JASA ARTICLE

Offline and online processing of acoustic cues to word stress in Papuan Malay

Constantijn Kaland[a)]

Institut für Linguistik, SFB-1252, Universität zu Köln, Luxemburger Straße 299, 50939 Köln, Germany

ABSTRACT:

The present study investigates to what extent acoustic cues to word stress facilitate both offline and online word processing in Papuan Malay. Previous production research has shown acoustic evidence for word-stress patterns in this language, counter to earlier predictions. A discussion of the literature on word stress perception and word stress in Papuan Malay is provided and complemented with reports of three word recognition tasks. The first two presented sequences of acoustically manipulated syllable dyads to native listeners in an offline word recognition task. This was done in order to investigate the individual contribution of each of the acoustic cues ($F0$, duration, intensity, spectral tilt) to the perception of word stress. $F0$ appeared a relevant cue when stimuli were embedded in a phrase, not in isolation. A follow-up reaction time experiment (online processing) investigated to what extent word recognition was facilitated when either an acoustically weak or an acoustically strong syllable was the cue to identify a word. Reaction times were shorter for strong syllables than for weak syllables. The outcomes suggest that Papuan Malay has a form of perceptually relevant word stress, which is particularly salient for irregular (ultimate) stress rather than for regular (penultimate) stress. © 2020 Acoustical Society of America. https://doi.org/10.1121/10.0000578

I. INTRODUCTION

The current study follows from acoustic analysis of word stress cues[1] in Papuan Malay (Kaland, 2019) and a preliminary investigation of its phonological status (Kaland *et al.*, 2019). Several acoustic correlates were found to be strong indicators of word stress in this language, despite the predictions that word stress is absent in (Eastern) Indonesian languages, including Papuan Malay (e.g., Goedemans and Van Zanten, 2014; Riesberg *et al.*, 2018, Himmelmann, 2018).

Crucially, the crosslinguistic empirical studies investigating word stress reach beyond acoustic analyses. Most of the acoustic, phonological, perceptual, and communicative aspects of word stress have been found to be present to a greater or lesser *degree* in a given stress language. This suggests that describing word stress as strictly present or absent is not a fruitful approach. Indeed, perception research found that French listeners have difficulties discriminating nonsense words that differ only in stress patterns (Dupoux *et al.*, 1997), suggesting that word stress has no function in French. However, the literature has not been clear on how to classify this language: it has "fixed stress" (Di Cristo, 1998, p. 196), "no primary word stress at all" (Van der Hulst, 2012, p. 1515), or is a language for which "the need for *w*-final accents is not hard to establish" (Gussenhoven, 2004, p. 258). Importantly, it has been argued that it would be possible to make French listeners sensitive to acoustic differences at the word level and that this sensitivity might be

used to detect word boundaries (Dupoux *et al.*, 1997), a function that has been shown for Dutch (Vroomen *et al.*, 1996), which has uncontroversially been described as a stress language.

Thus, it appears that perceptual research can complement existing production studies in a crucial way. It is insufficient to describe word stress only acoustically or only in terms of its lexical specification, as it interacts with different levels in the prosodic structure and might fulfill the same function in phonologically different languages. Concerning Papuan Malay, this study addresses the aforementioned issues by investigating both the perceptual relevance of acoustic cues (offline) and the potential communicative function (online) of Papuan Malay word stress. The remainder of this section discusses the status of the research on word stress in Papuan Malay and related languages in particular (Sec. I A) and the literature on word stress perception crosslinguistically (Sec. I B). Section I C presents the research questions and hypotheses.

A. Word stress in Papuan Malay and related languages

Papuan Malay is a variety of Trade Malay spoken in Indonesia's two easternmost provinces Papua and Papua Barat (Kluge, 2017). This section summarizes the current state of research on word stress in Papuan Malay and related languages.[2]

Papuan Malay is reported as having regular word stress on the penultimate syllable (Donohue and Sawaki, 2007; Kluge, 2017), except when that syllable contains /ɛ/ (which

[a)]Electronic mail: c.c.l.kaland@outlook.com

is claimed to reduce to schwa /ə/, but see Kaland, 2019 for the status of /ɛ/ in word stress). For example, [ˈbɐn.tu] "to help" and [pɛr.ˈgi] "to go." Some exceptions to the stress placement rules can be found. In the 1040 reported native Papuan Malay words (Kluge, 2017), 61 had penultimate stress when that syllable contained /ɛ/ (e.g., [ˈdɛ.pɐn], "front") and three words had ultimate stress when the penultimate syllable did not contain /ɛ/ (e.g., [kʊ.ˈmʊr], "to rinse mouth"). A more detailed investigation of possible phonological factors affecting stress position revealed that not only /ɛ/, but also /ɔ/ occurs mostly in unstressed positions compared to stressed positions (Kaland et al., 2019). Both /ɛ/ and /ɔ/ are the only vowels appearing in mid positions in the Papuan Malay acoustic space. This indicates that word stress tends to fall on the corner vowels /a/, /i/, and /u/, in line with crosslinguistic observations (Crosswhite, 2004). Furthermore, stressed syllables appeared to have generally higher vocalic portions (V, CV, VC) compared to unstressed syllables (CVC, CCV, CCVC), suggesting that word stress is sensitive to the syllable structure (i.e., vowel reduction; Kaland, 2019). For disyllabic Papuan Malay roots (loanwords excluded), consistent acoustic evidence was found to show that stressed syllables are more prominent than unstressed syllables (Kaland, 2019). The recorded material that was investigated consisted of spontaneous speech from 19 native speakers, elicited in a story-retelling task. A large number of acoustic cues were investigated in order to reveal to what extent they correlated with the presumed stress-patterns. The three most indicative acoustic correlates were segment duration (stressed syllables longer than unstressed ones), formant displacement (stressed syllables more displaced than unstressed ones), and spectral tilt (stressed syllables have more energy in higher frequency bands than unstressed ones; see also Kaland, 2018). The timing of $F0$ also appeared to be a relatively strong correlate, although this was likely related to the effect of duration. Furthermore, it appeared that, when compared to unstressed syllables, ultimate stress was realized with larger acoustic differences than penultimate stress. This finding confirms the rather exceptional (i.e., marked) status of ultimate stress in Papuan Malay. As further discussed below, recent outcomes indicating stress distinctions in Papuan Malay seem to contradict other findings on this language, other Trade Malay varieties, and Indonesian varieties.

A recent study compared prosodic annotations by native Papuan Malay speakers (Riesberg et al., 2018). Annotators indicated their perception of prominences and boundaries at the phrase level using rapid prosody transcription (RPT; Mo et al., 2008), after which the agreement between the annotations was computed. Results indicated mainly agreement for the boundaries and little agreement for the prominences. It was concluded that "prosodic prominence may not be a relevant category in PM [Papuan Malay]" and that the outcome would be in line with studies showing that word stress is largely absent in related Malay varieties (Riesberg et al., 2018). In a follow-up RPT study the Papuan Malay phrases were compared to a similar set of phrases in German, a

language in which prominence is used to mark information structure (e.g., Grice and Baumann, 2002). Both sets were annotated by speakers of each language and the results showed that Papuan Malay annotators were more consistent in indicating prominences for the German phrases than they were for the Papuan Malay phrases. This suggests that the Papuan Malay annotators perceived prominence to some extent. It remains to be seen, therefore, to what extent this group perceives prominence at the word level, as the RPT method mainly targets phrase prosody. Thus, these studies show that prosodic phenomena in Papuan Malay phrases are confined to boundaries. This outcome is in line with a study on phrase-final $F0$ movements (Kaland and Baumann, 2019), which found that the largest movements systematically occurred on the final two syllables in spontaneous Papuan Malay phrases. Furthermore, these movements were shown to correlate with penultimate word stress, as reported in Kluge (2017). That is, rising $F0$ movements occurred more often on penultimate stressed syllables than on ultimate unstressed syllables. This suggests an interaction between word stress and phrase prosody, explaining the common report of phrase-final rise-fall patterns (Himmelmann, 2018) in line with phrase prosodic accounts of word stress (Gordon, 2014).

For Manado Malay and Ambonese Malay, two other Trade Malay varieties, word stress has also been reported to occur regularly on the penultimate syllable (Stoel, 2007 and Van Minde, 1997, respectively). Although the acoustic information for Manado is limited to $F0$ contours, Ambonese has been re-analyzed using acoustic measures (Maskikit-Essed and Gussenhoven, 2016). It was shown that alleged minimal stress pairs actually consisted of segmental differences, leading to a re-analysis of the vowel inventory. In addition, no acoustic support was found to show that alleged stressed syllables were more prominent than unstressed ones. The few studies available on word stress in Trade Malay suggest that there might be essential differences among the varieties, an observation that also holds for other Indonesian languages. It was found, for example, that Javanese listeners had no preference for the location of the stressed syllable, whereas Toba Batak listeners did (Goedemans and Van Zanten, 2007). As further outlined in the next section, more stress perception research is required to complement the existing (mainly production-based) studies on languages in this area. Furthermore, perception research on these languages shed light on the strategies that listeners use to process speech, which are language-specific by nature (Cutler, 2012) and further discussed in the next Sec. I B.

B. Word stress perception

Research on stress perception has indicated which acoustic cues are relevant to signal stress in offline tasks (cue weighting) and how stress can contribute to the online processing of speech, in particular to the segmentation and identification of words. This section discusses a number of

these studies on diverse languages to provide an overview of which perceptual aspects of stress are relevant crosslinguistically.

1. Cue weighting

Concerning cue weighting, perception studies have focused on spectral, temporal, and amplitudinal aspects of the acoustic speech signal in order to investigate which (combination of these) contribute most to the perceived difference between stressed and unstressed syllables. For English, an investigation was initially carried out for duration, intensity, and $F0$, where $F0$ appeared to produce the strongest effects (Fry, 1958). However, the words used in the perception experiment were produced in focus (Fry, 1955), which is generally marked by (phrasal) $F0$ movements (i.e., pitch accents) in English, see also Beckman and Edwards (1994). In general, $F0$ has been claimed to be a better correlate of phrase level prosody than of word prosody (Sluijter and Van Heuven, 1996; Gordon, 2014). This claim is supported by a number of studies on English indicating that $F0$ is a weaker cue to word stress than duration (Adams and Munro, 1978; Isenberg and Gay, 1978), or intensity (Lieberman, 1959; Beckman, 1986). Still, studies on the (produced) acoustic correlates of word stress in other languages have confounded word-level and phrase-level phenomena (Roettger and Gordon, 2017), which in turn might have created difficulties for studies testing the perceptual relevance of these correlates.

An additional reason why the search for perceived stress cues did not result in a well-defined (universal) list is that languages differ substantially in the individual importance of the potential stress cues. It has been shown, for example, that $F0$ does not signal, or only weakly signals, stress in languages where this cue is used for tonal contrasts between words (i.e., tone languages; Gordon and Roettger, 2017). Furthermore, while the effects of $F0$ on the perception of English stress are disputed, in German word-level prominence perception this cue is stronger than others (Kohler, 2008; Niebuhr and Winkler, 2017).

Importantly, research has shown that word-stress cues are not necessarily found only in the prosody (i.e., suprasegmentals) of a language, but can also exist at the segmental level. This is best illustrated by studies on English, which found that vowel quality is a more reliable stress cue than $F0$, duration, or intensity (e.g., Fear et al., 1995). In Dutch, however, suprasegmental cues do play an important role in the perception of stress, in particular duration and a spectrally weighed measure of intensity (i.e., spectral tilt; Sluijter and Van Heuven, 1996). The difference between English and Dutch stress was confirmed in word recognition tasks (Cooper et al., 2002; Cutler et al., 2007). In these tasks, Dutch listeners outperformed English listeners as the identification responses of the former correlated better with suprasegmental cues (intensity, $F0$) than those of the latter. However, suprasegmental cues are not redundant for English listeners, who do use them when segmental cues are

controlled for, as shown in both offline (Cooper et al., 2002) and online (Jesse et al., 2017) word recognition tasks. This outcome illustrates that despite the availability of several cues, listeners do not always need to use all of them to successfully process speech (see also Connell et al., 2018).

2. Stress among different word recognition cues

The role of suprasegmentals relative to other cues in word recognition is often illustrated by minimal stress pairs, in which suprasegmental cues are the only cue available to distinguish between words (e.g., in Dutch: /ˈkaːnɔn/ and /kaːˈnɔn/, translating to "canon" in the musical and military sense, respectively). Despite their use in experimental paradigms, minimal word pairs where suprasegmental stress cues are the only distinguishing feature are crosslinguistically rare (Cutler, 2005). This predicts that suprasegmental stress cues are rarely decisive for word recognition. Indeed, contextual, lexical, and segmental cues often outweigh suprasegmentals in stress perception, as shown by a large body of psycholinguistic literature (e.g., Mattys et al., 2005, see Cutler, 2012; p. 144). Despite their marginal role compared with other cues, suprasegmental stress cues can still be useful in listeners' online speech processing as soon as they are available (Reinisch et al., 2010; Sulpizio and McQueen, 2012). It has been shown that listeners of languages with word-initial stress such as Slovak (Hanulíková et al., 2010) and Finnish (Suomi et al., 1997) used stressed syllables to detect word onsets. Also in languages with more variable stress patterns such as English and Dutch, there appears to be a tendency for stress to be located on the word-initial syllable (English: Cutler and Carter, 1987; Dutch: Van Heuven and Hagman, 1988). Listeners of these languages were also found to default to a word-initial stress pattern when reporting induced misperceptions (Cutler and Butterfield, 1992; Vroomen et al., 1996). Given that two-syllable words are relatively common crosslinguistically (e.g., Vihman and Croft, 2007), and given that stress is penultimate in the majority of the stress languages (see Gordon, 2014 for an account), it is not surprising to find word-initial stress as a helpful segmentation cue in languages that occupy different positions on the fixed-variable stress continuum.

3. Stress predictability and perception

Crosslinguistic differences regarding the functional role of stress in perception have been more clearly observed when taking into account the distribution of word stress in the lexicon, in particular, the ratio regular/irregular word-stress patterns (i.e., predictability of stress). This helps to explain why listeners of French, Finnish, and Hungarian have more difficulties recalling the correct stress placement in a two-syllable nonsense word compared to listeners of Spanish (Peperkamp et al., 2010). Word-stress placement in the first three languages has no exceptions, whereas in Spanish many exceptions to the default pattern exist. This explanation was further confirmed by listeners of Polish, a

J. Acoust. Soc. Am. **147** (2), February 2020

Constantijn Kaland   733

language with a small number of stress exceptions, who performed better than French, Finnish, and Hungarian listeners, but worse than Spanish listeners. That is, for a lexicon with predominantly predictable stress patterns, there is less need to store the stress information, and this need increases with increasing numbers of exceptions (Peperkamp *et al.*, 2010). In a similar vein, the realization of stress cues was reported to be acoustically weaker for languages with fixed stress than for languages with variable stress (Dogil, 1999). It is furthermore known that in Polish, listeners appear to be sensitive predominantly to the irregular stress pattern as opposed to the regular pattern (Domahs *et al.*, 2012). The exclusive sensitivity to irregular patterns was also found for languages such as Turkish (Domahs *et al.*, 2013) and Italian (Sulpizio and McQueen, 2012). Indeed, in all these languages, the number of irregular stresses is small, and the processing mechanism responsible for word recognition can efficiently default to the regular stress pattern whilst remaining sensitive to deviations from it (Sulpizio and McQueen, 2012). Similarly, the stress distribution of Papuan Malay shows a small number of irregular patterns, although it remains to be seen how this affects perception. The remaining issues investigated in this study are further outlined in Sec. I C.

C. Research questions and hypotheses

The current study hypothesizes that listeners of Papuan Malay are sensitive to word-stress cues. This hypothesis is further split into two sub-aspects of word stress: the relative importance of acoustic cues and the usefulness of these cues for speech processing.

As for the first, it is hypothesized that duration and spectral tilt are among the strongest stress cues, mirroring the acoustic realization. Vowel quality, which was also found to be a strong acoustic correlate, showed puzzling results with regard to the use of /ε/ (or schwa; see Kaland, 2019). Given that much of the stress distribution actually depends on the presence of /ε/ in Papuan Malay, a separate study should be devoted to the investigation of segmental cues. The current study is thus limited to the suprasegmental cues to word stress in this language.

Concerning the second, Papuan Malay could classify as having rather fixed stress (Sec. I A), for which it is not necessarily expected that stress cues play a role in word identification. Rather, given their fixed location, these cues tend to facilitate word segmentation (Sec. I B). There are, however, valid reasons to investigate the extent to which rather fixed stress patterns in Papuan Malay facilitate word identification. First, the current literature on word identification focuses predominantly on languages with variable stress patterns, leaving fixed stress patterns largely unexplored (Cutler, 2005, pp. 282–283). Furthermore, Papuan Malay does not appear to be a fixed-stress language for which acoustic cues are only weakly realized (cf. Kaland, 2019 and Dogil, 1999). That is, the acoustic signal provides the Papuan Malay listener with multiple consistent

suprasegmental cues (Kaland, 2019). The question, therefore, remains whether these cues are indeed exploited for word identification and, if they are, whether the regular pattern, irregular pattern, or both can help to identify the word. Given the limited number of irregular (ultimate) stress patterns in Papuan Malay, listeners are expected to be sensitive to at least these (Domahs *et al.*, 2012; Domahs *et al.*, 2013; Sulpizio and McQueen, 2012). The current study investigates only the potential word-identification effects of the regular pattern (penultimate stress in Papuan Malay). It appears that in spontaneous speech, word-initial stress has an advantageous effect on phoneme detection (Mehta and Cutler, 1988; McAllister, 1991). It can, therefore, be predicted that in Papuan Malay, the penultimate stress pattern in disyllabic words facilitates word identification. If that is the case, then the prediction follows that the facilitation effect is larger when acoustic differences between stressed and unstressed syllables are produced more clearly.

The hypotheses discussed above are investigated in three word recognition experiments. The relevance of the individual stress cues is assessed using acoustically manipulated syllable sequences in offline tasks (Experiment I and II). In Experiment I, syllable sequences were presented in a phrase to resemble natural speech. In Experiment II, syllable sequences were presented in isolation to minimize the interference of phrase prosody. The potential functionality of stress cues for online word recognition was tested in a reaction time experiment (Experiment III). Sections II, III, and IV describe each of the experiments and Sec. V provides an overall conclusion and general discussion.

II. EXPERIMENT I

Experiment I was designed to investigate the perceptual relevance of the acoustic correlates of word stress in Papuan Malay. The experiment consisted of a forced-choice word recognition task. Participants in the experiment listened to matrix phrases in which an acoustically manipulated two-syllable sequence was presented. The participants' task was to choose one out of two written two-syllable words that would fit best with the manipulated sequence. Their choice had to be made between a word with penultimate word stress and a word with ultimate word stress. Two-syllable words were chosen as this is the most frequent word length in Papuan Malay (Kluge, 2017).

A. Preparation of the stimulus material

Acoustic manipulations were carried out on the extracted syllable "ma" taken from a non-final unstressed position in a recorded matrix phrase in Kluge *et al.* (2014, "2417_makanya.wav"). The phrase was produced by a male speaker of Papuan Malay (see Kluge, 2017, pp. 62–63 for recording details). This syllable was chosen such that subsequent acoustic manipulations would not be interrupted. That is, in voiced segments (nasal and vowel in "ma") $F0$ is continuously present, unlike, for example, in voiceless stops or fricatives. Before manipulation of the word stress correlates,

Constantijn Kaland

the syllable "ma" was filtered using a sequence of second order filters in Praat [Boersma and Weenink, 2019, Praat manual page "PointProcess: To Sound (hum)…"]. This filtering procedure made the segmental content unintelligible by converting the formant frequencies into values that did not vary over time (see Fig. 1). The result resembles "hummed speech," which has been used in perception research before (e.g., Hart *et al.*, 1990). The advantage of this filtering method is the absence of segmental information whilst preserving the presence of formant frequencies. In earlier work (e.g., Van Bezooijen and Gooskens, 1999), experimental settings in which "prosody-only" was presented to participants, speech samples were low-pass filtered using a cut-off frequency of 350 Hz. This method is not suitable in the current study, as it does not allow for manipulation of spectral tilt, which requires high frequencies to be present in the signal. The acoustic properties of the filtered syllable, which was taken as the template for subsequent manipulations, are given in Table I.

Before actual word stress manipulations were generated, two versions of the filtered syllable were created. This was done to concatenate the two versions into a sequence that represented a two-syllable word. One version in which duration was not changed represented the first syllable. In another version, the duration was lengthened by a factor 1.06 (see respective duration values in Table I). This factor corresponded to the overall relative difference found between first (121 ms) and second syllables (128 ms) in the
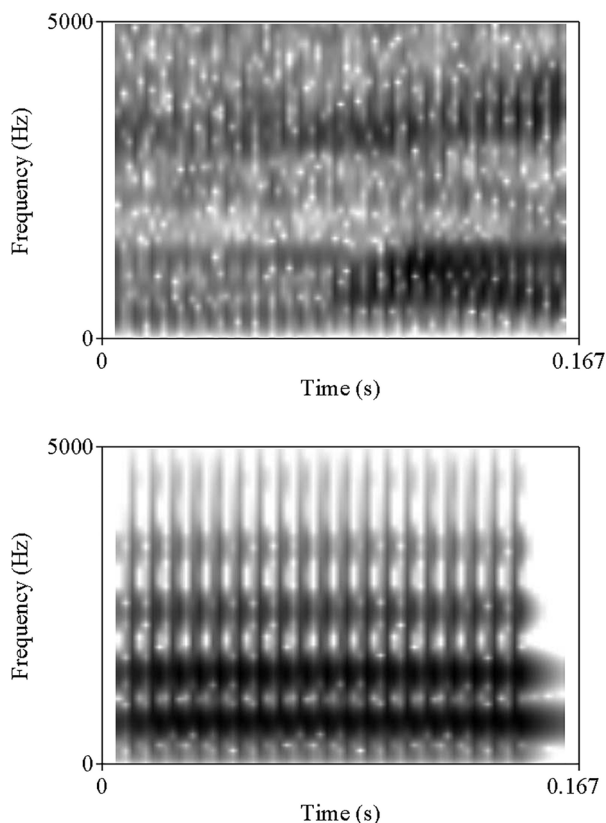
TABLE I. Acoustic properties and manipulation values before and after application of the manipulation to the filtered template syllable "ma" for each of the acoustic cues. Measures of overall intensity (O) in dB and of spectral tilt (T) in dB and H1-A2. Manipulation values based on Kaland (2019).

Cue	Before	Manipulation	After
$F0$	170 Hz	+ 3 ST rise	202 Hz
Duration	167 ms / 177 ms	+ 20 ms	187 ms / 197 ms
Intensity (O)	75.52 dB	+ 3 dB	78.52 dB
Intensity (T): overall	75.52 dB	+ 4 dB > 500 Hz	78.64 dB
Intensity (T): H1-A2	−9.59 dB		−5.85 dB

Papuan Malay words in Kaland (2019, Table VI). The duration manipulation was done in order to simulate final lengthening in the word domain, which occurs naturally and irrespective of possible word-stress cues.

From the recordings, a matrix phrase was selected in which the manipulated sequences were embedded (Kluge *et al.*, 2014; "1353_manfaat.wav"). The matrix phrase was produced by the same speaker who produced the "ma" syllable: *ko pu kata* [sequence] *itu, sa blum tau* (that word [sequence] of yours, I don't yet know it). In this matrix phrase, the manipulated syllable sequence was embedded in phrase-medial position, replacing the original target word. The phrase-medial position was chosen to avoid possible interference of phrase(-final) prosody (Kaland and Baumann, 2019). In order to provide a suitable acoustic context for the $F0$ manipulations of the syllable sequence, the $F0$ before and after the sequence was adjusted in the matrix phrase. This was done by applying a pitch increase of 13 Hz (157 to 170 Hz) before the sequence and a pitch decrease of 10 Hz (212 to 202 Hz) after the sequence (Table I). In this way, the difference between "ta" in *kata* and "i" in *itu* was 3 semitones (ST), matching the $F0$ manipulations described in Sec. II B.

B. Manipulation of word stress correlates

Four acoustic correlates of word stress in Papuan Malay were manipulated in the syllable sequences: $F0$ movement, duration, overall intensity, and spectral tilt. All manipulations closely resembled the differences observed in the production data (Kaland, 2019), where duration and spectral tilt were found to be the strongest correlates of word stress.

The $F0$ movements observed in the production data showed an average change of around 2 ST. The production data partially consisted of syllables in which not all segments were voiced. Because $F0$ was present throughout the filtered syllable (i.e., there were no voiceless segments) in the current manipulations, an $F0$ change of 2 ST may have been too small to cue word stress perceptually. Therefore, $F0$ change in the manipulation was defined as a 3 ST rise over the course of the entire syllable. Note that the $F0$ alignment effects in Kaland (2019) did not provide information with respect to the exact segmental anchoring of F0 movements. That is, little is known about the relevance of onsets and/or codas for the alignment of pitch movements in

FIG. 1. Spectrograms of the template syllable "ma" before (top) and after (bottom) formant conversion.

Papuan Malay syllables. It could, therefore, be the case that parts of the syllable structure are irrelevant to pitch changes due to word stress. For this reason, a 3 ST rise furthermore increases the likelihood that a sufficient pitch change is present in the part of the syllable that potentially cues word stress.

Duration was manipulated as a 10 ms lengthening per segment, corresponding to the observed lengthening in the production data. The duration manipulation thus resulted in a 20 ms lengthening for the filtered syllable.

Although spectral tilt appeared a strong correlate of word stress in the production data (Kaland, 2019), it remains unclear to what extent intensity differences were the result of an intensity increase across the entire spectrum (overall) or in higher frequencies only (spectral tilt). Research shows that overall intensity and spectral tilt manipulations both affect the perceived loudness of the syllable (Sluijter and Van Heuven, 1996). For this reason, a 3 dB increase was chosen as the target value for both manipulations. That is, the overall intensity was increased by 3 dB by multiplying sound pressure values in the original sound by 1.41 ($10^{3/20}$). In order to obtain the same 3 dB overall intensity increase, spectral tilt was manipulated by amplifying all frequencies above 500 Hz by 4 dB. As a reference, H1-A2 was measured before and after manipulation. H1-A2 values in the stimulus material yielded naturally observed values (cf. Table I and Kaland, 2019).

C. Design

All combinations of acoustic manipulations were generated following a $2 \times 2 \times 3$ design with $F0$ rise (present/absent), lengthening (present/absent), and amplification (overall, spectral tilt, absent) as predictors. Intensity was manipulated as a three-level predictor to distinguish possible perceptual effects of overall intensity from those of spectral tilt (see Sec. II B). The sequences were designed by applying acoustic manipulations on either the first or the second syllable. From the resulting set ($N = 24$), sequences were removed in which manipulations were either all present or all absent on *both* syllables. This was done because these syllable sequences do not reflect word prosody in Papuan Malay; i.e., consistent acoustic evidence was found for one most prominent syllable per word (Kaland, 2019). In this way, the resulting set consisted of 22 manipulated syllable sequences (Appendix A).

For the syllables in which no $F0$ rise was present (either the first or the second or both), an additional $F0$ manipulation was applied to maintain a smooth $F0$ transition between the first and second part of the matrix phrase and the embedded sequence (Fig. 2). This was done to avoid a further decrease in the naturalness of the stimuli. For the sequences in which there was an $F0$ change on the first syllable, the $F0$ of the entire second syllable was set to the end value of the rise (i.e., monotonously at 202 Hz). For the sequences in which there was an $F0$ change on the second syllable, the $F0$ of the entire first syllable was set to the start value of the rise (i.e., monotonously at 170 Hz). This was done to match the $F0$ levels in the parts of
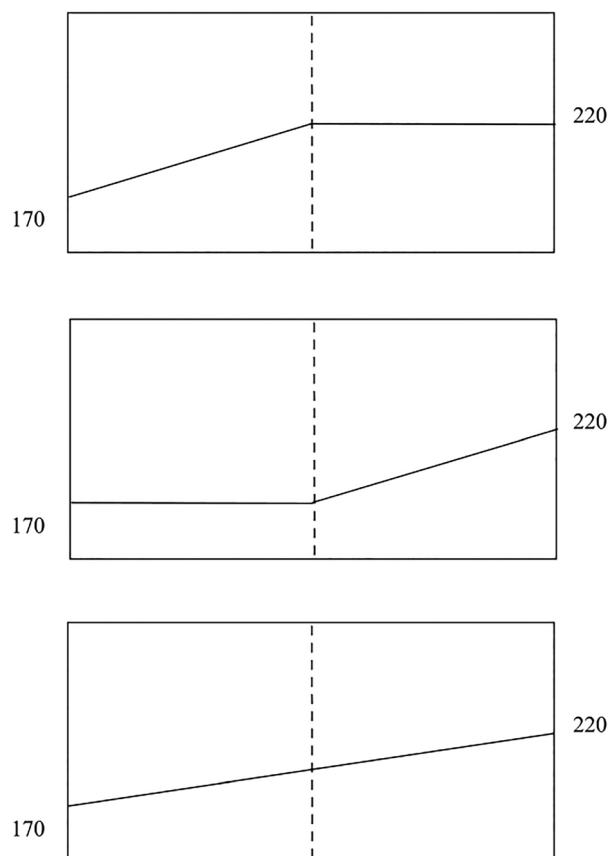

FIG. 2. Manipulation of an $F0$ rise (3 ST) on the first syllable only (top), the second syllable only (mid) or both (bottom). Numbers (Hz) and solid line represent $F0$; dashed line represents the syllable boundary.

the matrix phrase circumjacent to the sequence (170 and 202 Hz, respectively) and reflected the way the speaker uttered the elicited materials in Kluge (2017). For sequences in which neither the first nor the second syllable was specified for $F0$ change (i.e., both "absent," see Appendix A), a gradual 3 ST rise over the course of *both* syllables was applied (from 170 to 202 Hz). This was done to fit the $F0$ of the syllable sequence seamlessly within the overall contour of the matrix phrase. Although this manipulation still resulted in syllables with an $F0$ movement, the rise affected both syllables in a similar way and was therefore assumed to not highlight one specific syllable, as would be required to cue word stress.

After the sequences were created, they were embedded in the matrix phrase. To increase the perceptual salience of the sequences, a silent interval was added before and after the sequence of 75 and 100 ms, respectively. This supported the impression that the speaker had carefully articulated the word represented by the manipulated syllable sequences.

D. Participants

A total of 22 participants carried out Experiment I. All were students at the University of Papua; 13 male and 9 female participants ($M_{age} = 23.3$, age range 18–41), and were native speakers of Papuan Malay without hearing problems. They received a small present for their participation.

E. Setup and procedure

The word recognition task was designed using OpenSesame (Mathôt *et al.*, 2012). The experiment consisted of a script written in the programming language Python (Van Rossum and De Boer, 1991) and 22 wave files (Appendix A). In addition, a set of word pairs was selected which were presented as response options. The essential difference between the two words in a pair was the position of the stressed syllable following Kluge (2017): either on the penultimate or ultimate syllable. All words in the pairs consisted of two syllables to match the stimuli. Note that minimal word stress pairs are not reported for Papuan Malay. Therefore, the selection of word pairs was based on the difference in stress position plus one additional difference in the segmental makeup. In this way, a maximum of 42 near-minimal pairs could be selected from the Papuan Malay roots provided in Kluge (2017), as listed in Appendix B (e.g., *bebas* and *bekas* with penultimate and ultimate stress, respectively). The segmental difference could be a substitution, insertion, or deletion of either a vowel or a consonant. For each stimulus wave file, one randomly selected near-minimal word pair was taken to provide the response options. Stimuli were repeated five times within the experiment, resulting in a total of 110 presented stimuli. The order of the stimuli was random and different for each participant.

For each stimulus, the script generated a screen. The screen displayed a play button, a pair of words as buttons and a percentage counter. For each stimulus, participants were asked to indicate which word they thought would fit best with the manipulated syllable sequence (Fig. 3). They were instructed to pay close attention to the sound of the syllable sequence that represented the word. To play the stimulus, participants could click on the play button. The stimulus could be played as many times as needed and participants were explicitly instructed to listen multiple times. This was done to ensure careful listening to the stimuli. To choose one of the words, participants had to click on the word button displaying the word of their choice. The word buttons appeared at the bottom of the screen, only after the play button was clicked. This was done in order to make sure the participant had listened to the stimulus before making a choice. As soon as the participant made a choice, a screen
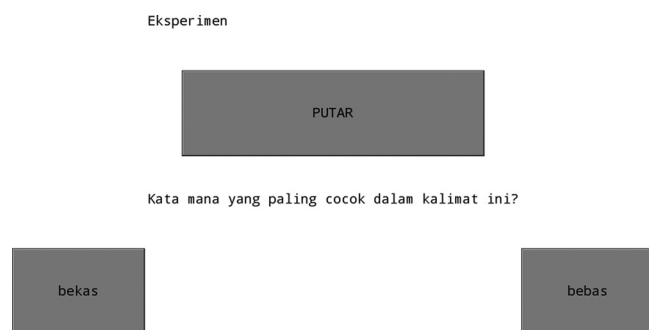

FIG. 3. Screen capture of Experiment I displaying the play button (*putar*), the question "Which word fits best in this phrase?" and two response options representing a near-minimal stress pair.

displaying the next stimulus would appear. Each of the word buttons appeared randomly on either the left or the right side of the screen, to balance the position of the words in the near-minimal pairs.

Before the start of the experiment participants received verbal instructions about the course of the tasks. Then, they took a seat behind a computer and completed the three subsequent parts of the experiment. First, participants entered their personal data. Second, they received instructions about their task both orally and written on the screen. To familiarize themselves with the task, participants completed a practice round consisting of five stimuli. At the end of the practice round, participants were asked whether they felt they needed to practice more or whether they were ready to start the actual task. When more practice was needed, participants were presented with additional stimuli. After each additional practice stimulus, participants could end the practice round. Third, when participants ended the practice session they were asked to start the actual task. Participants were instructed to switch off personal mobile devices and to use headphones during the entire experiment. After completing 50% of the actual task, participants were instructed to take a small break. The experiment lasted approximately 20 min. Responses were collected on the computer as 1 (correct) or 0 (incorrect). A response was considered correct when the member of the word pair was chosen that matched for stress position with the position of the manipulated cue(s) in the stimulus.

F. Statistical analysis

Statistical analyses were carried out using R (R Core Team, 2019) and the lme4 package (Bates *et al.*, 2015). One-sample Wilcoxon signed rank tests were carried out to investigate whether recognition choices were significantly different from chance level (0.50). This was done for all response values together, for the response values per syllable, and for the response values for each of the acoustic cues.

In addition, generalized linear mixed model (GLMM) analyses fit by maximum likelihood (Laplace approximation) were carried out on the response values (1 or 0). For each acoustic cue and syllable, a separate GLMM was carried out (six analyses). The acoustic cues were not put in the model together, as this would have computed their effects against a baseline in which none of them was manipulated (the intercept). The latter baseline condition did not exist in the stimulus set (see Sec. II C). Furthermore, the design of the cue manipulations (Appendix A) ensured that an effect of a given cue was computed against all combinations with/without other cues, eliminating the need to create a model where all cues are present. In addition, the model with all acoustic cues in interaction did not converge. Thus, in the final models, one of the following cues was added as a predictor: $F0$ rise (two levels: absent/present), duration (two levels: absent/present), or intensity (three levels: absent, overall, spectral tilt). To account for the potential additive effect of

TABLE II. Results of the Wilcoxon signed rank tests ($\mu = 0.50$) and the generalized linear mixed model analyses for the acoustic cues on each syllable in Experiment I.

Syllable	Cue	M correct		p
First	F0		0.52	n.s.
	Duration	0.50	0.47	n.s.
	Intensity (O)		0.45	= 0.06
	Intensity (T)		0.54	n.s.
Second	F0		0.58	<0.001
	Duration	0.56	0.54	<0.05
	Intensity (O)		0.53	n.s.
	Intensity (T)		0.59	<0.001

acoustic cues, a second predictor *number of cues* (3 levels: one, two, or three) was added to each of the final models with an acoustic cue as predictor.

A final GLMM on the correctness scores was run with the position of the manipulated cues (two levels: first, second) as a predictor. Participants and items (word pairs) were added as random slopes to all GLMMs.

G. Results

Overall, the mean of the response values was above chance level ($M = 0.53$, V $= 1\,550\,700$, $p < 0.01$). The Wilcoxon tests per syllable indicated chance level responses for the first syllable ($M = 0.50$, V $= 362\,690$, n.s.) and above chance level responses for the second syllable ($M = 0.56$, V $= 412\,950$, $p < 0.001$). The Wilcoxon tests per acoustic cue (Table II) indicated that for the first syllable overall intensity led to below chance level responses (trend), and that for the second syllable F0, duration, and spectral tilt led to above chance level responses.

In the GLMMs (Table III), F0 significantly improved correctness scores in both syllables, whereas duration significantly worsened the scores on the first syllable. The effects of the number of cues indicated that when duration was accompanied by both pitch and intensity cues in the first syllable, the correctness scores significantly increased compared to when duration was not manipulated. In the second syllable, correctness scores significantly decreased when F0 was

TABLE III. Results of the generalized linear mixed model analyses for each syllable and acoustic cue. Effects of number of cues (N cues) only reported when yielding significance.

Syllable	Cue	β	SE	z	p
First	F0	0.39	0.15	2.58	<0.05
	Duration	−0.35	0.15	−2.37	<0.05
	N cues (3)	0.39	0.21	1.85	= 0.06
	Intensity (O)	−0.20	0.17	−1.19	n.s.
	Intensity (T)	0.14	0.17	0.82	n.s.
Second	F0	0.35	0.16	2.24	<0.05
	N cues (3)	−0.47	0.22	−2.11	<0.05
	Duration	−0.24	0.15	−1.53	n.s.
	Intensity (O)	−0.26	0.18	−1.48	n.s.
	Intensity (T)	0.03	0.18	0.17	n.s.

accompanied by both duration and intensity cues compared to when F0 was not manipulated. The effect of cue position indicated that correctness scores were significantly higher for cues manipulated on the second syllable than those on the first syllable ($\beta = 0.28$, SE $= 0.08$, $z = 3.41$, $p < 0.001$).

H. Discussion

In general, participants scored above chance level, suggesting that the acoustic cues had a facilitating effect. Given the overall number of correct responses, however, this facilitation was minimal ($M = 0.53$). Participants scored above chance level only for the second syllable, due to all cues except overall intensity. An increase in overall intensity in the first syllable showed a trend in below chance level responses, suggesting that this cue is not a good correlate of word stress. This outcome is in line with the production data (Kaland, 2019). Only F0 played a role in both first and second syllables. This result is unexpected given the insignificant role of F0 movement as an acoustic correlate in the production of Papuan Malay word stress (Kaland, 2019).

First-syllable lengthening appeared to worsen the participants' ability to choose the correct word. A potential explanation may be found in the application of the two types of lengthening: final lengthening (on all second syllables) and lengthening as stress cue (on either the first or the second syllable), see Sec. II B. Thus, the duration difference between the first and the second syllable in the manipulated sequence was smaller when stress lengthening was applied to the first syllable than when stress lengthening was applied to the second syllable. Subsequently, the length contrast between the syllables in the sequence was smaller for penultimate stress than for ultimate stress. Although this asymmetry reflected the production data, it might have had a contradictory effect for penultimate stress in perception.

The experimental design required participants to evaluate only the syllable sequence within a phrase. Given the limited research on Papuan Malay phrase prosody, it cannot be entirely ruled out that phrase prosodic expectations of the participants affected the outcomes for F0. A task in which two different syllable sequences had been presented would have forced participants to evaluate the suitability of each sequence. Such a setup could reveal perceptual effects more closely related to word prosody, which remain hidden in a task in which participants did not evaluate acoustic differences. In Sec. III, a second experiment is reported that accounted for the two issues just discussed.

III. EXPERIMENT II

In Experiment II, participants were presented with a written target word and listened to two sequences of two acoustically manipulated syllables. The participants' task was to choose which of the two sequences corresponded to the target word. The syllable sequences were identical to those embedded in the matrix phrase in Experiment I (Sec. II A and II B). The setup and procedure differed in the ways discussed in Sec. III B.

A. Participants

A total of 21 participants carried out Experiment II. All were students at the University of Papua; 5 male and 16 female participants (M_{age} = 21.3, age range 18–33), and were native speakers of Papuan Malay without hearing problems. They received a small present for their participation. None of the participants of Experiment II participated in Experiment I.

B. Setup and procedure

In Experiment II, one stimulus consisted of one written target word and two wave files. The wave files corresponded to two different manipulated syllable sequences in isolation taken from the same set of stimuli used in Experiment I (Appendix A). The target word was taken from the list of 84 two-syllable words, half of which had penultimate and the other half ultimate word stress, according to word lists in Kluge (2017), see Appendix B. The two wave files differed as to which syllable had one or more acoustically manipulated cues. That is, for each stimulus, one wave file consisted of stress cue(s) on the first syllable and the other consisted of one or more stress cues on the second syllable. In this way, all possible combinations of stimulus pairs were presented to participants as auditory stimuli (11×11, see Appendix A). The total number of stimuli was 121, matching the number of wave-file pairs. The stimulus pairs were presented as response options for each target word. The target word was taken randomly from the word list for each stimulus. The order of the stimuli was random and different for each participant. Note that in this setup, some target words were used more than others due to the relatively small number of unique words with ultimate word stress (Appendix B) and due to the higher number of stimulus pairs (wave files) compared to the number of words in the list.

For each stimulus, the script generated a screen. The screen displayed the target word, two play buttons with selection boxes, a "next" button and a percentage counter. Participants were then asked to listen to each of the auditory stimuli by clicking the respective play buttons and indicate which of them corresponded with the target word (Fig. 4). The stimuli could be played as many times as needed. Participants were instructed to read the target word out loud if they faced difficulties identifying the corresponding auditory stimulus. This lets the participants activate their auditory memory for the word, including potential stress cues. To choose one of the auditory stimuli, participants had to tick the corresponding selection box and then click "continue." Only after both auditory stimuli in the pair had been played did the selection boxes and "continue" button appear on the screen. This ensured that the participant had listened to both auditory stimuli before making a choice. As soon as the participant clicked the "continue" button, the next stimulus would appear. The auditory stimuli were linked to the play buttons on either the left or the right side of the screen at random. This counter-balanced potential learning effects during the experiment. Responses were
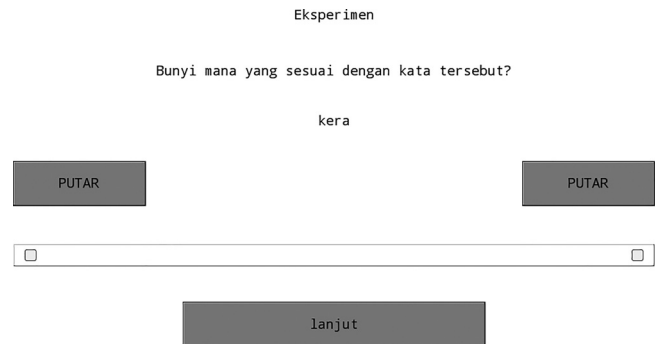

FIG. 4. Screen capture of Experiment II displaying the question "Which sound corresponds to the word?," the target word, two play buttons (*putar*) to play the wave files in the pair with accompanying selection boxes, and a button to continue (*lanjut*).

collected on the computer as 1 (correct) or 0 (incorrect). A response was considered correct when the position of the manipulated acoustic cue(s) in the selected wave file matched the stress position of the target word. All remaining aspects of the procedure were identical to those described for Experiment I (Sec. II F) and are not repeated in this section.

C. Statistical analysis

Statistical analyses were identical to the ones described for Experiment I (Sec. II F.)

D. Results

The results show that participants overall scored above chance level ($M = 0.53$, V = 1 707 000, $p < 0.01$). The mean correct responses for each stress position indicate that for penultimate stress the score was significantly below chance level ($M = 0.46$, V = 346 240, $p < 0.01$), whereas for ultimate stress, the scores were significantly above chance level ($M = 0.59$, V = 512 070, $p < 0.001$), see also Table IV.

Furthermore, the GLMM results (Table V) did not show a significant change in the correctness scores for any of the individual acoustic cues. Two trends could be observed: duration in the first syllable and $F0$ in the second syllable resulted in lower correctness scores when combined with

TABLE IV. Results of the Wilcoxon signed rank tests and the GLMM analysis on the acoustic cues in each word stress position in Experiment II.

Word stress	Cue	Wilcoxon ($\mu = 0.50$)	
		M correct	p
Penultimate	$F0$	0.46	0.44	<0.01
	Duration		0.46	<0.05
	Intensity (O)		0.45	<0.05
	Intensity (T)		0.45	<0.05
Ultimate	$F0$	0.59	0.60	<0.001
	Duration		0.57	<0.001
	Intensity (O)		0.60	<0.001
	Intensity (T)		0.59	<0.001

TABLE V. Results of the generalized linear mixed model analyses for each syllable and acoustic cue. Effects of number of cues (N cues) only reported when yielding significance.

Syllable	Cue	β	SE	z	p
First	$F0$	−0.14	0.14	−1.01	n.s.
	Duration	0.15	0.14	1.09	n.s.
	N cues (3)	−0.35	0.20	−1.76	= 0.08
	Intensity (O)	−0.01	0.16	−0.06	n.s.
	Intensity (T)	−0.02	0.16	−0.11	n.s.
Second	$F0$	0.22	0.14	1.60	n.s.
	N cues (2)	−0.24	0.14	−1.72	= 0.08
	Duration	−0.23	0.14	−1.64	n.s.
	Intensity (O)	0.02	0.16	0.14	n.s.
	Intensity (T)	−0.02	0.16	−0.10	n.s.

other cues. Higher correctness scores for ultimate stress were found when comparing the word-stress positions in the GLMM ($\beta = 0.58$, $SE = 0.09$, $z = 6.33$, $p < 0.001$).

E. Discussion

The results of Experiment II show that listeners benefit from acoustic cues to word stress only for ultimate syllables. For these syllables, participants score above chance level for all cues. Given that no effects of individual acoustic cues were found, the results seem to indicate a general perceptual sensitivity to (all) cues to ultimate stress. Acoustic cues to word stress on penultimate syllables showed the opposite effect in that participants scored below chance level. The latter result indicates that the acoustic cues on the penultimate syllable did not facilitate word-stress perception.

F. Experiments I and II

Experiments I and II revealed a crucial difference in the perception of the individual acoustic cues to word stress in Papuan Malay. That is, when the syllable sequences were presented in a phrase (Experiment I), $F0$ appeared to be a reliable cue to word stress on either syllable. However, when presented in isolation (Experiment II) the same syllable sequences did not show that listeners relied on specific acoustic cues. This outcome appears to confirm the presupposition formulated in the discussion of Experiment I (Sec. II H), namely, that phrase prosodic expectations may have affected the outcomes.

Overall, the facilitative effects found in Experiments I and II are minimal. That is, the maximum correctness score observed was 0.60 in Experiment II, suggesting that the acoustic cues to word stress do not play a crucial role in word recognition. This outcome is partially expected, given that Papuan Malay has a highly regular penultimate stress pattern with predictable deviations (Peperkamp et al., 2010). These factors suggest that in virtually all Papuan Malay words, the segments alone are enough to allow the word to be recognized. That would mean that word-stress cues in this language only have a marginal additive facilitation

effect, as predicted by psycholinguistic studies (e.g., Cutler, 2012).

Experiments I and II targeted suprasegmental cues only and no segmental information was available to the participants in either experiment. This could mean that promoting the salience of suprasegmental cues in each experiment did not reflect the perception of natural speech. In addition, the resynthesized syllable sequences may have also compromised the naturalness of the data in both Experiments I and II. Furthermore, Experiments I and II have shown evidence for listeners' sensitivity to the irregular (ultimate) stress pattern only. To what extent the facilitative effect of word-initial stress (e.g., Mehta and Cutler, 1988) holds for Papuan Malay listeners is investigated in Experiment III, as further discussed Sec. IV.

IV. EXPERIMENT III

Experiment III was designed to investigate word recognition latencies when the disambiguating cue to identify the word was either the stressed or the unstressed syllable. Stimuli consisted of phrases from a corpus in which target words were embedded (Kluge et al., 2014). The target words were embedded medially (1a) or finally (1b) in a matrix phrase, read by a male native speaker of Papuan Malay (the same speaker described in Sec. II A).

(1a) ko   pu   kata ___   itu,   sa   blum   taw
     2SG POSS word ___   D.DIST 1SG not.yet   know
     "that word ___ of yours, I don't yet know (it)"
(1b) sa   blum   taw   ko   pu   kata itu,   kata ___
     1SG not.yet know   2SG POSS word D.DIST word ___
     "I don't yet know that word of yours, the word ___"

A subset of the recordings was selected for use in the current experiment. The most frequently occurring syllable structure and stress pattern in Papuan Malay is 'CV.CV (Kluge, 2017). Therefore, only words with this pattern were selected. Given the considerable number of loanwords, only Papuan Malay roots were selected. Furthermore, a number of recordings were excluded because the intensity of the speaker's voice was low. The selected set of recordings consisted of 80 stimuli (half of type 1a, half of type 1b), each with a different target word.

The presence of stress cues in the selected stimuli was assessed by means of acoustic measures. All syllables in the target words were annotated using Praat textgrids (Boersma and Weenink, 2019). For each syllable, $F0$ movement (maximum $F0$–minimum $F0$) in semitones, duration in ms and average intensity in dB were measured. To obtain one measure of the acoustic difference between the first (stressed) and the second (unstressed) syllable, a difference score was computed. This was done by subtracting the measured value of the second syllable from the measured value of the first syllable for each acoustic cue. Averages are reported in Table VI and generally confirm that stressed syllables stand out acoustically compared to unstressed syllables, in particular with respect to duration (Kaland, 2019).

Constantijn Kaland

TABLE VI. Mean (*SD*) differences scores (Δ) for *F*0 movement (ST), duration (ms) and intensity (dB) in the target words (*N* = 80) in either phrase position (medial/final).

Measure	Medial	Final
Δ *F*0 movement	1.98 (3.68)	0.41 (2.91)
Δ Duration	20.17 (84.35)	20.92 (87.99)
Δ Intensity	0.88 (6.03)	4.36 (5.40)

A. Design

In the task, participants indicated as quickly as possible which word they heard whilst listening to a matrix sentence in which a target word was embedded (1a or 1 b). For each stimulus, they could choose between two visually presented response words; one correct (target) and one incorrect (distractor). Either the first or second syllable of the distractor was identical to the respective syllable in the target. This was done to control which syllable was the critical cue to identify the target word (cue syllable). The first (stressed) syllable would be the cue when target and distractor had identical second syllables. The second (unstressed) syllable was the cue when target and distractor had identical first syllables (see Table VII for examples). It should be noted that apart from stress cues, the segmental makeup of the cue syllable was strictly speaking sufficient to recognize the word (cf. *lida* and *lada*, Table VII). It has been shown that suprasegmental cues may nevertheless have an additional facilitative effect on word recognition (e.g., Cutler, 2005). Such an additional effect of suprasegmental cues can be tested when the effect of the segmental makeup is equal in target and distractor, as ensured in the current design. The cue syllable in the target always consisted of a different vowel than the distractor (in some cases also consonants), to ensure that the cue to recognition consisted of the most sonorous part of the syllable. For each combination of phrase position and cue syllable the same number of items was generated (20 in each condition).

Note that the number of words with ultimate stress in Papuan Malay is low. For this reason, no target-distractor pairs such as those in Table VII could be created with ultimate stress. The current design therefore only made use of words with penultimate stress. How this choice may have affected the outcomes of the experiment is further discussed in Sec. IV F.

B. Setup and procedure

The task was designed and run in OpenSesame using the legacy backend (Mathôt *et al.*, 2012). For each stimulus,

TABLE VII. Example stimuli (*English gloss*) with either stressed or unstressed syllable as recognition cue.

Cue syllable	Target	Distractor
Stressed (li)	*lida* (tongue)	*lada* (pepper)
Unstressed (bi)	*babi* (pig)	*bapa* (father)

the script generated a screen (Fig. 5). The screen showed "Kata mana yang Anda dengar?" (Which word did you hear?) and two buttons (1 and 0) with the response words (target/distractor) on either side. To choose one of the response words, participants had to press either 1 for the word on the left, or 0 for the word on the right. The response words were written underneath the respective buttons and were randomly assigned to each side of the screen for each stimulus to balance the effects of handedness (i.e., faster responses for preferred hand). The stimulus screen appeared for five seconds to let participants familiarize themselves with the response words. Three successive tones of 1 kHz sounded on the last three seconds of the familiarization time, indicating the upcoming stimulus. The stimulus screen was displayed until 2.5 s after participants had pressed "1" or "0" to ensure the stimulus had finished playing before continuing. Accidental key presses on other keys were not registered and did not affect the course of the experiment. After each stimulus, participants needed to press the space bar to proceed. This allowed them to set the pace of the experiment, which has been shown to lead to lower rates of missed responses and to improve participants' compliance (Krinzinger *et al.*, 2011). This aspect is crucial for participants in the current study, who had little to no familiarity with (reaction time) experiments. Reaction times were measured between either the start or end of the target word and the moment "1" or "0" was pressed. Commonly, stimulus onset latencies are reported in word recognition tasks, although stimulus offset latencies better account for differences in stimulus duration (Lipinski and Gupta, 2005). Given the varying durations of the target words in Experiment III, both onset and offset measures were taken. Half of the participants were presented with phrase-medial targets in the first part of the experiment and phrase-final targets in the second part. The other half of the participants were presented with the phrase-final target in the first part and phrase-medial targets in the second part. The presentation order of the stimuli within each part was random and different for each participant, in order to balance potential habituation effects (i.e., faster for stimuli presented later in the task).

Before the experiment started, participants received both oral and written instructions. To familiarize themselves with the task, participants completed a practice round consisting of five stimuli. Participants were instructed to switch off personal mobile devices and used headphones during the entire experiment. Participants were instructed to take a
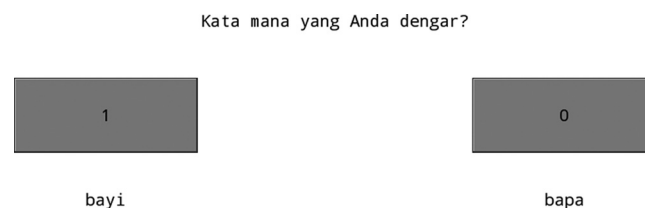

FIG. 5. Screenshot showing an example stimulus in Experiment III.

short break after completing half of the task. The experiment lasted approximately 20 min.

C. Participants

All 22 participants were students at the University of Papua, Manokwari; 5 male and 17 female participants (age: $M = 22.05$, range = 18–41), and were native speakers of Papuan Malay without hearing problems. Among the participants of Experiment III, 16 also participated in either Experiment I ($N = 2$) or Experiment II ($N = 14$). The order in which they completed the two experiments was random to balance out potential treatment effects.

D. Data processing and statistical analyses

Before statistical analysis, reaction times ($N = 184$) were removed from the data if participants had not correctly recognized the word ($N = 32$), reacted within 200 ms after the onset of the cue syllables ($N = 16$) or reacted later than two seconds after target word offset (outliers; $N = 126$). These reaction times are considered extreme and are not assumed to provide insight into word recognition processes. Removal of such times has been done in similar studies on word recognition (e.g., Balota *et al.*, 2013). The removed outliers constituted 7.16% of the data, which is within the suggested limits for cut-off points (Ratcliff, 1993; Baayen and Milin, 2010). After removal, 1576 reaction times remained for analysis.

GLMMs are proposed to be particularly suitable for reaction time measures (Lo and Andrews, 2015). These measures appear to have an inverse Gaussian distribution (Baayen and Milin, 2010; Lo and Andrews, 2015), which can be directly specified in a GLMM, thereby eliminating the need for prior transformation of the reaction times. It has been shown that transformations to obtain normally distributed reaction times could influence the effects that the predictors in a model might have (Balota *et al.*, 2013). The GLMM approach, however, is not suitable for negative reaction times, which were obtained in Experiment III when measured relative to the target word offset. That is, participants were able to identify the word correctly before the end of the target word in 79 cases. This happened in particular when the first syllable was the cue to identify the word. To avoid transformation and to avoid exclusion of negative reaction times, linear mixed effect models (LMM) were performed.

Two separate LMMs were performed for the target onset reaction times and the target offset reaction times using R (R Core Team, 2019) and the lme4 package (Bates *et al.*, 2015). In each LMM analysis, reaction time was the response, and the interaction between cue syllable (first, second) and phrase position (medial, final), the three difference scores taken from the acoustic cues ($F0$ movement, duration, intensity), and target word duration were predictors. Participants and items (target words) were added as random slopes. This structure represented the maximum structure for which the model converged. The inclusion of target

word duration as a predictor was done to account for expected differences between reaction times relative to the onset of the target word and those relative to the offset of the target word. It has been shown that only the latter take into account variation in stimulus length, thus providing a more accurate reaction time measure (e.g., Lipinski and Gupta, 2005). The predictor target word duration indeed showed an effect for the target onset measures ($\beta = 0.45$, $SE = 0.16$, $t = 2.72$, $p < 0.01$) and not for the target offset measures ($\beta = -0.25$, $SE = 0.19$, $t = -1.31$, n.s.). For this reason, the predictor duration of the target word was excluded from the final model applied to the target offset measures.

To analyze the relationship between the acoustic cues in the target words and the reaction times, Pearson correlation coefficients were computed between the mean reaction times and each of the acoustic difference scores for all stimuli ($N = 80$).

E. Results

Results indicate that participants were faster when the cue syllable was the first (stressed syllable) in the word (Fig. 6, Tables VIII and IX). This effect was found regardless of whether the target word occurred in phrase-medial or phrase-final position. Phrase position showed an effect in that participants were overall faster at recognizing the target word when it was presented phrase-finally than when it was presented phrase-medially. The LMM furthermore showed trends for the onset measures for the acoustic difference scores of duration and intensity, whereas for the offset measures only the duration difference score showed a significant effect. The latter indicated that participants took longer to react when the duration difference between stressed and unstressed syllables was larger. The correlations confirmed this effect, in particular for offset measures taken when the target appeared phrase-finally (Table X). The Pearson coefficients furthermore indicated positive trends for the correlation between the reaction times (onset and offset) and $F0$ movement difference scores taken from target words in phrase-final position.

F. Discussion

Experiment III showed that Papuan Malay listeners were faster when the stressed syllable was the syllable needed to recognize the word. However, the experimental design meant that the stressed syllable always occurred in word-initial position, making it difficult to disentangle a generic facilitative effect of first syllables from an effect originating in word-stress cues. A generic facilitative effect of the first syllable in strong-weak syllable patterns has been found in previous work (Cutler and Clifton, 1984). This effect disappeared when stimulus length was controlled for (see also Sec. I A 4). Stimulus length was also controlled for in the current study, making it unlikely that the facilitative effect of the word-initial syllable was only an artefact of the stimulus material (but see Sec. V for more discussion on this issue). It seems more reasonable, therefore, to interpret the

742   J. Acoust. Soc. Am. **147** (2), February 2020
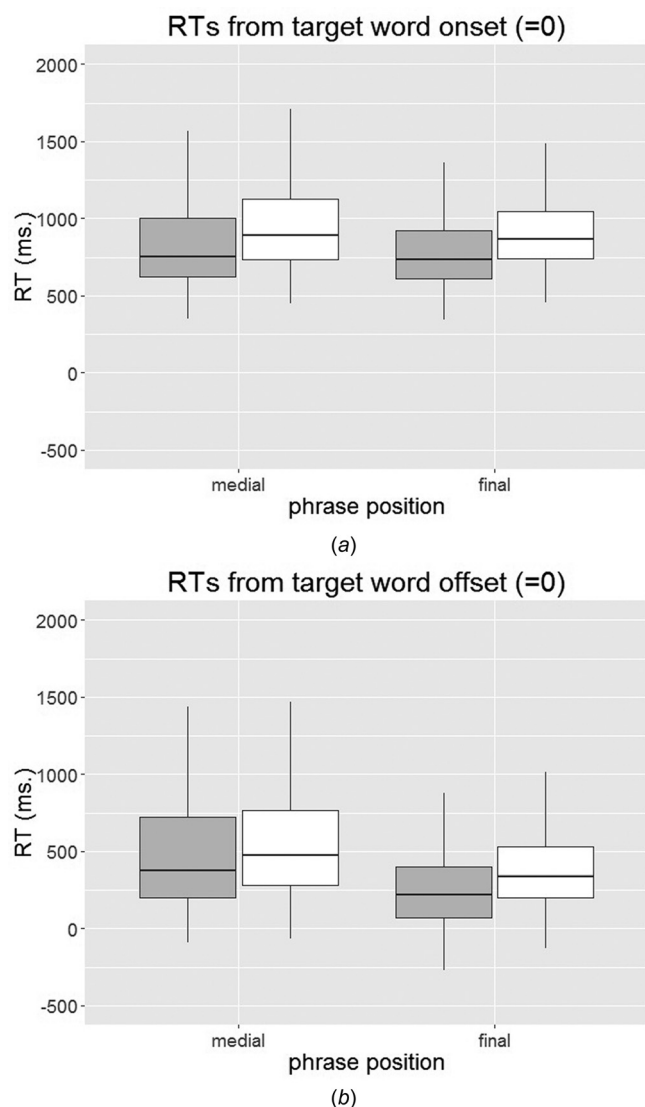
Constantijn Kaland

FIG. 6. Reaction times measured from target word onset (left) and offset (right) in phrase-medial and phrase-final position when the cue syllable was the first (grey) or second (white).

TABLE IX. Results of the LMMs performed on the target-word onset and offset reaction-time measures in Experiment III. Interactions not reported were not significant.

Response	Predictor	β	SE	t	p
Target onset	Cue syllable	110.87	22.09	5.02	<0.001
	Phrase position	225.48	26.53	8.50	<0.001
	$\Delta F0$ movement	4.18	2.60	1.61	n.s.
	Δ Duration	193.56	104.77	1.85	= 0.07
	Δ Intensity	3.22	1.64	1.96	= 0.05
	Target word duration	0.45	0.16	2.72	<0.01
Target offset	Cue syllable	116.89	25.81	4.53	<0.001
	Phrase position	310.15	30.63	10.13	<0.001
	$\Delta F0$ movement	3.52	3.04	1.16	n.s.
	Δ Duration	307.05	113.38	2.71	<0.01
	Δ Intensity	0.85	1.90	0.45	n.s.

differences between stressed and unstressed syllables were larger. If larger acoustic differences are assumed to provide stronger cues to word stress, longer reaction times seem to suggest a counter-facilitative effect. Still, it is pertinent that larger duration differences indicate that the stressed syllable was (proportionally) longer compared to the unstressed syllable. Longer stressed syllables may take more time to process, thereby lowering participants' speed of recognition. Although this explanation might hold for the positive (as opposed to negative) correlation between duration difference scores and reaction times, it also indicates that caution is needed when both predictor and response depend on timing measures, as was the case in Experiment III. Nevertheless, an effect of target-word duration on the outcomes can be ruled out, as this predictor was taken into account for both the analysis of target onset and target offset reaction times (Sec. IV D). Furthermore, if the positive correlation found for duration difference scores indeed reflects a generic processing cost due to longer syllables, similar positive correlations in phrase-medial position and for target onset measures might have been expected. The latter effects were, however, not found (Table X). Thus, regardless of whether longer syllables reversed the facilitation effect in reaction times or not, the current results seem to indicate that listeners were more sensitive to the duration in phrase-final position than to other acoustic cues.

To conclude this section, the results of Experiment III must be treated with caution in two regards. First, although it is highly likely that the facilitative effect of the first syllable on word recognition latencies in Papuan Malay results

effect of cue syllable in relation to word stress. Phrase-final duration differences between stressed and unstressed syllables appeared to correlate positively with reaction times when these were taken from the target word offset. However, whether this effect can be interpreted as a facilitative effect of word-stress cues on word recognition is unclear. That is, participants were slower when duration

TABLE VIII. Mean (SD) reaction times in Experiment III measured from target word onset and offset, split by phrase position (medial/final) and cue syllable (1/2).

	Phrase position			
	Medial		Final	
Measure	Cue syllable 1	Cue syllable 2	Cue syllable 1	Cue syllable 2
Target onset	966.22 (492.80)	1067.71 (477.62)	831.96 (344.32)	949.34 (349.00)
Target offset	547.06 (493.58)	602.46 (474.48)	290.04 (346.90)	416.29 (353.91)

TABLE X. Pearson correlation coefficients (and p-values) between each of the acoustic difference scores in the stimuli ($N = 80$) and the respective reaction time measures in Experiment III ($df = 38$).

Acoustic cue	Target onset		Target offset	
	Medial	Final	Medial	Final
Δ F0 movement	0.01 (n.s.)	0.29 (= 0.08)	−0.04 (n.s.)	0.29 (= 0.08)
Δ Duration	0.07 (n.s.)	0.17 (n.s.)	0.16 (n.s.)	0.43 (<0.01)
Δ Intensity	0.30 (= 0.06)	0.27 (n.s.)	0.15 (n.s.)	0.06 (n.s.)

from word-stress differences, a direct relationship was difficult to establish on the basis of the current results. Second, Experiment III did not investigate the recognition of words with ultimate stress. It could be hypothesized, based on Experiments I and II, that ultimate stress patterns facilitate word recognition to a larger extent than penultimate stress, as listeners appeared more sensitive to acoustic cues on the ultimate syllable in these experiments. This hypothesis could not be tested using the design of Experiment III.

V. CONCLUSION AND GENERAL DISCUSSION

Experiment I showed that $F0$ movements cue word stress, although this effect was not replicated in Experiment II. Experiment III showed that Papuan Malay listeners recognize words faster when the penultimate syllable provides crucial cues. Taken together, the outcomes of the three experiments indicate that Papuan Malay word stress is a perceptually relevant phenomenon, although suprasegmental cues play a marginal role.

The outcomes are challenging to interpret in the light of existing work on the acoustic correlates of word stress in Papuan Malay. Among the investigated acoustic cues, $F0$ movements were not necessarily expected to affect word stress perception. Overall, $F0$ movements appeared to be a weak correlate of word stress in production and were of smaller size on penultimate stressed syllables than on ultimate stressed syllables (Kaland, 2019). Apart from the phrase-prosodic influences that might explain the results of $F0$ in Experiment I, it is not fully understood how speech production and perception relate. Studies have shown, for example, that $F0$ peak timing differences can affect the perception of prominence. That is, Dutch listeners perceived peaks with delayed timing (longer onset) as more prominent than earlier peaks, even though they had the same absolute peak height acoustically (Gussenhoven et al., 1997). This finding shows that, as far as $F0$ is concerned, its relation to perceived prominence is much more complex than can be captured by a rise/fall distinction or a difference in $F0$ level. Even less is known about how $F0$ is integrated with other acoustic cues to perceive the difference between two phonetic categories (e.g., Repp, 1983). Thus, the effect of $F0$ in Experiment I should be interpreted with caution. Furthermore, $F0$ appears to operate differently from the other acoustic cues tested in Experiments I and II, given the (marginal) effects found for the number of cues. That is, in

both experiments, $F0$ cues combined with other acoustic cues on the second syllable lowered the correctness scores of the participants. This seems to indicate that $F0$ operates on a different level than word stress, most plausibly the phrase level.

The contribution of word-stress cues to the recognition of words in Papuan Malay is small overall, as shown by the correctness scores (Experiments I and II) and the limited number of (consistent) acoustic effects on the reaction times (Experiment III). Although this result is not surprising in a language in which word prosody follows a fixed pattern (e.g., Dogil, 1999; and Peperkamp et al., 2010), the current study has been able to show that the penultimate syllable can be helpful for word recognition. This appeared in particular in two-syllable words where the penultimate (stressed) syllable was the word-initial one. This was shown by the consistently shorter reaction times when the penultimate stressed syllable provided crucial information to recognize the word (Experiment III). In this respect, it is important to note that the facilitation effect did not correlate with the strength of the suprasegmental cues, which may indicate a rather generic facilitation effect instead of one related to specific stress cues (Sec. IV F). Such an effect could be explained by the recently reported longer reaction times for later segmental uniqueness points (UP; Tucker et al., 2019), such as the stimuli in Experiment III for which the identification cue was the second syllable. Another more general processing benefit could also be at work. That is, for phrase-medial targets in Experiment III, participants were required to react whilst they received ongoing auditory input. This could have slowed down their responses more than when the auditory input had finished, as was the case for phrase-final targets.

The absence of effects in this study could furthermore be related to the lack of vowel reduction as a cue in the design. As discussed in Sec. I C, vowel reduction, or vowel quality in general, may be a more effective stress cue in Papuan Malay. Even though word-initial syllables may have a generic privileged status in word recognition, it is likely that the occurrence of stress cues on this syllable facilitates this status at least partially (Mehta and Cutler, 1988; McAllister, 1991). Related to this conclusion, the limited effects of the individual word-stress cues as found in this study were also the result of the asymmetry between penultimate and ultimate syllables. This asymmetry is analogous to that found in earlier studies on languages with mostly fixed stress patterns and has been called a form of "stress-deafness" (Domahs et al., 2012; Domahs et al., 2013). Although the concept of stress-deafness (see also Dupoux et al., 1997) suggests that listeners cannot perceive certain stress patterns, such a conclusion does not hold for Papuan Malay. The outcomes of the three experiments presented here rather suggest that listeners by default perceive the regular stress pattern, meaning there is little demand for its acoustic salience (e.g., Sulpizio and McQueen, 2012, for a similar account). An explanation along the same lines can be advanced for why so few direct effects of the acoustic

cues were found in Experiment III, even though the acoustic signal provided sufficient cues that correlated with the production of penultimate stress. In this line of reasoning, one could claim that there is no communicative need to mark the unmarked, which is the default penultimate stress that speakers and listeners expect. However, deviations from this pattern need to be salient enough such that the words that fail to meet the default expectations (ultimate stress) can be communicated successfully. This strategy is precisely that which is reflected in both the production of ultimate stress (Kaland, 2019) and its perception.

The current findings also directly shed light on how phrase prosody affects the perception of word-stress patterns. Experiment III showed a facilitation effect of phrase-final words over phrase-medial words. There is currently no investigation of how Papuan Malay word stress is acoustically realized in phrase-final position. However, there is preliminary evidence that $F0$ movements are largest in this position and that they correlate with word stress to some extent (Kaland and Baumann, 2019). In this respect, it seems reasonable to assume that the current results fit with the proposal that word prosody across languages reflects phrase-edge prosody (Gordon, 2014). However, it is too early to draw this conclusion as little is known about the function of phrase-final $F0$ movements, which could be a pitch accent, boundary tone, or a hybrid form (Kaland and Baumann, 2019). More research is needed in order to fully understand the nature of this relationship. This research is currently being undertaken and the results can be expected in the near future.

The current study has shown that Papuan Malay stress patterns do contribute to some extent to word recognition. Taking the existing acoustic work and the current results together, it can be concluded that word stress in Papuan Malay is present in the acoustic signal and has some perceptual relevance, in particular for deviant patterns.

ACKNOWLEDGMENTS

APPENDIX A

Schematic overview of the F(0 movement), D(uration), and I(ntensity) manipulations on the first and second syllables (- = not manipulated, O = overall intensity, T = spectral tilt) in the syllable sequences.

Sequence nr.	First syllable			Second syllable		
	F	D	I	F	D	I
1	F	-	-	-	-	-
2	-	D	-	-	-	-
3	-	-	O	-	-	-
4	F	D	-	-	-	-
5	F	-	O	-	-	-
6	-	D	O	-	-	-
7	F	D	O	-	-	-
8	-	-	T	-	-	-
9	F	-	T	-	-	-
10	-	D	T	-	-	-
11	F	D	T	-	-	-
12	-	-	-	F	-	-
13	-	-	-	-	D	-
14	-	-	-	-	-	O
15	-	-	-	F	D	-
16	-	-	-	F	-	O
17	-	-	-	-	D	O
18	-	-	-	F	D	O
19	-	-	-	-	-	T
20	-	-	-	F	-	T
21	-	-	-	-	D	T
22	-	-	-	F	D	T

APPENDIX B

Overview of the near-minimal word stress pairs, with position of word stress, English gloss, and the additional segmental differences indicated for the words of each pair (s = substitution, i/d = insertion/deletion, c = consonant, v = vowel).

Pair no.	Penultimate stress		Ultimate stress		Add. difference	
	Papuan Malay	English	Papuan Malay	English	Operation	C/V
1	bebas	*be free*	bebang	*burden*	s	c
2	bebas	*be free*	bekas	*trace*	s	c
3	bengkok	*be crooked*	bengkak	*be swollen*	s	v
4	besi	*metal*	bersi	*be clean*	i/d	c
5	nekat	*to determine*	dekat	*to near*	s	c
6	depang	*front*	dengan	*with*	s	c
7	enak	*be pleasant*	enam	*six*	s	c
8	gedi	*aibika*	geli	*tickle*	s	c
9	kewa	*dance party*	kena	*hit*	s	c
10	kintal	*yard*	kental	*be fluent*	s	v
11	bera	*defecate*	kera	*ape*	s	c
12	kewa	*dance party*	kera	*ape*	s	c
13	kira	*think*	kera	*ape*	s	v
14	mera	*be red*	kera	*ape*	s	c
15	kutuk	*to curse*	ketuk	*to knock*	s	v
16	sumur	*well*	kumur	*to rinse mouth*	s	c
17	kukus	*to steam*	kuskus	*cuscus*	i/d	c
18	lama	*be long (duration)*	lema	*be weak*	s	v

J. Acoust. Soc. Am. **147** (2), February 2020

Constantijn Kaland   745

(*Continued*)

Pair no.	Penultimate stress		Ultimate stress		Add. difference	
	Papuan Malay	English	Papuan Malay	English	Operation	C/V
19	lima	*five*	lema	*be weak*	s	v
20	lomba	*contest*	lemba	*valley*	s	v
21	lebar	*be wide*	lembar	*sheet*	i/d	c
22	lempar	*throw*	lembar	*sheet*	s	c
23	memang	*indeed*	menang	*to win*	s	c
24	minang	*propose*	menang	*to win*	s	v
25	minta	*to request*	menta	*be uncooked*	s	v
26	munta	*to vomit*	menta	*be uncooked*	s	v
27	pecis	*light bulb*	pedis	*be spicy*	s	c
28	pasang	*pair*	pesang	*to order*	s	v
29	pisang	*banana*	pesang	*to order*	s	v
30	ribut	*to trouble*	rebut	*to race each other*	s	v
31	gedi	*aibika*	sedi	*be sad*	s	c
32	sapi	*cow*	sepi	*be quiet*	s	v
33	sarang	*suggestion*	serang	*to attack*	s	v
34	semang	*outrigger*	serang	*to attack*	s	c
35	tepu	*to clap*	tedu	*be calm*	s	c
36	tugas	*duty*	tegas	*be firm*	s	v
37	tukang	*craftsman*	tekang	*to press*	s	v
38	memang	*indeed*	temang	*friend*	s	c
39	semang	*outrigger*	temang	*friend*	s	c
40	tandang	*banana plant stem*	tendang	*to kick*	s	v
41	tepu	*to clap*	tepung	*flour*	i/d	c
42	tatap	*to gaze at*	tetap	*be unchanged*	s	v

[1]In this study, stress in Papuan Malay is investigated as a feature that manifests itself primarily in the word domain. The literature has not always kept the word level separate from the phrase level and, consequently, the term "stress" has been used to refer to syllables that were most prominent due to both word- and phrase-level phenomena (e.g., Fry, 1955). However, it has been shown to be more useful to explicitly distinguish between word prosody and phrase prosody in work on the acoustics (Sluijter and Van Heuven, 1996) and typology (e.g., Gordon, 2014) of stress. Unless otherwise indicated, "stress" and "word stress" are used interchangeably in this paper to refer to the word domain only.

[2]For an extensive overview of research on word stress in Trade Malay varieties, see Kaland (2019).

Adams, C., and Munro, R. R. (**1978**). "In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English," Phonetica **35**, 125–156.

Baayen, H. R., and Milin, P. (**2010**). "Analyzing reaction times," Int. J. Psychol. Res. **3**, 12–28.

Balota, D. A., Aschenbrenner, A. J., and Yap, M. J. (**2013**). "Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations," J. Exp. Psychol. **39**, 1563–1571.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**, 1–48.

Beckman, M. (**1986**). Stress and non-Stress Accent (Netherlands Phonetic Archives, Foris, Dordrecht), p. 239.

Beckman, M., and Edwards, J. (**1994**). "Articulatory evidence for differentiating stress categories," in Papers in Laboratory Phonology (Cambridge University Press, Cambridge, UK), Vol. 3, pp. 7–33.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer [computer program]," http://www.praat.org/ (Last viewed 11/7/2019).

Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, H., and Tremblay, A. (**2018**). "English learners' use of segmental and

suprasegmental cues to stress in lexical access: An eye-tracking study: Linguistic cues to stress in lexical access," Lang. Learn. **68**, 635–668.

Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang. Speech **45**, 207–228.

Crosswhite, K. M. (**2004**). "Vowel reduction," in Phonetically Based Phonology, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, Cambridge, UK), pp. 191–231.

Cutler, A. (**2005**). "Lexical stress," in Handbook of Speech Perception, edited by D. B. Pisoni and R. E. Remez (The Blackwell Publishing Ltd, Oxford, UK), pp. 264–289.

Cutler, A. (**2012**). Native Listening: Language Experience and the Recognition of Spoken Words (The MIT Press, Cambridge, MA), p. 555.

Cutler, A., and Butterfield, S. (**1992**). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," J. Mem. Lang. **31**, 218–236.

Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.

Cutler, A., and Clifton, C. (**1984**). "The use of prosodic information in word recognition," in Attention and Performance X: Proceedings of a Symposium, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ), pp. 183–196.

Cutler, A., Wales, R., Cooper, N., and Janssen, J. H. (**2007**). "Dutch listeners' use of suprasegmental cues to English stress," in Proceedings of the 16th International Congress of the Phonetics Sciences, August 6–10, Saarbrücken, Germany, pp. 1913–1916.

Di Cristo, A. (**1998**). "Intonation in French," in Intonation Systems: A Survey of Twenty Languages, edited by D. Hirst and A. Di Cristo (Cambridge University Press, Cambridge, UK), pp. 195–218.

Dogil, G. (**1999**). "The phonetic manifestation of word stress in Lithuanian, Polish, German, and Spanish," in Word Prosodic Systems in the Languages of Europe, edited by H. van der Hulst (De Gruyter, Berlin), pp. 273–311.

Domahs, U., Genc, S., Knaus, J., Wiese, R., and Kabak, B. (**2013**). "Processing (un-)predictable word stress: ERP evidence from Turkish," Lang. Cogn. Process. **28**, 335–354.

Domahs, U., Knaus, J., Orzechowska, P., and Wiese, R. (**2012**). "Stress 'deafness' in a Language with fixed word stress: An ERP study on Polish," Front. Psychol. **3**, 439.

Donohue, M., and Sawaki, Yusuf. (**2007**). "Papuan Malay pronominals: Forms and functions," Oceanic Ling. **46**, 253–276.

Dupoux, E., Pallier, C., Sebastian, N., and Mehler, J. (**1997**). "A destressing 'deafness' in French?," J. Mem. Lang. **36**, 406–421.

Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang. Speech Lond. **1**, 126–152.

Goedemans, R. W. N., and van Zanten, E. (**2007**). "Stress and accent in Indonesian," in Prosody in Indonesian Languages, LOT Occasional Series, edited by V. J. van Heuven and E. van Zanten (LOT, Netherlands Graduate School of Linguistics, Utrecht), pp. 35–62.

Goedemans, R., and van Zanten, E. (**2014**). "No stress typology," in Above and Beyond the Segments, edited by J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller, and E. van Zanten (John Benjamins, Amsterdam, the Netherlands), pp. 83–95.

Gordon, M. (**2014**). "Disentangling stress and pitch-accent: A typology of prominence at different prosodic levels," in Word Stress, edited by H. van der Hulst (Cambridge University Press, Cambridge, UK), pp. 83–118.

Gordon, M., and Roettger, T. (**2017**). "Acoustic correlates of word stress: A cross-linguistic survey," Linguist. Vanguard.

Grice, M., and Baumann, S. (**2002**). "Deutsche Intonation und GToBI" ("German intonation and GToBI"), Linguistische Berichte, Helmut Buske Verlag, Hamburg.

Gussenhoven, C. (**2004**). The Phonology of Tone and Intonation (Cambridge University Press, Cambridge, UK), p. 355.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., and Terken, J. (**1997**). "The perceptual prominence of fundamental frequency peaks," J. Acoust. Soc. Am. **102**, 3009–3022.

Hanulíková, A., McQueen, J. M., and Mitterer, H. (**2010**). "Possible words and fixed stress in the segmentation of Slovak speech," Q. J. Exp. Psychol. **63**, 555–579.

Hart, J. T., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge, UK), p. 232.

Himmelmann, N. P. (**2018**). "Some preliminary observations on prosody and information structure in Austronesian languages of Indonesia and East Timor," in *Perspectives on information structure in Austronesian Languages*, edited by S. Riesberg, A. Shiohara, and A. Utsumi (Language Science Press, Berlin, Germany), pp. 347–374.

Isenberg, D., and Gay, T. (**1978**). "Acoustic correlates of perceived stress in an isolated synthetic disyllable," J. Acoust. Soc. Am. **64**, S21–S21.

Jesse, A., Poellmann, K., and Kong, Y.-Y. (**2017**). "English listeners use suprasegmental cues to lexical stress early during spoken-word recognition," J. Speech Lang. Hear. Res. **60**, 190–198.

Kaland, C. C. L. (**2018**). "Spectral tilt as a correlate of Papuan Malay word stress," in *Proceedings of the 9th International Conference on Speech Prosody 2018*, June 18–20, Poznan, Poland, pp. 339–343.

Kaland, C. C. L. (**2019**). "Acoustic correlates of word stress in Papuan Malay," J. Phon. **74**, 55–74.

Kaland, C., and Baumann, S. (**2019**). "Different functions of phrase-final F0 movements in spontaneous Papuan Malay," in *Proceedings of the 19th International Congress of Phonetic Sciences*, August 5–9, Melbourne, Australia, pp. 1312–1316.

Kaland, C. C. L., Himmelmann, N. P., and Kluge, A. (**2019**). "Stress predictors in a Papuan Malay random forest," in *Proceedings of the 19th International Congress of Phonetic Sciences*, August 5–9, Melbourne, Australia, pp. 2871–2875.

Kluge, A. (**2017**). *A Grammar of Papuan Malay* (Language Science Press, Berlin, Germany), p. 880.

Kluge, A., Rumaropen, B. A. W., and Aweta, L. (**2014**). "Papuan Malay data—Word list," SIL International, Dallas, TX, https://www.sil.org/resources/archives/59649 (Last viewed 4/19/2019).

Kohler, K. J. (**2008**). "The perception of prominence patterns," Phonetica **65**, 257–269.

Krinzinger, H., Koten, J. W., Hennemann, J., Schueppen, A., Sahr, K., Arndt, D., and Konrad, K. (**2011**). "Sensitivity, reproducibility, and reliability of self-paced versus fixed stimulus presentation in an fMRI study on exact, non-symbolic arithmetic in typically developing children aged between 6 and 12 years," Dev. Neuropsychol. **36**, 721–740.

Lieberman, P. (**1959**). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. **31**, 844.

Lipinski, J., and Gupta, P. (**2005**). "Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration," J. Mem. Lang. **52**, 171–192.

Lo, S., and Andrews, S. (**2015**). "To transform or not to transform: Using generalized linear mixed models to analyse reaction time data," Front. Psychol. **6**, 1171.

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

Mathôt, S., Schreij, D., and Theeuwes, J. (**2012**). "OpenSesame: An open-source, graphical experiment builder for the social sciences," Behav. Res. **44**, 314–324.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework," J. Exp. Psychol. General **134**, 477–500.

McAllister, J. (**1991**). "The processing of lexically stressed syllables in read and spontaneous speech," Lang. Speech **34**, 1–26.

Mehta, G., and Cutler, A. (**1988**). "Detection of target phonemes in spontaneous and read speech," Lang. Speech **31**, 135–156.

Mo, Y., Cole, J., and Lee, E.-K. (**2008**). "Naive listeners' prominence and boundary perception," *Proceedings of the 4th International Conference on Speech Prosody*, May 6–9, Campinas, Brazil, pp. 735–738.

Niebuhr, O., and Winkler, J. (**2017**). "The relative cueing power of F0 and duration in German prominence perception," in *Proceedings of Interspeech 2017*, August 20–24, Stockholm, Sweden, pp. 611–615.

Peperkamp, S., Vendelin, I., and Dupoux, E. (**2010**). "Perception of predictable stress: A cross-linguistic investigation," J. Phon. **38**, 422–430.

R Core Team (**2019**). "R: The R project for statistical computing," https://www.r-project.org/ (Last viewed 11/7/2019).

Ratcliff, R. (**1993**). "Methods for dealing with reaction time outliers," Psychol. Bull. **114**, 510–532.

Reinisch, E., Jesse, A., and McQueen, J. M. (**2010**). "Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately," Q. J. Exp. Psychol. **63**, 772–783.

Repp, B. H. (**1983**). "Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization," Speech Commun. **2**, 341–361.

Riesberg, S., Kalbertodt, J., Baumann, S., and Himmelmann, N. P. (**2018**). "On the perception of prosodic prominences and boundaries in Papuan Malay," in *Perspectives on Information Structure in Austronesian Languages*, edited by S. Riesberg, A. Shiohara, and A. Utsumi (Language Science Press, Berlin, Germany), pp. 389–414.

Roettger, T., and Gordon, M. (**2017**). "Methodological issues in the study of word stress correlates," Linguist. Vanguard.

Sluijter, A. M. C., and van Heuven, V. J. (**1996**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.

Stoel, R. B. (**2007**). "The intonation of Manado Malay," in *Prosody in Indonesian Languages, LOT Occasional Series*, Vol. *9*, edited by V. van Heuven and E. van Zanten (LOT, Netherlands Graduate School of Linguistics, Utrecht), pp. 117–150.

Sulpizio, S., and McQueen, J. M. (**2012**). "Italians use abstract knowledge about lexical stress during spoken-word recognition," J. Mem. Lang. **66**, 177–193.

Suomi, K., Mcqueen, J. M., and Cutler, A. (**1997**). "Vowel harmony and speech segmentation in Finnish," J. Mem. Lang. **36**, 422–444.

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., and Sims, M. (**2019**). "The massive auditory lexical decision (MALD) database," Behav. Res. **51**, 1187–1204.

Van Bezooijen, R., and Gooskens, C. (**1999**). "Identification of language varieties: The contribution of different linguistic levels," J. Lang. Social Psychol. **18**, 31–48.

Van der Hulst, H. (**2012**). "Deconstructing stress," Lingua **122**, 1494–1521.

Van Heuven, V., and Hagman, P. J. (**1988**). "Lexical statistics and spoken word recognition in Dutch," in *Linguistics in the Netherlands*, edited by P. Coopmans and A. Hulk (Foris, Dordrecht, the Netherlands), pp. 59–68.

Van Minde, D. (**1997**). *Malayu Ambong: Phonology, Morphology, Syntax* (Research School CNWS, Leiden, the Netherlands), p. 420.

Van Rossum, G., and De Boer, J. (**1991**). "Interactively testing remote servers using the Python programming language," CWI Quart. **4**, 283–304.

Vihman, M., and Croft, W. (**2007**). "Phonological development: Toward a 'radical' templatic phonology," Linguistics **45**, 683.

Vroomen, J., Van Zon, M., and de Gelder, B. (**1996**). "Cues to speech segmentation: Evidence from juncture misperceptions and word spotting," Mem. Cogn. **24**, 744–755.