



## Research Article

## Acoustic correlates of word stress in Papuan Malay

Constantijn Kaland

Institute of Linguistics, University of Cologne, Germany



## ARTICLE INFO

## Article history:

Received 26 April 2018

Received in revised form 12 February 2019

Accepted 20 February 2019

## Keywords:

Word stress

Acoustic correlates

Papuan Malay

Prosody

## ABSTRACT

The current study investigates acoustic correlates of word stress in Papuan Malay. This language is claimed to exhibit word stress, although empirical evidence is lacking. Moreover, related languages have been shown not to have stress, counter to earlier claims. Studies on stress in Austronesian languages have furthermore struggled to separate phrase level intonation phenomena from word level stress. The current study investigates a set of twelve potential acoustic correlates of stress, covering spectral, temporal and amplitudinal aspects of the speech signal. The measurements are taken from spontaneous Papuan Malay narratives. A subsequent comprehensive acoustic analysis was carried out and designed to avoid common pitfalls in the study of word stress. Results indicate that measures of duration, formant displacement and spectral tilt provide consistent evidence for the production of word stress in Papuan Malay.

© 2019 Elsevier Ltd. All rights reserved.

## 1. Introduction

Papuan Malay is considered a Trade Malay variety, spoken in the Indonesian provinces Papua and Papua Barat. Other Trade Malay varieties include Ambonic (Ambonese, Banda), Kupang, Larantuka and Manadoic (Manado, North Moluccan), see [Paauw \(2008\)](#) for a comparative overview. The relatively little research that has been carried out on these languages consists mostly of descriptive grammars, with little to no attention paid to prosody. Most of the Trade Malay varieties are claimed to have word level stress. In some varieties stress occurs at variable locations and might be phonemic (i.e. the sole distinction between two otherwise identical words), while in others the stress location is rather fixed and non-phonemic. In the current study, both types are henceforth referred to as (word) stress; i.e. the single acoustically most prominent syllable in a word. The goal of the present study is to investigate whether there is acoustic evidence for word stress patterns in Papuan Malay. The communicative function of potential word stress patterns is beyond the scope of the current study, although some discussion is provided in the final section.

As the majority of the stress claims for Trade Malay varieties have remained without empirical support, it has been questioned to what extent prosodic descriptions were influenced

by the (mostly Western) background of the authors ([Himmelman, 2018](#)). In fact, recent work on Ambonese Malay did not find acoustic evidence for word stress, counter to earlier claims ([Maskikit-Essed & Gussenhoven, 2016](#)). Studies on varieties of Indonesian indicated that stress, if present at all, has a different status in the prosody compared to well-studied languages such as English. The current study on word stress correlates in Papuan Malay provides a necessary empirical investigation. In addition, it sheds light on an under-researched language with potential implications for the study of prosody in related languages.

The current study investigates to what extent acoustic correlates support earlier claims on the existence of word stress in Papuan Malay ([Kluge, 2017](#)). Crucially, acoustic measures from all aspects of the speech signal are taken into account; broadly categorised as either spectral, temporal or amplitudinal. Spectral measures are defined as correlates that relate to frequency aspects of the signal, such as pitch and formants. Temporal aspects relate to the duration of speech. Amplitudinal measures in this study represent the acoustic power in (parts of) the speech signal, such as intensity and spectral tilt. Furthermore, possible interference of phrase intonation phenomena is avoided in the current study. This is done both by focusing on a specific subset of the available data as well as by the calculation of relative acoustic measures. In this way, pitfalls of earlier work are avoided and the results constitute

E-mail address: [ckaland@uni-koeln.de](mailto:ckaland@uni-koeln.de)

more robust evidence concerning the existence of word stress in Papuan Malay.

In the next (sub)sections an overview is given of word stress claims in all Trade Malay varieties (Section 1.1). Thereafter, studies on word stress in Indonesian, a closely related language, are discussed (Section 1.2). Furthermore, an overview of common acoustic correlates found crosslinguistically is given in Section 1.3 and the methodological concerns in previous work on word stress are listed in Section 1.4. Finally, Section 1.5 states the research questions.

### 1.1. Stress in Trade Malay

In this section, all Trade Malay varieties distinguished in Paauw (2008) are discussed to the extent that word stress claims have been made. Phrase level prosody is discussed to a limited extent, focusing mainly on its relation to word level phenomena. Table 1 gives a summary of the aspects of the literature on Trade Malay, which are most relevant to the current study. These include the position of word stress, whether it is phonemic, whether acoustic measures were carried out, whether the collected samples consisted of free speech (e.g. spontaneous unscripted) or constrained speech (due to scripting or prompting) and the context in which the samples appeared (isolation, phrase-medial, phrase-final or uncontrolled)

#### 1.1.1. Papuan Malay

Kluge (2017) carried out extensive fieldwork on 44 speakers, predominantly from the Sarmi region in the Papua province. Observations based on a list of 1116 words consisting of Papuan Malay roots revealed that word stress is located on the penultimate syllable in 90% of the cases. The remaining 10% show stress on the ultimate syllable. Many of the cases in which stress is located on the ultimate syllable, the penultimate syllable contains the vowel /ɛ/. It cannot be concluded that /ɛ/ rejects word stress, as 7% of the words with penultimate stress contains /ɛ/ (Kluge, 2017, p. 96). The other reported Papuan Malay vowels are /i/, /a/, /ɔ/ and /u/. Although the observations in Kluge (2017) are mainly based on the author's auditory impressions, a preliminary acoustic analysis of spectral tilt as correlate of word stress confirmed the stress claims to a large extent (Kaland, 2018). Furthermore, no minimal stress pairs are available in the word lists in Kluge (2017). It deserves to be noted that near-minimal stress pairs tend to consist of /ɛ/ in the penultimate syllable, for example: ['bɛ.bas] 'be free' and [bɛ.'ban] 'burden', ['bɛŋ.kək] 'be crooked' and [bɛŋ.'kak]

'be swollen', ['ɛ.nak] 'be pleasant' and [ɛ.'nam] 'six', and ['mɛ.maŋ] 'indeed' and [mɛ.'naŋ] 'to win'. Also in older and smaller descriptions of Papuan Malay exist, word stress has been undisputedly assumed to occur regularly on the penultimate syllable (e.g. Donohue & Sawaki, 2007). As for phrase prosody, the few studies available suggest that marking of information structure using pitch movements (i.e. pitch accents) is limited to phrase boundaries in Papuan Malay (Kaland & Himmelmann, 2019; Riesberg, Kalbertodt, Baumann, & Himmelmann, 2018).

#### 1.1.2. Ambonese Malay

In a description of Ambonese Malay, Van Minde (1997) assumed irregular word stress. No indications are given about the (default) location of word stress and, crucially, no report on the existence of schwa is provided (Paauw, 2008). The main reason for irregularity is the phonemic nature of stress; hence the minimal pair /ba.rat/ 'west' and /ba.'rat/ 'heavy'. However, a re-evaluation and acoustic measurements were carried out on read speech samples from four speakers (Maskikit-Essed & Gussenhoven, 2016). The samples were comparable to the ones in Van Minde (1997) and led to the conclusion that word stress does not exist in Ambonese Malay. The vowel /a/ that occurs in syllables described as unstressed by Van Minde (1997) was seen as a different phoneme (*a-caduc*) in Maskikit-Essed and Gussenhoven (2016) on the basis of its distinct spectral characteristics. In addition to the reanalysis of the phoneme inventory, Maskikit-Essed and Gussenhoven (2016) did not find reliable acoustic evidence for stress in duration, F0 peak alignment and spectral tilt for the examples provided in Van Minde (1997). Furthermore, no support was found for the use of pitch accents in this language.

#### 1.1.3. Kupang Malay

Limited work has been carried out on Kupang Malay. Claims on stress were made by Steinhauer (1983) on the basis of one speaker (the author's wife) and no acoustic investigation. It was assumed that word stress falls on the penultimate syllable unless it contains schwa or a central, mid-low and unrounded vowel. However, schwa did not appear in the vowel inventory in a Kupang Malay dictionary by Jacob and Grimes (2003). It has not been investigated what this means for the analysis of word stress placement in Kupang.

#### 1.1.4. Larantuka Malay

Word stress in Larantuka Malay was described in Kumanireng (1993) and cited in Paauw (2008). Historically,

Table 1

Overview of stress claims for native lexical roots in studies on Trade Malay varieties. For varieties not listed no stress claims have been made. Abbreviations: Penult/Ultimate/Irregular word stress; Free/Constrained speech samples in Isolated/Medial/Final/Uncontrolled (phrase) contexts; n.a. = information not available or not reported.

Variety	Source(s)	Word stress position	Phonemic	Acoustic measures	Sample		Context			
					F	C	I	M	F	U
Papuan	Donohue and Sawaki (2007)	P, alt. U	n.a.	No	x	x	x			x
	Kluge (2017)	P, alt. U	No	No	x	x	x	x	x	x
Ambonese	Van Minde (1997)	I	Yes	No		x	n.a.			
	Maskikit-Essed and Gussenhoven (2016)	-	-	Yes		x		x	x	
Kupang	Steinhauer (1983)	P, alt. U	Yes	No	n.a.		n.a.			
Larantuka	Kumanireng (1993)/Paauw (2008)	P	n.a.	No	n.a.		n.a.			
Manado	Stoel (2007)	P, alt. U	Yes	No	x	x		x	x	x
Tidore	Van Staden (2000)	P, alt. U	No	No		x				x
Ternate	Litamahuputty (2012)	P, alt. U	Yes	No	x					x

Larantuka did not lose schwa and therefore developed regular rather than phonemic word stress (Paauw, 2008). The default location, like in other Trade Malay varieties, is the penultimate syllable. The literature does not report acoustic analyses of word stress in Larantuka Malay.

#### 1.1.5. Manado Malay

Stoel (2007) analysed the intonation of Manado Malay and described word stress as being located regularly on the penultimate syllable. In addition, some words have stress on the ultimate syllable and a limited amount of minimal pairs indicate the existence of phonemic stress. However, it has to be noted that the minimal pairs provided concern proper names and loanwords (Stoel, 2007; p. 118), which might not reliably reflect native stress patterns. The data was obtained in elicitation tasks and further analysed for intonation. Stoel (2007) provided examples of the prosodic marking of focus in Manado Malay. In this analysis, pitch accents were aligned with stressed syllables. These examples were illustrated with pitch contours, whereas for word stress proper, no acoustic evidence was provided.

#### 1.1.6. Tidore (North Moluccan Malay)

A grammar on Tidore described stress placement as regularly occurring on the penultimate syllable (Van Staden, 2000). In some cases, stress occurred on the final syllable if this syllable was heavy (CVC or CVV). Because no minimal stress pairs were provided, word stress was assumed not to be phonemic. The data was taken from read texts recorded during extensive fieldwork.

#### 1.1.7. Ternate (North Moluccan Malay)

A corpus of short stories and anecdotes was collected for Ternate and extensively described in a grammar (Litamahuputty, 2012). Stress was claimed to fall on the penultimate syllable regularly. Stress occurred on the ultimate syllable when the penultimate syllable would have consisted of schwa in other Malay varieties. Ternate was assumed not to have schwa in its vowel inventory and therefore to have developed phonemic stress. No acoustic analyses were provided for word stress in Ternate.

#### 1.1.8. Summary Trade Malay

The general claim made in descriptive work on Trade Malay varieties is that word stress occurs on the penultimate syllable unless that syllable contains schwa or a vowel that can be historically linked to schwa. Crucially, the only acoustic analysis available showed no support for the existence of word stress or pitch accents in Ambonese Malay (Maskikit-Essed & Gussenhoven, 2016). These results contrast with the analysis of Manado Malay in which stress and pitch accents could co-occur (Stoel, 2007). In sum, although stress claims are highly similar for all Trade Malay varieties, at least two issues remain unresolved. First, it is by no means granted that impressionistic stress claims find acoustic support. Second, if stress indeed does exist, it cannot be assumed that all Trade Malay varieties have similar stress placement. Thus, existing stress claims need to be complemented with in-depth acoustic investigations of the Trade Malay languages. The discrepancies between the outcomes of descriptive work and those of experimental inves-

tigations are already reported in the literature on word stress in Indonesian, which has been studied more extensively. The current state of the work for Indonesian therefore illustrates the potential added value of (more) empirical investigations of word prosody in otherwise under-researched languages. For this reason, and because Indonesian has a considerable influence on Trade Malay varieties (Paauw, 2008) a summary of work on word stress in Indonesian is given in the next section.

### 1.2. Stress in Indonesian

Various studies have investigated to what extent word stress exists in Indonesian. Early work claimed that stress occurs regularly on the penultimate and only moves to the ultimate syllable when the penultimate contains schwa (Alieva, Arakin, Ogioblin, & Sirk, 1991; Teeuw, 1984). In other work, the ultimate syllable has been claimed the default location for word stress (Samsuri, 1971). An overview of mainly impressionistic work on stress in Indonesian is provided in Odé (1994, p. 41), followed by a perception experiment on phrasal prominence. The remainder of this subsection, however, is devoted to a discussion of experimental work on Indonesian word stress.

Elaborate acoustic investigations on Indonesian were carried out by Halim (1981) on spontaneous speech from 13 speakers. Analyses indicated that pitch peaks and syllable duration are more important correlates of word stress than intensity peaks. Overall, all acoustic correlates showed higher values for the penultimate syllable than for the ultimate syllable.

Pitch was also found to be the main word stress correlate by Laksman (1994), occurring predominantly on the penultimate syllable. Data consisted of read speech from one speaker from Jakarta, assumed to be representative of standard Indonesian with little to no regional influences. The acoustic analysis furthermore showed that syllables containing schwa could be stressed as much as other vowels, counter to earlier claims.

In a perception experiment, Goedemans and Van Zanten (2007) examined two Indonesian languages: Toba Batak and Javanese. One speaker per language read carrier phrases including four syllable target words. For both languages, pitch movements at the right phrase edge were observed and analysed as part of phrase level prosody rather than of word level prosody. Word stress correlates (pitch, duration and intensity) in the target words were manipulated such that stress occurred once on each syllable in turn. Native listeners in a subsequent perception experiment indicated the acceptability of the stimulus words. The acceptability judgments indicated clear preferences for penultimate stress in Toba Batak, and no preferences for Javanese.

These results are in line with an earlier study on the perception of manipulated pitch movement timings by Toba Batak and Javanese speakers (Van Heuven & Van Zanten, 1997). It was predicted that the accuracy of locating pitch movements within word boundaries would be better for Toba Batak speakers, because of word stress in this language. Locating pitch movement across word boundaries was predicted to be similarly accurate for both languages, as both show evidence for the use of phrase accents. It was found that speakers of Toba

Batak were more accurate in indicating the location of pitch movements than speakers of Javanese, regardless of whether pitch movements were within or across word boundaries. This outcome was taken as evidence that the timing sensitivity of Toba Batak speakers at the syllable level (within word), shaped their sensitivity at the higher level. Javanese listeners were presumably lacking this sensitivity at the lower level and therefore it would not emerge at higher levels at all (Van Heuven & Van Zanten, 1997).

Another acoustic study compared Toba Batak with Betawi Malay (Roosman, 2007). For each variety, four speakers read carrier phrases which were analysed by several spectral and temporal measures as well as by expert listeners. Target words occurred in and out of focus as well as in phrase medial and final position. Toba Batak was observed to have contrastive word stress, with the penultimate syllable as default location. Word stress in Toba Batak was signalled by a clear pitch movement, regardless of focus condition. In Betawi Malay, no evidence for word stress was found. This finding is in line with another study on Betawi Malay that found highly variable pitch movements on the penultimate syllable (Van Heuven, Roosman, & Van Zanten, 2008). The pitch movements were mainly found on the ultimate syllable in cases where the penultimate consisted of schwa or when it occurred in phrase final position, indicating a privileged status of schwa and a crucial role of phrase level factors determining the F0.

To sum up, the work on word stress in Indonesian reveal a number of important points for the study of Papuan Malay. First, the language diversity in Indonesia deserves crucial attention, as there exist clear differences in prosody between language varieties in Indonesia. Given the large area in which Trade Malay varieties are spoken, these languages are also expected to show individual differences. Second, it has not always been clear to what extent word prosody and phrase prosody were (possible to keep) separate in Indonesian varieties (e.g. Van Heuven et al., 2008). In a survey of stress patterns from many languages of the world, Goedemans and Van Zanten (2014) point out that Malay varieties spoken in Indonesia might have lost stress through contact with other languages. Therefore, it was predicted that future investigations are more likely to reveal the absence of stress, in particular in Malayo-Polynesian languages. The observed prosodic characteristics in these languages should then be interpreted as a reflection of phrase level phenomena (Goedemans & Van Zanten, 2014). Third, the studies on Indonesian discussed above highlight the importance of complementing impressionistic word stress claims with acoustic investigations. Thanks to those investigations, a more detailed, reliable and comprehensive understanding of word stress can be achieved.

### 1.3. Word stress correlates in a crosslinguistic context

Word stress' most common acoustic correlates include duration, pitch (F0), formant frequencies (F1 and F2), intensity and spectral tilt (see Gordon & Roettger, 2017 for an overview of 110 studies on 75 languages). Although duration appeared to be the most consistent correlate of word stress across languages, not all acoustic correlates are equally reliable. For example, some of the studies on Indonesian word stress found pitch to be the most important correlate (Halim, 1981;

Laksman, 1994; Roosman, 2007). While these outcomes hold for the Indonesian varieties that were studied, work on Germanic languages generally found pitch as a correlate of phrase level prosody (i.e. Bolinger, 1958; Sluijter & Van Heuven, 1996). Moreover, many studies on word stress have failed to keep word and phrase level phenomena apart (Roettger & Gordon, 2017). In particular, words read in isolation as well as the lack of control for phrase accents in the experimental design have confounded the two levels. It could therefore be problematic to rely on pitch as a main correlate of word stress, in particular if little is known about a language's prosody at other levels.

Formant frequencies, especially F1 and F2 (vowel quality), were shown to correlate with word stress (e.g. Crosswhite, 2004). In stressed syllables, vowels tend to gravitate away from the centre of the vowel space, as defined by the open-close (F1) or front-back/rounded-unrounded (F2) dimensions. However, the number of studies taking vowel quality into account is small, and effects often did not hold for all vowels, nor could they reliably be distinguished from phrase level prosodic phenomena (Gordon & Roettger, 2017).

The overall intensity of a syllable has been shown to correlate with word stress. It has to be noted that the languages for which intensity appeared to be a reliable correlate, are likely to be tone languages. That is, the availability of pitch as a word stress correlate in these languages is limited compared to languages that do not use tone (Gordon & Roettger, 2017). In addition, overall intensity is the most reliable when the distance between the mouth of the speaker and the microphone is kept constant. That is, intensity levels fall when the sound source is further away from the microphone. With relatively small distances (e.g. using a table top microphone), head movements of the speaker could have significant effects on the intensity measures. This could therefore make overall intensity a potentially unreliable correlate. In addition, other acoustic measures depend on intensity such that for recordings with low loudness levels F0 and formants are harder to track and measured values could falsely indicate attenuation processes (i.e. phonological reduction). Controlling for microphone distance could demand experimental settings that decrease the naturalness of the speech. While taking relative intensity measures (e.g. Remijsen & Van Heuven, 2005) can largely overcome these issues, it is not clear to what extent studies on word stress have done this (e.g. Vogel, Athanasopoulou, & Nadia, 2016).

In a limited number of studies, frequency sensitive intensity measures appear as stress correlates (i.e. spectral tilt, spectral balance, spectral emphasis, spectral slope). The correlation with word stress is defined by the intensity reduction towards higher frequencies, which is smaller in stressed syllables (shallow downward tilt) than in unstressed syllables (steep downward tilt). Although there are various ways in which this correlate has been measured (Heldner, 2003), it has been taken as a reliable indicator of whether word stress exists or not, as shown for a variety of languages of the world (Gordon & Roettger, 2017); among other languages, Dutch (Sluijter & Van Heuven, 1996), Catalan (Ortega-Llebaria & Prieto, 2010) and Ambonese Malay (Maskikit-Essed & Gussenhoven, 2016). However, in some studies, frequency sensitive intensity measures did not reliably correlate with word stress, even though languages in these studies were



shown to make use of word stress. For example, in American English (Campbell & Beckman, 1997) and Swedish (Heldner, 2003) these measures rather correlated with focal accent.

To sum up, correlates of word stress can be found in all possible aspects of the acoustic signal (temporal, spectral and amplitudinal). Their reliability differs depending on how the measure is taken and on language specific issues. Concerning the latter, none of the acoustic correlates discussed here corresponds exclusively to word stress. This makes any acoustic definition of word stress highly dependent on how these correlates are used on other levels of prosody in a given language. In the most extreme analysis, word stress is completely independent from phrase level intonation (e.g. Lindström & Remijsen, 2005). In this case, there is hardly any overlap between word level and phrase level acoustic correlates. In other languages, accents at the phrase level can only occur on stressed syllables and therefore some acoustic properties of word stress and pitch accents overlap (e.g. Sluijter & Van Heuven, 1996). As for Papuan Malay, little is known about phrase level prosody. Therefore, the interpretation of the results of the current study is to a large extent bound to existing claims on word stress and to comparisons with other (related) languages. These comparisons predict little as to which correlates could be particularly indicative of word stress in Papuan Malay. While mean F0 was argued to be the main correlate of word stress in Indonesian (Halim, 1981; Laksman, 1994), this finding appeared more nuanced in later work (Goedemans & Van Zanten, 2007, 2014). As for Ambonese Malay, no consistent acoustic correlates were found (Maskikit-Essed & Gussenhoven, 2016). In the current study, the focus is therefore on a broad set of acoustic measures to investigate which of them are potential cues to word stress in Papuan Malay.

#### 1.4. Methodological considerations

In addition to the above mentioned issues concerning acoustic measurements, there are some methodological pitfalls frequently observed in studies on word stress. That is, little or no attention has been given to the importance of the formality of speech. On a continuum of formality, read aloud laboratory speech and unscripted interactive talking would constitute either extreme end. It was observed in Roettger and Gordon (2017) that the majority of studies on word stress make use of scripted laboratory speech. To determine the existence of stress it is preferred to investigate spontaneous and interactive speech; for two reasons. First, languages of the world tend to have an oral tradition in most cases. If a writing tradition exists, it would have commonly developed from orality (Ong, 1982). In other words, read-aloud speech would compromise the extent to which a given speech sample represents natural language. Although previous work made use of elicited readings (e.g. Maskikit-Essed & Gussenhoven, 2016), it has to be noted that most Trade Malay varieties in particular have a limited writing tradition (Paauw, 2008). Therefore, there is considerable discrepancy to expect between scripted and spontaneous speech in these languages. Second, the realisation of stress differs significantly between read and spontaneous speech samples (Howell & Kadi-Hanifi, 1991), although it is possible that this finding was influenced by phrase level phe-

nomena as well. Throughout the literature on word stress in general (Roettger & Gordon, 2017) and on Trade Malay prosody in particular (Table 1), reading tasks have commonly been used. Here it is argued that, in the investigation of word stress crosslinguistically, unscripted spontaneous speech leads to the most representative samples and is preferred over read speech. Given the variable nature of unscripted spontaneous speech, more conservative decisions need to be made to select a corpus of samples suitable for acoustic analysis (see Section 2.3).

Furthermore, studies on Trade Malay have relied on a small number of speakers (e.g. Maskikit-Essed & Gussenhoven, 2016; Steinhauer, 1983). The sample size appears to be a common shortcoming in studies on word stress. That is, most of the studies relied on speech data taken from 10 or fewer speakers, with a considerable amount of work based on the data of only one speaker (Roettger & Gordon, 2017). As a consequence, the number of items in experimental paradigms was limited, possibly resulting in a lack of statistical power that could indicate effects (Roettger & Gordon, 2017). Moreover, small sample sizes could cast doubt on the extent to which the outcomes are representative of a speaker population of a given language. Future work on word stress, including the current study, should therefore take these methodological considerations into account in order to sample more representative amounts of speakers.

#### 1.5. Research questions

Considering the discussion of the literature above, several issues remain to be investigated. First, although many stress claims have been made for Trade Malay varieties, little empirical work has been carried out. As has been shown for Ambonese Malay and Indonesian; impressionistic stress claims need acoustic validation. Second, investigating the correlates of word stress provides more insight into the prosody of under-researched languages. And third, knowing and avoiding pitfalls of previous work on word stress provides more valid results in future work.

To investigate these issues, the current study focuses on Papuan Malay. Word stress in Papuan Malay is claimed to occur regularly on the penultimate syllable (Kluge, 2017). This claim was tested by means of acoustic analysis in the current study. Two specific questions were investigated for Papuan Malay. First, to what extent is there acoustic evidence for word stress? Second; if there is, which acoustic measures predict word stress best? Spectral, temporal and amplitudinal measures were computed, both in raw form and in derived form. The derived measures were taken to avoid acoustic effects that were unrelated to the word stress hypothesis. Linear mixed model analyses were carried out to investigate these questions (Section 2.4). Crucially, the investigated speech data consists of spontaneous narratives.

## 2. Methodology

This section describes the data collection task and the acoustic analysis carried out on Papuan Malay speech data.

### 2.1. Data collection procedure

Speech was collected in a storytelling task. In this task speakers were instructed to watch a short video clip and tell what they had seen to an interlocutor who had not seen the video. The video clip showed a small story about a man picking pears. The actors in the video clip did not use any speech. The video clip has been previously used in cross-linguistic studies on narrative production (Pear Film; Chafe, 1980). Recordings were made at the Center for Endangered Languages Documentation (CELD) in Manokwari, West Papua (Riesberg & Himmelmann, 2012, 2014). Participants received instructions about the experimental procedure before the start of the task. Prior to the retelling, participants watched the video clip on a laptop computer, which lasted six minutes. Thereafter, participants were introduced to their interlocutor and retold the story they had seen. The participants and interlocutor were seated next to each other during the retelling. The interlocutor was allowed to ask clarification questions during the participant's retelling. This happened up to three times per participant.

No soundproof or silent rooms were available at the recording location. Therefore, recordings were made outside, behind a building where background noise was as minimal as possible. The recordings were made using a Sony ECM-MS957 unidirectional stereo microphone connected to a Sony HDR-SR11 portable video camera. The microphone was placed in front of the participant and interlocutor and recorded the speech of both. The experimenter supervised the entire recording procedure. The duration of the collected recordings ranged between two and five minutes.

### 2.2. Participants

All participants were students at the University of Papua. There were 10 male and 9 female participants ( $M_{\text{age}} = 22$ , age range = 20–28). All were native speakers of Papuan Malay without speech problems, as assessed by a language background questionnaire. All participants were also speakers of Indonesian, the country's standard language. Seven participants had basic knowledge of another language (2 Javanese, 2 Biak, 1 Mpur, 1 Abun and 1 English).

### 2.3. Data processing and selection

Audio-tracks were extracted from the recordings on the portable video camera and converted to 48 kHz, 16 bit, mono wave-files. Native speakers of Papuan Malay transcribed the participant's speech and segmented it into intonation units (Chafe, 1994). Thereafter, a group of six labellers annotated all words and syllables produced by the participants for each wave file using Praat textgrids (Boersma & Weenink, 2017). All labellers received phonetic training to set label boundaries by auditory and visual inspection of the wave-form and they were familiar with the syllable structure of Papuan Malay.

A subset of the labelled syllables was selected on the basis of the following criteria. Syllables in utterances that were interrupted or cut off, were omitted. Reduplicated words (e.g. *tiba-tiba*) were omitted, whereas the syllables of single occurrences (e.g. *tiba*) were taken into account. This was done because reduplication could affect acoustic properties of the word being

reduplicated and make that word less comparable to single occurrences. The syllables of words produced with hesitation or that were unidentifiable due to laughter, severe speech reduction (i.e. mumbling) or background noise were also omitted. Words containing double vowel sequences were also omitted, because they allow for two ways of syllabification (either as VV or V.V, see Kluge, 2017). These sequences appeared to lead to inconsistent annotations in the current study.

Furthermore, syllables occurring in final or pre-final phrase position were omitted. This was done to avoid possible interference with phrase level intonation. That is, syllables in these positions are commonly assumed to be locations for possible phrase accents in Trade Malay and Indonesian varieties (e.g. Goedemans & Van Zanten, 2014). In addition, this selection avoided effects of phrase-final lengthening, which mainly concerns the final syllable and possibly the pre-final syllable (Cambier-Langeveld, 2000; Shattuck-Hufnagel & Turk, 1998). The syllables were taken from phrases with varying lengths (3–25 syllables).

The Papuan Malay lexicon has a considerable number of loanwords, mainly originating from Indonesian. It is known from previous work that Indonesian varieties differ with respect to their use of word stress (Section 1.2). In order to exclude possible influences of loanwords on the realization of word stress, only syllables of words that were classified as native roots (Kluge, 2017) were selected. Given the common occurrence of affixation in Malay varieties, the selected forms were unaffixed (roots). Furthermore, single syllable words were omitted as their acoustic characteristics cannot be compared to other syllables in the word. This comparison is required to investigate possible stress differences between syllables within the same word. In addition, the most common word length in Papuan Malay is two syllables (Kluge, 2017), as confirmed by the collected data in the current study. In order to have a consistent dataset, words with a length other than two syllables were omitted.

After applying the selection criteria just described, a total of 3437 syllables were left for acoustic analysis (20% of the data). They were all labelled as 'stressed' or 'unstressed' by taking over the stress indications in wordlists in Kluge (2017). This resulted in 1682 syllables labelled stressed and 1755 syllables labelled as unstressed, occurring in 187 unique words. An overview of how the vowel nuclei are distributed among syllables labelled for stress, is given in Table 2a. An overview of the syllable structures and stress positions is given in Table 2b. The most common stress position was the penultimate syllable.

**Table 2a**

Vowel counts in words with presumed penultimate and ultimate stress in syllables labelled as stressed and unstressed in collected data. For readability purposes the position of the syllable in the word is indicated (pre-final/final).

Vowel	Penultimate stress		Ultimate stress		Total
	Stressed (pre-final)	Unstressed (final)	Stressed (final)	Unstressed (pre-final)	
/i/	361	371	59	–	791
/ɛ/	74	141	–	103	318
/a/	1003	728	33	–	1764
/ɔ/	52	15	–	–	67
/u/	170	324	3	–	497
Total	1660	1579	95	103	3437

Table 2b

Counts of syllables labelled as stressed and unstressed in different syllable structures. In *italic*: absolute standardized residuals for each of the two chi-square tests described in Section 2.3. For readability purposes the position of the syllable in the word is indicated (pre-final/final).

Structure	Penultimate stress			Ultimate stress			Total
	Stressed ( <i>pre-final</i> )	$\chi^2$ [st.res]	Unstressed ( <i>final</i> )	Stressed ( <i>final</i> )	$\chi^2$ [st.res]	Unstressed ( <i>pre-final</i> )	
V	313	12.57	72	–	–	–	385
CV	1110	0.68	1038	13	9.10	79	2240
VC	63	2.27	38	–	–	1	102
CVC	168	10.94	389	82	9.10	21	660
CCV	5	4.62	32	–	–	–	37
CCVC	1	2.80	10	–	–	–	11
Total	1660		1579	95		103	3437

ble, and when the ultimate syllable was stressed, the unstressed (penultimate) syllable always contained /*ɛ*/. In a number of cases, however, /*ɛ*/ can occur in stressed penultimate position. This reflects the distributions reported in Kluge (2017) and, crucially, does not mean that /*ɛ*/ is the only factor explaining the mobility of word stress.

In order to investigate to what extent there is a relationship between stress and syllable structure, chi-square tests (R Core Team, 2017) were performed on the counts in Table 2b for each of the stress positions separately (penultimate and ultimate). The tests revealed significant differences between the observed values and values that would be expected on the basis of a chance-level distribution, both for penultimate stress [ $\chi^2(5, N = 3239) = 272.36, p < 0.001$ ] and ultimate stress [ $\chi^2(1, N = 198) = 80.16, p < 0.001$ ]. The differences between observed and expected are given in Table 2b as absolute standardized residuals. This allows to assess the difference between observed and expected values across syllable structures, and accounts for the differences in row and column totals (Agresti, 2007). For penultimate stress, the largest differences between observed and expected values were found for V and CVC syllables. For ultimate stress, the only available syllable structures for the chi-square test (CV and CVC) both show high residual values. Note that for CVC syllables, these results mainly indicate that this syllable structure is most frequent in word-final position. When stress position and syllable position are compared for CVC syllables, the distribution is not significantly different from chance level [ $\chi^2(1, N = 660) = 3.05, p = 0.08$ ]. This sheds a different light on the potential relation between word stress and syllable structure. It appears from Table 2b that the most common word structure is CV.CVC. Given that penultimate stress is the most common stress pattern, it is challenging to disentangle word stress and syllable structure. However, a more general observation can be made, in that syllables with higher vocalic portions (V, CV, VC) are more often stressed than syllables with low vocalic portions (CVC, CCV, CCVC); [ $\chi^2(1, N = 3437) = 79.17, p < 0.001$ ], even when CV and CVC syllables are left out [ $\chi^2(1, N = 535) = 86.45, p < 0.001$ ]. Thus, the distributions of word stress among syllable structures indicate some sensitivity for syllable structure. It remains unclear, however, to what extent this sensitivity indicates a phonological distinction between “heavy” and “light” syllables. To date, however, there is no known source on Papuan Malay that explicitly distinguishes light and heavy syllables for word stress. The only context in which heavy syllables are mentioned in Papuan Malay is (partial) reduplication (Kluge, 2017, chap. 4). Furthermore, ultimate

stress in the selected dataset only occurred on syllables containing /*i*/, /*a*/ or /*u*/. While this could be a reflection of their frequency of occurrence, it is unclear whether there is an additional phonological reason for this distribution. For example, Papuan Malay word stress could be attracted to the most extreme vowels in the acoustic space.

#### 2.4. Acoustic measures

All acoustic measures were taken using Praat (Boersma & Weenink, 2017). The measurements were taken automatically using scripts. Measures were categorised according to which aspect of the acoustic signal they describe; spectral, temporal and amplitudinal measures. Both raw acoustic measures and derived acoustic measures were taken. The raw (unconverted) measures were not expected to reliably correlate with word stress crosslinguistically (Gordon & Roettger, 2017). None of the raw measures were therefore included in the statistical analyses (Section 2.5). For reference purposes and because until now it is unclear how word stress could potentially be realised in Papuan Malay, the descriptive values of the raw measures were nevertheless reported (Table 6). The derived measures were either a correction, conversion or alternative calculation of the raw measures, in order to improve their accuracy and to minimize the possible effects of acoustic processes that were not part of word stress (e.g. phrase prosody). That is, relative measures of F0 and intensity were calculated on the basis of the difference between syllables labelled as stressed and syllables labelled as unstressed within each word, as described in more detail in the subsequent sections. The acoustic measures were taken from different (sub)intervals relative to the annotated syllables (Fig. 1); either the entire syllable (syllable), a voiced subinterval (svoiced) or a voiced subinterval around the intensity peak (intpeaks). The different subintervals were used to ensure accurate spectral measures (Table 3). The syllable level intervals (syllable) were obtained from the manual annotation procedure (Section 2.3). The two subintervals were extracted in an automatized way in two subsequent rounds. In the first round, syllable subintervals for which Praat (Boersma & Weenink, 2017) was able to detect periodicity were taken (svoiced). Syllables for which F0 tracking errors occurred were omitted. In a second round, portions were taken from the voiced subintervals where formants showed stable values. Stable formant values guarantee accurate frequency measurements and were mostly found close to where intensity levels reached their peak. Therefore, boundaries of these stable portions were calculated relative to the

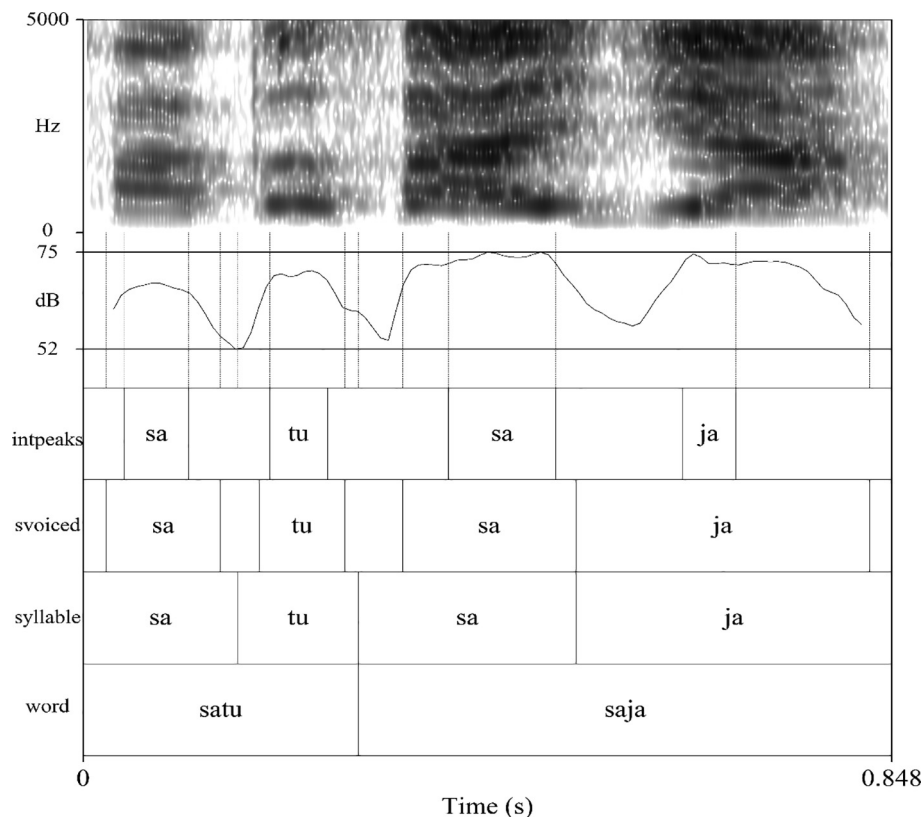


Fig. 1. Example phrase “satu saja” (‘only one’) with different annotation levels (bottom), intensity curve (mid) and spectrogram (top) from which different acoustic measures were taken.

Table 3

Overview of all acoustic measures, units and intervals from which the measures were taken.

	Measure	Unit	Interval
Spectral	Raw F0	ST	svoiced
	Rel. F0	ST	svoiced
	F0 minima/maxima	ST/ms	svoiced
	F1	Bark	intpeaks
	F2	Bark	intpeaks
Temporal	Raw duration	ms	syllable
	Duration per phoneme	ms	syllable
	Duration deviation	ms	syllable
Amplitudinal	Raw intensity	dB	syllable
	Rel. intensity	dB	syllable
	H1-A2	dB	intpeaks
	H1*-A2*	dB	intpeaks
	H1-A3	dB	intpeaks
	H1*-A3*	dB	intpeaks

intensity peak (intpeaks). This was done by taking the time stamps at which the intensity dropped 4% relative to the peak intensity on either side. The 4% margin was chosen after manual inspection of the spectrogram in order to find an optimum between stability of formant detection and interval length. That is, smaller margins resulted in intervals that were too short for (reliable) formant measures, whereas larger margins more often resulted in intervals where formants were harder to distinguish from each other due to higher amounts of variability in the detected frequency values. When the 4% drop on either side of the intensity was not found within the boundaries of the syllable, the syllable boundary was set as the boundary for the subinterval with stable formant values. This occurred

in a minority of the cases, and if so, the syllable mostly started or ended with a vowel. All measures taken from this subinterval (intpeaks) accounted for gender differences in the frequency range of speech. Maximum formant frequencies were set to 5000 Hz for male speakers and to 5500 Hz for female speakers.

#### 2.4.1. Spectral measures

The following spectral measures were distinguished; the F0 measures raw F0, relative F0 and F0 minima/maxima, and the formant measures F1 and F2. F0, F1 and F2 measures were expressed on logarithmic scales (F0 in semitones, F1 and F2 in Bark) to account for how pitch is perceived and, in the case of F0 movements, to abstract over gender differences (Traunmüller & Eriksson, 1994). F0 minima and F0 maxima were measured for both their level in ST and their timestamp in ms. From the minima/maxima values the measures F0 movement, rise/fall ratio and movement onsets/offsets were derived. Raw F0 was measured as the mean F0 value per syllable.

Relative F0 was computed to abstract over absolute F0 levels and the declination effect (Breckenridge, 1977). This was done to obtain a more accurate measure that took into account both syllables labelled as stressed and syllables labelled as unstressed. Thus, relative F0 was computed by subtracting the mean F0 of the syllable labelled as unstressed from the mean F0 of the syllable labelled as stressed in the same word. Positive relative F0 values therefore indicated that the syllable labelled as stressed had a higher F0 than the syllable labelled as unstressed.



F0 movement was measured by subtracting the minimum F0 from the maximum F0 in the voiced subinterval in the syllable (svoiced). Measures of F0 change have been shown to be more reliable indicators of word stress than static F0 measures in Estonian, Italian and Thai (Gordon & Roettger, 2017).

Labels for direction of movement (rise or fall) were derived from the F0 minima and maxima timestamps. That is, whenever the F0 minimum occurred before the F0 maximum, the direction was labelled “rise” and whenever the F0 minimum occurred after the F0 maximum, the direction was labelled “fall”. Note that this labelling does not take the size of the movement into account. Thus, labels were given even when minimum and maximum would be perceptually indistinguishable. Although perceptual assessment of the stress correlates is beyond the scope of the current study, the movement labels together with the measure F0 movement could provide insight into the relevance of the size and direction of the F0 movement for the acoustic realisation of stress. Furthermore, the number of rises and falls were used to calculate a rise/fall ratio, making the comparison across conditions easier to interpret with varying numbers of observations. Rise/fall ratio was calculated by dividing the number of rises by the number of falls. In general, the direction labels provide additional information to the shape of the F0 movements, which could provide further insight into the relationship between word level and phrase level prosody. That is, some languages have been shown to align phrase level F0 accents with stressed syllables (e.g. Sluijter & Van Heuven, 1996). In addition, F0 has been reported as the main cue for word stress in Indonesian (Halim, 1981; Laksman, 1994).

The onset and offset of the F0 movements were measured in milliseconds (ms) relative to the midpoint of the syllable. Note that segment-level annotations were not available, which could have allowed for syllable internal alignment points. The midpoint was defined as the point from which either syllable boundary was equally far away, and chosen to minimize the effect of syllable structure. For example, taking the left syllable boundary as alignment point is not comparable between CV and V syllables, whereas the midpoint represents (part of) the syllable nucleus in either syllable structure. The onset and offset measures of the F0 movements were taken separately for rises and falls, such that onsets coincided with the F0 minimum in rises and with the F0 maximum in falls, and offsets coincided with the F0 maximum in rises and with the F0 minimum in falls. Onsets generally occurred before the midpoint of the syllable, yielding negative values, whereas offset generally occurred after the midpoint of the syllable, yielding positive values.

F1 and F2 were measured in the voiced portions of syllables where formants showed stable values (intpeaks). Thereafter, the Euclidean distance between the formant measures of each vowel and the centre of the vowel space was calculated (Harrington, 2010). This was done to obtain one measure of formant displacement relative to the centre of the vowel space. The centre was defined as the overall mean F1 and the overall mean F2. In this way, the centre measure took into account the natural distribution of vowels and stresses and could therefore deviate from the visual centre of the acoustic space when expressed in a triangular shape (e.g. front vowels in the Papuan Malay inventory occurred more often in the data than back vowels).

#### 2.4.2. Temporal measures

Three temporal measures (expressed in milliseconds) were taken: raw duration, duration per phoneme, and duration deviation. Raw duration was measured by taking the absolute length of the entire syllable. It is known that word stress in some languages is attracted by phonologically heavy syllables. The weight of the syllable commonly depends on its segmental structure, which is not accurately captured by absolute syllable duration. Although Kluge (2017, chap. 4) makes reference to heavy syllables in relation to reduplication processes, it is unclear to what extent there is a morpho-phonological preference for word stress placement in Papuan Malay. In order to account for this possibility, the two other temporal measures took into account the segmental makeup of the syllable. Duration per phoneme was computed by taking the raw syllable duration and dividing it by the number of phonemes in the syllable. It has to be noted that the relationship between number of phonemes and syllable duration is not linear. That is, final lengthening generally affects codas more than onsets (e.g. Campbell & Isard, 1991). Number of segments has nevertheless been a reliable and commonly used correlate in the modelling of syllable duration (Campbell & Isard, 1991; Fletcher & McVeigh, 1993). Furthermore, in some languages, consonant durations correlated better with word stress than vowel duration (Gordon & Roettger, 2017). In order to account for the actual vocalic and consonantal makeup of the syllable, duration deviation was computed separately for each of the syllable structures in the data (CV, CVC, V, VC, CCV, CCVC). This was done by subtracting the mean duration of all syllables with a particular structure from the absolute duration of the syllable with that structure. The outcome represented the deviation in milliseconds of a particular syllable relative to the mean duration. Positive duration deviation was reflective of long syllables, whereas negative duration deviation was reflective of short syllables. Thus, both duration per phoneme and duration deviation take into account the segmental makeup of the syllable and were expected to be more accurate correlates of word stress compared to raw duration.

#### 2.4.3. Amplitudinal measures

Six amplitudinal measures (expressed in decibels; dB) were taken; raw intensity, relative intensity and two measures of spectral tilt, each in uncorrected and corrected form. Spectral tilt was grouped under the amplitudinal measures as it contributes to the perceived loudness of speech and is expressed in dB in the current study. Raw intensity was measured by taking the mean intensity from the entire syllable. It has to be noted that raw intensity is not expected to be a reliable correlate of stress as the recording procedure did not control for the distance between the mouth of the speaker and the microphone (see Section 1.3). For this reason, relative intensity was calculated in the same way as was done for relative F0. That is, the intensity of the syllable labelled as unstressed was subtracted from the intensity of the syllable labelled as stressed. Positive relative intensity values therefore indicate that the syllable labelled as stressed had a higher intensity than the syllable labelled as unstressed. In this way, possible head movements, changes in speaking style (i.e. excited louder speech) or phrase level effects are better controlled for.

As for spectral tilt, the method of Sluijter, Shattuck-Hufnagel, Stevens, and Van Heuven (1995) and Stevens and Hanson (1994) was adopted for half of these measures. That is, intensities of the second formant (A2) or the third formant (A3) were subtracted from the intensity of the F0 (first harmonic; H1). H1 and A2 were corrected for the effect of F1, and A3 was corrected for the effect of both F1 and F2. The overall mean F1 and overall mean F2 were used as reference values for the corrections, following the formulas in Hanson (1995, pp. 113–115). Spectral magnitude correction (Iseli, Shue, & Alwan, 2007) has been applied in previous work to account for the increased intensity around formant frequencies, in particular F1 and F2 (Sluijter et al., 1995; Stevens & Hanson, 1994). Note that other work did not apply this correction and argued that the corrected measure is not able to reliably separate word stress effects on vowel quality from those on spectral tilt, as the two are highly correlated (see Ortega-Llebaria & Prieto, 2010 for Spanish). Furthermore, spectral magnitude correction has been argued to result in less accurate formant intensity measures (Caballero & Carroll, 2015). In the current study, therefore, both the uncorrected (H1-A2 and H1-A3) and corrected (H1\*-A2\* and H1\*-A3\*) were taken as approximations to the intensity slope (tilt) of the spectrum. H1, A2 and A3 intensities were extracted using a Praat script based on Mayer (2014) from the voiced portions of syllables where formants showed stable values (intpeaks). Additional measures of spectral tilt in Papuan Malay are discussed in Kaland (2018).

## 2.5. Statistical analysis

Statistical analyses were carried out using R (R Core Team, 2017). Depending on the computation of the acoustic measure, different types of analyses were carried out. The used R packages are mentioned for each analysis.

As for relative F0 and relative intensity, linear mixed model analyses (LMMs) fit by maximum likelihood (using Satterthwaite approximations to degrees of freedom to calculate *p*-values) were carried out using the “lme4” package (Bates, Maechler, Bolker, & Walker, 2015). A separate analysis was done for each acoustic measure as response, with stress position (two levels: penultimate, ultimate) as predictor and with subjects (speakers) and items (words) as random intercepts. In these models effects of stress were indicated by the intercept (i.e. difference from 0) and obtained by alternating the reference level of stress position.

With regard to F0 movement, movement onset/offset, formant displacement, duration per phoneme, duration deviation, H1-A2, H1\*-A2\*, H1-A3 and H1\*-A3\*, LMMs fit by maximum likelihood (using Satterthwaite approximations to degrees of freedom to calculate *p*-values) were carried out using the “lme4” package (Bates et al., 2015). A separate analysis was done for each acoustic measure as response, with stress (two levels: stressed, unstressed) and stress position (two levels: penultimate, ultimate) as predictors and with subjects (speakers) and items (words) as random intercepts. As for movement onset/offset, also direction (two levels: rise, fall) was added as predictor to analyse timing differences between rises and falls. As for movement onset/offset and

duration per phoneme, syllable structure (six levels: V, CV, VC, CVC, CCV, CCVC) as additional random intercept was added to account for timing differences due to the respective syllable structures.

Rise/fall ratio was response in a generalized linear mixed effect model analysis using the “lme4” package (Bates et al., 2015). Stress (two levels: stressed, unstressed) and stress position (two levels: penultimate, ultimate) were predictors and subjects (speakers) and items (words) were included as random intercepts.

Concerning formant displacement, additional post-hoc pairwise comparisons using Tukey HSD test (Bonferroni corrected) were performed on the interactions between the fixed factors stress (two levels: stressed, unstressed) and vowel (five levels: /i/, /e/, /a/, /ɔ/ and /u/) with subjects (speakers) and items (words) as random factors. This was done to test for displacement effects due to stress for each vowel separately (see Section 1.3) using the “multcomp” package (Hothorn, Bretz, & Westfall, 2008).

To investigate the extent to which each of the acoustic measures could predict word stress, generalised linear mixed model (GLMM) analyses were carried out with model fit comparisons (“lme 4” package; Bates et al., 2015). A full model was created with stress (two levels: stressed, unstressed) as response, and with F0 movement, rise/fall ratio, movement onset, movement offset, formant displacement, duration per phoneme, duration deviation, H1-A2, H1\*-A2\*, H1-A3 and H1\*-A3\* as predictors, and with subjects (speakers) and items (words) as random intercepts. Note that the acoustic measures relative F0 and relative intensity were not included in the model as these measures expressed word stress relative to neighbouring syllable(s) and had therefore no predictive value for the dichotomous dependent variable word stress (see Section 2.4). In the full model, some of the variables were highly correlated, in particular the temporal measures and the measures of spectral tilt (Table 4). Therefore, the contribution of the independent variables was assessed in a stepwise manner using LRTs between a null model (with only the random variables subjects and items) and a model in which one independent variable was added to the null model. As an estimate of how well a model fitted the data, Akaike's An Information Criterion (AIC; Akaike, 1973) was computed for each model in the comparisons. Although the absolute AIC values have no predictive value, AIC values of the models relative to each other can be used to obtain a ranking, with the lower AIC values indicating better model fit (Burnham & Anderson, 2002).

**Table 4**

Moderate to (very) strong correlations (>0.40; Evans, 1996) between predictors in the full GLM model as measured by Pearson *r* (*N* = 3395, *p* < 0.001). Other predictors did not correlate or correlated only weakly.

Correlating predictors		<i>r</i>
dur. per phoneme	dur. deviation	0.86
H1-A2	H1*-A2*	0.84
H1-A2	H1-A3	0.58
H1-A2	H1*-A3*	0.41
H1*-A2*	H1-A3	0.48
H1*-A2*	H1*-A3*	0.59
H1-A3	H1*-A3*	0.70

### 3. Results

This section reports the (G)LMM and LRT effects (Table 5) and means (Table 6), mean formant values and their pairwise comparisons (Table 7), and the results of the model fit comparisons (Table 8).

#### 3.1. Spectral measures

Relative F0 deviated from zero for both penultimate and ultimate stress (Fig. 2). In the case of penultimate stress, the deviation was negative, indicating overall lower F0 for syllables labelled as stressed than for syllables labelled as unstressed. In the case of ultimate stress, the deviation was positive, indicating overall higher F0 for syllables labelled as stressed than for syllables labelled as unstressed. The deviation from zero was only significant in the case of ultimate stress. For penultimate stress, a marginally significant difference from zero was found. As for position, a significant effect indicated that relative F0 was larger for ultimate stress than for penultimate stress.

F0 movements showed larger values for syllables labelled as stressed than for syllables labelled as unstressed (Fig. 3). This difference was, however, not significant. A marginally significant effect was found for stress position, in that F0 movements were larger in the case of penultimate stress when compared to ultimate stress. A marginally significant interac-

tion effect showed that stress differences were larger in ultimate position than in penultimate location.

Overall, rise/fall ratios indicated more falls than rises (Table 6). The ratios differed significantly as a function of stress, in that higher ratios (increase in rises/decrease in falls) were found for syllables labelled as stressed than for syllables labelled as unstressed (Table 5). No (interaction) effect involving position was found.

As for the timing of the F0 movements (Fig. 4), onsets occurred significantly later (i.e. closer to the syllable midpoint) in syllables labelled as stressed than in syllables labelled as unstressed. The significant effect of position indicated that F0 movement onsets occurred later for ultimate stress than for penultimate stress. The effect of direction indicated that onsets occurred earlier for rises than for falls. The interaction between stress and direction indicated that the effect of stress was larger for rises than for falls.

Offsets occurred significantly later in syllables labelled as unstressed than in syllables labelled as stressed. The effect of direction indicated that offsets occurred earlier for rises than for falls. The interaction between stress and position indicated that the effect of stress was larger for ultimate stress than for penultimate stress. The interaction between stress and direction indicated that the effect of stress was larger for rises than for falls. The three-way interaction indicated that offsets were significantly affected by stress, position and direction such that

**Table 5**

Overview of effects for all acoustic measures. Interactions not reported were not significant.

	Measure	Factor	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t/z</i>	<i>p</i>		
Spectral (ST/Bark)	Relative F0	penultimate	−0.26	0.15	105	−1.72	= 0.09		
		ultimate	1.64	0.56	175	2.95	<0.01		
		position	1.89	0.58	169	3.29	<0.01		
	F0 movement	stress	0.05	0.08	3307	0.06	n.s.		
		position	−0.50	0.27	215	−1.87	= 0.06		
		stress*position	0.62	0.32	3299	1.94	= 0.05		
	Rise/fall ratio	stress	0.27	0.07	1	3.66	<0.001		
		position	−0.03	0.28	1	−0.12	n.s.		
	Movement	Onset	stress	8.17	1.52	3301	5.39	<0.001	
			position	16.69	5.18	462	3.22	<0.01	
			direction	−9.05	1.66	3376	−5.45	<0.001	
		Offset	stress*direction	5.18	2.30	3380	2.25	<0.05	
			stress	3.12	1.47	3229	2.12	<0.05	
			position	−5.44	4.82	344	−1.13	n.s.	
			direction	−19.14	1.61	3383	−11.89	<0.001	
			stress*position	19.05	6.10	3375	3.12	<0.01	
			stress*direction	24.81	2.23	3386	11.12	<0.001	
			stress*position*direction	−26.80	9.34	3377	−2.87	<0.01	
			Formant displacement	stress	0.22	0.02	3256	9.15	<0.001
				position	0.00	0.12	242	0.03	n.s.
				stress*position	−0.29	0.10	3258	−2.87	<0.01
Temporal (ms)			Duration per phoneme	stress	2.14	0.91	3414	2.36	<0.001
				position	−5.10	3.42	363	−1.49	n.s.
	stress*position	9.51		3.71	3414	2.56	<0.05		
	Duration deviation	stress	5.22	1.60	3307	3.26	<0.01		
		position	−10.02	6.41	301	−1.57	n.s.		
		stress*position	25.50	6.68	3303	3.82	<0.001		
Amplitudinal (dB)	Relative intensity	penultimate	0.43	0.26	119	1.65	n.s.		
		ultimate	2.94	0.84	218	3.50	<0.001		
		position	2.51	0.87	210	2.88	<0.01		
	H1-A2	stress	−2.17	0.33	3268	−6.51	<0.001		
		position	−0.67	1.69	263	−0.40	n.s.		
	H1 <sup>+</sup> -A2 <sup>+</sup>	stress	−0.84	0.31	3280	−2.70	<0.01		
		position	−0.84	1.35	276	−0.62	n.s.		
	H1-A3	stress	−0.77	0.30	3276	−2.58	<0.01		
		position	−1.57	1.45	276	−1.08	n.s.		
	H1 <sup>+</sup> -A3 <sup>+</sup>	stress	−0.25	0.32	3307	−0.78	n.s.		
		position	0.03	1.30	322	0.02	n.s.		

**Table 6**

Means and SDs for all acoustic measures according to stress and stress position.

	Measure	Penultimate		Ultimate	
		Stressed	Unstressed	Stressed	Unstressed
Spectral (ST/ms/Bark)	Raw F0	23.49 (5.17)	23.74 (5.49)	24.82 (5.70)	22.68 (5.62)
	Rel. F0	−0.38 (3.60)		2.11 (3.35)	
	F0 movement	1.98 (2.22)	1.92 (2.21)	2.12 (3.11)	1.37 (1.61)
	Rise/fall ratio	0.80	0.64	1.07	0.47
	Rise				
	Onset	−24.92 (31.30)	−39.96 (32.74)	−28.62 (35.35)	−21.62 (30.76)
	Offset	42.24 (29.79)	16.96 (33.55)	43.63 (35.59)	18.40 (30.09)
	Fall				
	Onset	−22.10 (29.29)	−31.54 (37.48)	−17.03 (40.86)	−10.03 (30.43)
	Offset	36.40 (29.32)	35.76 (32.35)	58.32 (34.35)	31.19 (30.76)
Temporal (ms)	Formant displacement	1.65 (0.82)	1.43 (0.71)	1.47 (0.68)	1.53 (0.68)
	Raw duration	122.12 (50.08)	126.30 (53.56)	163.86 (47.97)	109.06 (45.07)
	Duration per phoneme	66.80 (29.54)	58.65 (27.71)	57.61 (17.22)	49.70 (16.41)
	Duration deviation	2.82 (46.30)	−2.52 (51.22)	12.42 (46.15)	−18.20 (38.62)
Amplitudinal (dB)	Raw intensity	63.91 (5.07)	63.53 (5.01)	63.52 (4.18)	60.55 (4.24)
	Rel. intensity	0.23 (4.41)		3.12 (3.23)	
	H1-A2	7.23 (11.52)	9.38 (10.27)	9.27 (12.21)	11.72 (9.29)
	H1 <sup>+</sup> -A2 <sup>+</sup>	6.02 (10.36)	6.90 (9.68)	5.89 (10.34)	7.26 (10.00)
	H1-A3	12.89 (10.37)	13.66 (9.62)	12.58 (10.52)	13.09 (8.93)
	H1 <sup>+</sup> -A3 <sup>+</sup>	10.98 (11.38)	11.32 (10.59)	10.65 (10.11)	12.34 (9.96)

**Table 7**

Mean F1 and F2 (Bark) and SDs per vowel for syllables labelled as stressed and unstressed, and effects of stress (Tukey HSD) on formant displacement from/towards the vowel space centre.

Vowel	F1		F2		Displacement	
	Stressed	Unstressed	Stressed	Unstressed	Re. centre	<i>p</i>
/i/	4.86 (1.06)	5.13 (1.09)	13.34 (1.08)	13.17 (1.04)	from	n.s.
/ɛ/	5.38 (0.84)	5.45 (0.93)	12.47 (0.97)	12.87 (1.09)	towards	=0.06
/a/	6.42 (1.02)	5.94 (0.90)	11.8 (1.01)	11.94 (1.10)	from	<0.001
/ɔ/	5.83 (0.81)	5.89 (1.24)	10.33 (0.91)	10.57 (1.36)	from	n.s.
/u/	5.16 (1.18)	5.32 (0.92)	11.31 (1.60)	11.59 (0.89)	from	<0.001
Centre	5.72 (1.15)		12.20 (1.28)			

**Table 8**Results of the GLMM analysis on the full model (*b*, *SE*, *z*) and LRT model fit comparisons with null model (AIC,  $\chi^2$ , *p*). Ranks of null model comparisons are indicated in superscript. Null model AIC given in heading.

Acoustic measure	GLMM (full)			Added to null model		
	<i>b</i>	<i>SE</i>	<i>z</i>	AIC (4711.22)	$\chi^2$	<i>p</i>
F0 movement	0.04	0.02	2.44	4711.31 <sup>10</sup>	1.92	n.s.
Rise/fall ratio	0.34	0.08	4.44	4699.41 <sup>7</sup>	13.82	<0.001
Movement onset	1.50	1.38	7.59	4633.08 <sup>2</sup>	80.15	<0.001
Movement offset	7.53	1.38	5.45	4609.00 <sup>1</sup>	104.22	<0.001
Formant displacement	0.35	0.05	7.01	4649.25 <sup>4</sup>	63.97	<0.001
Duration per phoneme	27.34	3.56	7.68	4642.75 <sup>3</sup>	70.47	<0.001
Duration deviation	−0.01	0.00	−5.65	4697.34 <sup>6</sup>	15.89	<0.001
H1-A2	−0.06	0.01	−6.51	4680.66 <sup>5</sup>	32.57	<0.001
H1 <sup>+</sup> -A2 <sup>+</sup>	0.05	0.01	4.90	4706.25 <sup>8</sup>	6.97	<0.01
H1-A3	0.03	0.01	3.68	4709.29 <sup>9</sup>	3.94	<0.05
H1 <sup>+</sup> -A3 <sup>+</sup>	−0.02	0.01	−3.48	4712.03 <sup>11</sup>	1.19	n.s.

the latest offsets were found for falls in syllables labelled as stressed when stress was penultimate and the earliest offsets were found for rises in syllables labelled as unstressed when stress was ultimate (see also Table 6). Stress differences were largely absent for fall offsets when stress was penultimate.

Formant displacement relative to the centre of the vowel space was overall significantly larger for syllables labelled as stressed than for syllables labelled as unstressed (Table 5). No significant difference in formant displacement was found between the respective stress positions. The interaction between stress and position was significant in that the formant displacement effect due to stress was mainly present in the case of penultimate stress. The post-hoc pairwise compar-

isons (Table 7) revealed significant displacement from the centre of the vowel space for /a/ and /u/, marginally significant displacement for /ɛ/ and no significant displacement for /i/ and /ɔ/. Crucially, the displacement of /ɛ/ was *towards* the centre of the vowel space, while other vowels were displaced *away from* the centre (Fig. 5).

### 3.2. Temporal measures

Duration per phoneme showed larger values for syllables labelled as stressed than for syllables labelled as unstressed (Fig. 6); a difference that appeared to be significant. The effect of position was only found significant in interaction with stress,



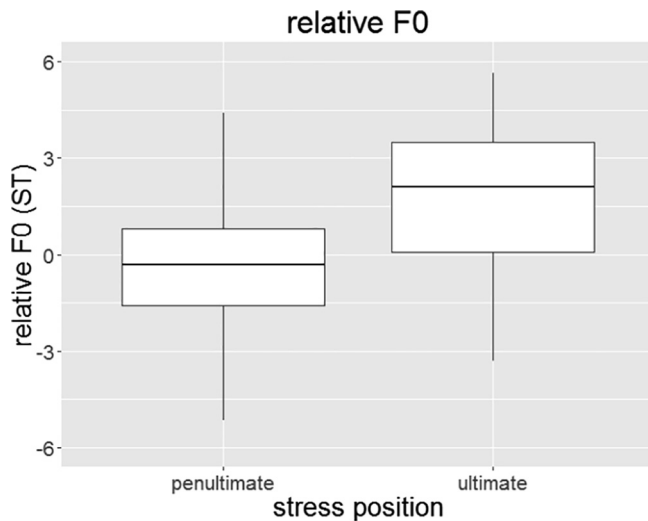


Fig. 2. Boxplot of relative F0 (ST) as a function of stress position.

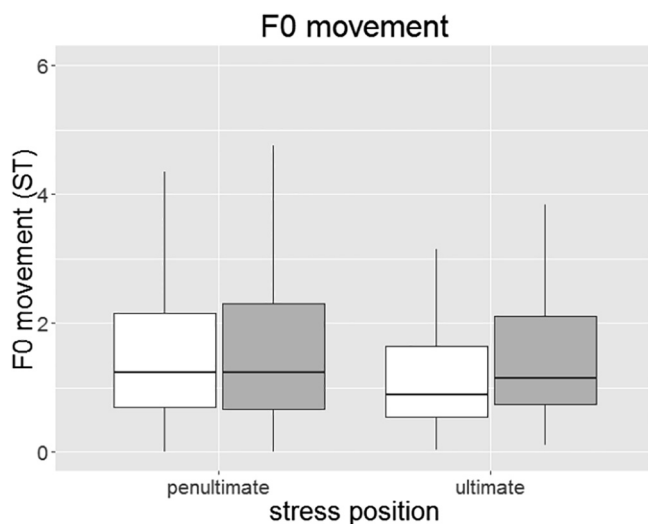


Fig. 3. Boxplot of F0 movement (ST) as a function of stress (grey = stressed, white = unstressed) and stress position.

indicating that the effect of stress (around 8 ms increase in either stress position) occurred on overall longer phoneme durations for penultimate stress than for ultimate stress.

Overall positive values of duration deviation (Fig. 7) were found for syllables labelled as stressed and negative values were found for syllables labelled as unstressed. This effect was significant. The effect of position was only significant in interaction with stress, indicating that duration deviation was generally larger for ultimate stress than for penultimate stress.

### 3.3. Amplitudinal measures

Relative intensity generally showed positive deviations from zero (Fig. 8). This deviation was, however, only significant in the case of ultimate stress. The effect of position indicated that relative intensity values were larger for ultimate stress than for penultimate stress.

As for H1-A2, syllables labelled as stressed showed significantly lower values compared to syllables labelled as

unstressed (Fig. 9). No effect of position was found. The same results were obtained for H1\*-A2\* (Fig. 9).

H1-A3 values were significantly lower in syllables labelled as stressed than in syllables labelled as unstressed (Fig. 10). There was no significant effect of position. H1\*-A3\* revealed no significant effects involving stress or position (Fig. 10).

### 3.4. Predictors of word stress

Table 8 reports the relevant statistical outcomes for each of the acoustic measures in the GLMM analysis. As for F0 movement, the null model comparison indicated no significant differences when adding this variable. Crucially, the AIC value indicated slightly better fit for the null model, where F0 movement was not included. Adding rise/fall ratio significantly improved the null model, with AIC values showing a small increase. Movement onset as well as movement offset significantly improved the null model, with the offset showing the best model fit obtained in this analysis (lowest AIC) and onset showing the second best model fit. Adding formant displacement significantly improved the model fit and this variable was ranked relatively high, as indicated by the AIC scores. The model comparisons showed the third largest AIC improvement for duration per phoneme, which appeared to be significant. Duration deviation significantly improved the model fit, with a moderately ranked AIC value. H1-A2 contributed significantly to the model fit when added to the null model. The AIC values indicate that this contribution was moderate. H1\*-A2\* showed significant effects, with relatively low AIC rankings. As for H1-A3, a significant effect was found when adding this variable. AIC rankings showed a low contribution to the model fit for H1-A3. With regard to H1\*-A3, no significant effect was found and the AIC value worsened when including this variable in the model.

## 4. Conclusions and discussion

This study has analysed a set of possible acoustic correlates of word stress in order to investigate to what extent they provide evidence for the existence of word stress in Papuan Malay. The acoustic measures were categorised as spectral, temporal and amplitudinal. The extent to which these measures provide evidence for word stress is discussed in this section.

### 4.1. Spectral evidence for word stress

Significant effects involving stress were found for relative F0 (only for ultimate stress), rise/fall ratio, movement onset, movement offset and formant displacement. Support for the hypothesis that Papuan Malay has stress was found in that syllables labelled as stressed had higher F0 than the unstressed syllable in that word when stress was penultimate. As for F0 movement, the significant interaction between stress and position indicated that mainly in the case of ultimate stress, larger F0 movements were found on syllables labelled as stressed than on syllables labelled as unstressed. Furthermore, syllables labelled as stressed were more likely to show a rising F0 movement. While syllables labelled as stressed showed generally later onsets and offsets than syllables labelled as

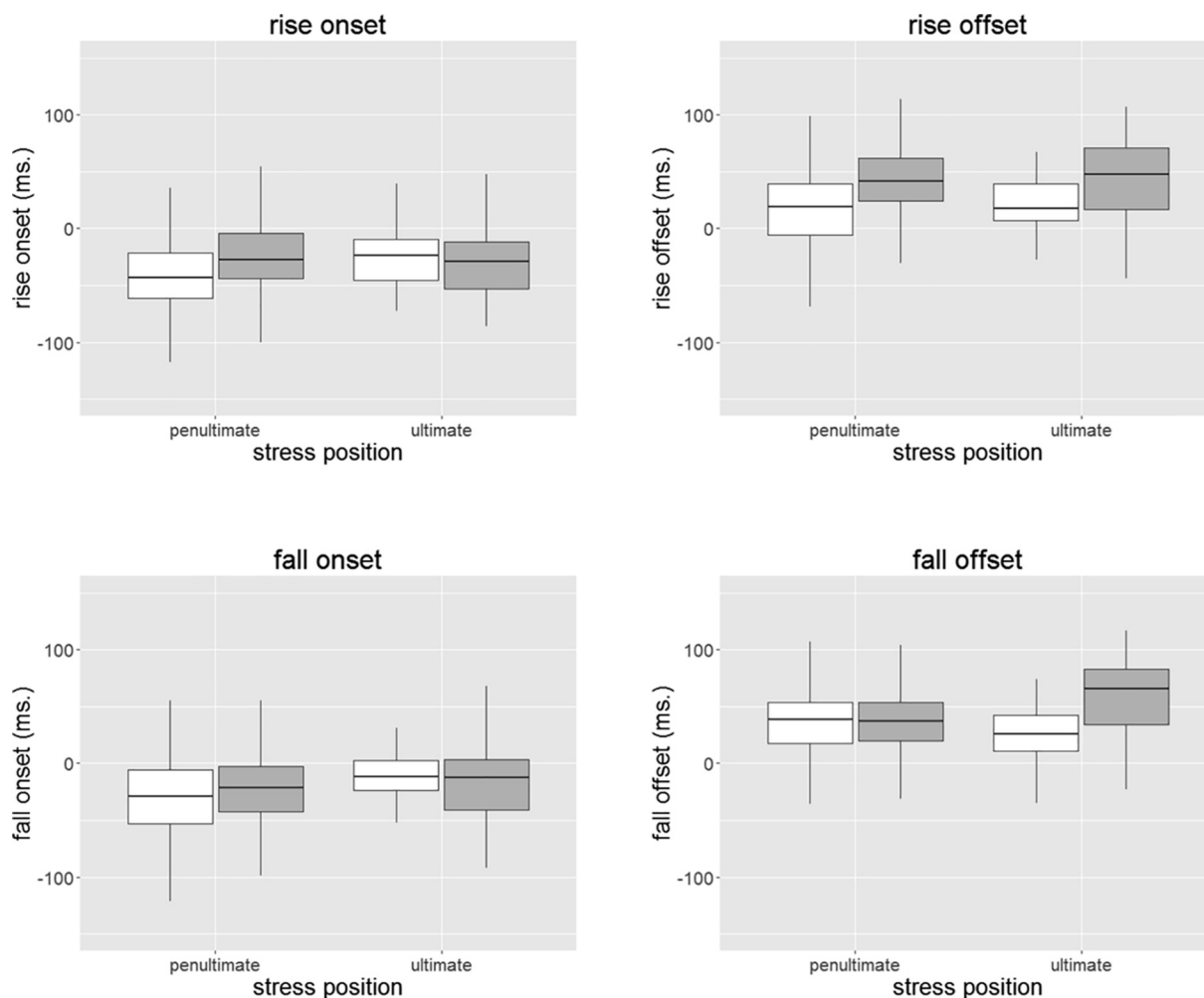


Fig. 4. Boxplots of onsets (left) and offsets (right) of rises (top) and falls (bottom) as a function of stress (grey = stressed, white = unstressed) and stress position.

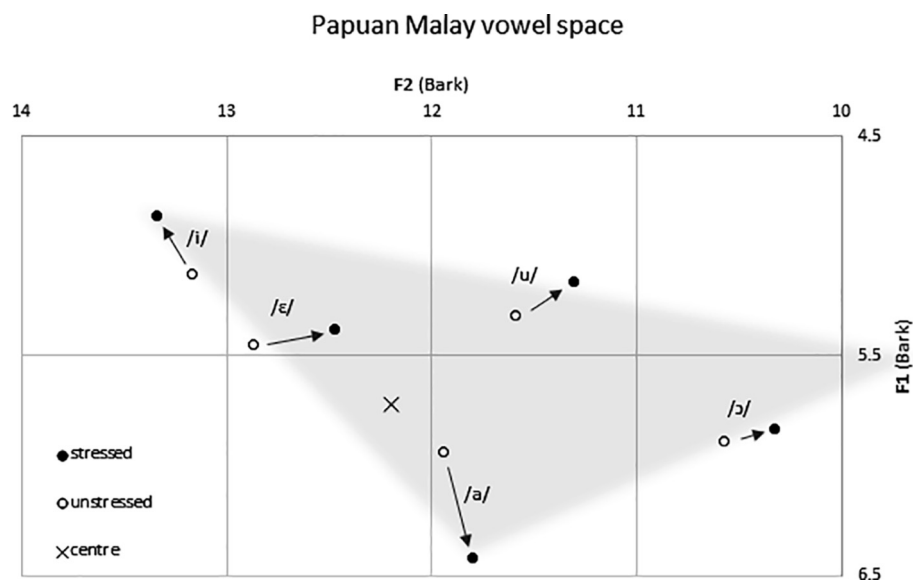


Fig. 5. Papuan Malay vowels in syllables labelled as stressed and unstressed. The grey area represents the estimated vowel space, with the centre mark indicating the overall mean formant values (Table 7).

unstressed, this was mainly true for penultimate stress. For ultimate stress, onsets occurred earlier and offsets occurred later in syllables labelled as stressed than in syllables labelled

as unstressed. Furthermore, rises on syllables labelled as stressed generally had a later offset compared to syllables labelled as unstressed, regardless of stress position. Falls,

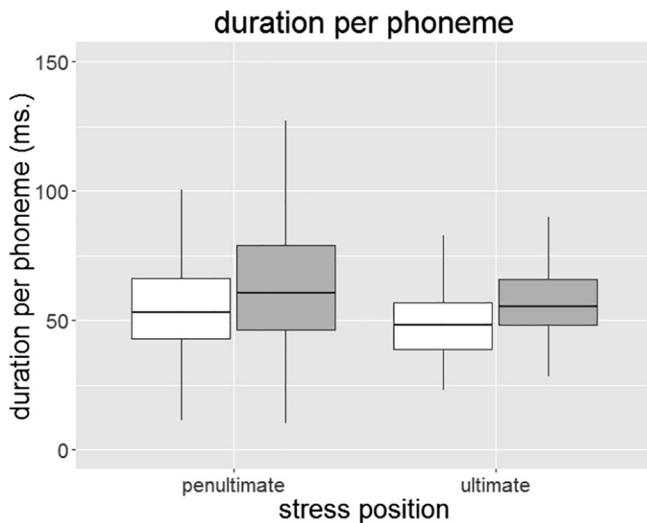


Fig. 6. Boxplot of duration per phoneme (ms.) as a function of stress (grey = stressed, white = unstressed) and stress position.

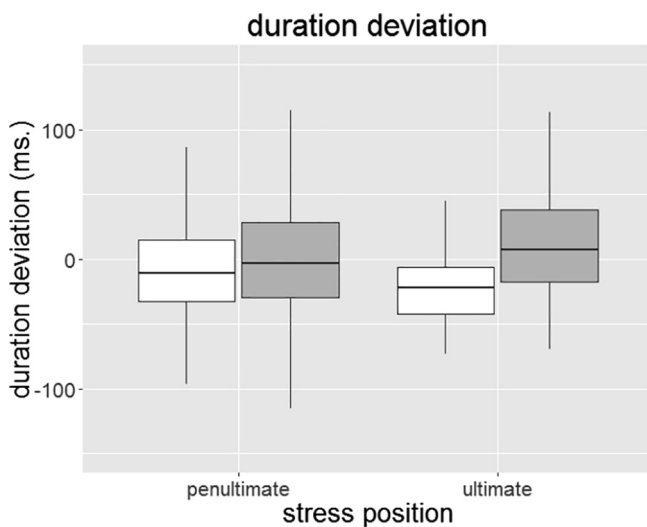


Fig. 7. Boxplot of duration deviation (ms.) as a function of stress (grey = stressed, white = unstressed) and stress position.

however, mainly showed later offsets in syllables labelled as stressed when stress was ultimate and a minimal difference when stress was penultimate. These differences are reflected in the significant three-way interaction between stress, position and direction for movement offsets.

Displacement relative to the centre of the vowel space provided evidence for stress for /a/ and /u/ (Table 7, Fig. 5). The displacement of /i/ and /ɔ/ was too small to be significant. Remarkably, /ɛ/ was realised closer to the centre in syllables labelled as stressed. This result is counter to the expectation that stressed vowels are realised further away from the centre to maximize (perceptual) vowel contrasts (i.e. Liljencrants & Lindblom, 1972; Flemming, 2004). It can furthermore be observed that the phonetic realisation of vowels in Papuan Malay do not always match their phonemic representation. That is, /u/ is realised relatively central, similar to [ʊ], and /ɛ/ is realised relatively high, similar to [e]. While this could be the result of allophonic variation, none of these realisations

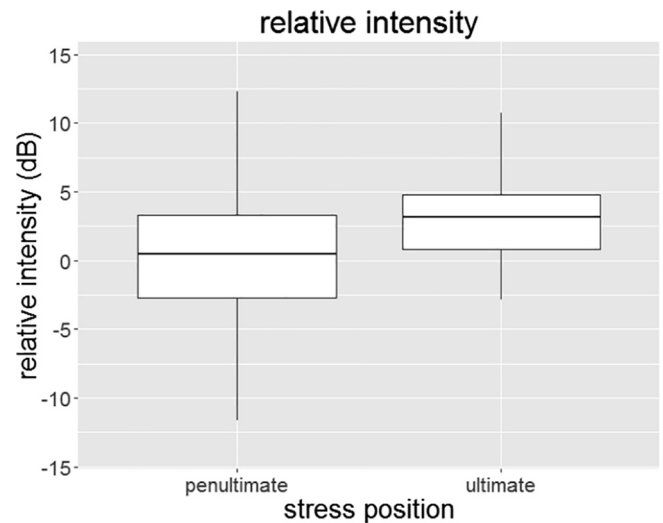


Fig. 8. Boxplot of relative intensity (dB) as a function of and stress position.

match with the allophones reported in Kluge (2017). This is particularly noteworthy in the case of /ɛ/, of which its displacement is counter to the stress hypothesis. One could therefore question to what extent a reanalysis of the vowel inventory based on acoustic measures would still be compatible with the one in Kluge (2017). While this investigation is beyond the scope of the current study, a reanalysis of the Ambonese Malay vowel inventory shed a different light on earlier stress claims (Maskit-Essed & Gussenhoven, 2016). In sum, the formant displacement measure provided some support for word stress when individual differences between vowels are taken into account. Further discussion of the formant displacement of /ɛ/ is provided in Section 4.5.2.

#### 4.2. Temporal evidence for word stress

Consistent support for the stress hypothesis was obtained for duration per phoneme and duration deviation. Both measures were significantly affected by word stress in that longer durations were found for syllables labelled as stressed than for syllables labelled as unstressed. Therefore, both duration per phoneme and duration deviation can be taken as correlates indicating acoustic evidence in favour of word stress in Papuan Malay. Both temporal measures also showed significant interactions between stress and position, indicating that there was more lengthening of syllables labelled as stressed when stress was ultimate than when stress was penultimate. It has to be noted that the measures did not account for the tendency of word final syllables to be longer (word final lengthening). These effects are likely, given the mean raw durations (Table 6). Therefore, the stress-position interactions found for both temporal measures could have been affected by word-lengthening.

#### 4.3. Amplitudinal evidence for word stress

The relative intensity measure confirmed the word stress hypothesis for ultimate stress cases. The effect of position further confirmed differences between penultimate and ultimate stress, in that significant intensity differences between syllables

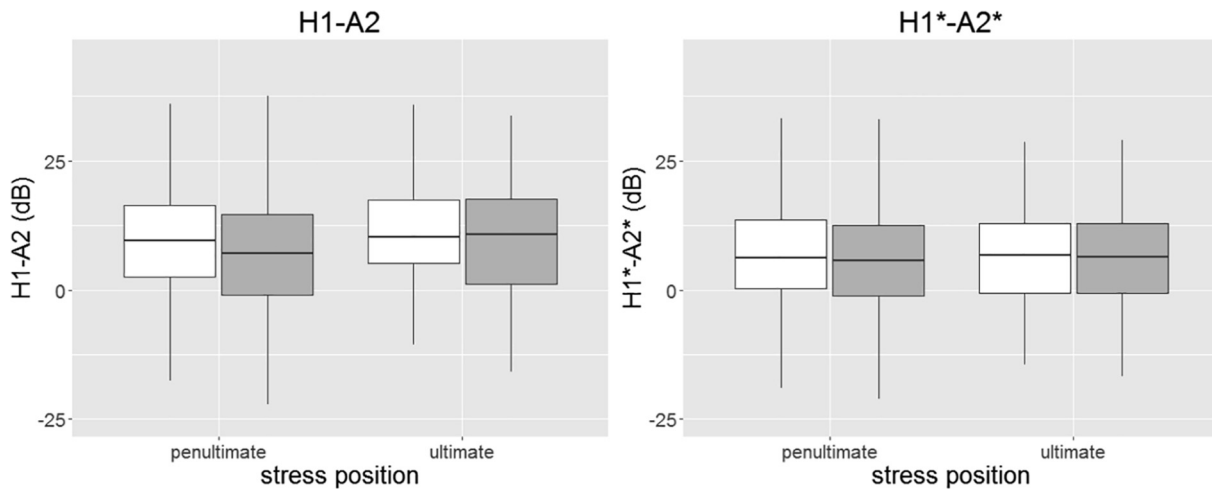


Fig. 9. Boxplots of H1-A2 (left) and H1\*-A2\* (right) (dB) as a function of stress (grey = stressed, white = unstressed) and stress position.

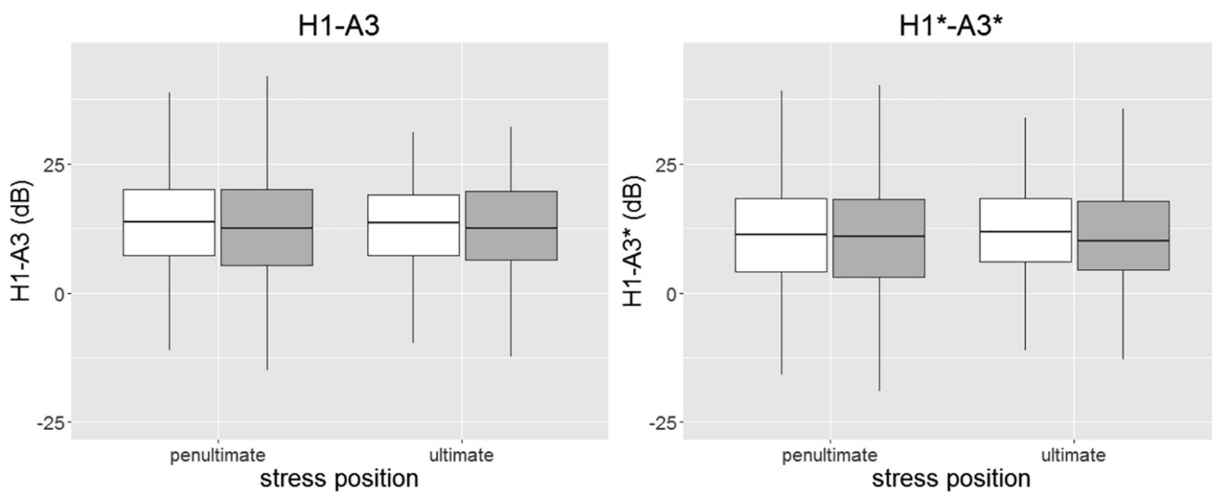


Fig. 10. Boxplots of H1-A3 (left) and H1\*-A3\* (right) (dB) as a function of stress (grey = stressed, white = unstressed) and stress position.

bles labelled as stressed and syllables labelled as unstressed were only found for ultimate stress. All spectral tilt measures showed shallower slopes in syllables labelled as stressed than in syllables labelled as unstressed. H1-A2 and H1\*-A2\* were more indicative than H1-A3 and H1\*-A3\* respectively, indicating the intensity of the second formant relative to the first harmonic was particularly informative of stress differences. Furthermore, the uncorrected H1-A2 and H1-A3 were more indicative of word stress than their corrected versions. In particular, H1\*-A3\* appeared of no contribution to word stress in the current study. It has to be noted that the intensity levels decrease towards higher frequencies, regardless of word stress effects. Due to this natural roll-off, possible stress differences could be smaller at the third formant than at the second formant, and could therefore fail to reach significance.

#### 4.4. Relative contribution of the acoustic measures to word stress

From the acoustic measures that were tested for their contribution to word stress, the measures movement onset/offset and duration per phoneme were most indicative correlates of word stress in Papuan Malay. Other relatively strong correlates of word stress were formant displacement and H1-A2. Duration

deviation, H1\*-A2\* and H1-A3 correlated with word stress in a weaker manner. F0 movement and H1\*-A3\* did not predict word stress at all. Overall, the results of the GLMMs and model fit comparisons provide additional insight into the relative contribution of the possible acoustic correlates of word stress. The results of the model comparisons were furthermore in line with those of the LMMs on each of the acoustic measures.

#### 4.5. Discussion

The results of the current study showed acoustic support for the Papuan Malay stress claims. In particular, the timing of F0 movements and duration per phoneme were shown to be strong correlates of word stress. It has to be noted that the timing measures were potentially affected by lengthening effects. Given that duration per phoneme was a good predictor of stress in the model comparison (ranked third), the onset/offset timing effects need to be interpreted in this context. That is, lengthening due to stress could have caused F0 movements to start or end later because segments were longer on average in stressed syllables (i.e. the effect of duration per phoneme). More insightful for the interpretation of the F0 movement timings are the interaction effects for the movement onsets/off-



sets. These indicate longer F0 movements (earlier onsets, later offsets) only in the case of ultimate stress. Given the potential effect of lengthening, earlier onsets in particular are a strong indication of the importance of F0 movement timing in ultimate stress.

Thus, the F0 movement timings strongly hint at crucial differences between the realisation of stress in penultimate and ultimate position. It appears from the means and the reported effects that generally acoustic differences are more clearly marked when stress is ultimate. This conclusion is further supported by (interaction) effects involving position for most other acoustic measures, except rise/fall ratio and the spectral tilt measures. In particular, the relative measures of F0 and intensity indicate significant deviations from zero (hence larger difference between the respective syllables), only in the case of ultimate stress.

#### 4.5.1. Word stress position

A couple of factors need to be taken in account when interpreting the stress position difference just observed. First, due to the asymmetry in the distribution of words with presumed penultimate (95%) or ultimate stress (5%), statistical tests could falsely indicate that there was no difference in stress realisation in the respective positions (type II error). If this error indeed applies, the current data does not allow to either confirm nor to reject the mobility of word stress to the ultimate syllable. As a consequence, the acoustic analysis would then only confirm penultimate stress. It has to be noted that the penultimate/ultimate asymmetry in the current dataset reflects the presumably natural distribution of word stress in Papuan Malay (Kluge, 2017). For this reason, increasing the sample size of words with presumed ultimate stress in order to reduce the type II error rate, would yield unrealistic word stress distributions. In addition, inspection of the most indicative correlates reveals that mean acoustic differences between syllables labelled as stressed and syllables labelled as unstressed were often larger for penultimate stress as opposed to ultimate stress (Table 6). These data suggest that the existence of ultimate stress is realistic despite the asymmetric distribution.

Second, the asymmetrical distribution of word stress in Papuan Malay hints at a marked (i.e. exceptional) status for ultimate stress. When ultimate stress is indeed the marked stress position – as supported by most acoustic measures – the question remains to what extent stress is acoustically marked in penultimate position. Although there is clear acoustic evidence for stress across the respective positions, it could be the case the penultimate stress is the default pattern which does not require large acoustic differences. In this view, the notion of “stress” in terms of markedness rather applies to ultimate positions. Such an asymmetry between the respective stress positions would be explained by the acoustic data investigated here for Papuan Malay. As a consequence, the notion of stress marking is more constrained to position in the word than the attribution of a specific amount of acoustic prominence to a single syllable in a word would imply. The literature has observed that in languages where there is little variation in the position of the stressed syllable (i.e. “fixed stress” languages), the acoustic difference between stressed and unstressed syllables is small (Cutler, 2005; Dogil, 1999). This can be explained when considering that speakers and listeners

of this type of languages expect the most frequent pattern by default and that only deviations from this pattern need to be acoustically salient enough to warrant successful communication. Indeed, EEG studies on languages where the stressed syllable has limited mobility, have shown that listeners are particularly sensitive for non-default stress patterns (Domahs et al., 2012, 2013). Given the current results, the acoustic realisation of word stress in Papuan Malay would fit the asymmetry observed for other languages. While the communicative importance of a deviant stress pattern could explain much of the acoustic asymmetry, the position of the stressed syllable in the prosodic structure of Papuan Malay needs to be considered as well. That is, ultimate syllables are particularly suitable to stand out as acoustically more prominent due to their final position. Word final lengthening, which appeared to be present in the current data (Section 4.2), could thus have had an additional, strengthening effect to ultimate word stress. In other words, longer syllables give more room for acoustic cues to be realized more saliently compared to shorter ones. It is likely that the penultimate stress patterns found in the current study exhibit such a combined effect of word stress and final lengthening, as hinted at by the effects of position reported in Table 5.

Third, while most measures find small acoustic differences for penultimate stress and large acoustic differences for ultimate stress, some measures find hardly any acoustic differences for penultimate stress. In particular, relative F0, F0 movement and relative intensity fit this pattern. Given that two of these measure concern F0, it has to be noted that the role of F0 in word stress has not always been clear. Significantly lower F0 values have been found for syllables labelled as stressed compared to syllables labelled as unstressed in Italian (Eriksson et al., 2016) and Lahore speakers of Urdu (Hussain, 1997). It is unclear how these opposing results should be interpreted. If F0 is indeed a main correlate of phrase level prosody, the unexpected values might be an artefact of the weak correlation with stress in general. However, this interpretation is problematic from an articulatory acoustic point of view. That is, prominence at any prosodic level is generally associated with greater articulatory effort (e.g. Streefkerk, 2002). The physical relationship between spectral, temporal and amplitudinal aspects of the speech signal predict that for larger F0 movements, more time and more vocal energy is needed. This seems to be the pattern in the current results for Papuan Malay ultimate stress. Thus, F0 is used to mark mainly a deviation from the default (penultimate) stress pattern.

#### 4.5.2. Formant displacement of /ɛ/

As for F1 and F2, consistent formant displacement was observed. For most vowels this meant displacement from the centre of the acoustic space. An exception was found for /ɛ/, which was centralised in syllables labelled as stressed (Fig. 5). This is counter to the crosslinguistically common observation that vowels are centralised, or likely to be realised as schwa in unstressed positions. It has to be noted, however, that schwa can be stressed in languages where this sound is part of the phoneme inventory (e.g. Romanian; Chitoran, 2001). In Papuan Malay, /ɛ/ appears more often in unstressed positions than the other Papuan Malay vowels (Table 2a). It is

therefore not immediately clear how the formant displacement of /ɛ/ should be interpreted.

Given the research question of the current study, two types of explanations are discussed. In the first, the exceptional behaviour of /ɛ/ is not interpreted as a reflection of word stress. Rather, the formant values are taken as an indication that the vowel inventory of Papuan Malay includes /ɛ/, either as an addition to or a replacement of /e/ (see also Section 4.1). Such an approach would be similar to the reanalysis of the vowel inventory proposed for Ambonese Malay, with the crucial difference that no acoustic evidence for word stress was found for Ambonese (Maskikit-Essed & Gussenhoven, 2016). Thus, in a stress-less reanalysis of the vowel inventory of Papuan Malay, it would be unclear how the regular formant displacement observed in the other vowels would have to be interpreted. In addition, the consistent evidence for word stress from most other acoustic measures in the current study would also have to be reanalysed.

In the second explanation, the formant displacement of /ɛ/ is interpreted in accordance with the claim that Papuan Malay has word stress (Kluge, 2017). Although it is unclear in this view how exactly the realisation of stress could lead to vowel centralisation, the position of schwa in Papuan Malay might provide further insight. That is, Papuan Malay /ɛ/ often corresponds to schwa in Indonesian (Donohue, 2003). Like for Romanian, it has been claimed that schwa can be stressed in Jakarta Indonesian (Laksman, 1994), see Section 1.2. Although there is debate about the existence of word stress in Indonesian (e.g. Goedemans & Van Zanten, 2014), regular word level prosody has been observed for some of the Indonesian languages (e.g. Toba Batak; Goedemans & Van Zanten, 2007). Crucially, the large influence of (Jakarta) Indonesian on Papuan Malay is undisputed (Kluge, 2017). Therefore, the centralised realisation of stressed /ɛ/ can be explained by assuming that /ɛ/ is underlyingly treated as schwa, however only when stress is applied (i.e. like in Jakarta Indonesian). In this way, stress on /ɛ/ is defined relative to schwa in the centre of the vowel space, not relative to unstressed /ɛ/. As a consequence, stressed /ɛ/ is realised as a “displaced schwa”, and located between unstressed /ɛ/ and the position of schwa. The Papuan Malay vowel space allows for this midway location and ensures that stressed /ɛ/ remains acoustically distinct from both unstressed /ɛ/ and schwa. Such an unusual application of word stress would be a phonologically costly process, which explains why stressed /ɛ/ is relatively rare. That is, /ɛ/ occurs more often in syllables labelled as unstressed (>80%) than in syllables labelled as stressed. This distribution is different from the other Papuan Malay vowels, which occur more equally frequent among syllables labelled as stressed and syllables labelled as unstressed (Table 2a).

Additional support to explain the formant displacement of /ɛ/ within the claim that word stress is present, is the potentially limited communicative function of word stress in Papuan Malay. This is particularly clear from the lack of minimal stress pairs, suggesting that the contribution of word prosody to distinguish word meanings is minimal. Due to this limited function, the hypothetical influence of the Indonesian stressed schwa just outlined is most likely harmless for word recognition in Papuan Malay. The adoption of stressed schwa could rather reflect the influence of Indonesian in formal settings, such as

interactions with Western researchers described in Kluge (2017, p. 18). The recording procedure described in Section 2.1 could have caused such a level of formality for the participants in the current study. In addition to possible influences of register, it has been shown that language contact situations can be reflected in the variation of a single phoneme (Kaland, Galatà, Spreafico, & Vietti, 2019). Similarly for North Moluccan and specifically Manado Malay, schwa was originally lacking, but does occur “in words which might represent code-switching to Indonesian” (Paauw, 2008, p. 95).

Although the two views just described could provide directions for future research, they remain speculative and leave the current study inconclusive on how to interpret formant displacement of /ɛ/. Two issues remain open: how formant displacement fits within the other acoustic evidence in favour of word stress and to what extent a reanalysis of the Papuan Malay vowel inventory is needed.

#### 4.5.3. Outlook

Overall, it can be concluded that Papuan Malay shows consistent acoustic evidence for the word stress claim by Kluge (2017) to the extent that it is non-phonemic and regularly located on the penultimate syllable. Acoustic evidence for word stress is found in all aspects of the speech signal; mainly F0 movement timing, duration per phoneme, formant displacement and spectral tilt. The results are in line with earlier work showing that duration measures were among the strongest correlates of word stress crosslinguistically (Gordon & Roettger, 2017). In light of stress claims for Trade Malay varieties, the outcome of the current study is not evident. It has been shown that earlier impressionistic stress claims (Van Minde, 1997) were not supported for Ambonese Malay (Maskikit-Essed & Gussenhoven, 2016). Therefore, the current study shows that Trade Malay varieties might differ in terms of word level prosody, although some essential differences in methodology have to be taken into account (Section 1.4). Future work should concentrate on the acoustic correlates of word stress in other Trade Malay varieties to further complete the empirical literature on these languages. Of special interest are varieties that did not lose schwa (Kupang, Larantuka, Ternate), which could shed further light on the role of schwa in Trade Malay word prosody.

The current results deserve further caution on several aspects. First, the stress hypothesis tested in the current study was taken over from Kluge (2017). Although consistent acoustic evidence was found, a more unbiased hypothesis would have been based on the basis of purely acoustic criteria. These criteria would then have to correspond with a specific expectation of the acoustic properties of a stressed syllable. While the outcome of such an investigation would constitute stronger empirical support on word stress, the approach would face a number of difficulties. Foremost, the acoustic realisation of word stress differs crosslinguistically (Section 1.3), which makes it challenging to find the acoustic criteria needed to call a syllable “stressed”. This is particularly complicated for languages like Papuan Malay, for which little to no acoustic research is available. The investigation outlined in this paper furthermore suggests that acoustic realisations differ depending on where in the word stress is realised. A second word of caution concerns Papuan Malay phrase prosody, of which

to date little is known. While recent work suggests that this language does not make use of pitch accents (Kaland & Himmelmann, 2019; Riesberg et al., 2018), a direct assessment of the functions of F0 in Papuan Malay phrases is lacking. While the influence of phrase level phenomena was kept to a minimum in the current study, it cannot be fully excluded that some of the words in the data subset were subject to acoustic processes resulting from phrase prosody. Whether or not these processes originated from accent placement cannot be concluded at this stage. The interaction between word level and phrase level prosody is left for future research. Third, it remains to be investigated to what extent the word stress correlates found here are perceptually relevant. Minimal stress pairs are not reported for Papuan Malay (Kluge, 2017) and the acoustic differences found in the current study could therefore turn out to be irrelevant for listeners. An investigation of the extent to which the acoustic cues facilitate word recognition is planned in a follow-up study.

## Acknowledgements

This research is funded by the Deutsche Forschungsgemeinschaft (DFG), project SFB-1252 “Prominence in Language”. As for (1) data collection (funded by the Volkswagen Foundation within its DoBeS program), (2) translation/transcription of intonation units and (3) annotation/segmentation of words and syllables the author thanks: Apriani Arilaha (1,2), Maria Bardaji i Farré (3), Lisa Barz (3), Christoph Bracks (3), Pascal Coenen (3), Nikolaus P. Himmelmann (1,2), Jan-Niklas Linnameier (3), Sonja Riesberg (1,2), Anna Rumaikew (2), Emanuel Tutuop (2), Volker Unterladstetter (2), Boas Wabia (2), Kristian Walianggan (1) and Katja Wiesner (3). The author is furthermore grateful to Rob Goedemans, Nikolaus Himmelmann, Vincent van Heuven and two anonymous reviewers for valuable comments on previous versions of this paper, to Maximilian Hörl for statistical advice and to Sarah Agius for proofreading. Parts of this work are based on an earlier study (Kaland, 2018).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.wocn.2019.02.003>.

## References

- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed). New York: John Wiley & Sons.
- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Alieva, N. F., Arakin, V. D., Ogloblin, A. K., & Sirk, Yu. H. (1991). *Bahasa Indonesia: Deskripsi dan teori [The Indonesian language: Description and theory]*. Yogyakarta: Penerbit Kanisius.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer. Version 6.0.28. (<http://www.praat.org/>).
- Bolinger, D. (1958). A theory of pitch accents in English. *Word*, 14, 109–149.
- Breckenridge, J. (1977). The declination effect. *Journal of the Acoustical Society of America*, 61(S1), S90.
- Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretical approach* (2nd ed.). New York: Springer.
- Caballero, G., & Carroll, L. (2015). Tone and stress in Choguita Rarámuri (Tarahumara) word prosody. *International Journal of American Linguistics*, 81, 459–493.
- Cambler-Langeveld, C. M. (2000). Temporal marking of accents and boundaries. PhD Dissertation. LOT dissertation series 32.
- Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. Proceedings of the ISCA workshop on intonation: theory, models, and applications, 67–70. Athens, Greece.
- Campbell, W. N., & Isard, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics*, 19, 37–47.
- Chafe, W. L. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Chafe, W. L. (1994). *Discourse, consciousness, and time*. Chicago: The University of Chicago Press.
- Chitoran, I. (2001). *The phonology of Romanian: A constraint-based approach*. Mouton de Gruyter: Berlin & New York.
- Crosswhite, K. (2004). Vowel reduction. In B. Hayes, D. Steriade, & R. Kirchner (Eds.), *Phonetically based phonology* (pp. 191–231). New York: Cambridge University Press.
- Cutler, A. (2005). Lexical Stress. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of speech perception*. Oxford: Blackwell.
- Dogil, G. (1999). The phonetic manifestation of word stress in Lithuanian, Polish, German, and Spanish. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 273–311). New York: Mouton de Gruyter.
- Domahs, U., Genc, S., Knaus, J., Wiese, R., & Kabak, B. (2013). Processing (un-) predictable word stress: ERP evidence from Turkish. *Language and Cognitive Processes*, 28(3), 335–354.
- Domahs, U., Knaus, J., Orzechowska, P., & Wiese, R. (2012). Stress “deafness” in a language with fixed word stress: an ERP study on Polish. *Frontiers in psychology*, 3, 439.
- Donohue, M., & Sawaki, Y. W. (2007). Papuan Malay pronominals: Forms and functions. *Oceanic Linguistics*, 46(1), 253–276.
- Donohue, M. (2003). Papuan Malay. Singapore: National University of Singapore. <[http://papuan.linguistics.anu.edu.au/Donohue/downloads/Donohue\\_2003\\_PapuanMalay.pdf](http://papuan.linguistics.anu.edu.au/Donohue/downloads/Donohue_2003_PapuanMalay.pdf)> Accessed 2018-03-02.
- Eriksson, A., Bertinetto, P. M., Heldner, M., Nodari, R., & Lenoci, G. (2016). The acoustics of lexical stress in Italian as a function of stress level and speaking style. In *INTERSPEECH 2016, San Francisco, USA, September 8–12, 2016* (pp. 1059–1063). International Speech Communication Association.
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole Publishing.
- Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 232–276). Cambridge, U. K.: Cambridge University Press.
- Fletcher, J., & McVeigh, A. (1993). Segment and syllable duration in Australian English. *Speech Communication*, 13(3–4), 355–365.
- Goedemans, R. W. N., & Van Zanten, E. (2007). Stress and accent in Indonesian. In V. J. van Heuven & E. van Zanten (Eds.), *Prosody in Indonesian languages* (pp. 35–62). Utrecht: LOT.
- Goedemans, R. W. N., & Van Zanten, E. (2014). No stress typology. In J. Caspers, Y. Chen, W. F. L. Heeren, J. Pacilly, N. Schiller, & E. van Zanten (Eds.), *Above and beyond the segments: Experimental linguistics and phonetics* (pp. 83–95). Amsterdam: Benjamins.
- Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, 3(1). <https://doi.org/10.1515/lingvan-2017-0007>.
- Halim, A. (1981). *Intonation in relation to syntax in Bahasa Indonesia*. Jakarta: Djambatan.
- Hanson, H. M. (1995). *Glottal characteristics of female speakers* Ph.D. dissertation. Cambridge, MA: Harvard University.
- Harrington, J. (2010). *Phonetic analysis of speech corpora*. West Sussex: John Wiley and Sons.
- Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics*, 31, 39–62.
- Himmelmann, N. P. (2018). Some preliminary observations on prosody and information structure in Austronesian languages of Indonesia and East Timor. In S. Riesberg, A. Shiohara, & A. Utsumi (Eds.), *A cross-linguistic perspective on information structure in Austronesian languages* (pp. 347–374). Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.1402555>.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363.
- Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, 10(2), 163–169. [https://doi.org/10.1016/0167-6393\(91\)90039-V](https://doi.org/10.1016/0167-6393(91)90039-V).
- Hussain, S. (1997). *Phonetic correlates of lexical stress in Urdu*. Northwestern University PhD Dissertation.
- Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America*, 121(4), 2283–2295.
- Jacob, J., & Grimes, C. E. (2003). *Kamus Pengantar Bahasa Kupang* [Introductory Dictionary of the Kupang Language]. Paradigma Centre for Regional Studies, Series B, No. 10). Kupang: Artha Wacana Press.
- Kaland, C.C.L., Galatà, V., Spreafico, L., & Vietti, A. (2019). Which Language R You Speaking? /r/ as a Language Marker in Tyrolean and Italian Bilinguals. *Language and Speech*, 62(1), 137–163. <https://doi.org/10.1177/0023830917746551>.
- Kaland, C.C.L., & Himmelmann, N. P. (2019). Repetition reduction revisited: The prosody of repeated words in Papuan Malay. *Language and Speech*, 1–25. <https://doi.org/10.1177/0023830918820044>.
- Kaland, C.C.L. (2018). Spectral tilt as a correlate of Papuan Malay word stress. Proc. 9th International Conference on Speech Prosody 2018, pp. 339–343, doi: 10.21437/SpeechProsody.2018-69.
- Kluge, A. (2017). A grammar of Papuan Malay. Studies in Diversity Linguistics 11. Berlin: Language Science Press. doi: 10.17169/langsci.b78.35.
- Kumanireng, Th. Y. (1993). *Struktur Kata dan Struktur Frasa Bahasa Melayu Larantuka* [Word structure and Phrase Structure in Larantuka Malay]. Ph.D. dissertation. University of Indonesia.



- Laksman, M. (1994). Location of stress in Indonesian words and sentences (Semaian 9). In C. Odé & V. J. van Heuven (Eds.), *Experimental studies of Indonesian prosody* (pp. 108–139). Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Leiden University.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839–862.
- Lindström, E., & Remijsen, B. (2005). Aspects of prosody of Kuot, a language where intonation ignores stress. *Linguistics*, 43, 839–870.
- Litamahuputty, B. (2012). *Temate Malay: Grammar and texts* LOT Dissertation Series 307. Utrecht: LOT.
- Maskikit-Essed, R., & Gussenhoven, C. (2016). No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay. *Phonology*, 33, 353–389. <https://doi.org/10.1017/S0952675716000154>.
- Mayer, J. (2014). Spectral\_profile.txt. Version 2014/05/02. <http://praatpfanne.lingphon.net/>.
- Odé, C. (1994). On the perception of prominence in Indonesian (Semaian 9). In C. Odé & V. J. van Heuven (Eds.), *Experimental studies of Indonesian prosody* (pp. 27–107). Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Leiden University.
- Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. London: Methuen.
- Ortega-Llebaria, M., & Prieto, P. (2010). Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, 54(1), 73–97. <https://doi.org/10.1177/0023830910388014>.
- Paauw, S. H. (2008). *The Malay contact varieties of Eastern Indonesia: A typological comparison* PhD Dissertation. Buffalo: State University of New York.
- R Core Team. (2017). R: A language and environment for statistical computing. Computer program, v. 3.4.0, retrieved April 21st 2017 from <https://www.r-project.org>.
- Remijsen, B., & Van Heuven, V. (2005). Stress, tone, and discourse prominence in the Curacao dialect of Papiamentu. *Phonology*, 22, 205–235.
- Riesberg, S., & Himmelmann, N. P. (2012–2014). Papuan Malay. Summits-page collection. DoBeS Archive MPI Nijmegen, <http://www.mpi.nl/DOBES/>.
- Riesberg, S., Kalbertodt, J., Baumann, S., & Himmelmann, N. P. (2018). On the perception of prosodic prominences and boundaries in Papuan Malay. In S. Riesberg, A. Shiohara, & A. Utsumi (Eds.), *A cross-linguistic perspective on information structure in Austronesian languages* (pp. 389–414). Berlin: Language Science Press. doi: 10.5281/zenodo.1402559.
- Roettger, T., & Gordon, M. (2017). Methodological issues in the study of word stress correlates. *Linguistics Vanguard*, 3(1). <https://doi.org/10.1515/lingvan-2017-0006>.
- Roosman, L. (2007). Melodic structure in Toba Batak and Betawi Malay word prosody. In V. J. van Heuven & E. van Zanten (Eds.), *Prosody in Indonesian Languages* (pp. 89–115). Leiden: Leiden University Center for Linguistics.
- Samsuri, J. (1971). *Ciri-ciri prosodi dalam kalimat Bahasa Indonesia* [Prosodic features in the Indonesian sentence]. Flores: Nusa Indah.
- Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of final-lengthening in English. *Proceedings of 16th international congress on acoustics and 135th meeting Acoustical Society of America* (2, pp. 1235–1236).
- Sluijter, A. M. C., Shattuck-Hufnagel, S., Stevens, K. N., & Van Heuven, V. J. (1995). Supralaryngeal resonance and glottal pulse shape as correlate of stress and accent in English. *Proceedings of ICPhS*, 630–633, Stockholm.
- Sluijter, A. M. C., & Van Heuven, V. J. (1996). Spectral balance as an acoustic correlate for linguistic stress. *Journal of the Acoustical Society of America*, 100, 2471–2485.
- Steinhauer, H. (1983). Notes on the Malay of Kupang. In J. T. Collins (Ed.), *Studies in Malay dialects: Part II (NUSA – Linguistic Studies of Indonesian and other Languages in Indonesia 17)* (pp. 42–64). Jakarta: Badan Penyelenggara Seri NUSA, Universitas Katolik Atma Jaya.
- Stevens, K. N., & Hanson, H. M. (1994). Classification of glottal vibration from acoustic measurements. In O. Fujimura & M. Hirano (Eds.), *Vocal fold physiology: Vocal quality control* (pp. 147–170). San Diego: Singular Publishing Group.
- Stoel, R. B. (2007). The intonation of Manado Malay. In V. J. van Heuven & E. van Zanten (Eds.), *Prosody in Indonesian Languages* (pp. 117–150). Utrecht: LOT.
- Streefkerk, B. M. (2002). *Prominence: acoustic and lexical/syntactic correlates*. LOT (Vol. 58).
- Teeuw, A. (1984). *Leerboek Bahasa Indonesia* [Indonesian textbook] (4th ed.). Groningen: Wolters-Noordhoff.
- Traunmüller, H., & Eriksson, A. (1994). *The frequency range of the voice fundamental in the speech of male and female adults* Unpublished manuscript. S-106 91 Stockholm, Sweden: Institutionen för lingvistik, Stockholms universitet.
- Van Heuven, V. J., Roosman, L., & Van Zanten, E. (2008). Betawi Malay word prosody. *Lingua*, 118, 1271–1287.
- Van Heuven, V. J., & Van Zanten, E. (1997). Effects of substrate language on the localization and perceptual evaluation of pitch movements in Indonesian. In C. Odé & W. Stokhof (Eds.), *Proceedings of the 7th International Conference on Austronesian Linguistics* (pp. 63–79). Amsterdam/Atlanta: Rodopi.
- Van Minde, D. (1997). *Malayu Ambong: Phonology, morphology, syntax* PhD dissertation. Leiden University.
- Van Staden, M. (2000). *Tidore: A linguistic description of a language of the North Moluccas* Ph.D. Dissertation. Leiden, The Netherlands: Leiden University.
- Vogel, I., Athanasopoulou, A., & Nadia, P. S. (2016). Prominence, contrast, and the functional load hypothesis: An acoustic investigation. In J. Heinz, R. Goedemans, & H. van der Hulst (Eds.), *Dimensions of phonological stress* (pp. 123–167). Cambridge: Cambridge University Press.