GESTURE AND PROSODIC PROMINENCE IN AMBONESE INDONESIAN

Alessa Farinella[a], Constantijn Kaland[b] & Daniel Kaufman[c]

[a]University of Massachusetts, Amherst, USA; [b]University of Cologne, Germany; [c]Queens College & The Graduate Center, CUNY, USA
afarinella@umass.edu; ckaland@uni-koeln.de; daniel.kaufman@qc.cuny.edu

ABSTRACT

Studies on co-speech gesture have revealed a strong tendency for manual gesture to be aligned to prosodic prominence in many languages. In this study, we investigate co-speech gesture in a variety of Standard Indonesian as spoken in Ambon, an island in the eastern part of Indonesia. Prior studies on both Indonesian and Ambonese Malay, the local variety of Malay spoken in Ambon, disagree as to whether either of these varieties shows evidence of word stress. The current study provides a new perspective through gesture on the way prosody in these languages is organized. While we find some evidence that gesture-aligned syllables are prosodically prominent, this does not appear to always be the case. These findings suggest that Ambonese Indonesian may possess prosodic structure that is not always acoustically marked, and highlights the importance of collecting evidence from a variety of methodologies, including gesture alignment, in the study of word stress.

Keywords: Gesture, prosody, Indonesian, stresslessness, multimodality

1. INTRODUCTION

1.1. Indonesian prosody

In this study, we examine the alignment of manual gesture to speech in Ambonese Indonesian to shed light on the prosody of Indonesian. Previous studies on word level prosody in Indonesian have come to conflicting conclusions about the status of stress in the language. It has been variously claimed that Indonesian is a language with regular penultimate stress that shifts to the ultimate syllable when the penult contains schwa [1], [2] and that Indonesian has regular ultimate stress [3]. More recent studies argue that Indonesian lacks evidence of stress at the word level entirely [4], [5], [6]. However, the area in which Indonesian is spoken is quite vast and linguistically diverse and, as a result, differences between varieties has been reported. A study by [7], for example, found that the prosody of Indonesian is influenced by a speaker's native language.

This study focuses on the variety of Indonesian spoken in Ambon in eastern Indonesia. While work on the prosody of the languages spoken in this area is lacking, studies on word level prosody have been conducted on Ambonese Malay, a regional variety of Malay. These studies also come to contradictory conclusions. In a grammar of Ambonese Malay, [8] describes stress as being lexical and unpredictable, evidenced by the existence of minimal pairs. However, a more recent acoustic investigation failed to find evidence of word level stress in Ambonese Malay [9]. The only other language in this region for which an acoustic investigation of the stress system has been undertaken is Papuan Malay, a regional Malay variety spoken in close geographic proximity to Ambonese Malay. In contrast to the finding for Ambonese Malay, it was concluded that this variety possesses regular penultimate stress, with a small class of words with ultimate stress [10], [11], [12].

Even for closely related varieties spoken in close geographic proximity, then, we might expect to find prosodic differences. However, there are also clear methodological differences between the acoustic study on Ambonese Malay and that on Papuan Malay. While [9] relied on scripted, lab based speech for Ambonese Malay, the acoustic study in [10] used spontaneous narratives and a much larger amount of data. Differences in methodology may also be in part responsible for the different results for Ambonese and Papuan Malay. Against this conflicting background, we use gesture alignment as a window into the prosody of Indonesian as spoken in Ambon.

1.2. Gesture & prosody

Studies on the temporal alignment of gesture to speech have demonstrated the close coordination of manual gesture and prosodic prominence for a growing number of languages [13], [14], [15]. The ma-

jority of these studies have been conducted on languages with word level stress, and find that the apex, defined as the point of maximal extension of the gesture, tends to align to pitch accents.

However, non-stress and tonal languages appear to differ from canonical stress languages in their patterns of gesture alignment. In French, for which pitch accents have been analyzed as serving a demarcative rather than prominence-marking function, gesture tends broadly to align to pitch accents but less consistently than in stress languages [16]. In Medumba, a tonal Grassfields Bantu language, it has been shown that gesture does not preferentially align with high toned syllables [17]. As high tones mark lexical contrasts in this language, this possibly supports the idea that gesture targets linguistic prominence and not just acoustic salience or pitch accents per se. However, the same study finds that gesture-aligned syllables are longer and louder than in non-aligned syllables, showing that there remains a correspondence between manual and acoustic prominence even though gestures are not anchored to prosodically determined prominent/stressed syllables.

Most relevant for the present work, a study by [18] on gesture alignment in two varieties of Indonesian found that for the variety spoken in eastern Indonesia, gesture tended to align to the penultimate syllable, while for the variety spoken in western Indonesia, gesture tended to align to the ultimate syllable. While it's not yet entirely clear how gesture aligns to speech in this variety, one of the aims of the present work is to uncover patterns in gesture alignment (if any) in this variety. Results of prior studies support the idea that gesture alignment patterns may reflect prosodic properties.

1.3. Research questions

As is apparent from the overview in Section 1.1, there is little agreement as to the status of word stress in Indonesian, and evidence that varieties may differ in their prosody. What is clear is that evidence for word stress should be sought from a variety of sources. Here, we employ a novel methodology of using manual gesture as a unique source of evidence for word level prosody in Indonesian. In light of the findings discussed above that gesture is tightly coordinated with prosodic prominence across a number of languages, the aim of the present study is to use gesture as a window into the word prosody of a variety of Indonesian. In doing so, we ask the following questions: i) is there a systematic tendency to align gesture to a particular syllable in the word and ii) do gesture-aligned syllables also exhibit prosodic prominence?

There are several possibilities we might expect for Ambonese Indonesian:

1. If gesture does not target a particular position in the prosodic word, it would be consistent with the claim that Ambonese Indonesian lacks a word level prominence distinction, on the basis of the findings in [16] for French.
2. If gesture targets a consistent position in the prosodic word *and* the syllable with gesture exhibits prosodic prominence, it would be consistent with the claim that Ambonese Indonesian has word level prosodic prominence.
3. If gesture targets a consistent position in the prosodic word and the gesture-aligned syllable does *not* exhibit prosodic prominence, it would be consistent with the claim that Ambonese Indonesian possesses word prosodic structure that lacks consistent acoustic correlates.

2. METHODOLOGY

2.1. Data and annotation

The data come from video recordings of two preachers (S1, female and S2, male) from Ambon, in eastern Indonesia, who are both speaking Standard Indonesian. In the video excerpts, both are speaking in a spontaneous style.

The apex, or target, of the gesture was annotated in ELAN [19] without audio. While many studies on gesture and prosody include only beat gestures, which are by definition non-referential, no effort was made here to exclude referential gestures, as they have also been shown to be coordinated in time with prosodic prominence [20]. However, only gestures with clear starting and stopping points were included in this study.

The recordings were transcribed in Praat [21] by a native speaker and segmented by syllable on a separate tier by a coder with knowledge of the syllable structure of Indonesian. Words were marked as phrase final or non-final impressionistically using common crosslinguistic correlates to boundaries (pauses, final lengthening, pitch reset). We then confirmed that mean duration by phoneme count was significantly greater for phrase final words than non-final words. Words whose status as final or non-final was unclear were excluded, as were monosyllables.

2.2. Acoustic measures

Duration per phoneme was measured for all syllables in the data set by taking the total syllable dura-

Table 1: Number of syllables aligned to gesture by word position (Position column) and phrase position (Phrase column).

Spkr	Position	Phrase	Frequency
S1	Penult	Medial	115
		Final	41
	Ultimate	Medial	29
		Final	5
S2	Penult	Medial	41
		Final	26
	Ultimate	Medial	14
		Final	10

tion divided by the number of phonemes in the syllable (similar to [10]). This way of measuring duration accounts to some extent for differences in syllable structure. Vowel quality was also measured by taking the frequencies of the first and second formants converted to Bark scale. These acoustic measures were chosen as they were found to be the strongest correlates of word stress in [10] for Papuan Malay.

3. RESULTS

3.1. Gesture alignment

Only penultimate and ultimate syllables were included in the analysis, as prior studies on Indonesian have claimed that either the penult or ultimate syllable (or neither) is prosodically prominent. In total, 521 syllables for S1, 191 of which were aligned to gesture, and 387 syllables for S2, 94 of which were aligned to gesture, were annotated.

Table 1 shows the frequency of gesture aligned to penultimate and ultimate syllables by phrase position. For both speakers, there is a clear tendency to align gesture to penultimate syllables in both phrase final and non phrase final position, though this tendency is slightly stronger for S1 than for S2. A Chi-square test revealed a significant relationship between presence of gesture and syllable position for S1 ($\chi^2 = 127.4, df = 1, p < .001$) and S2 ($\chi^2 = 25.7, df = 1, p < .001$). The fact that alignment is unaffected by phrase position suggests that gesture is targeting a particular position in the word.

3.2. Acoustics

3.2.1. Duration

Our second research question concerns how well gesture lines up with prosodic prominence, as measured by the acoustic correlates discussed in Section 2.2. We separated words that were phrase final from those that were in medial position, as words in final position are expected to be affected by final lengthening. Looking first at the words in phrase medial position, we find that penultimate and ultimate syllables aligned with gesture tend to be longer than those that are unaligned. This difference is larger for S1 than S2.

Table 2: Duration per phoneme (ms)

Spkr	Position	Phrase medial		Phrase final	
		Gest	No gest	Gest	No gest
S1	Penult	104	81	114	82
	Ult	88	66	119	103
S2	Penult	95	83	94	103
	Ult	95	80	174	144

Turning to phrase final words, for S1, gesture-aligned syllables are longer than those that are unaligned, and this difference is smaller for ultimate syllables in phrase final words than non final words. This is presumably due to the effect of final lengthening, which is expected to affect ultimate syllables to a greater degree than penultimate syllables. Support for this comes from the fact that, unlike in phrase medial position, ultimate syllables are on average longer than penultimate syllables.

For S2, phrase final words display a different pattern. While ultimate syllables aligned with gesture are longer than those that are not aligned, penultimate syllables aligned with gesture are in fact *shorter* than those that are not aligned. This is distinct from what was observed for S1, and on the face of it is quite puzzling. We interpret this finding as indicative that the penultimate position is prominent (and therefore targeted by gesture) for this speaker even in the absence of prosodic prominence as signaled by duration. We return to a discussion of these results in Section 4.

3.2.2. Vowel quality

Values for F1 and F2 were converted into Bark and compared for vowels in syllables that are aligned with gesture and vowels in syllables that are not aligned with gesture. Crosslinguistically, there is a tendency for vowels in stressed syllables to be more peripheral than vowels in unstressed syllables.

The results are shown in Figures 1 and 2. Small font letters are vowels aligned to gesture, large font are unaligned vowels. For both speakers, it is indeed the case that vowels in gesture-aligned syllables tend to be more peripheral than vowels in syllables that are not aligned with gesture.

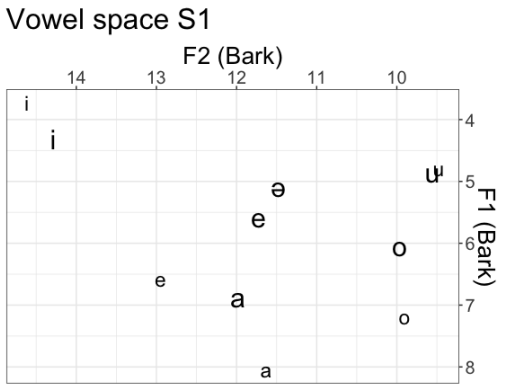

Figure 1: Vowel space for S1. Small font = vowels in syllables aligned to gesture, large font = vowels in unaligned syllables.

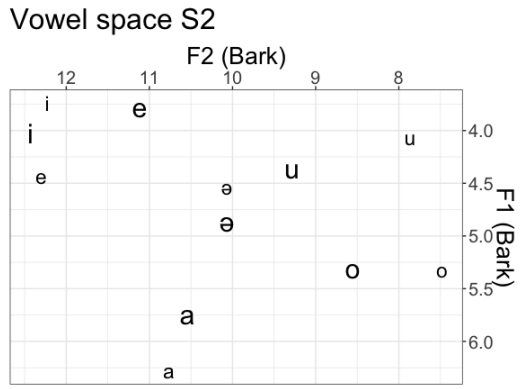

Figure 2: Vowel space for S2. Small font = vowels in syllables aligned to gesture, large font = vowels in unaligned syllables.

4. DISCUSSION

The two speakers in this study showed a strong tendency to align gesture to penultimate syllables in both phrase final and non-final words. Gesture-aligned syllables tended to show prosodic prominence through greater duration compared to syllables not aligned to gesture, and more peripheral vowels. The fact that gesture-aligned syllables tend to exhibit some degree of prosodic prominence *and* that gesture tends to align to penultimate syllables indicates that penultimate syllables are prominent in Ambonese Indonesian.

However for S2, penultimate syllables do not always show evidence of prosodic prominence. For this speaker, it's clear that penultimate syllables are targeted for gesture even in the absence of durational prominence in phrase final position. It's possible that final lengthening on final syllables is an important cue to a boundary, and therefore prevents lengthening in penultimate syllables. However, gesture still targets penultimate syllables at a greater rate than ultimate syllables in phrase final position even in the absence of prosodic prominence.

Of the three possible outcomes in Section 1.3, gesture consistently targets the penultimate syllable, and gesture aligned syllables tend to be longer and have more peripheral vowels than syllables not aligned with gesture. These findings are in line with what has been found for languages with word level stress (and our prediction 2). However, the fact that the syllable aligned to gesture is not always prosodically prominent (for S2) is in line with an explanation of abstract word level prominence (and with our prediction 3). These findings are what would be expected if there's regular penultimate word level prominence that may not be consistently marked acoustically, precisely what was found by [10] for Papuan Malay. This raises the possibility that prosodic structure in this language may differ in its acoustic marking from European languages.

This study leaves open several questions for future investigation. One of these relates to the direction of the relationship between gesture and prosody. While many studies have demonstrated that gesture seeks a candidate that is already prosodically prominent, is it also possible that alignment with gesture triggers acoustic prominence on syllables already bearing abstract prominence. A study by [22] on Dutch found that gesture can enhance acoustic prominence in much the same way as pitch accents. It's possible that gesture selects candidates that bear abstract prominence (which the results of this study would suggest are penultimate syllables) and makes them more acoustically prominent, except perhaps where that would interfere with cues to phrasing (as discussed in the results for S2).

Our findings should not be taken as conclusive evidence that Indonesian as spoken in Ambon has word level stress. However, they are consistent with (mostly) penultimate word level stress, and surprising if this variety has no word level prosodic structure. This is unexpected in light of the study by [9] on Ambonese Malay, which claimed that that variety lacked evidence of stress. While this study looked at Standard Indonesian, it's not clear where prominence distinctions in Ambonese accented Indonesian would originate from if both Standard Indonesian and Ambonese Malay are described as stressless. As it's been shown that regional varieties/languages may influence the word level prosody of Indonesian [7], it's likely that the patterns observed here originate from Ambonese Malay and other local languages. Work on prosody in these varieties should be investigated with a variety of different methodologies, of which we report on just one.

5. ACKNOWLEDGMENTS

This work is funded by the Deutsche Forschungsgemeinschaft (DFG), Project-ID 281511265, CRC 1252 "Prominence in Language."

6. REFERENCES

[1] A. C. Cohn, "Stress in Indonesian and bracketing paradoxes," *Natural Language and Linguistic Theory*, vol. 7, no. 2, pp. 167–216, May 1989.

[2] R. Stoel, "The intonation of Manado Malay," *LOT Occasional Series*, vol. 9, pp. 117–150, 2007.

[3] Samsuri, *Ciri-ciri prosodi dalam kalimat Bahasa Indonesia [Prosodic features in the Indonesian sentence]*, Flores: Nusa Indah, 1971.

[4] C. Odé, V. van Heuven, and E. van Zanten, *Experimental studies of Indonesian prosody*. Vakgroep Talen en Culturen van Zuidoost-Azie en Oceanie, Rijksuniversiteit, 1994, vol. 9.

[5] R. W. Goedemans and E. van Zanten, "No stress typology." Benjamins Amsterdam, 2014, pp. 83–95.

[6] A. Athanasopoulou, I. Vogel, and N. Pincus, "Prosodic prominence in a stressless language: An acoustic investigation of Indonesian," *Journal of Linguistics*, vol. 57, no. 4, pp. 695–735, Nov. 2021.

[7] R. Goedemans and E. van Zanten, "Chapter three Stress and accent in Indonesian," *LOT Occasional series*, vol. 9, pp. 35–62, 2007.

[8] D. Van Minde, *Malayu Ambong: phonology, morphology, syntax*. Leiden: Research School CNWS, 1997.

[9] R. Maskikit-Essed and C. Gussenhoven, "No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay," *Phonology*, vol. 33, no. 2, pp. 353–389, Aug. 2016.

[10] C. Kaland, "Acoustic correlates of word stress in Papuan Malay," *Journal of Phonetics*, vol. 74, pp. 55–74, May 2019.

[11] ——, "Offline and online processing of acoustic cues to word stress in Papuan Malay," *The Journal of the Acoustical Society of America*, vol. 147, no. 2, pp. 731–747, Feb. 2020.

[12] C. Kaland, A. Kluge, and V. J. van Heuven, "Lexical analyses of the function and phonology of Papuan Malay word stress," *Phonetica*, vol. 78, no. 2, pp. 141–168, Apr. 2021.

[13] D. P. Loehr, "Temporal, structural, and pragmatic synchrony between intonation and gesture," *Laboratory Phonology*, vol. 3, no. 1, Jan. 2012.

[14] S. Jannedy and N. Mendoza-Denton, "Structuring information through gesture and intonation," *Interdisciplinary studies on information structure: ISIS; working papers of the SFB 632*, no. 3, pp. 199–244, 2005.

[15] S. Shattuck-Hufnagel and A. Ren, "The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech," *Frontiers in Psychology*, vol. 9, p. 1514, Sep. 2018.

[16] P. L. Rohrer, P. Prieto, and E. Delais-Roussarie, "Beat gestures and prosodic domain marking in french," in *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, VIC, eds S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Canberra, ACT: Australasian Speech Science and Technology Association Inc)*, 2019, pp. 1500–1504.

[17] K. Franich and H. Keupdjio, "The influence of tone on the alignment of speech and co-speech gesture," in *Proceedings of Speech Prosody*, 2022, pp. 2022–63.

[18] D. Kaufman and A. Farinella, "Gesture alignment in a "stressless" language," in *Proceedings of the Twenty-Eighth Meeting of the Austronesian Formal Linguistics Association (AFLA)*, 2022, pp. 29–46.

[19] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "Elan: A professional framework for multimodality research," in *5th international conference on language resources and evaluation (LREC 2006)*, 2006, pp. 1556–1559.

[20] N. Esteve-Gibert and P. Prieto, "Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 3, pp. 850–864, Jun. 2013.

[21] P. Boersma, "Praat: doing phonetics by computer [computer program]," *http://www. praat. org/*, 2011.

[22] E. Krahmer and M. Swerts, "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception," *Journal of Memory and Language*, vol. 57, no. 3, pp. 396–414, Oct. 2007.