GESTURE ALIGNMENT IN A "STRESSLESS" LANGUAGE*

Daniel Kaufman
Queens College, CUNY & ELA
dkaufman@qc.cuny.edu

Alessa Farinella
University of Massachusetts Amherst
afarinella@umass.edu

1. Introduction

Varieties of Indonesian and Malay have been variously described as showing both penultimate stress (Cohn 1989; Stoel 2007; Halim 1974) and final stress (Samsuri 1971; Alisjahbana 1964). At the same time, it has been noted from the earliest descriptions of Malay onwards that word stress, wherever it was located, was not equivalent to the type more familiar to European grammarians. A growing body of work builds on the old intuition that all syllables (excluding those with a schwa nucleus) receive “even stress” and claims that Indonesian is a language without word stress at all (Odé and van Heuven 1994; Tadmor 2000; Zanten et al. 2003, 2010; Goedemans and van Zanten 2014; Maskikit-Essed and Gussenhoven 2016; Athanasopoulou et al. 2021).[1] This line of investigation emphasizes the influence of both the speaker’s as well as the linguist’s mother tongue in the production and perception of stress.

Despite considerable differences across regional varieties of Malay, some claims of stresslessness have been sweeping. For instance, Zanten et al. (2010, 103) argue for a language-wide pattern of stresslessness despite attested areal effects: “Even if some varieties of Indonesian will reflect the stress pattern of regional substrates [...] Indonesian as a language does not have stress as a linguistic property.” It is unclear how the regional varieties alluded to can both reflect substrate stress patterns while simultaneously lacking stress on the language level. This question is of special importance in the context of Indonesian, which constitutes a classic case of diglossia. Standard Indonesian is only used in formal settings, e.g. broadcasting, lectures, and writing, while dozens of local varieties are used throughout the archipelago in casual settings. Virtually no Indonesian speaks only Standard Indonesian just as virtually no one raised in the Arab world speaks only Classical (*fusħa*) Arabic. Regional varieties of Indonesian, referred to as Malay (i.e. *Melayu Manado/Ambon/Kupang*, etc.) are now most often acquired simultaneously with one or more local languages with Standard Indonesian being learned in school and through the media.

The few studies explicitly focused on the prosody of regional varieties of Indonesian/Malay do not yet reveal strong areal patterns. Western varieties, like the Betawi dialect spoken in Jakarta, have been observed to lack stress (Roosman 2006; van Heuven et al. 2008) while eastern varieties, represented by only a handful studies, come to different conclusions: Stoel (2007) claims that Manado Malay displays penultimate stress; Kaland (2019) shows similar findings for Papuan Malay; Maskikit-Essed and Gussenhoven (2016), on the other hand, claim that Ambon Malay is stressless and that pitch prominence is anchored only to phrasal boundaries.

If differences in local Indonesian prosodies are rooted in substrate effects, what do areal

*We thank the audience at AFLA 28 for helpful suggestions and especially Amalia Suryani, Nikolaus Himmelmann and Kristine Yu. All usual disclaimers apply.

[1] See Himmelmann and Kaufman (2020) for a recent summary of the literature.

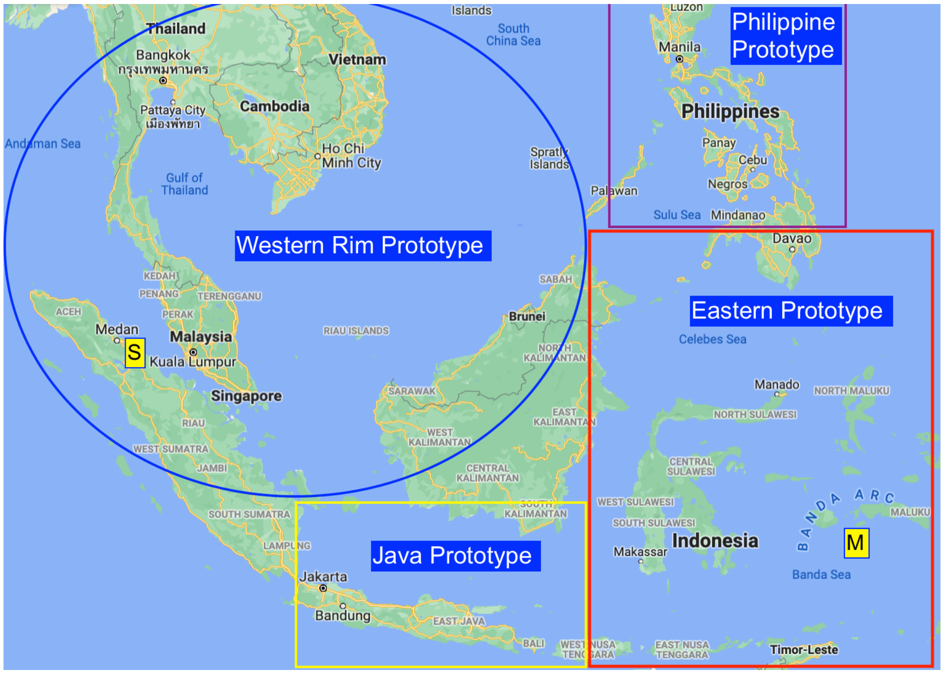


Figure 1: Areal prosodic typology (based on Kaufman and Himmelmann forthcoming)

patterns across the languages of Indonesia lead us to expect? Kaufman and Himmelmann (forthcoming) propose an areal typology of Austronesian prosody in which stresslessness is centered in Java, an iambic pattern is found on the western rim of Borneo and coastal Sumatra, and a penultimate prominence pattern covers much of eastern Indonesia extending outwards all the way to Polynesia (see, for example Zuraw et al. 2014, for the details of one such pattern in Samoan). There is a close match between this areal typology and the patterns reported for local varieties of Indonesian/Malay. Here, we examine a portion of this typology from the novel perspective of manual gesture alignment and prosody in a genre of declamatory speech across two dialects, that of the Asahan/Riau region in northern Sumatra versus that of Ambon in the east.

2. Background

2.1. Austronesian prosodic typology

Kaufman and Himmelmann's (forthcoming) areal typology for Austronesian prosody, based on existing descriptions, is shown in Figure 1. The characteristics of the four areal prototypes are summed up as follows:

- **Western Rim prototype**: Final prominence either on the word or phrase level.
- **Philippine prototype**: Phonemic vowel length distinction in open penultimate syllables. Both initial and final phrase edges are tonal targets, with long vowels in penultimate position in a phrase attracting (intonational) edge tones. Suffixes but not clitics shift length rightwards.
- **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.

- **Eastern prototype**: No phonemic length/stress distinctions. Predictable penultimate prominence on the phrase level in unmarked declarative statements, commonly shifting to final position if 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 included in the stress window.

The dialects of Indonesian investigated here are located in Figure 1 with the initials of the two speakers. The S south of Medan in north Sumatra is located within the Western Rim Prototype area and the M located in Ambon lies near the southern range of the Eastern Prototype area.

2.2. Beat gestures and prosody

There is a strong, cross-linguistic tendency for manual gestures to be temporally aligned to prosodic prominence in speech (Kendon 1980; McNeill 1992). An increasing number of studies using observations from naturalistic speech data as well as experimental methods have shown this to be the case for a variety of languages (English: Loehr 2004, 2012, Leonard and Cummins 2011, Yasinnik et al. 2004, Jannedy and Mendoza-Denton 2005 a.o.; Italian: Esposito et al. 2007; Catalan: Esteve-Gibert and Prieto 2013; Dutch: Krahmer and Swerts 2007; Brazilian Portuguese: Rochet-Capellan et al. 2008). The majority of studies investigating the alignment of co-speech gestures focus on a type of non-referential gesture known as a beat gesture (McNeill 1992). Beat gestures are discrete, biphasic gestures involving abrupt, rhythmic movements and clear targets. The non-referential nature of beat gestures are distinguished from deictic or iconic gestures (those that physically resemble an object or an idea) in bring less anchored to particular semantic content. Rather, it has been shown that prosodic prominence determines the alignment of beat gestures to speech. The present study therefore focuses exclusively on the alignment of beat gestures to syllables.

Investigations of beat gesture alignment have focused almost exclusively on languages with word level stress. While there have been studies of gesture alignment in stressless languages, these studies are mainly concerned with iconic gestures and the words they are semantically associated with (Chui 2005; Ferré 2014) or the alignment of gesture to prosodic focus (Fung and Mok 2018; Roustan and Dohen 2010). A rare exception is Rohrer et al. (2019), who investigate the alignment of gesture to pitch accents in French, a “non-stress accent" language in Beckman’s (1986) typology, where pitch accents are associated with the phrase and not a particular prominent position within a word (Post 2000; Jun and Fougeron 2002). Rohrer et al. (2019) find that while there is a tendency to align beat gestures to H pitch accents, alignment is both less precise and less consistent in comparison to what has been observed for stress languages. From this, we might expect that languages with word level stress exhibit strict, consistent alignment of gestures to prosodically prominent syllables, while languages in which prominence is strictly a phrasal phenomenon exhibit more variable alignment that does not target particular syllables.

3. Manual gesture as a natural experiment in prosodic structure

The gesture studies discussed above investigated the alignment of beat gestures to syllables in languages whose prosodic structure has been well-described, and thus provided evidence of a cross-linguistic tendency for manual gestures to be temporally aligned to prosodically prominent syllables. The goal of the present study is to investigate the alignment of gesture to syllables as an indirect way of assessing prosodic prominence in Indonesian, a language whose stress system, or lack thereof, has generated considerable debate.

3.1. Data

The data used in this study come from two publicly available video recordings of religious sermons. Religious sermons were used because this genre typically instantiates lively, animated speech, in which the speaker employs multiple modalities for dramatic effect.[2] The first video is a studio-recorded sermon by Nadia Manuputty-Jambormias (henceforth Manuputty), who, from all available data, was born around 1978 and raised in Ambon, east Indonesia. The second video analyzed is a lecture by Abdul Somad Batubara (henceforth Somad), an Islamic preacher of Malay descent from Asahan province in northern Sumatra, born in 1977.

The speech genres in both videos are similar but not identical. Manuputty's sermon appears to be well-rehearsed and is presented directly to the camera. Somad's lecture, which is far longer, was recorded at a public event in a mosque and is more spontaneous, containing many asides and digressions. Both Manuputty and Somad are speaking in regionally inflected varieties of the national language, Indonesian. They are *not* speaking in the local Malay varieties (Riau Malay in the case of Somad and Ambon Malay in the case of Manuputty). Impressionistically, however, the prosody of their Indonesian appears clearly rooted in these Malay varieties. Thus, while the language they employ is at least doubly removed from any non-Malay substrate, the speakers are still easily identified as speaking in a Sumatran inflected style (in the case of Somad) and a generally eastern inflected style (in the case of Manuputty).

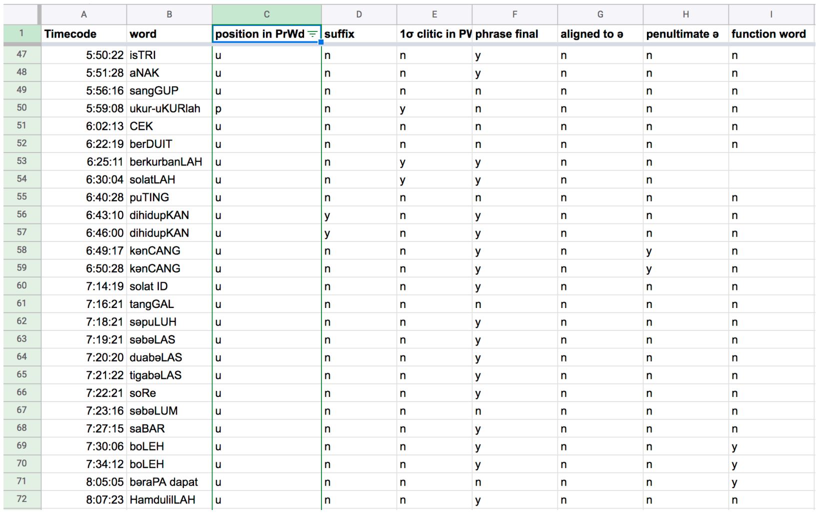
3.2. Method and procedure

Despite increasing work on gesture alignment, there remains much variation in methodology and a lack of consensus as to which aspects of the speech signal and the gesture should be treated as anchors for purposes of measurement (Shattuck-Hufnagel et al. 2007). See Shattuck-Hufnagel et al. (2007); Leonard and Cummins (2011); Loehr (2012); Pouw and Dixon (2019) for recent overviews of the literature. Here, we use the gesture apex as the anchor as it is easier to locate it in a single point in time compared to the peak velocity or the stroke. This decision is additionally supported by the results of Esteve-Gibert and Prieto (2013) and Leonard and Cummins (2011), who find that, among other possible anchor points, the apex bears the strongest correlation with prosodic prominence.

For each video, the location of the apex of the gesture was annotated graphically in Adobe Premiere. Audio was muted during the entire coding process so as not to influence the positioning of the annotations. In addition to the apex, the location of peak effort was also marked in cases where it did not coincide with the apex. For each annotated gesture apex, the following five data points were tabulated:

1. the location of the apex within the word (U = ultimate, P = penultimate, A = antepenultimate)
2. whether the word contained a suffix (Y/N)

[2] Declamatory speech may differ systematically with casual conversational speech in having a higher density of pitch accents and other prominences (Im et al. 2018).



	A	B	C	D	E	F	G	H	I
1	Timecode	word	position in PrWd	suffix	1σ clitic in PV	phrase final	aligned to ə	penultimate ə	function word
47	5:50:22	isTRI	u	n	n	y	n	n	n
48	5:51:28	aNAK	u	n	n	y	n	n	n
49	5:56:16	sangGUP	u	n	n	n	n	n	n
50	5:59:08	ukur-uKURlah	p	n	y	n	n	n	n
51	6:02:13	CEK	u	n	n	n	n	n	n
52	6:22:19	berDUIT	u	n	n	n	n	n	n
53	6:25:11	berkurbanLAH	u	n	y	y	n	n	
54	6:30:04	solatLAH	u	n	y	y	n	n	
55	6:40:28	puTING	u	n	n	n	n	n	n
56	6:43:10	dihidupKAN	u	y	n	y	n	n	n
57	6:46:00	dihidupKAN	u	y	n	y	n	n	n
58	6:49:17	kənCANG	u	n	n	y	n	y	n
59	6:50:28	kənCANG	u	n	n	y	n	y	n
60	7:14:19	solat ID	u	n	n	y	n	n	n
61	7:16:21	tangGAL	u	n	n	n	n	n	n
62	7:18:21	səpuLUH	u	n	n	y	n	n	n
63	7:19:21	səbəLAS	u	n	n	y	n	n	n
64	7:20:20	duabəLAS	u	n	n	y	n	n	n
65	7:21:22	tigabəLAS	u	n	n	y	n	n	n
66	7:22:21	soRe	u	n	n	y	n	n	n
67	7:23:16	səbəLUM	u	n	n	n	n	n	n
68	7:27:15	saBAR	u	n	n	y	n	n	n
69	7:30:06	boLEH	u	n	n	y	n	n	y
70	7:34:12	boLEH	u	n	n	y	n	n	y
71	8:05:05	bəraPA dapat	u	n	n	n	n	n	y
72	8:07:23	HamdulilLAH	u	n	n	y	n	n	n

Figure 2: Coding of data

3. whether the word contained a monosyllabic enclitic (Y/N)
4. whether the word contained a schwa in the penultimate syllable (Y/N)
5. whether the word was phrase final (Y/N)[3]

A total of 282 gestures were annotated from the Somad recording, and 279 gestures were annotated from the Manuputty recording. Of these, ambiguous tokens and those that were too unclear to transcribe were excluded, as were monosyllabic words. Additionally, tokens with pre-antepenultimate or penultimate alignment were excluded, as they made up less than 3% of the total for Somad and less than 8% for Manuputty (although we return to these below in §5.1.2). In cases where rapid gestures produced multiple apices in a single word, each apex was treated independently. For instance, in one of the recordings (Somad, 18:39:24), each of the three syllables in the word *terapi* 'therapy' coincides with a gesture apex and thus corresponds to three data points (with an antepenultimate, penultimate and ultimate alignment). This yielded a total of 230 tokens for Somad and 244 tokens for Manuputty. A sample of the coding can be seen in Figure 2.

Three frames from one of Somad's gesture sequences are shown in Figure 3. Here, he raises a book above the table and taps it down forcefully. A typical gesture sequence from Manuputty is shown in Figure 4, with the last frame showing the peak.

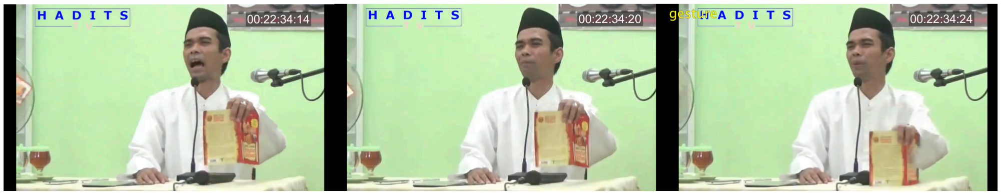


Figure 3: Somad sequence

[3] A word was identified as phrase final if it was followed by a clear prosodic break, as indicated by the presence of a pause and pre-boundary lengthening

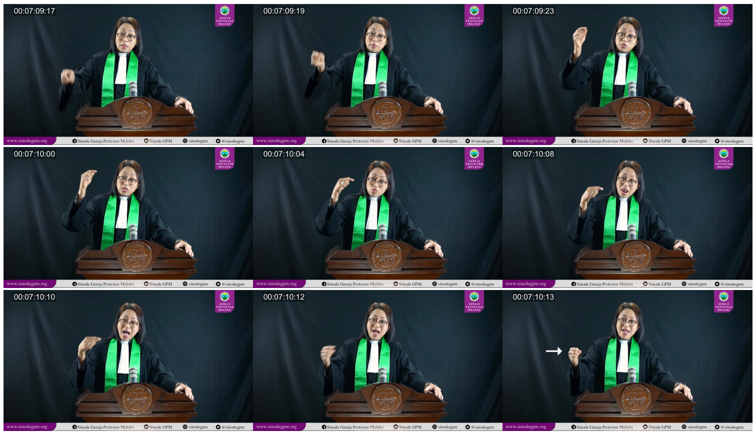


Figure 4: Manuputty sequence

4. Results

The data from both speakers were analyzed separately and subject to a basic within-speaker statstical analysis in R (R Core Team 2021). The two speakers showed strikingly different patterns of gesture-to-speech alignment. Further, the alignment patterns shown by Manuputty, but not those of Somad, are indiciative of a system with word-based prosodic prominence, on the basis of the prior studies discussed above. We discuss the findings in detail below.

4.1. Alignment patterns

Somad's beat gestures show a strong tendency to coincide with the final syllable within a word, with approximately 84% of his total gestures following this pattern. In contrast, approximately 77% of Manuputty's gestures coincided with the penultimate syllable, a near mirror image of Somad's pattern, as seen in Figure 5.

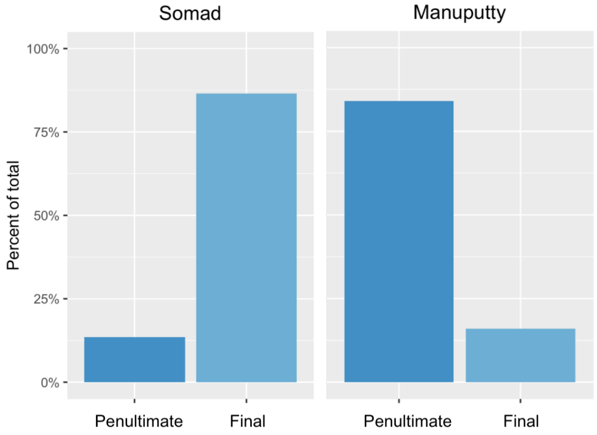


Figure 5: Location of gesture within word

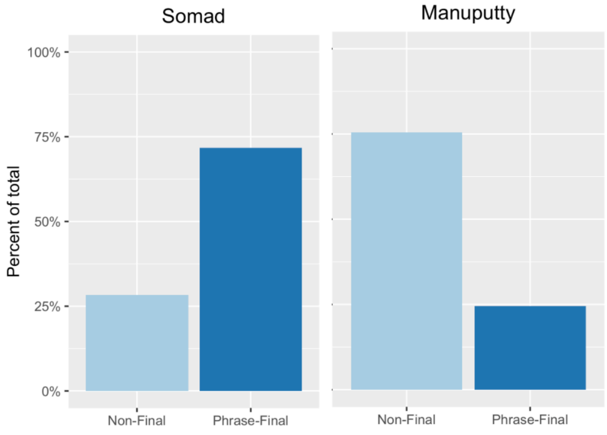


Figure 6: Location of gesture within phrase

In addition to differences in alignment within the word, the effect of phrase position differs significantly for these two speakers. In Somad's case, gestures occurred on phrase-final words over 70% of the time. Again, the pattern is the mirror image for Manuputty; only 25% of gestures occurred on words that were phrase-final, and the majority occurred on phrase-medial words. This is shown in Figure 6.

The fact that phrase-final words are significantly more likely to be a target for a gesture in Somad's production suggests that prominence is crucially linked to phrase edges rather than the word level in his variety of Indonesian (cf. Himmelmann 2010). At the same time, the strong tendency towards realizing this phrasal prominence on the final syllable accords well with Kaufman and Himmelmann's (forthcoming) Western Rim prototype.

Manuputty's data suggest something entirely different; 75% of her gestures occurred in phrase-initial or phrase-medial words. Thus, in contrast to Somad, Manuputty's gestures do not tend to coincide with phrase edges. Instead, gesture appears to track a prominent position within the word, the penultimate syllable, regardless of the word's position within the phrase. The fact that gesture is more common phrase medially than phrase finally may even indicate an avoidance of prominence marking too close to phrase edges, and a separation of boundary marking and prominence marking.[4]

We also examined the proportion of gesture alignment to final syllables and the proportion of alignment to penultimate syllables out of the total for each phrase condition (final and non-final) in order to determine whether the position of the word within the phrase co-varies with the alignment of the gesture within the word. These results are shown in Figure 7. For Somad, there is an overall tendency for final alignment in both phrase final and non-final positions. Importantly, though, this tendency is weaker in non phrase-final words, as can be seen by the smaller difference between the proportion of final and penultimate alignment in the non phrase-final condition. Put differently, phrase-final words are more likely than non-final words to have a gesture aligned to their final syllable, and the difference is significant ($\chi 2$ (1, N = 230) = 17.44, p < .01). This suggests that phrasal prominence, as opposed to word-internal prosodic prominence, may in part be responsible for final syllable alignment. The greater variability in alignment in the non-final condition may again reflect a lack of word based prosody, as was found by Rohrer et al. (2019) for French.

In contrast, Manuputty shows no significant difference in syllable alignment for words in final vs. non-final position. The majority of gestures occur on penultimate syllables and the distribution of word-final gestures and penultimate gestures is almost identical for both phrase conditions. In contrast to the findings for Somad, phrase position and gesture-syllable alignment within the word are independent ($\chi 2$ (1, N = 244) = .29, p > .01). If manual gestures target prominent syllables, Manuputty's overwhelmingly penultimate gesture alignment in non-final conditions suggests that the penultimate syllable of the word in her variety bears prominence and thus functions as an anchor.

[4] One potential basis for the greater coincidence of gesture at phrase boundaries in Somad's speech is that his speech contains more phrase breaks. We examined this possibility and concluded that it is an unlikely explanation. Average pauses per minute were calculated for each speaker to obtain an approximation for the number of phrase breaks per minute. Although this is not a perfect measure - as pauses might signal disfluencies rather than phrase breaks and phrase breaks might not always be accompanied by a pause - this calculation resulted in an average of 26 pauses per minute for Somad and an average of 24 pauses per minute for Manuputty. This difference seems far too small to be the cause of Somad's strong phrase-final tendency in comparison to Manuputty.

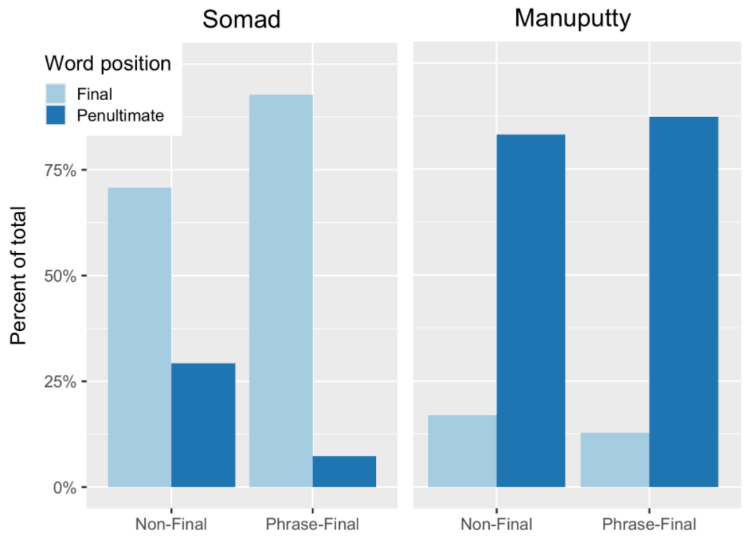


Figure 7: Alignment in the word by phrasal position

5. Alignment of gesture and pitch

Although we have not made an exhaustive analysis of pitch movements in these two recordings, we note that there appears to be a very close relation between manual gesture and pitch prominence, specifically, upwards pitch excursions that would likely indicate either an H or LH pitch accent in autosegmental-metrical terms. The annotation of a representative example, shown in (1), is given in Figure 8. Peak alignments are indicated with a B in the lower tier. The red boxes encapsulate significant upwards pitch excursions, which we take to be the key realization of prosodic prominence on syllables. In the corresponding glossed examples, boldface indicates gesture alignment and underlining indicates the alignment of pitch excursions.

(1)	itu	bər-**ar**ti	bahwa	kua**li**tas	**hi**dup	dan	kəi**ma**nan
	that	AV-meaning	COMP	quality	life	and	faith

‘That means that the quality of life and faith...’

Strikingly, every content word in (1) shows both an upwards pitch excursion and a manual gesture aligned to their penultimate syllables. More typically, manual gestures will coincide with only a subset of the pitch excursions in an utterance, as in (2). There are two other noteworthy points in this example. The clitic in *umumnya* (umum=ɲa) and the suffix in *meletakkan* (mə-lətak-kan) appear to be included in the stress window that determines the penultimate anchor for both prosody and gesture, although this is not entirely regular (see §5.1). Second, compound words like *oraŋ-tua* (person-old) ‘parent’ and *taŋguŋ-ʤawab* (guarantee-answer) ‘responsibility’ receive prosodic prominence on the penultimate syllable of their rightmost member only (which corresponds to the modifier rather than the head, as Indonesian compounds are left-headed).

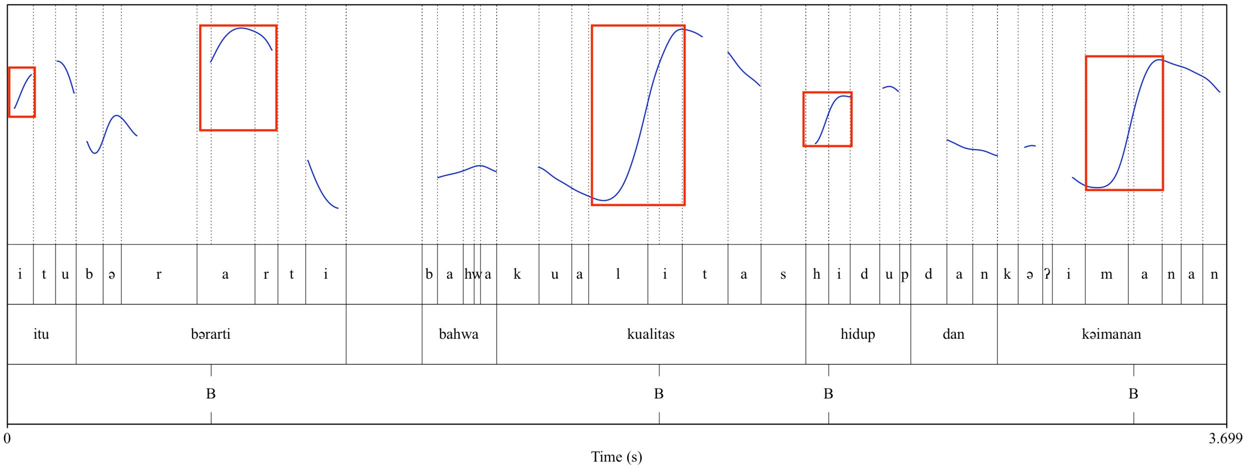


Figure 8: Pitch track and gesture alignment for ex. (1) [Manuputty]

(2) umum=ɲa oraŋ‿tua haɲa mə-lə**ta**k-kan taŋguŋ‿dʒawab mə**ŋa**dʒar dan mən-**di**dik
general=3S.GEN parent only AV-place-APPL responsibility AV-teach and AV-educate
'In general, parents only place the responsibility to teach and educate'

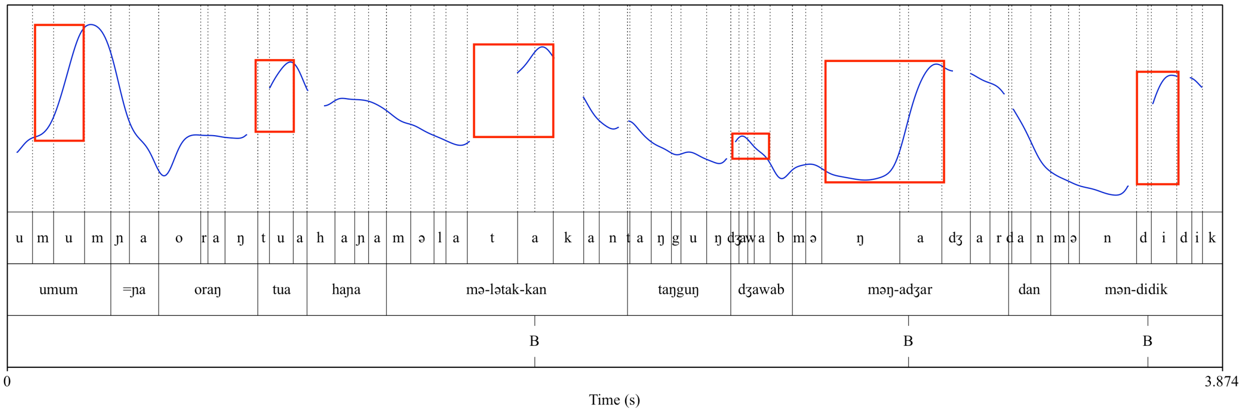


Figure 9: Pitch track and gesture alignment for ex. (2) [Manuputty]

In (3), we see that reduplicated nouns like *bilik-bilik* can show prosodic prominence and gesture alignment on the penultimate syllable of both the base and the reduplicant, providing strong evidence for word-based rather than strictly phrase-based prominence in this variety. Note, furthermore, that here, too, function words such as *dari* 'from', *bahwa* COMP, *dan* 'and', *haɲa* 'only', avoid prominence across modalities.

(3) di-mu**la**i dari **bi**lik-**bi**lik
PV-begin from PL-chamber
'it began from chambers'

As a final exemplification of the predictability of prosodic prominence and systematicity of gesture alignment in Manuputty's speech, we observe many of the same patterns as in (4),

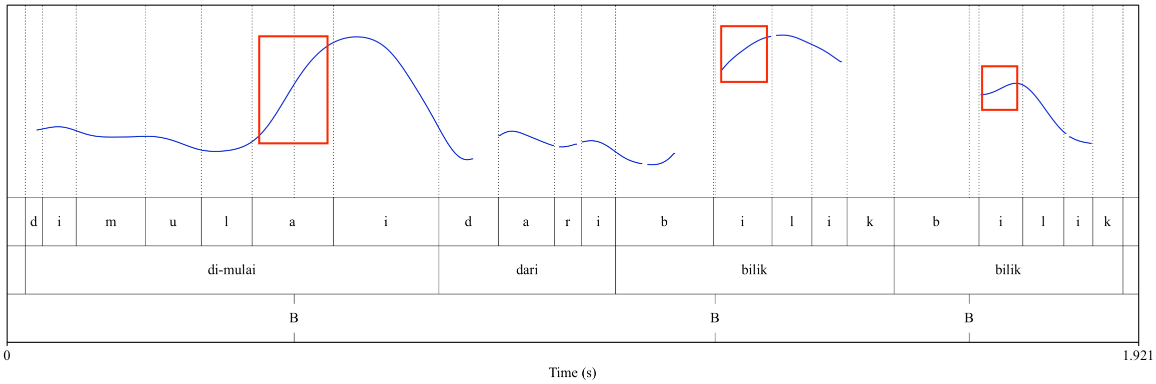


Figure 10: Pitch track and gesture alignment for ex. (3) [Manuputty]

below. Every word receives prosodic prominence on the penultimate syllable excluding *oraŋ*, the first member of a compound, and possibly the two function words *ini* 'this' and *untuk* 'for'. The alignment of beat gestures corresponds perfectly to a subset of these prominent syllables.

(4) ini **so**al bagai**ma**na konsis**ten**si oraŋ_tua untuk **ha**dir
this issue how consistent parent for present
'This is an issue of how consistent the parents are in being present.'

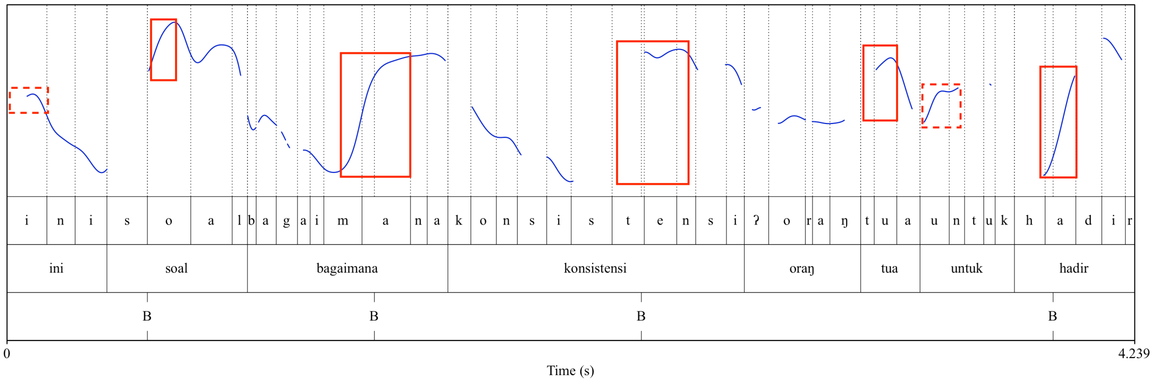


Figure 11: Pitch track and gesture alignment for ex. (4) [Manuputty]

For Somad, there is also a clear pattern of correspondence between manual gesture and prosodic prominence, as can be seen from the representative example in (5) and Figure 12, where arrows highlight the intersection of the gesture peak with the pitch track. Here, we see prominence, both in terms of pitch and duration, falls on the final syllables of each phonological phrase (as demarcated by commas in [5]). Final lengthening is highly salient in Somad's speech and the penultimate syllable, which serves as such a prominent anchor for Manuputty, is often reduced.

(5) reski a**da**, duit a**da**, umur pan**dʒaŋ**, badan se**hat**, tak mau bər-kor**ban**
fortune EXT money EXT age long body healthy NEG want AV-sacrifice
'(One) has fortune, has money, a long life, a healthy body, and doesn't want to sacrifice.'

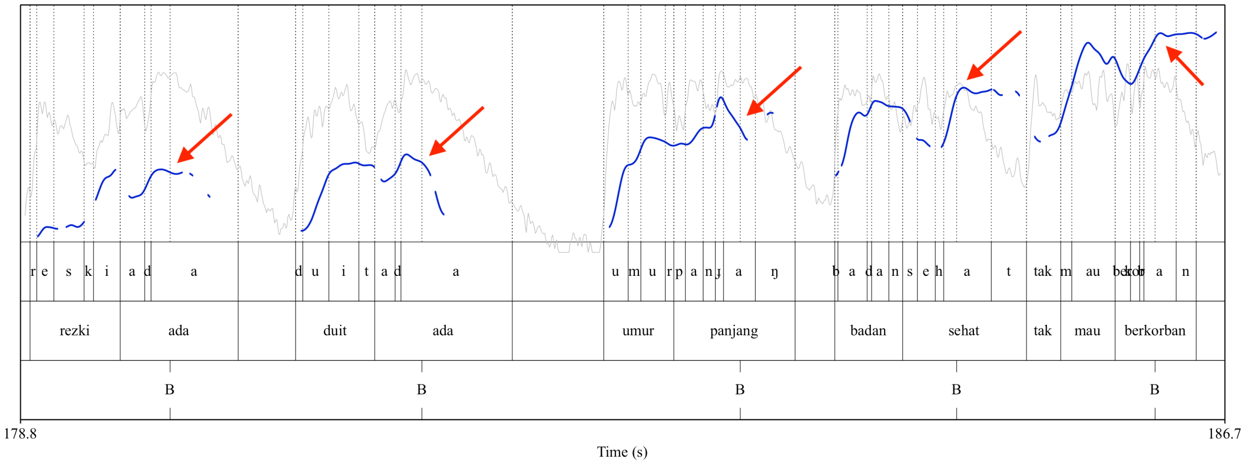


Figure 12: Pitch track and gesture alignment for ex. (5) [Somad]

The coordination between pitch and gesture is further supported by sequences like the one shown in Figure 13, where variable productions of the same phrase, in this case, *makan=lah* (eat=EMPH) ‘Eat!’, shift both prosodic prominence and gestural alignment in lockstep. In this case, the two examples were produced consecutively, with the second being a more emphatic pronunciation of the first (and containing a salient rise-fall pattern on the first syllable).

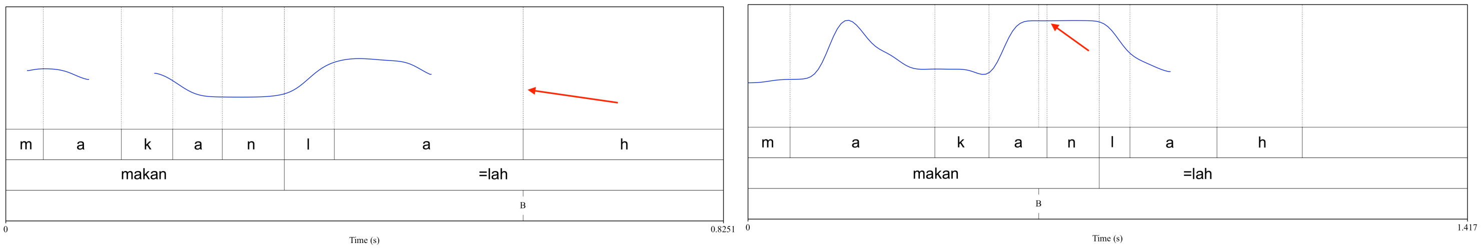


Figure 13: Consecutive productions by Somad of *makan=lah*

Despite Somad’s general alignment of prosodic prominence and gestural peaks to final syllables, especially phrase-final syllables, there are also cases of gesture peaks being aligned to syllables which receive no prosodic prominence, as in the second half of (6), whose pitch track is shown in Figure 14. This appears to be extremely rare in Manuputty’s speech.

(6)	kala**u**	jaŋ	məŋ-aʤi	tu	anak-a**nak**,	ki**ta**	i**ŋat**	anak	**ki**ta	di	ru**mah**
	if	RELT	AV-recite	that	PL-child	1PL	remember	child	1PL	at	home

‘If the ones reciting are children, we remember our child at home’

Another phenomenon that appears common for Somad but vanishingly rare for Manuputty is the alignment of a gestural peak with the right edge of a phrase preceding deaccented material. Deaccentuation, which is realized by a salient lowering and flattening of pitch, is indicated by italics in (7), whose pitch track is shown in Figure 15.

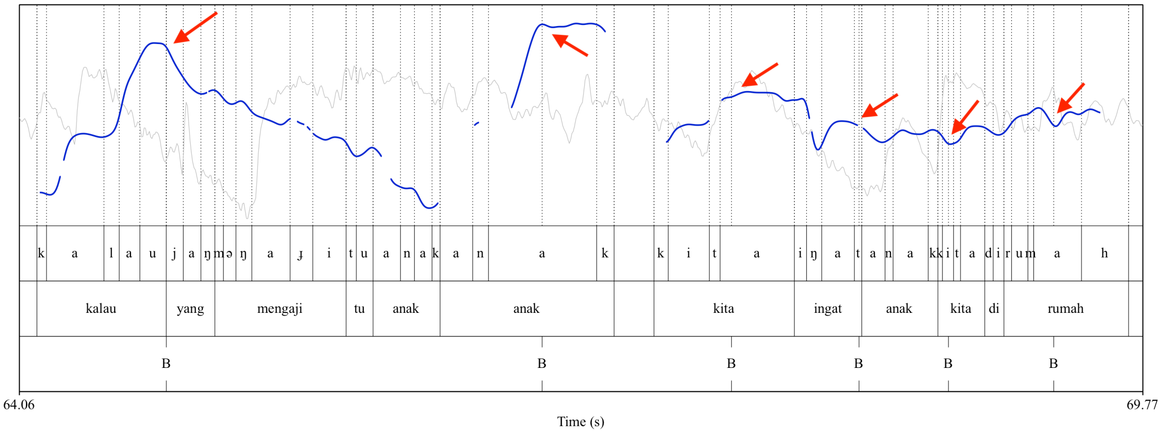


Figure 14: Pitch track and gesture alignment for ex. (6) [Somad]

(7) tuŋgu=**lah** *kə-hancur-an,* han**cur** *leher sapi*
wait=EMPH STA-break-STA break neck cow
'Wait for it to be shattered, the cow's neck is shattered.'

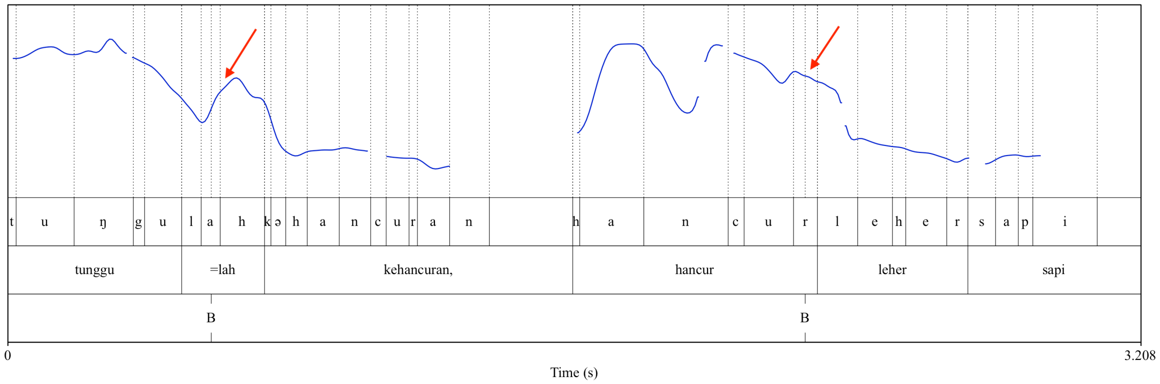


Figure 15: Pitch track and gesture alignment for ex. (7) [Somad]

A similar marking of final phrase edges is seen in (8), where the first gesture is aligned to the end of a clause and the second one is aligned to the right edge of the subject, with both edges corresponding to higher prosodic boundaries.

(8) baca-an=a biasa sa**ʤa**, lipstik=**ɲa** dua inci
read-NMLZ=3S.GEN average only lipstick=3S.GEN two inch
'Her reading is just average, (but) her lipstick is two inches (thick).'

Thus, while there is a systematic pattern to Somad's gestural alignment (i.e. attraction to final syllables and final phrase edges) as well as a general correspondence to prosodic prominence, the match is not nearly as tight as it is for Manuputty. We also do not see regular prosodic prominence on the word level in Somad's speech. Importantly, though, we find an attraction of gesture

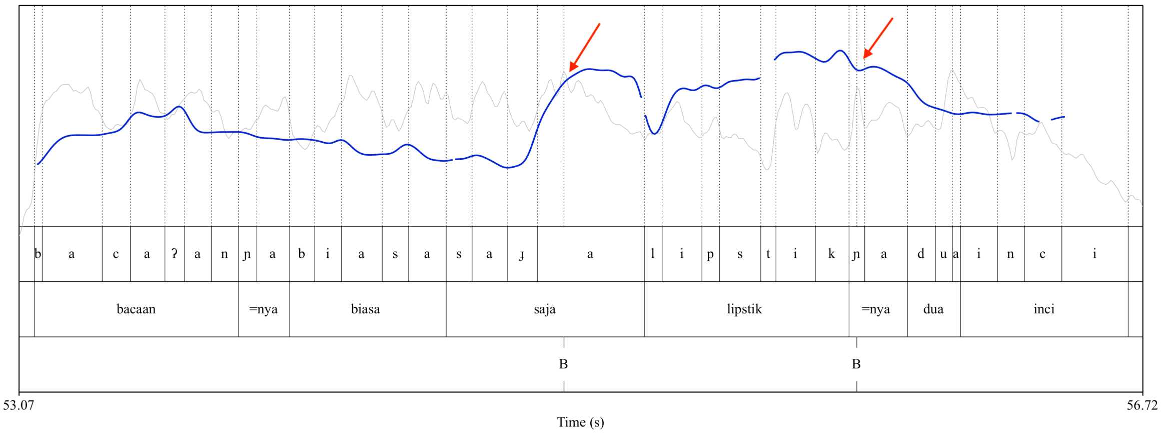


Figure 16: Pitch track and gesture alignment for ex. (8) [Somad]

peaks to final syllables even in the absence of prosodic prominence, something that was not found at all for Manuputty.

5.1. Exceptions

While significant patterns were noted for both Somad and Manuputty, it is also clear that there are numerous exceptions to these patterns. In the following, we take a magnifying class to these exceptions to find possible sub-regularities.

5.1.1. Somad

For Somad, beat gestures were aligned to the final syllable in 86.5% of tokens, and to the penultimate syllable in 13.5% of tokens. These tokens with penultimate gesture alignment can be considered exceptions, as they do not conform to the majority pattern of final alignment.

Potentially mitigating factors that were considered were the presence or absence of a monosyllabic enclitic, the presence or absence of a suffix, and the position of the word within the phrase. While there was a greater likelihood for exceptional penultimate alignment in non-phrase final position, as noted above, the presence or absence of a suffix or monosyllabic enclitic did not correlate at all with location of the gesture within the word. We thus understand penultimate alignment to be part of a natural pattern of variation for Somad.

5.1.2. Manuputty

In words coinciding with a gesture in Manuputty's data set, 84% had a gesture aligned to the penultimate syllable, while the remaining 16% had a gesture aligned to the final syllable. As for the Somad data, an effort was made to identify variables correlating with exceptional final alignment. The factors examined were the same as those for Somad (phrase position and the presence of a monosyllabic enclitic and/or a suffix), but this time also included the presence or absence of a schwa in the penultimate syllable, as it has been widely noted that pentultimate schwas shift prominence to the final syllable in a number of Indonesian languages. If this description is accurate for Manuputty's variety, we would expect the presence of a penultimate schwa to correlate with

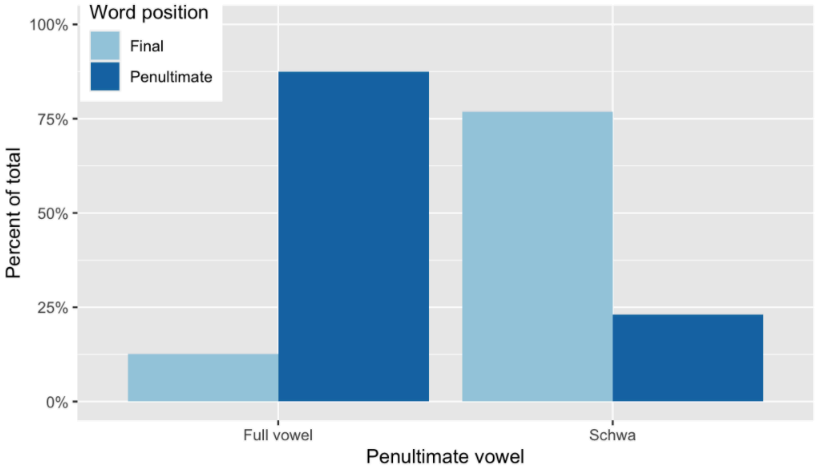


Figure 17: Alignment to schwa vs. full vowel [Manuputty]

gesture alignment to the final syllable. This prediction is indeed borne out, and we find a regular shift of both prosodic prominence and gesture alignment to the final syllable when the penultimate syllable contains a schwa. This is shown in Figure 17, where tokens without a penultimate schwa are represented by the bars on the left, and tokens containing a penultimate schwa by the bars on the right, as a proportion of the total number of tokens for this speaker. The relationship between the position of gesture within the word and presence of penultimate schwa is significant ($\chi 2$ (1, N = 17) = 33.3, p < .01).

An example is shown in Figure 18, where the penultimate schwa of the content word in *pandəmik ini* (pandemic this) causes the pitch and gesture anchor to shift to the final syllable.

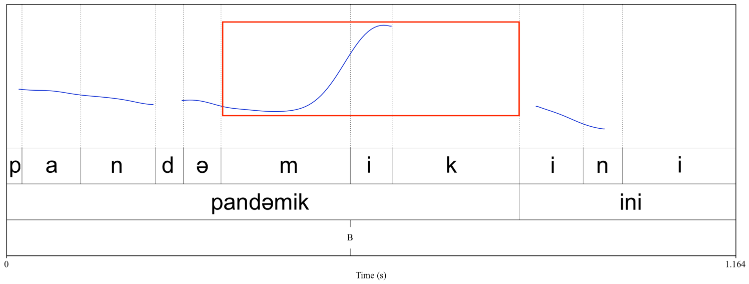


Figure 18: Pitch track and gesture alignment for *pandəmik ini*

However, the presence of a penultimate schwa only accounts for 25% of exceptions. Of the remaining tokens with exceptional final alignment, in 58% of cases the apex of the gesture coincided with the final syllable, but the peak effort coincided with the penultimate syllable. Peak effort is also a commonly used anchor point in gesture studies (see references in §3.2) and these particular cases may thus not constitute true exceptions, as a salient part of the gesture is targeting the penultimate syllable.[5]

Allowing for peak effort (as measured by velocity), as a secondary anchor, we are left with only ∼7% of tokens with unexplained final alignment, compared to 15% of the tokens in Somad

[5] Note that peak effort in manual gesture may also be a better cross-modal fit with the acoustic cue we have used to assess prosodic prominence: maximum pitch rise rather than pitch extrema.

with unexplained penultimate alignment.

Expanding our view of exceptional alignments in Manuputty to those we excluded above, we find a small minority consisting of 21 tokens of pre-penultimate alignment. These are shown in Figure 19, where the exceptions are categorized into three classes.

a.	məm-pro**jik**si-kan	mə-**ru**pa-kan	pən-**di**dik-an	men-ʤa**di**-kan=ɲa
	AV-project-APPL	AV-appear-APPL	NMLZ-educate-NMLZ	AV-happen-APPL=3S.GEN
	kə-**kris**tən-an	mə-**li**hat=ɲa		
	NMLZ-Christian-NMLZ	AV-see=3S.GEN		
b.	ma**ʃa**rakat=ɲa	**fak**ta=ɲa	**po**sitif	ti**mo**teus
	society=3SG.GEN	fact=3S.GEN	positive	Timothy
c.	**mə**-laku-kan	**məŋ**-hadap-i	**bər**-hasil	**kə**-dua
	AV-do-APPL	AV-face-APPL	AV-result	ORD-two

Figure 19: Exceptional alignments in Manuputty

In group (a), the antepenultimate alignment can be accounted for by variable integration of suffixes and clitics into the penultimate stress window. In an example like men-ʤa**di**-kan=ɲa the suffix would be included in the stress window but not the enclitic *=ɲa*. Group (b) is made up of loan words from Arabic and English and the gesture is aligned with the stress-bearing syllable in the source language. The remaining exceptions, shown in group (c), all display word-initial alignment. Given the relatively small number of exceptions it is yet unclear if the word-initial edge is a relevant anchor for prominence. Strikingly, exceptional alignment of manual gestures still match up strongly with prosodic prominence, although we cannot include examples for reason of space.

6. Conclusion

This paper presents a novel methodology using multimodal evidence to investigate prosody in spontaneous speech. It was shown that applying the findings of close temporal alignment of gesture to prosodic prominence for well studied languages can provide new insights into the prosody of Indonesian, a language whose prosodic system has resisted simple categorization. While recent acoustic analyses describe Indonesian as a stressless language, clear differences in gesture alignment were found for speakers of two different regional varieties. These findings shed light on the variation in prosodic systems in Indonesia, and contribute to understanding of the regional prosodic typology.

Specifically, it was found that in the western variety examined, gesture tended to coincide with final syllables, and most often occurred on phrase final words, but also displayed a considerable amount of variation. Variability in gesture to syllable alignment in this variety may indicate that prosodic prominence is phrasal rather than word based, and is not targeting any particular syllable within the word. This would be consistent with the finding in Rohrer et al. (2019) that gesture to syllable alignment was more variable in French, a stressless language with phrasal pitch accents, when compared to stress languages.

In contrast, in the eastern variety, gesture tended to coincide with penultimate syllables regardless of the word's position in the phrase. While instances of final alignment were found here, they were shown to be mostly systematic, and gesture alignment consistently occurred with prosodic prominence even in the exceptional cases. Consistent alignment of gesture to penultimate syllables, regardless of a word's position in the phrase, is indicative of word level penultimate prominence in this variety of Indonesian. This finding is at odds with the results of Maskikit-Essed and Gussenhoven (2016), which concluded that Ambonese Malay lacks word and phrase level prominence.[6] We tentatively suggest that the differences between our findings and theirs are based in methodology rather than variation across idiolect or sub-varieties. Elicited speech is more likely to evince stresslessness while naturalistic genres, especially declamatory speech, are more likely to reveal regular patterns of prosodic prominence.

Although we have argued here that Ambonese Indonesian, and most probably Ambonese Malay, in fact, do not constitute an exception to Kaufman and Himmelmann's (forthcoming) areal typology, we do find one aspect in which both the eastern and western varieties validate Zanten et al.'s (2010) statement, cited earlier: "Even if some varieties of Indonesian will reflect the stress pattern of regional substrates [. . .] Indonesian as a language does not have stress as a linguistic property." Namely, even the eastern variety, with its more regular penultimate pattern, shows significant variation in the integration of suffixes and enclitics. This type of variation seems to present an interesting difference between local varieties of eastern Malay and substrate languages, which often have patterns that strictly include or exclude various sets of suffixes and enclitics.

Importantly, our findings suggest that manual gesture tracks prosodic prominence very closely in languages with predictable stress but less closely in stressless languages. If this pattern can be further corroborated, stresslessness would not be a simple matter of lacking prominence realization over more familiar prosodic structures but would rather indicate a more deep-rooted difference in prosodic structure itself. Studies on other, more clearly stressless Austronesian languages such as Javanese should shed light on this question.

References

Alisjahbana, Sutan Takdir. 1964. *Tatabahasa baru Melayu/Indonesia [new grammar of Malay/Indonesian]*. Kuala Lumpur: Zaman Baru Limited.

Athanasopoulou, Angeliki, Irene Vogel, and Nadya Pincus. 2021. Prosodic prominence in a stressless language: An acoustic investigation of Indonesian. *Journal of Linguistics* 1–41.

Beckman, Mary. 1986. *Stress and non-stress accent*. Dordrecht: Foris Publications.

Chui, Kawai. 2005. Topicality and gesture in Chinese conversational discourse. *Language and Linguistics* 6:635–654.

Cohn, Abigail C. 1989. Stress in Indonesian and bracketing paradoxes. *Natural Language and Linguistic Theory* 7:167–216.

Esposito, A., D. Esposito, M. Refice, M. Savino, and S. Shattuck-Hufnagel. 2007. A preliminary investigation of the relationship between gestures and prosody in Italian. In *Fundamentals of verbal and nonverbal communication and the biometric issue*, ed. Anna Esposito, Maja Bratanić, Eric Keller, and Maria Marinaro. IOS Press.

[6] Note that our study investigated a more formal genre of Indonesian as spoken in Ambon (what could be called Ambonese Indonesian) rather than Ambonese Malay itself. However, the standard language as heard on national radio and television emanates from Java, the heart of the stressless region in the areal typology presented above, and is thus not expected to contribute a particular stress pattern to a stressless substrate. On the contrary, the investigations of the standard Indonesian cited earlier conclude that it is stressless while Ambonese Malay sits squarely within a region dominated by penultimate stress patterns.

Esteve-Gibert, Núria, and Pilar Prieto. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *The Journal of Speech, Language, and Hearing Research* 56:850–864.

Ferré, Gaëlle. 2014. A multimodal approach to markedness in spoken French. *Speech Communication* 57:268–282.

Fung, Holly Sze Ho, and Peggy Pik Ki Mok. 2018. Temporal coordination between focus prosody and pointing gestures in cantonese. *Journal of Phonetics* 71:113–125.

Goedemans, Rob, and Ellen van Zanten. 2014. No stress typology. In *Above and beyond the segments: Experimental linguistics and phonetics*, ed. Johanneke Caspers, Yiya Chen, Willemijn Heeren, Jos Pacilly, Niels O. Schiller, and Ellen van Zanten, 83–95. Amsterdam and Philadelphia: John Benjamins.

Halim, Amran. 1974. *Intonation in relation to syntax in Bahasa Indonesia*. Jakarta: Djambatan.

van Heuven, Vincent J., Lilie M. Roosman, and Ellen van Zanten. 2008. Betawi Malay word prosody. In *Trends in prosodic phonology*, ed. Colin J. Ewen, Nancy C. Kula, and Harry van der Hulst, volume 118–9 of *Lingua*, 1271–1287. Elsevier.

Himmelmann, Nikolaus. 2010. Notes on Waima'a intonational structure. In *East Nusantara: typological and areal analyses*, ed. Michael C. Ewing and Marian Klamer, volume 618, 47–70. Canberra: Pacific Linguistics, ANU.

Himmelmann, Nikolaus P., and Daniel Kaufman. 2020. Austronesia. In *The Oxford handbook of language prosody*, ed. Carlos Gussenhoven and Aoju Chen, chapter 25, 370–383. Oxford University Press.

Im, S., J. Cole, and S. Baumann. 2018. The probabilistic relationship between pitch accents and information status in public speech. In *Proceedings of Speech Prosody 9*, 508–511.

Jannedy, Stefanie, and Norma Mendoza-Denton. 2005. Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure* 199–244.

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

Kaland, Constantijn. 2019. Acoustic correlates of word stress in Papuan Malay. *Journal of Phonetics* 74:55–74.

Kaufman, Daniel, and Nikolaus P. Himmelmann. forthcoming. Suprasegmental phonology. In *The Oxford guide to the western Austronesian languages*, ed. Alexander Adelaar and Antoinette Schapper. Oxford: Oxford University Press.

Kendon, Adam. 1980. Gesture and speech: two aspects of the process of utterance. In *Nonverbal communication and language*, ed. Mary R. Key, 207–227. The Hague: Mouton.

Krahmer, Emiel, and Marc Swerts. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57:396–414.

Leonard, Thomas, and Fred Cummins. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes* 26:1457–1471.

Loehr, Daniel P. 2004. Gesture and intonation. Doctoral Dissertation, Georgetown University.

Loehr, Daniel P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3:71–89.

Maskikit-Essed, Raechel, and Carlos Gussenhoven. 2016. No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay. *Phonology* 33:353–389.

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

Odé, Cecilia, and V.J. van Heuven. 1994. *Experimental studies of Indonesian prosody*, volume 9 of *Semaian*. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Leiden University.

Post, Brechtje. 2000. *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

Pouw, Wim, and James A Dixon. 2019. Quantifying gesture-speech synchrony. In *Proceedings of the 6th Gesture and Speech in Interaction Conference*, 75–80. Universitaetsbibliothek Paderborn.

Rochet-Capellan, Amélie, Coriandre Vilain, Marion Dohen, Rafael Laboissière, and Jean-Luc Schwartz. 2008. Does the number of syllables affect the finger pointing movement in a pointing-naming task. In *8th International Seminar on Speech Production (ISSP 2008)*, 257–260.

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

Roosman, Lilie. 2006. Phonetic experiments on the word and sentence prosody of Betawi Malay and Toba Batak. Doctoral Dissertation, Leiden University, Utrecht.

Roustan, B., and M. Dohen. 2010. Gesture and speech coordination: The influence of the relationship between manual gesture and speech. In *11th Annual Conference of the International Speech Communication Association 2010 (Interspeech 2010).*. Makuhari, Japan.

Samsuri. 1971. *Ciri-ciri prosodi dalam kalimat Bahasa Indonesia [prosodic characteristics in indonesian sentences]*. Flores: Nusa Indah.

Shattuck-Hufnagel, Stefanie, Yelena Yasinnik, Nanette Veilleux, and Margaret Renwick. 2007. A method for studying the time alignment of gestures and prosody in American English: 'hits' and pitch accents in academic-lecture-style speech. In *Fundamentals of verbal and nonverbal communication and the biometric issue*, ed. Anna Esposito, Maja Bratanić, Eric Keller, and Maria Marinaro. IOS Press.

Stoel, Ruben. 2007. The intonation of Manado Malay. In *Prosody in Indonesian languages*, ed. Vincent J. van Heuven and Ellen van Zanten, 117–150. Leiden: Leiden University.

Tadmor, Uri. 2000. Rekonstruksi aksen kata bahasa Melayu (the reconstruction of word accent in Malay). In *Pelbba 13 (pertemuan linguistik (pusat kajian) bahasa dan budaya atma jaya ketiga belas)*, ed. Bambang Kaswanti Purwo and Yassir Nasanius, 153–167. Jakarta: Unika Atma Jaya.

Yasinnik, Yelena, Margaret Renwick, and Stefanie Shattuck-Hufnagel. 2004. The timing of speech-accompanying gestures with respect to prosody. In *Proceedings of the International Conference: From sound to sense*, volume 50, 10–13.

Zanten, Ellen van, Rob Goedemans, and Jos Pacilly. 2003. The status of word stress in Indonesian. In *The phonological spectrum II: Suprasegmental structure*, ed. Jeroen van de Weijer, Vincent J. van Heuven, and Harry van der Hulst, 151–175. Amsterdam/Philadelphia: John Benjamins.

Zanten, Ellen van, Ruben Stoel, and Bert Remijsen. 2010. Stress types in Austronesian languages. In *A survey of word accentual patterns in the languages of the world*, ed. Harry van der Hulst, Rob Goedemans, and Ellen van Zanten, 87–112. Berlin: De Gruyter.

Zuraw, Kie, Kristine M. Yu, and Robyn Orfitelli. 2014. The word-level prosody of Samoan. *Phonology* 31:271–327.