in f0 between penultimate and ultimate syllables in the word in phrase-final position appears negligible for all speakers. The difference is also negligible in phrase non-final position for Maspaitella, but there is a slightly overall higher f0 on the penultimate syllable in the word for Atock and a substantially higher f0 on the penultimate syllable in the word for Manuputty. While Himmelmann (2023) predict all three eastern speakers to favor the penultimate syllable in the word for f0 prominence, only for Manuputty do penultimate syllables in the word show a difference in f0, being higher in phrase non-final position, and highest when aligned to gesture in this position. A similar result is merely suggestive for Atock. There is a, however, a clear difference between the eastern speakers and Bhante, from Java, as he is the only speaker among the four to show higher f0 on the ultimate syllable in the word in phrase non-final position while otherwise maintaining an overall level f0 across four categories.

With regard to correlations between f0 and gesture alignment, we also find differences between Bhante and the three eastern speakers. For all three eastern speakers, if there was a non-negligible difference between gesture-aligned and non-gesture-aligned syllables, the gesture-aligned syllables always coincided with *higher* pitch in comparison to the unaligned syllables. However, for Bhante, in both phrase non-final and phrase-final position, penultimate syllables in the word that were aligned with gesture showed a *lower* F0 than their unaligned counterparts. If this pattern is indeed general, it could be that gesture frequently targets an underlying L or LH tone in this variety, as opposed to the eastern varieties. As an anonymous reviewer points out, it could also be that, for the speaker of this variety, gestures are timed with respect to different parts of the syllable (e.g., the beginning or the end), which could affect how they align to f0. Differences in gesture alignment within the syllable have not been explored in this study. However, the fact that gesture tends to coincide with lower f0 may also be a segmental effect, as the historically voiced stops in Javanese (and Javanese-inflected Indonesian, see Adisasmito-Smith, 2004) are distinguished from their voiceless counterparts by breathy or slack voice (Kenstowicz, 2021, and references therein), which has been shown at least in some cases to induce f0 lowering on following vowels.

In addition to looking at level f0 values, we also calculated the proportions of perceptible rises and falls within a syllable out of all gesture-aligned syllables for a

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

Table 3. Proportion of F0 rises and falls on gesture-aligned syllables per speaker

		Phrase non-final		Phrase final	
Speaker	F0	Penult	Ultima	Penult	Ultima
Maspaitella	Rise	.14	.07	.18	.18
	Fall	0	0	.07	.18
Manuputty	Rise	.56	.14	.61	0
	Fall	.14	.28	.12	.40
Atock	Rise	.17	.20	.16	.19
	Fall	.07	.20	.07	.31
Bhante	Rise	.11	.17	.12	.31
	Fall	.11	.15	.14	.17

speaker, shown in Table 3. For all speakers, the majority of syllables occurred with level f0, and those syllables are omitted from the counts in the table.

From the table, it is clear that for most speakers, only a minority of syllables coinciding with gesture contain f0 rises or falls (with the notable exception of Manuputty). For two of the eastern speakers, Maspaitella and Atock, the majority of gesture-aligned syllables did not coincide with any perceptible pitch movements (and, as is clear from the table, Maspaitella's speech showed less f0 movement overall). There are, however, two points to note from the results in Table 3. In Manuputty's speech, a significant portion of penultimate gesture-aligned syllables contain an f0 rise. For word-ultimate syllables, falls were more frequent, reflecting the fact that pitch rises are consistently located on the penultimate syllable in her speech and are followed by a dip in f0. The second point shown in Table 3 is that for Bhante, the speaker from Java, rises occurred on 31% of word-ultimate syllables in phrase-final position. While the data are clearly not conclusive, it is worth noting that this is the only speaker with a non-negligible amount of rises on word ultimate syllables in phrase-final position, suggesting that this may be a feature of Javanese inflected Indonesian and possibly inherited from Javanese prosody itself. This conjecture is supported by Stoel (2006), who notes that intermediate 'Accentual Phrases' in Banyumas Javanese regularly end in a High edge tone (H%).

4. Discussion

While the results presented here only represent a first approach, certain findings corroborate an emerging picture of regional variation in the prosody of Indonesian. Across each part of the investigation, the Javanese speaker grouped separately from the three eastern speakers, despite a good deal of variation among the three eastern speakers.

Perhaps, the most striking finding is that the three eastern speakers showed a rather clear pattern of aligning beat gestures to penultimate syllables in the word, while the Javanese speaker showed no consistent alignment preference at the word level in phrase non-final position but did show significantly greater ultimate alignment in phrase-final position. In contrast, the two eastern speakers from Ambon showed a roughly equal or stronger preference for the penultimate syllable in the word in phrase non-final position when compared to phrase-final position. This suggests that, at least for this speaker of Javanese Indonesian, gesture targets a position in the phrase, rather than the word. The third eastern speaker, Atock, also

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

showed greater alignment with the ultimate syllable in the word in phrase-final position than in phrase non-final position but showed significantly greater alignment with the penultimate syllable in phrase non-final position. This could signal that gesture aligns at both the word and phrase level for this speaker.

Generally speaking, for the eastern varieties, it appears that gesture targets a particular position in the word, as evidenced by the strong preference for alignment with the penultimate syllable in the word in phrase non-final position. More generally, if these patterns are truly representative of dialect differences, the alignment of gesture to phrase edges in the Javanese variety but to the word level in eastern varieties accords well with the predictions of Himmelmann (2023). What is perhaps unexpected given this typology is that only one of the three eastern speakers, Manuputty, displayed clear acoustic correlates that might provide evidence for word-based penultimate prominence. The two factors examined, duration and f0, showed little to no evidence for such prominence in the two other speakers *despite their use of the word-penultimate syllable as a gestural anchor*.

This finding presents the tantalizing possibility of gesture coordination uncovering word-level prosodic structure even when it is inaudible (either because it is inactive or occluded by higher level prosody). Fleshing out this idea a bit further, some of the disagreement in the literature on the prosody of Indonesian and Malay varieties, may be due to a split that is rare in more familiar languages. Prosodic structure on the word level is present and may even contain (syllabic or moraic) heads that function as anchor points, but there is a paucity of obligatory rules that refer to those heads. In those varieties that contain such word-level structure, the anchors may only emerge clearly in emphatic speech or in manual gesture. Under typical elicitation conditions, and especially in reading tasks, these anchors may go completely unused, giving the impression of a total lack of prosodic structure at least as far as suprasegmental correlates are concerned. The absence of evidence should thus not be taken as evidence of absence under these conditions (cf. Maskikit-Essed & Gussenhoven, 2016, who argue for the lack of word-based prominence in Ambonese Malay on such grounds). Rather, it is only unpredictable *variation* in anchoring gestures and pitch accents in phrase non-final position that truly supports the view that certain varieties may lack prosodic word structure altogether or at least do not make reference to such structure in their higher level prosody. The relatively weak difference between the coordination of gestures to penultimate versus ultimate syllables in non-phrase-final position in the Javanese Indonesian data examined here compares well to the variation found in French by Rohrer et al. (2019). This provides a small hint toward a deeper lack of word-based anchors in these types of stressless languages, although far more work is needed to verify this claim.

Another point which deserves emphasis here is that only for the speaker who showed regular evidence of acoustically prominent penultimate syllables, Manuputty, did we find regular lengthening of gesture-aligned syllables. This seems to suggest that the more prominent stress is, the more likely it is to be tightly linked to gesture, with gesture and acoustic prominence equally re-enforcing each other, even in the exceptional cases where both acoustic prominence and gestural anchoring occur on the ultimate syllable in the word in a generally word-penultimate pattern.[2]

[2]See Kaufman and Farinella (2022) for possible explanations of exceptional alignment to non-penultimate syllables in the word in Manuputty's speech.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

Our findings cannot be used to argue conclusively for word-level stress in the eastern varieties examined here. However, they are certainly compatible with a notion of penultimate prominence at the word level and difficult to account for under the assumption that this variety lacks word-level anchors entirely. Further, the differences in gesture alignment and duration between the speakers of the eastern and western varieties are difficult to explain under the assumption that both varieties are stressless. However, it should be noted that stress, if it does exist as a phonological category in the eastern variety, may be different when compared to more familiar languages in terms of its acoustic properties, as well as the way it is used by speakers and the way it interacts with intonation (Himmelmann, 2023). Rather, it may be the case that word-level prominence exists in an abstract sense but is not manifested in familiar ways.

5. Conclusion and further directions

The study on gesture alignment presented here centered on varieties of Indonesian as spoken in Java, which has been argued to lack word-level stress, and the eastern part of the country, which has been claimed to lack word-level stress by some authors and to show penultimate word-level prominence by others. On one hand, we have shown a surprising level of word-based alignment for gesture, even in the absence of clear acoustic correlates of word-based prominence. On the other hand, we have shown that the speaker of the Javanese variety differed systematically from the speakers of the eastern varieties with respect to anchoring and lack of anchoring. In addition to the need for expanding the dataset with more material and more speakers, there are many points that merit further investigation.

Firstly, the f0 levels in Figure 7 and the rises and falls in Table 3 are evaluated on a single syllable. To assess whether gesture coincides with phrasal f0 events, as has been suggested for European languages such as English, the coincidence of gesture and f0 rises or falls across syllables should also be examined. This could be done through manual prosodic annotation, such as Rapid Prosody Transcription (Cole & Shattuck-Hufnagel, 2016) or PoLaR (Ahn et al., 2021), rather than relying on the acoustic measure of f0 taken here.

There have also been claims made for other languages that, in addition to aligning with prosodic prominence, gesture is also timed to co-occur with prosodic phrase edges. Specifically, Kendon (1972, 1980) claims that the gestural phrase, which consists of the preparation phase, the stroke and the retraction, aligns with the intonational phrase in English, while Loehr (2012) claims that in English, gesture phrases are timed to align with intermediate phrases. Esteve-Gibert and Prieto (2013) furthermore find that gesture is sensitive to prosodic phrasing in Catalan. It is worth exploring the possibility of phrase-based anchoring using more fine-grained techniques than the simple phrase-final/phrase non-final dichotomy employed here.

It may also be informative to look further into the instances where gesture aligned to pre-antepenultimate and antepenultimate syllables in the word. These cases were not analyzed in this study, as these syllables were small in number for all speakers. The variability observed here across speakers, combined with the small amount of data, would make it difficult to interpret any patterns for these syllables. However, a possibility to be explored is that gesture alignment may be affected by morphological structure (for instance, if there are affixes or clitics). See also Kaufman and Farinella (2022) for a discussion of this point.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

An additional issue to address concerns the precise relation between the gestural apex and the prosodic anchor. Kendon (1980) makes the claim that the gesture stroke occurs at or *just before* the stressed syllable (similar to how certain languages display a systematic delayed pitch peak with regard to the anchor). What this would mean for our data is that some gestures that we have categorized as aligning with the penultimate syllable in the word are actually targeting the ultimate syllable in the word. We hope to address these and other issues in following work.

Acknowledgments. The authors also wish to thank Amalia Suryani, Nikolaus Himmelmann and Kristine Yu.

Funding statement. Research for this paper was funded by the German Research Foundation (DFG) – Project-ID 281511265 – SFB 1252.

References

Adisasmito, N. (1993). Syllable structure and the nature of schwa in Indonesian. *Studies in Linguistic Sciences*, 23, 1–19.

Adisasmito-Smith, N. (2004). *Phonetic and phonological influences of Javanese on Indonesian* [Ph.D. thesis]. Cornell University, Ithaca, NY.

Ahn, B., Veilleux, N., Shattuck-Hufnagel, S., & Brugos A. (2021). Embarking on PoLaR explorations: A framework for intonational annotation and analysis. ms. https://ling.auf.net/lingbuzz/006269.

Alisjahbana, S. T. (1964). *Tatabahasa baru Melayu/Indonesia [New grammar of Malay/Indonesian]*. Zaman Baru Limited.

Athanasopoulou, A., Vogel, I., & Pincus, N. (2021). Prosodic prominence in a stressless language: An acoustic investigation of Indonesian. *Journal of Linguistics*, 57(4), 695–735.

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.23. http://www.praat.org/

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, 14(2–3), 109–149.

Cohn, A. (1989). Stress in Indonesian and bracketing paradoxes. *Natural Language and Linguistic Theory*, 84, 395–415.

Cole, J., and Shattuck-Hufnagel, S. (2016) New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1–29.

Esposito, A., Esposito, D., Refice, M., Savino, M., & Shattuck-Hufnagel, S. (2007). A preliminary investigation of the relationship between gestures and prosody in Italian. *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, Vol. 18, p. 65. IOS Press.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 56, 850–864.

Farinella, A., Kaland, C., & Kaufman, D. (2023). Gesture and prosodic prominence in Ambonese Indonesian. In: R. Skarnitzl & J. Volín (Eds.) *Proceedings of the 20th International Congress of Phonetic Sciences – ICPhS 2023*. International Phonetic Association.

Franich, K., Keupdjio, H. (2022) The Influence of Tone on the Alignment of Speech and Co-Speech Gesture. *Proc. Speech Prosody 2022*, 307–311, https://doi.org/10.21437/SpeechProsody2022–63

Goedemans, R., & van Zanten, E. (2007). Stress and accent in Indonesian. *LOT Occasional Series*, 9, 35–62.

Goedemans, R., & van Zanten, E. (2014). No stress typology. In J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller & E. Van Zanten (Eds.), *Above and Beyond the Segments* (pp. 83–95). John Benjamins.

Halim, A. (1981). *Intonation in relation to syntax in Indonesian. Pacific linguistics, D-36.* Department of Linguistics, RSPAS, Australian National University, Canberra.

Hart, J. T., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Himmelmann, N. (2023). On the comparability of prosodic categories: Why 'stress' is difficult. *Linguistic Typology*, 27(2), 341–361.

Himmelmann, N., & Kaufman, D. (2020). Prosodic systems: Austronesia. In C. Gussenhhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 370–383). Oxford University Press.

Jannedy, S., & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure: ISIS; Working Papers of the SFB*, 632(3), 199–244.

Jun, S. A. (2005). Prosodic typology. In S. A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430–458). Oxford University Press.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, 14, 147–172.

Kaland, C. (2019). Acoustic correlates of word stress in Papuan Malay. *Journal of Phonetics*, 74, 55–74.

Kaland, C. (2020). Offline and online processing of acoustic cues to word stress in Papuan Malay. *The Journal of the Acoustical Society of America*, 147(2), 731–747.

Kaland, C. (2021). The perception of word stress cues in Papuan Malay: A typological perspective and experimental investigation. *Laboratory Phonology*, 12(1), 1–33.

Kaland, C., & Baumann, S. (2020). Demarcating and highlighting in Papuan Malay phrase prosody. *The Journal of the Acoustical Society of America*, 147(4), 2974–2988.

Kaufman, D., & Farinella, A. (2022). Gesture alignment in a "stressless" language. In T. Clark, J. Dussere, & C. Ting (Eds.), *Proceedings of the Twenty-Eighth Meeting of the Austronesian Formal Linguistics association (AFLA)* (pp. 29–46). University of Western Ontario.

Kaufman, D., & Himmelmann, N. (2024). Suprasegmental phonology. In A. Adelaar & A. Schapper (Eds.), *Oxford guide to the Malayo-Polynesian languages of Southeast Asia*. Oxford: Oxford University Press.

Kendon, A. (1972). Some relationships between body motion and speech. *Studies in Dyadic Communication*, 7(177), 90.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication*, the Hague: Mouton (pp. 207–227).

Kenstowicz, M. J. (2021). Phonetic correlates of the Javanese voicing contrast in stop consonants. *NUSA: Linguistic Studies of Languages in and Around Indonesia*, 70, 1–37.

Kluge, A. (2017). *A grammar of Papuan Malay*. Language Science Press.

Ladd, D. R., & Arvaniti, A. (2023). Prosodic prominence across languages. *Annual Review of Linguistics*, 9, 171–193.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457–1471.

Loehr, D. (2004). *Gesture and intonation* [PhD thesis]. Georgetown University.

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3(1), 71–89.

Maskikit-Essed, R., & Gussenhoven, C. (2016). No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay. *Phonology*, 33, 353–389.

McClave, E. Z. (1991). *Intonation and gesture* [PhD thesis]. Georgetown University.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Odé, C., & van Heuven, V. J. (1994). *Experimental studies of Indonesian prosody. Volume 9. Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië*. Leiden University.

Paauw, S. H. (2008). *The Malay contact varieties of eastern Indonesia: A typological comparison* [Ph.D. thesis]. State University of New York at Buffalo.

Post, B. (2000). *Tonal and phrasal structures in French intonation* (Vol. 34). Thesus.

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: A labelling proposal. In *Proceedings of the 9th international conference on Speech Prosody* (pp. 201–205).

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Roettger, T., & Gordon, M. (2017). Methodological issues in the study of word stress correlates. *Linguistics Vanguard*, 3(1), 1–11.

Rohrer, P. L., Prieto, P., & Delais-Roussarie, E. (2019). Beat gestures and prosodic domain marking in French. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, (pp. 1500–1504).

Samsuri. (1971). *Ciri-ciri prosodi dalam kalimat Bahasa Indonesia [Prosodic characteristics in indonesian sentences]*. Nusa Indah.

Shattuck-Hufnagel, S., & Ren, A. (2018). The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Frontiers in Psychology*, 9, 1514.

Steinhauer, H. (1983). Notes on the Malay of Kupang (Timor). Studies in Malay dialects. Part II. *NUSA: Linguistic Studies in Indonesian and Languages in Indonesia*, 17, 42–64.

https://doi.org/10.1017/langcog.2025.10043 Published online by Cambridge University Press

Stoel, R. (2005). *Focus in Manado Malay: Grammar, particles, and intonation* (Vol. 134). Leiden University Press.
Stoel, R. (2006). The intonation of Banyumas Javanese. In *Proceedings of Speech Prosody 2006*, (pp. 827–830).
Stoel, R. (2007). The intonation of Manado Malay. *LOT Occasional Series*, 9, 117–150.
van Heuven, V. J., Roosman, L., & van Zanten, E. (2008). Betawi Malay word prosody. *Lingua*, 118(9), 1271–1287.
van Minde, D. (1997). *Malayu Ambong: Phonology, morphology, syntax*. Research School CNWS.
van Zanten, E., Stoel, R., & Remijsen, B. (2010). Stress types in Austronesian languages. In H. van der Hulst,R. Goedemans, and E. van Zanten (Eds.), *A survey of word accentual patterns in the languages of the world* (pp. 87–112). De Gruyter.
Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209–232.
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.
Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. (2004). The timing of speech-accompanying gestures with respect to prosody. *Proceedings of the International Conference: From sound to sense*, 50, 10–13.

Cite this article: Farinella, A., Kaland, C., & Kaufman, D. (2025). Gesture, prosodic prominence and stresslessness in Indonesian, *Language and Cognition*, 17, e92, 1–22. https://doi.org/10.1017/langcog.2025.10043