Research Article

Constantijn Kaland* and Matthew K. Gordon

The role of f0 shape and phrasal position in Papuan Malay and American English word identification

https://doi.org/10.1515/phon-2022-2022
Published online August 19, 2022

Abstract: The prosodic structure of under-researched languages in the Trade Malay language family is poorly understood. Although boundary marking has been uncontroversially shown as the major prosodic function in these languages, studies on the use of pitch accents to highlight important words in a phrase remain inconclusive. In addition, most knowledge of pitch accents is based on well-researched languages such as the ones from the Western-Germanic language family. This paper reports two word identification experiments comparing Papuan Malay with the pitch accent language American English, in order to investigate the extent to which the demarcating and highlighting function of prosody can be disentangled. To this end, target words were presented to native listeners of both languages and differed with respect to their position in the phrase (medial or final) and the shape of their f0 movement (original or manipulated). Reaction times for the target word identifications revealed overall faster responses for original and final words compared to manipulated and medial ones. The results add to previous findings on the facilitating effect of pitch accents and further improve our prosodic knowledge of underresearched languages.

Keywords: acoustic phonetics; auditory word recognition; phonetic typology; prosody; speech perception

1 Introduction

Suprasegmental properties of phrases such as f0, duration or intensity (henceforth phrase prosody) serve several functions. The literature traditionally distinguished

***Corresponding author: Constantijn Kaland**, Institute of Linguistics, Universität zu Köln – SFB 1252, Luxemburger Straße 299, 50939 Cologne, Germany, E-mail: ckaland@uni-koeln.de
Matthew K. Gordon, Department of Linguistics, University of California, Santa Barbara, USA

at least two main functions of phrase prosody; either indicating important information (highlighting), or grouping information in successive chunks, such as phrases (demarcating). The distinction between highlighting and demarcating has been maintained in influential models of prosody and intonation such as autosegmental metrical (AM) accounts (e.g. Silverman et al. 1992). In these accounts, f0 movements are central to the description of intonation. These movements come in specific shapes (pitch accents) that mark the most important word(s) within an utterance or occur at edges indicating how words are grouped into phrases (boundary tones). The development of AM-theories departed from work on Western-Germanic languages (see e.g. Ladd 2008, p. 43) and shaped later descriptions of typologically more diverse languages (e.g. Jun 2005, 2014) as well as research addressing the role of prosody in language comprehension (Frazier et al. 2006). However, many languages remain underresearched regarding both the documentation of crosslinguistic variation in prosody (Himmelmann and Ladd 2008) and the investigation of the speech processing mechanisms involved in phrase prosody (Cutler 2012, ch. 7). A growing body of literature that describes prosody in less well-studied languages refines traditional views on prosodic structure, in particular the interpretation of f0 movements near the edges of phrases (e.g., Gordon 2014).

Of particular interest to the distinction between highlighting and demarcating is the Trade Malay language family spoken in Eastern Indonesia. Languages in this family are often described as showing mainly f0 rise-fall movements in phrase-final positions (Kaufman and Himmelmann n.d). However, it remains unclear to what extent phrase final prosody is used for highlighting purposes, for demarcating purposes or a combination of both. Research so far indicated that Manado Malay (Stoel 2007) has both highlighting and demarcating f0 movements, whereas Ambonese Malay only makes use of demarcating ones (Maskikit-Essed and Gussenhoven 2016). Papuan Malay is the Trade Malay language for which prosody is best researched to date. However, the functions of phrase (final) prosody in this language are still poorly understood. The available studies do not rule out that highlighting and demarcating might be two sides of the same coin in this language (e.g. Kaland and Baumann 2020). The challenge remains, therefore, whether it is possible to fully disentangle these prosodic functions in Papuan Malay. This challenge is taken up in this paper by means of an experimental comparison between Papuan Malay and a well-studied language: American English. The comparison thus allows us to interpret the Papuan Malay results in a typological context with respect to the demarcating and highlighting roles of f0.

1.1 Highlighting and demarcating in prosodic theory

The most influential approaches to prosody and intonation have made a general distinction between prosodic events that mark important information in the phrase and those that signal the edge(s) of phrases. For example, autosegmental-metrical

(AM) accounts (ToBI, Silverman et al. 1992), adopt an inventory of discrete tones (*pitch accents* and *boundary tones*) as the building blocks of intonation contours with the core assumption that particular combinations of tones encode specific meanings (e.g. Pierrehumbert and Hirschberg 1990). The AM framework has been used to describe the prosodic structure of many languages (e.g. Ladd 2008 for an overview), including in the transcription of underresearched languages. For example, Rapid Prosody Transcription (Cole and Shattuck-Hufnagel 2016) makes use of the amount of agreement among native listeners regarding the location of *prominences* and *boundaries* to evaluate the importance of these prosodic events in a given language. Typological accounts of prosodic structure based on the AM model (Jun 2005, 2014) assume a general split between languages based on whether prominence occurs only at phrase edges in the form of boundary or phrasal tones, as in edge-marking languages like Korean and French, or also on prosodic heads in the form of pitch accents, as in head-marking languages like English and Dutch, or both (Jun 2014). Pitch accents, which serve a highlighting function, are assigned in bottom-up fashion to one or more syllables that are metrically strong (i.e. stressed) at the word level. Boundary tones, which have a demarcating role, are linked to phrasal units of various sizes; these units include the universal and relatively large Intonational Phrase (IP) and, on a language-specific basis, smaller units like the Accentual phrase (AP) and/or intermediate phrase (ip) kept apart at different levels of prosodic theory. Crucially, both pitch accents and boundary tones are realized primarily through f0 properties unlike word stress, which is characteristically associated with a constellation of properties that may or may not include f0 (Gordon and Roettger 2017).

When analysing the prosodic system of an understudied language, it is often difficult to determine whether a f0 prominence near a domain edge is attributed to a prosodic head or to a phrase edge, an ambiguity that has led to many languages that traditionally have been described as head-marking being reanalysed more recently as edge-marking, e.g. French (Jun and Fougeron 2002), Hindi, Bengali, Tamil (Fery 2010), Mongolian (Karlsson 2014), West Greenlandic (Arnhold 2014), and Turkish (Özçelik, 2012). Gordon (2014), in fact, suggests that both types of marking may ultimately be rooted in f0 properties found in words occurring at phrase boundaries. He distinguishes between languages on the basis of whether they allow pitch accents to occur in close proximity to boundary tones or not. In some languages, they may co-occur on the same syllable, reflecting tonal crowding, whereas in others the pitch accent may shift away from the boundary tone, reflecting edge repulsion. Under his account, word stress originates from the extension of phrasal prosody to the word level, in keeping with a widely held view that phrase-level phonetic properties often drive word-level phonology (Blevins 2004, 2006). Crucially, this view of prosody attaches primary importance to f0 movements at phrase edges as the driving force behind prominence not only in edge-marking languages but also in head-marking ones. To sum up, although theoretical models assume a distinction between highlighting prominence associated with prosodic heads and demarcating prominence attributed to boundary tones, the two sources of prominence are often difficult to tease apart since they both are conveyed at phrase edges through f0 properties.

1.2 Prosodic functions in speech processing

Perception studies have also informed our understanding of the highlighting and demarcating function of prosody. Research has shown that both the shape and the phrasal position of f0 movements affect listeners' language comprehension. As for shape, studies have shown that removing f0 excursions from a phrase (i.e., making the f0 contour "flat"), reduced its intelligibility (Hillenbrand 2003; Laures and Weismer 1999). There are likely multiple factors involved in this effect. For instance, a flat f0 contour removes the cues that listeners need to detect the most important content words (highlighting), at least in languages that make use of this prosodic function (e.g., pitch accent languages such as English). This effect has been demonstrated by an experiment in which listeners needed to detect the first phoneme of a target word in a phrase as fast as possible (Cutler and Foss 1977). The target word was either pitch accented in a natural way or deaccented by shifting the main pitch accent to another word in the phrase. Results showed that listeners were significantly faster in detecting the word-initial phoneme when the target had a pitch accent than when it did not. Furthermore, a flat f0 affects how listeners perceive speech tempo. This was shown in particular by experiments in which the f0 movements at phrase boundaries were flattened (Rietveld and Gussenhoven 1987). These indicated that perceived tempo was much higher when tonal boundary marking was absent than when it was present, even though both phrases had identical duration. Thus, these outcomes taken together reconfirm that f0 contributes to both highlighting (f0 shape) and demarcating (f0 position).

Concerning the position of f0 movements, studies have furthermore shown that phoneme detection is faster for pitch accents in phrase-final position compared to those in earlier positions (Shields et al. 1974). This effect was explained by rhythmical expectations of the listener, who anticipates upcoming speech by means of prosodic cues. The more cues available (i.e., towards the end of the phrase), the more accurate these predictions. A similar explanation in terms of semantic predictability was given for phrase-final words. That is, towards the end of the phrase the listener has processed more semantic context to be able to predict upcoming words compared to earlier phrase positions (Foss 1969). In addition, it should be noted that the final syllables of phrases are generally lengthened (final-lengthening, e.g. Shattuck-Hufnagel and Turk, 1998), which provides more time for f0 movements to be realised. Thus, f0 and duration are tightly coupled at phrase boundaries (see also Rietveld and Gussenhoven 1987). Crucially, the effect of phrase position on the perception of f0 movements was found in read speech, but not in spontaneous speech (Mehta and Cutler 1988). It was argued that spontaneous speech is much more fragmented (i.e., consists of shorter phrases) than read speech, thus providing the listener with sufficient prosodic cues to anticipate boundaries across contexts. Research has furthermore shown that in infant-directed speech, focused words are positioned more often phrase-finally and with wider f0 excursions compared to adult speech, ostensibly playing a facilitating role in language acquisition (Fernald and Mazzie 1991). This outcome corroborates earlier psychoacoustic work on the perception of tone sequences showing that listeners discriminate final stretches of tone sequences more accurately than initial ones (Watson et al. 1975). This was interpreted as the result of the silent period directly after the final tones, giving listeners more opportunity to process them.

To sum up, the literature has shown that the shape and position of f0 movements play crucial roles in the perception of speech, in particular around phrase boundaries. Perception studies offer support for a more nuanced division between highlighting and demarcating in keeping with recent work in prosodic typology. There appears to be a natural tendency to perceive final elements in a (speech) signal as more salient, primarily due to f0 properties, whether attributed to a head marking (highlighting) or edge-marking (demarcating) function. This observation as well as typological work on prosody (Gordon 2014) puts boundaries in a position central to the explanation of variation in the prosodic systems attested across languages of the world. Consistent with the universal salience of boundaries, a study of four unrelated languages (Himmelmann et al. 2018) showed that native listeners and listeners unfamiliar with those languages agreed on where they perceived a prosodic boundary. These results led to the hypothesis that “all natural languages make use of the same kinds of phonetic cues for IPs and these cues can be perceived by speaker-hearers even in unfamiliar languages” (Himmelmann et al. 2018, p. 30). In contrast, only head-marking languages recruit f0 in a highlighting capacity. The relationship between f0 used in a demarcative versus highlighting capacity is the target of the present study, which compares the role of prosody in word identification in two languages instantiating different prosodic types. One language, American English, is a head-marking language that clearly uses f0 in the marking of prominence via pitch accents, while the other language, Papuan Malay, may employ f0 primarily, and perhaps exclusively, to signal prosodic boundaries. In particular, we focus on how prosodic differences in the use of f0 between the two languages impact word recognition. The role of language specificity of prosodic type in word processing tasks is vastly understudied from a broad typological perspective. For example, the brief overview of processing studies in this paragraph mainly concerned results on relatively well studied pitch accent languages such as English (Cutler and Foss 1977; Laures and Weismer 1999; Mehta and Cutler 1988; Shields et al. 1974) or Dutch (Rietveld and Gussenhoven 1987) that are not typologically representative of languages of the world (see also Henrich et al. 2010). The current study aims to counter this bias by investigating a lesser studied language that appears to rely less on highlighting prominence than other languages on which our knowledge of speech processing is based.

1.3 Phrase prosody in Trade Malay languages

Of particular interest to the question of the extent to which highlighting and demarcating are essentially separate prosodic functions is the Trade Malay language family spoken in Eastern Indonesia. Although phrase prosody has not been studied extensively for these languages, the available work suggests that they have different prosodic structures compared to Western-Germanic languages. That is, Ambonese Malay has been analysed as a language without pitch accents and with only phrase-final boundary tones (Maskikit-Essed and Gussenhoven 2016). These boundary tones were reported to have a loose temporal alignment in the final two syllables of the phrase (see Figure 1). The loose alignment was explained as the result of the lack of word stress, i.e. no particular syllable at the word level to which the phrase final f0 movements could align. This explanation was furthermore supported by measurements failing to show a typical stress pattern acoustically.

The prosody of Manado Malay is analysed differently in that this language has pitch accents that align to stressed syllables (Stoel 2007). Stress is reported to be penultimate regularly, and ultimate or variable in a small number of exceptions. Two prosodic levels are distinguished, the IP and the phonological phrase (PhP; roughly corresponding to the ip in Jun 2014). One IP may consist of more PhP's if no more than one PhP is pitch accented. Each PhP has edge tones; commonly a low tone phrase-initially and a high tone phrase-finally. Focus is marked by a high tonal target on the focused PhP (e.g., subject-, object-, verb- or predicate-focus). The intonational analysis was supported by visualisations of f0 contours in this study (see examples in Figure 2).

To date, Papuan Malay has been studied to a larger extent than Ambonese Malay or Manado Malay. Production and perception studies provided clear indications that this language has regular penultimate word stress (e.g., Kaland

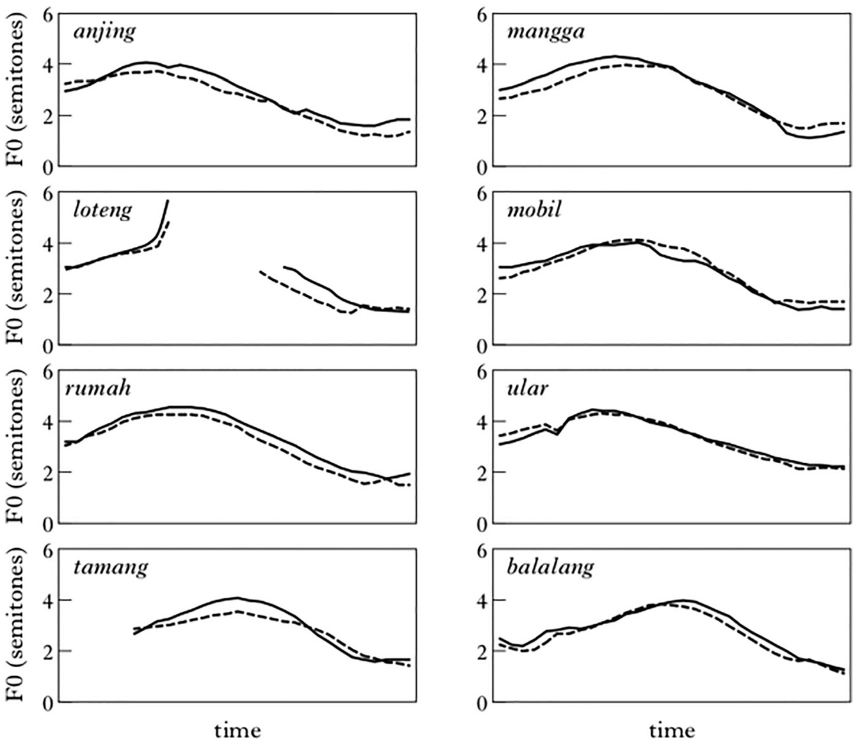

Figure 1: Ambonese Malay words in sentence-final position with corrective declarative focus (solid lines) and neutral declarative focus (dashed lines). Normalised time scale. Taken from Maskikit-Essed and Gussenhoven (2016).

2019, 2020), confirming the claim in Kluge (2017). As for phrase prosody, final f0 movements are the largest ones and have the tendency to show a rise on the stressed syllable. A tentative autosegmental analysis of phrase-final f0 movements was proposed in which f0 movements align to stressed syllables only in phrase-final positions (Kaland and Baumann 2020). Contrastive focus is not marked by means of particular f0 movements (Kaland and Himmelmann 2020). Systematic elicitations of picture descriptions (noun phrases) such as *sapi biru* 'blue cow' in neutral, noun or adjective focus, in different phrase positions (medial or final) and phrase types (ANTecedent or TARget), see examples in (1-a) and (1-b)), invariably showed a rise on the pre-final syllable and either a (shallow) rise or a fall on the final syllable (see Figure 3). The latter movement signalled continuation or finality respectively and therefore can be interpreted as a boundary tone, which is used in similar ways in many languages (Jun 2005, 2014). However, the status of the

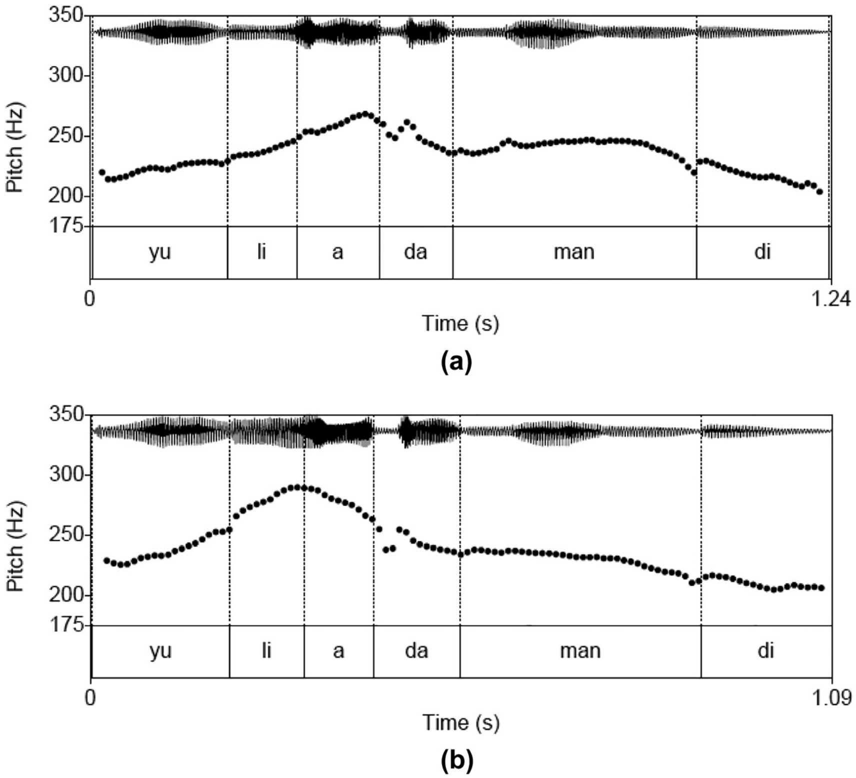

Figure 2: Manado Malay phrase *Yulia da mandi* 'Yulia is bathing' with (a) predicate focus (H* on *man*) and (b) subject focus (H* on *li*). In both phrases *Yulia* forms a PhP; either with H edge tone (a) or with L edge tone (b). Taken from Stoel (2007).

pre-final rise is unclear. In particular, it remains unclear whether it belongs to the boundary tone movement, making that movement bitonal such as in Ambonese Malay, or whether it should be interpreted as a pitch accent, aligned to the stressed syllable as in Manado Malay. The studies on Papuan Malay phrase prosody were entirely based on production data. To complement this research, perception studies are needed to shed light on how phrase prosody affects speech processing in this language. This is done in the current study and the research questions are outlined in the following subsection.

(1) a. Saya liat [ANT] di sebla kiri, dang saya liat [TAR] di sebla kanang.
I see [ANT] on the left side, but I see [TAR] on the right side.
b. Di sebla kiri saya liat [ANT], dang di sebla kanang saya liat [TAR].
On the left side I see [ANT], but on the right side I see [TAR].

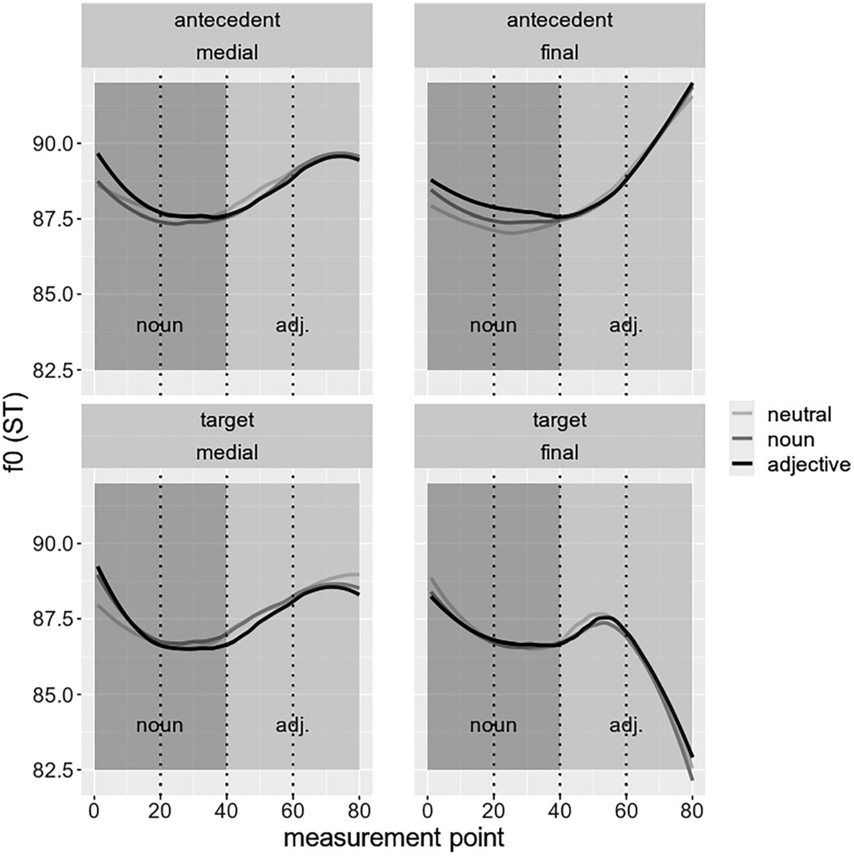

Figure 3: Papuan Malay noun phrases in neutral focus and with contrastive focus on either the noun or adjective in different phrase positions (medial, final) and phrase types (antecedent, target). Taken from Kaland and Himmelmann (2020).

1.4 Research questions

As the analyses for Ambonese Malay and Manado Malay differ fundamentally, it is a challenge to interpret the Papuan Malay results accurately. Boundaries appear to be marked by f0 movements, although a crucial underlying question is whether this language also makes use of pitch accents. In this respect, the phrase pre-final movements would be potential candidates due to their excursion size and alignment tendencies (Kaland and Baumann 2020). More research is therefore needed to understand the function(s) of phrase-final f0 movements in Papuan Malay, in particular concerning the shape and the position of f0 movements. This is carried out in the current study by investigating identification latencies for words that have unmanipulated f0 movements or manipulated ones (shape) in phrase-medial and phrase-final context (position). These two aspects could therefore shed more light on the extent to which highlighting (shape) and/or demarcating (position) respectively are prosodic functions that can be disentangled for Papuan Malay. Perception research is needed to investigate the extent to which f0 movements affect speech processing, as research in this area has been mainly conducted on well-studied languages (Cutler 2012). Exactly because well-studied languages provide an informative reference to interpret the results of understudied languages in a crosslinguistic context, the Papuan Malay results in the current study are compared to those of American English, a pitch accent language. Such a comparison also allows to separate potential task-related effects from actual linguistic differences. This is done in the current study by means of an identical word identification experiment carried out with participants from both languages.

The research question addressed in this study is therefore how the shape and phrase position of f0 movements affect word identification in Papuan Malay and American English. With two languages and two factors that can each have an effect or not there are in total eight theoretical outcomes ($2 \times 2 \times 2$). However, based on the literature it can be hypothesized that at least phrase position affects word identification latencies in both languages. This assumption is based on the literature promoting boundary marking as the main underlying (Gordon 2014) and potentially universal (Himmelmann et al. 2018) function of prosody. This reduces the most plausible hypotheses to the ones below (H1 and H0), where H1 predicts an effect of word identification on word recognition in at least American English. The two versions of H1 differ in their predictions for Papuan Malay. H1a predicts that f0 manipulation affects word recognition only in American English and not in Papuan Malay. Confirmation of H1a would corroborate the view that American English is a pitch accent language (e.g., Jun 2005; Ladd 2008; Silverman et al. 1992) and would favour an interpretation that Papuan Malay does not make use of specific pitch accents, similar to Ambonese Malay (Maskikit-Essed and Gussenhoven 2016) and in line with the lack of prosodic marking of contrastive focus in Papuan Malay (Kaland and Himmelmann 2020). H1b predicts that word recognition is affected by f0 manipulation in both languages. This would confirm that highlighting and demarcating can be distinguished in both languages. In this way, Papuan Malay would be similar to Manado Malay for which demarcating and highlighting were distinguished (Stoel 2007) and confirming the tentative highlighting analysis for Papuan Malay in Kaland and Baumann (2020).

H1a: F0 shape does not affect word identification in Papuan Malay, but it does in American English.

H1b: F0 shape affects word identification in both Papuan Malay and American English.

H0: F0 shape affects word identification in neither Papuan Malay nor American English.

2 Methodology

A reaction time (RT) experiment was set up to investigate native listeners' word identification latencies in Papuan Malay and American English phrases. The target words had an unmanipulated f0 contour or a manipulated (flat) f0 contour and appeared in either phrase-medial or phrase-final position (2 × 2 design for each language).

2.1 Recordings

As for the Papuan Malay experiment, recordings were used with words embedded in a matrix clause, read by a male native speaker (Kluge et al. 2014). The target words in the Papuan Malay recordings were selected based on the most frequent syllable structure and stress pattern in this language ('CV.CV). The matrix clause was constructed in such a way that the target word [T] appeared either in phrase-medial (2-a) or in phrase-final position (2-b), see Appendix A for the lists of target words. From the recordings a subset was selected for use in the current experiment. Because Papuan Malay makes use of a considerable number of loanwords, only native Papuan Malay roots were selected based on the word lists in Kluge (2017). Furthermore, recordings that were unclear due to the low intensity of the speaker's voice were not used in the current study.

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

b. sa blum taw ko pu kata itu, kata [T]
1SG not.yet know 2SG POSS word D.DIST word [T]
'I don't yet know that word of yours, the word [T].'

As for American English, recordings were made to match the Papuan Malay matrix phrases as closely as possible, see (3-a) and (3-b). These matrix sentences were read by a male native speaker of American English (36 years old). The speaker was instructed to read the sentences in a clear and natural way.

(3) a. 'The word [T] there, I don't know.'
b. 'I don't know that word there, the word [T].'

In order to minimize overall acoustic differences among the stimuli, the selected phrases for each language were scaled for their intensity using Praat (Boersma and Weenink 2019) such that the average intensity in each matrix phrase including the target word was 70 dB SPL.

2.2 Design

The RT experiment was designed such that participants chose as fast as possible which word they heard when listening to a matrix phrase with the target word in either medial or final position (i.e., a stimulus). They chose between two words written on the screen, of which one matched the target (correct response), and the other was a distractor (incorrect response). The distractor was chosen such that it partially matched the target word. The second syllable of the distractor was identical to the second syllable in the target. This was done to guarantee that the crucial cue to identify the target word for the participants was the first (stressed) syllable. Specifically, the distractor was chosen such that the difference with respect to the target was the vowel in the first syllable. This was done to make sure that the most sonorous part of the stressed syllable would always contribute to the identification of the target word. Note that this part carries the f0 contour. For some distractor words the consonant in the first syllable was also different from the target (due to the limited number of suitable words). For example, when the stimulus was *sa blum taw ko pu kata itu, kata laki* 'I don't yet know that word of yours, the word husband', the distractor was *hoki* (plant stem). As for American English, a word list of targets and distractors was created that closely matched the Papuan Malay ones. This also included matching whether the initial consonant was the same or different between target and distractor and whether the initial consonant was voiced or unvoiced in the target and the distractor. Note that word frequency affects word identification (Taft and Hambly 1986), which could not be accounted for in the selected target words. Although frequency information is available for American English, this is not the case for Papuan Malay. The Papuan Malay words, however, were elicited in conversation about every day life topics (Kluge 2017) and believed to be representative of relatively common words in this language. The selected set consisted of 20 recordings with phrase-medial target words and 20 recordings with phrase-final target words in each language. Appendix A provides a list of all target and distractor words in each language.

2.3 F0 manipulation

For each recorded matrix phrase two versions were created in which f0 was manipulated using Time-Domain Pitch-Synchronous Overlap-and-Add (TD-PSOLA; Moulines and Charpentier 1990) as implemented in Praat (Boersma and Weenink 2019), see Figure 4. It has been shown that the naturalness of speech due to TD-PSOLA resynthesis is somewhat decreased compared to unmanipulated speech (Mixdorff and Mehnert 1999). Therefore, in one version, the f0 contour was only stylized using a frequency resolution of two semitones. The stylised f0 contour closely followed the one in the original recording. In this way, the recording would undergo TD-PSOLA resynthesis without a change in the trajectory of the f0 contour. This was done to decrease the naturalness of the stimuli to a level comparable to the other version, which underwent TD-PSOLA resynthesis for the purpose of f0 manipulation. The rationale behind this procedure is that potential side-effects on participants' response latencies due to naturalness were balanced. In the other version, the stylised f0 contour and the contour of the target word were manipulated such that no pitch excursions occurred within the target word. This was done differently for phrase-medial target words than for phrase-final target words. For phrase-medial words, the original f0 level at the start of the first syllable was maintained throughout the entire word. In the Papuan Malay phrases, a rise that originally occurred on the first syllable of the medial target word was then shifted onto the first syllable of the next word in the matrix sentence (*i* in *itu*), see Figure 4 bottom left. In this way, the f0 in the remaining part of the matrix sentence would follow its original contour. In the American English phrases, the rise-fall on the medial target word was shifted onto the next word as well (on *there*). The rise-fall movement on the final target in both languages was flattened by maintaining the f0 at the level at the start of the first syllable with a declination towards the end of the word (equals phrase end). The declination was determined by taking the original f0 end point in the phrase. This was done to minimize the impression that the contour was manipulated.

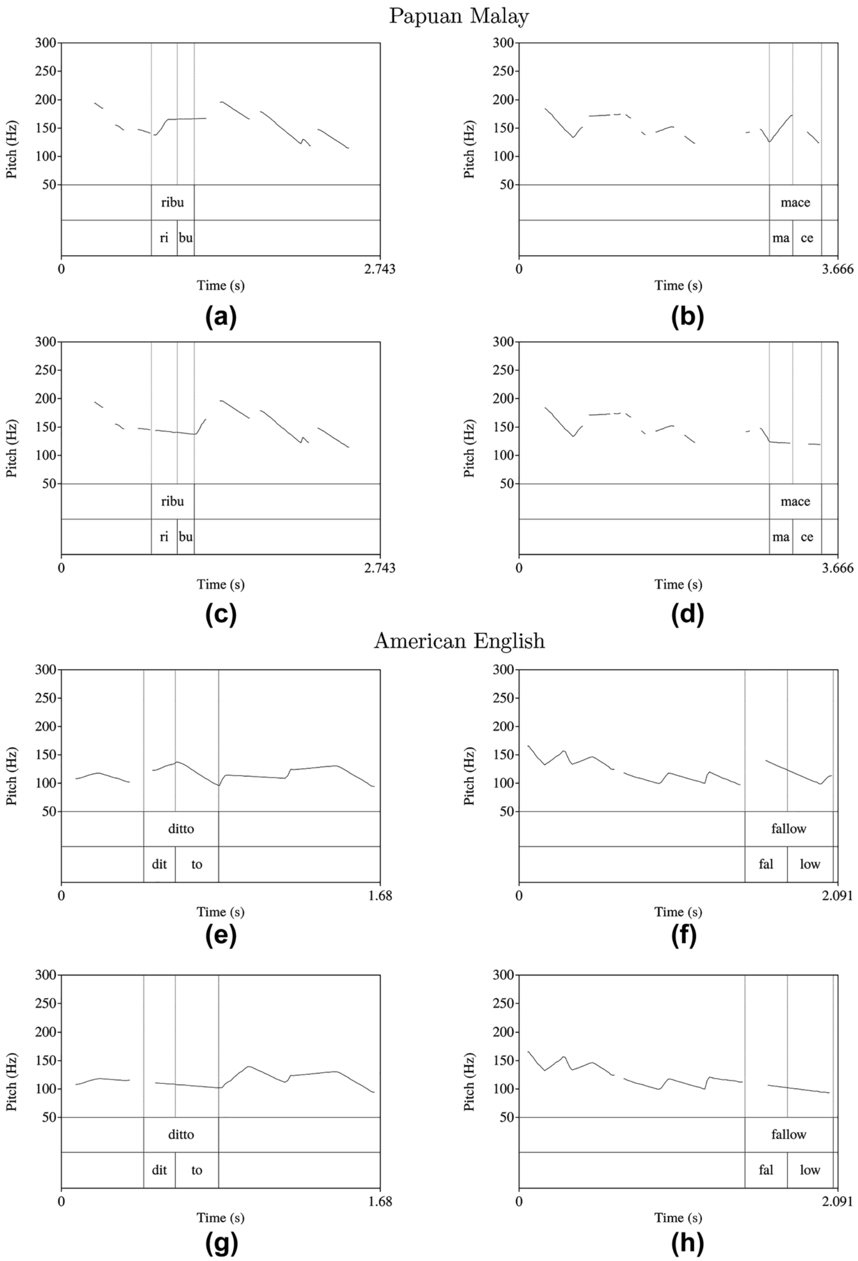

Figure 4: Stylized f0 contours on Papuan Malay (a–d) and American English (e–h) target words following the original (a–b, e–f) and manipulated (c–d, g–h) contour in phrase medial (left) and phrase final (right) positions. Target words are segmented on the word level (top tier) and syllable level (bottom tier).

2.4 Procedure

The word identification task for Papuan Malay was built using OpenSesame (Mathôt et al. 2012), which uses a Python (Van Rossum and De Boer 1991) script to present the stimuli on a computer. For American English, the task was built using PsyToolkit (Stoet 2010, 2017), which provides an online environment to run experiments remotely. The procedure was identical for each language and is described in the following.

The experiment presented 80 wave files (medial position: 20 original, 20 manipulated, final position: 20 original, 20 manipulated; henceforth stimuli). For each stimulus, a screen displayed the question *Kata mana yang Anda dengar?* (Papuan Malay) or *Which word did you hear?* (American English) and two buttons (Figure 5). On each of the buttons the corresponding key was shown that should be pressed to choose one of the response words: either "1" for the word on the left, or "0" for the word on the right. The "1" or "0" keys were aligned on the left and right side of the keyboard respectively (above the main alphabetical panel). Participants were instructed to use both their left and right index finger for the respective keys. The response words were written underneath the respective buttons on the screen. Target and distractor were randomly assigned as left or right word on the screen, differently for each participant. The stimulus screen was displayed for 5 s in order to let participants familiarize themselves with the two response options and to prepare them to hear the stimulus. During the last 3 s participants heard three successive tones of 1 kHz that cued the upcoming stimulus. The first two were 250 ms in length and the last one lasted for the entire final second before the stimulus was played. The stimulus screen was displayed until participants had pressed "1" or "0". After making their choice participants needed to press space bar to initiate the next stimulus. The space bar was chosen so that participants could

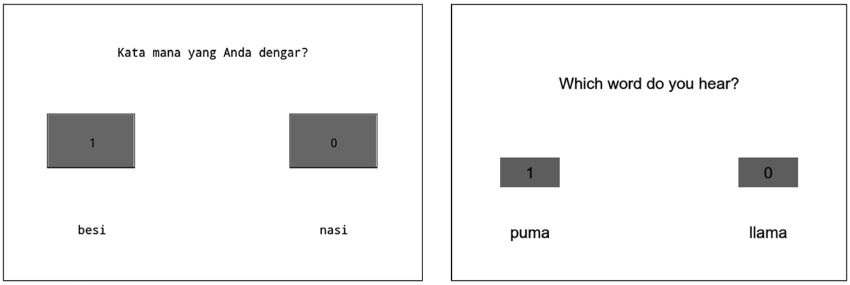

Figure 5: Screenshots of an example stimulus in Papuan Malay (left) and American English (right).

keep their hands on the keyboard during the entire experiment. The time participants saw the two response options prior to making a choice (stimulus familiarisation time) was fixed to 5 s to make sure all participants underwent the same procedure. Note that the time between successive stimuli was participant-initiated to allow participants 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 who had little to no familiarity with RT experiments.

RTs were measured from the uniqueness point (UP; Taft and Hambly 1986) in the word. For the target words in this study, the UPs would lie either after the first segment (e.g. Papuan Malay: *jata* vs. *kita*; American English: *llama* vs. *puma*) or after the second segment, which was the final segment of the first syllable (e.g. Papuan Malay: *tali* vs. *tuli*; American English: *taller* vs. *teller*), see also Appendix. Taking the reaction times relative to the UP of the target words therefore accounts for potential language differences in the segmental makeup of the target words, e.g. differences in the relative duration of first and second syllable, for the moment at which participant received unambiguous cues to recognize the target word. Thus, RTs were measured between the UP and the moment the participant had pressed "1" or "0". The experiment was divided into two parts, separated by a short break. Phrase positions were balanced across the two parts of the experiment. That is, one half of the participants was presented with phrase-medial targets in the first part and phrase-final targets in the second part. The other half was presented with phrase-final targets in the first part and phrase-medial targets in the second part. The presentation order of the stimuli was random and different for each participant. This was done to balance potential effects of handedness (faster with preferred hand), as well as other side-effects potentially associated with a fixed order (e.g. learning effects).

Before the start of the experiment participants received instructions about the course of the tasks. They were instructed to press the corresponding button on their keyboard as quickly as possible when they heard one of the words displayed on the screen. Then, they took a seat behind a computer and completed two subsequent parts of the experiment. First, they received written instructions on the screen about their task. To familiarize themselves with the task, participants completed a practice round consisting of five stimuli, which were the same for each participant and not part of the experiment. 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 additional stimuli. After each additional practice stimulus, participants could end the practice round. Second, when participants ended the practice session, they were asked to start the actual word identification task. After completing 50% of the actual identification task, participants were instructed to take a short break. The experiment lasted approximately 25 min. Results were collected digitally.

2.5 Participants

In total, 20 native speakers of Papuan Malay and 23 native speakers of American English without hearing or vision problems completed the experiment for course credit (Papuan Malay: 20 participants; mean age: 21.2; age range: 18–41. American English: 23 participants; mean age: 20; age range: 18–23).

2.6 Statistical analysis

RTs shorter than 200 ms after target onset (Papuan Malay: $N = 3$, American English: $N = 1$) and RTs longer than 2 s after target offset were discarded (Papuan Malay: $N = 1$, American English: $N = 41$). These latencies were considered to be erroneous, either due to accidental keypresses or attention problems. The RTs were log-transformed to obtain a normal distribution of the residuals (see Figure 6). Thereafter, outliers were removed from the data, following a procedure outlined in Baayen and Milin (2010). This involved removal of data points with absolute standardized residuals exceeding two standard deviations. This resulted in a better model fit ($R^2 = 0.76$) compared to when outliers were not removed or when RTs were untransformed (both $R^2 = 0.64$). The total remaining RTs for analysis was 3,169 (Papuan Malay: 1,467, American English: 1,702). Table 1 and Figure 7 report descriptive statistics of the RTs.

Linear mixed modelling (LMM) was carried out using R (R Core Team 2019) and the lme4 package (Bates et al. 2015). LMM fit by maximum likelihood (using

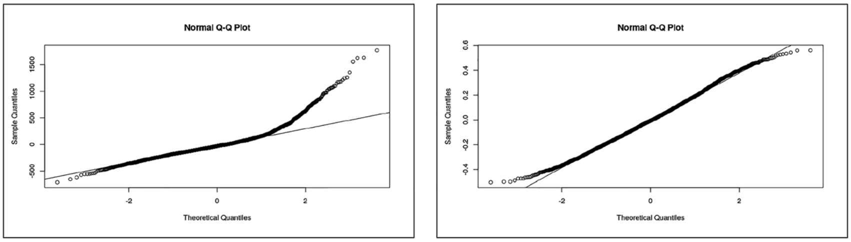

Figure 6: Quantile-quantile plots for the maximum model with untransformed reaction times (left) and for the model with log-transformed reaction times after outlier removal (right).

Table 1: Mean reaction times (in ms and log-RT) and standard deviations (italic) from target word UP in each language and each experimental condition.

Language	Phrase position	f0 manipulation	RT (ms)	log-RT
Papuan Malay	Medial	Original	656.37 (*178.27*)	6.45 (*0.26*)
		Manipulated	677.26 (*181.49*)	6.48 (*0.26*)
	Final	Original	655.09 (*180.66*)	6.45 (*0.26*)
		Manipulated	665.62 (*183.02*)	6.47 (*0.26*)
American English	Medial	Original	932.90 (*415.46*)	6.75 (*0.43*)
		Manipulated	967.57 (*440.23*)	6.78 (*0.45*)
	Final	Original	901.17 (*399.38*)	6.71 (*0.43*)
		Manipulated	913.03 (*417.41*)	6.72 (*0.43*)

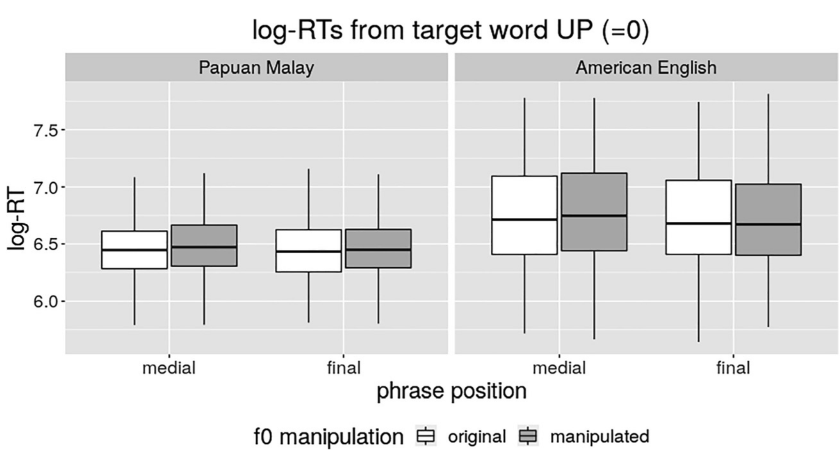

Figure 7: Boxplots of the reaction times (log-RT) for Papuan Malay and American English correct target word identifications.

Satterthwaite approximations to degrees of freedom to calculate *p*-values) was done on the offset RTs with the following predictors in the maximum model: the interaction of language (2 levels: Papuan Malay, American English), f0 manipulation (2 levels: original, manipulated) and phrase position (2 levels: medial, final), duration of segments before UP, UP position (2 levels: first segment, second segment), trial number and participant age. Participant and item (target word) were included as random intercepts with f0 manipulation as by-item random slope (the maximal converging model). UP position, participant age, and the interactions of language, f0 manipulation and phrase position did not have significant effects and were removed from the maximal model without decreasing the model fit (R^2 = 0.76). For the resulting final model collinearity was checked by means of correlation coefficients for all the predictors. The maximum Pearson correlation coefficient value of −0.04 was found and is considered a "very weak" correlation (Evans 1996). Variable inflation factors of the predictors as calculated using the 'car' package (Fox and Weisberg 2019) did not exceed 5, which is also considered a sign of weak correlation (James et al. 2013). The results of this final model are reported below.

3 Results

The statistical analysis on the reaction times showed that participants were significantly faster for Papuan Malay than for American English (Table 2, further discussion in Section 4). F0 manipulation showed a significant effect in that participants were faster recognizing the original target word than recognizing the acoustically manipulated one. Phrase position had a significant effect in that participants identified the target word faster when it occurred in final position than when it occurred in medial position. The duration of the segments until the UP had a significant effect in that reaction times were longer for longer pre-UP durations. Stimulus number had a significant effect in that participants were faster for stimuli occurring later in the experiment.

4 Discussion and conclusion

The results of the two reaction time experiments with Papuan Malay and American English listeners showed effects of f0 shape and phrase position on word recognition. In addition, an unexpected main effect of language was found

Table 2: Effects of the LMM on the log-RTs.

predictor	*β*	*SE*	*df*	*t*	*p*
(Intercept)	6.67	0.07	52.48	97.98	<0.001
Language: Papuan Malay	−0.29	0.09	43.14	−3.07	<0.01
f0 manipulation: manipulated	0.02	0.01	77.76	2.22	<0.05
Phrase position: medial	0.04	0.02	75.84	2.57	<0.05
Segment duration until UP	0.72	0.12	75.81	6.00	<0.001
Stimulus number	0.00	0.00	3,019.23	−9.46	<0.001

revealing that word recognition was overall faster in Papuan Malay than in American English. Further discussion of this effect is provided in this section.

The effect of f0 manipulation was expected and in line with previous research showing processing benefits for pitch accented words (e.g. Cutler and Foss 1977). In this context it is interesting that there was no interaction effect between language and f0 manipulation. This effect would have been expected if American English and Papuan Malay differed in their use of pitch accents. Hypothesis H1b was therefore confirmed by the current results, suggesting that both languages use f0 for highlighting as well as demarcation, as altering f0 shape affects word recognition. It is important to note that the type of task as well as the acoustic realisation of a pitch accent differed between the study by Cutler and Foss (1977) and the current one. That is, in Cutler and Foss (1977) phoneme detection was tested, which targets sublexical speech processing. In the current study, lexical processing was more directly targeted by means of word identification tasks. The task difference is relatively small, as both provide reliable indications of word processing. The acoustic difference between the presence or absence of a pitch accent was, however, larger in Cutler and Foss (1977) than in the current study. In Cutler and Foss (1977) a native speaker of American English produced the version with and without pitch accent, such that acoustic differences in the target words with and without pitch accent were to be found in duration, intensity and f0. In the current study, the only acoustic difference concerned f0, as this is assumed to be the main correlate of a pitch accent (e.g. Ladd 2008). The results therefore confirm that the shape of the f0 movement is important in both languages.

As for the effect of phrase position, this study has shown that in both Papuan Malay and American English words in final position are identified faster than in medial position, in line with previous findings (e.g. Shields et al. 1974). This effect cannot be attributed to predictability on the basis of previous context (e.g. Shields et al. 1974), since in the current study the matrix phrases did not provide contextual cues to the upcoming target word. It is more likely that the effect of phrase position relates to a general processing benefit for phrase-final words (see also Section 1.2). This processing benefit might give rise to the supposedly linguistically universal function of prosody to demarcate (Himmelmann et al. 2018) and the central role of phrase-edges in shaping prosodic structure (Gordon 2014).

There is another aspect related to the pitch accent facilitation effect reported in Cutler and Foss (1977) that is relevant to the interpretation of the current results. In Cutler and Fodor (1979) it was found that phoneme monitoring was facilitated for words occurring in a phrase position where listeners expected a pitch accent, regardless of whether a pitch accent was acoustically realised or not. It was therefore concluded that the accentuation effect should be interpreted as a focus effect, which is primarily dependent on phrase position and secondarily on acoustic realisation. Although the current results did show larger RT differences between original and manipulated targets in medial positions than in final positions, no interaction between manipulation and position was found. In this regard, the effects of f0 manipulation and phrase position should be seen as separate ones, confirming that highlighting and demarcating play a role in both languages (H1b).

The effect of language was unexpected due to the identical design of the experiments. Note, however, that the software used to measure the reaction times differed per language. For Papuan Malay, OpenSesame (Mathôt et al. 2012) was used, which measures reaction times on the computer it is run. For American English, however, reaction times were measured in the online web-interface PsyToolkit (Stoet 2010, 2017). There is no reason to assume that the measurement accuracy of either software was compromised or in any way imprecise. Studies have further shown that an online or laboratory setting does not necessarily lead to different reaction time results (Hilbig 2016; Kim et al. 2019). However, in the current study the different settings might have had an influence on the audio quality of the stimuli. That is, American English listeners were instructed to use their own headphones, which might have been in some cases of a lower quality than the one used consistently across participants in the Papuan Malay experiment, thereby leading to slower RTs in American English. Studies have also shown that a laboratory setting (cf. remote participation) could make participants focused more on performance accuracy than on performance speed, as suggested in Slote and Strand (2016). This would, however, predict faster RTs in American English, counter to the current results. It is therefore unclear whether the language effect is only the result of differences in experimental settings or also reflects genuine language differences.

An additional post-hoc analysis was carried out to investigate whether the stimulus material might have differed in acoustic clarity. For example, speakers providing the stimuli might have differed in speech tempo or in overall salience of the f0 movements. To explore this possibility, the duration and f0 range (maximum f0 – minimum f0) in the unmanipulated target words were measured. Recall that the intensity was scaled (Section 2.1) and therefore not expected to cause differences in word identification latencies between Papuan Malay and American English. The results showed that the duration of the target words did not differ significantly between the languages (Papuan Malay: $\mu = 478.10$, SD = 84.66; American English: $\mu = 485.09$, SD = 71.73; Wilcoxon signed-rank test: $V = 732$, n.s.). The results of the f0 range measurements (in semitones) revealed a significantly smaller range in Papuan Malay than in American English (Papuan Malay: $\mu = 5.49$, SD = 2.13; American English: $\mu = 8.84$, SD = 7.31; Wilcoxon signed-rank test: $V = 559$, $p < 0.05$).

Whether and how exactly the f0 range difference affects reaction times remains speculative. It has been shown, for example, that languages differ in the range they use for f0 movements (e.g. Dutch and British English; De Pijper 1983). It is conceivable that the f0 range used in the Papuan Malay experiment more closely aligned with the range encountered in natural speech relative to the American English materials. Further work on the production and perception side would be needed, however, to explore this hypothesis.

To conclude, the current study reconfirmed that (original) f0 shape and (phrase-final) positions facilitate speech processing (Section 1.2). Both Papuan Malay and American English thus distinguish a highlighting and demarcating role of f0, a finding that coheres with prosodic theories that separate these two functions of prosody (e.g. Jun 2005, 2014). With regard to the processing of f0 as a cue to phrase prosody, the current study has shown the importance of crosslinguistic experimental comparisons, an area in which relatively little research has been done (Cutler 2012).

Acknowledgements: Research for this paper was funded by the German Research Foundation (DFG) – Project-ID 281511265 – SFB 1252. The authors thank the staff of the Center for Endangered Languages Documentation (CELD, Manokwari, West-Papua) and Bob Kennedy (Department of Linguistics at the University of California, Santa Barbara; UCSB) for facilitating the Papuan Malay and American English experiments respectively and for participant recruitment. The authors are furthermore grateful to Michael Fiddler for providing the American English materials and to two anonymous reviewers for valuable comments on an earlier version of this paper.
Research funding: The research for this paper has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252 Prominence in Language.
Author contributions: CK: concept and design, experimental setup, running PM experiment, analysis and interpretation of data, drafting and revising manuscript. MG: recording AE stimuli, running AE experiment, interpretation of data, revising manuscript.
Conflict of interest statement: The authors have no conflicts of interest to declare.
Statement of ethics: The experiments reported in this paper have been conducted following protocols and informed consent practices in compliance with the Helsinki Declaration. Prior approval of the protocols and informed consent procedures was granted by the CELD and the human subject committee of the UCSB for the Papuan Malay and American English experiment respectively (Project Number: 55, Keycode: LING-GO-MA-007, Protocol Number: 55-21-0302).

Appendix A

Target and distractor words in Papuan Malay and English as used in the different phrase positions in the experiments.

Position	Papuan Malay				American English	
	Target	Gloss	Distractor	Gloss	Target	Distractor
medial	jata	*allotment*	kita	1PL	llama	puma
	laki	*husband*	hoki	*plant stem*	really	holy
	ruma	*house*	lama	*to be long (of duration)*	needy	moody
	tali	*cord*	tuli	*to be deaf*	taller	teller
	ribu	*thousand*	rubu	*to collapse*	roller	ruler
	satu	*one*	situ	L.MED	sully	silly
	tiga	*three*	juga	*also*	fuzzy	dizzy
	tuju	*seven*	maju	*to advance*	fishy	bushy
	buru	*to hunt*	baru	*to be new*	burrow	borrow
	butu	*to need*	jatu	*to fall*	lucky	rocky
	kira	*to think*	tara	*to be matching*	psycho	taco
	luru	*to chase after*	biru	*to be blue*	worry	berry
	pili	*to choose*	kali	*river*	tuna	china
	pisa	*to be separate*	hosa	*to pant*	polo	cello
	suka	*to enjoy*	nika	*to marry officially*	silly	Jolly
	bisa	*to be able*	basa	*to be wet*	butter	batter
	gila	*to be crazy*	mala	*even*	ditto	ghetto
	kaco	*to tell off*	koco	*to tell off*	petty	pity
	mara	*to feel angry*	mera	*to be red*	moody	muddy
	lusa	*day after tomorrow*	masa	*to be impossible*	limo	memo
final	besi	*metal*	nasi	*cooked rice*	body	lady
	duri	*thorn*	diri	*self*	belly	bully
	gaba	*unhulled paddy*	tiba	*to arrive*	baggy	foggy
	gigi	*tooth*	pagi	*morning*	beady	tidy
	gora	*water apple*	gara	*to irritate*	runny	rainy
	lida	*tongue*	lada	*pepper*	lousy	lazy
	mace	*woman*	cece	*great-grandchild*	naggy	piggy
	mati	*to die*	meti	*low tide*	pillow	polo
	paku	*nail*	suku	*ethnic group*	kitty	fatty
	puri	*anchovy-like fish*	kiri	*left*	picky	caky
	rawa	*swamp*	kewa	*dance party*	buggy	soggy
	sala	*to be wrong*	hela	*to haul*	kilo	Halo
	subu	*very early morning*	tubu	*body*	fallow	shallow
	tipu	*to cheat*	tepu	*to clap*	phoney	funny
	tugu	*monument*	lagu	*song*	sorry	marry

(continued)

Position	Papuan Malay				American English	
	Target	Gloss	Distractor	Gloss	Target	Distractor
	bera	*to defecate*	para	*to be in serious condition*	data	pita
	cebo	*to wash after defecating*	bobo	*nipah palm fruit schnapps*	sumo	limo
	gale	*to dig up*	bule	*white person*	nutty	meaty
	taru	*to put*	tiru	*to imitate*	curry	carry
	malu	*to feel embarrassed*	bulu	*body hair*	mono	rhino

References

Arnhold, Anja. 2014. Prosodic structure and focus realization in West Greenlandic. In Sun-Ah Jun (ed.), *Prosodic typology II*, 216–251. Oxford: Oxford University Press.

Baayen, Harald R. & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28.

Bates, Douglas, Martin Mächler, Bolker Ben & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2). 117–166.

Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer*. Available at: http://www.praat.org/.

Cole, Jennifer & Stefanie Shattuck-Hufnagel. 2016. New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology* 7(1). 1–29.

Cutler, Anne. 2012. *Native listening: Language experience and the recognition of spoken words*. OCLC: 816479318. Cambridge, Mass: The MIT Press.

Cutler, Anne & Jerry A. Fodor. 1979. Semantic focus and sentence comprehension. *Cognition* 7(1). 49–59.

Cutler, Anne & Donald J. Foss. 1977. On the role of sentence stress in sentence processing. *Language and Speech* 20(1). 1–10.

De Pijper, J. Roelof. 1983. *Modelling British English intonation: An analysis by resynthesis of British English intonation*. Dordrecht: Foris Publications.

Evans, James D. 1996. *Straightforward statistics for the behavioral sciences. English*. Pacific Grove: Duxbury Press.

Fernald, Anne & Claudia Mazzie. 1991. Prosody and focus in speech to infants and adults. *Developmental Psychology* 27(2). 209–221.

Fery, Caroline. 2010. Indian languages as intonational ’phrase languages. In S. Imtiaz Hasnain & Shreesh Chaudhary (eds.), *Problematizing language studies: Cultural, theoretical and applied perspectives—Essays in honor of Rama Kant Agnihotri*, 288–312. Delhi: Aakar Books.

Foss, Donald J. 1969. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior* 8(4). 457–462.

Fox, John & Sanford Weisberg. 2019. *An R companion to applied regression*, 3rd edn. Thousand Oaks CA: Sage.

Frazier, Lyn, Katy Carlson & Charles Cliftonjr. 2006. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10(6). 244–249.

Gordon, Matthew. 2014. Disentangling stress and pitch-accent: A typology of prominence at different prosodic levels. In Harry vander Hulst (ed.), *Word stress*, 83–118. Cambridge, UK: Cambridge University Press.

Gordon, Matthew & Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1). 20170007.

Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2–3). 61–83.

Hilbig, Benjamin E. 2016. Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods* 48(4). 1718–1724.

Hillenbrand, James M. 2003. Some effects of intonation contour on sentence intelligibility. *Journal of the Acoustical Society of America* 114(4). 2338.

Himmelmann, Nikolaus P. & D. Robert Ladd. 2008. Prosodic description: An introduction for fieldworkers. *Language Documentation & Conservation* 2(2). 244–274.

Himmelmann, Nikolaus P., Meytal Sandler, Strunk Jan & Volker Unterladstetter. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2). 207–245.

James, Gareth, Daniela Witten, Trevor Hastie & Robert Tibshirani. 2013. *An introduction to statistical learning*, vol. 103 (Springer Texts in Statistics). New York, NY: Springer New York.

Jun, Sun-Ah (ed.). 2005 *Prosodic typology: The phonology of intonation and phrasing (Oxford linguistics)*. Oxford, New York: Oxford University Press.

Jun, Sun-Ah (ed.). 2014 *Prosodictypology II: The phonology of intonation and phrasing (Oxford linguistics)*. Oxford: Oxford University Press.

Jun, Sun-Ah & Cécile Fougeron. 2002. Realizations of accentual phrase in French intonation. *Probus* 14(1). 147–172.

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

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

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

Kaland, Constantijn & Nikolaus P. Himmelmann. 2020. Time-series analysis of F0 in Papuan Malay contrastive focus. In *Proceedings of the 10th international conference on speech prosody 2020*, 230–234. Tokyo: University of Tokyo.

Karlsson, Anastasia M. 2014. The intonational phonology of Mongolian. In Sun-Ah Jun (ed.), *Prosodic typology II*, 187–215. Oxford: Oxford University Press.

Kaufman, Daniel & Nikolaus P. Himmelmann. n.d. Suprasegmental phonology. In Adelaar Alexander & Antoinette Schapper (eds.), *The Oxford guide to the western Austronesian languages*. Oxford, UK: Oxford University Press.

Kim, Jonathan, Ute Gabriel & Gygax Pascal. 2019. Testing the effectiveness of the internet- based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based

(e-Prime3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLoS One* 14(9). e0221802.
Kluge, Angela. 2017. *A grammar of Papuan Malay*. Berlin, Germany: Language Science Press.
Kluge, Angela, Benny Rumaropen & Lodowik Aweta. 2014. *Papuan Malaydata-Wordlist*. Dallas, TX: SIL International.
Krinzinger, Helga, Jan Willem Koten, Julia Hennemann, André Schueppen, Katleen Sahr, Dominique Arndt, Kerstin Konrad & Klaus Willmes. 2011. Sensitivity, reproducibility, and reliability of self-paced versus fixed stimulus presentation in an fMRI study on exact, non-symbolic arithmeticin typically developing children aged 664 between 6 and 12 years. *Developmental Neuropsychology* 36(6). 721–740.
Ladd, D. Robert. 2008. Intonational phonology, 2nd edn. (Cambridge studies in linguistics). Cambridge, NewYork: Cambridge University Press.
Laures, Jacqueline S. & Gary Weismer. 1999. The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research* 42(5). 1148–1156.
Maskikit-Essed, Raechel & Carlos Gussenhoven. 2016. No stress, nopitchaccent, noprosodic focus: The case of ambonese Malay. *Phonology* 33(2). 353–389.
Mathôt, Sebastiaan, Daniel Schreij & Theeuwes Jan. 2012. Open Sesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324.
Mehta, Gita & Anne Cutler. 1988. Detection of target phonemes in spontaneous and read speech. *Language and Speech* 31(2). 135–156.
Mixdorff, Hansjörg & Dieter Mehnert. 1999. Exploring the naturalness of several German high-quality-text-to-speech systems. In *Sixth European Conference on speech communication and technology*. Budapest: International Speech Communication Association.
Moulines, Eric & Francis Charpentier. 1990. Pitch-synchronous wave form processing techniques for text-to-speech synthesis using diphones. *Speech Communication*. Neuropeech ’89 9(5). 453–467.
Özçelik, Öner. 2012. Redefining the prosodic hierarchy. *McGill Working Papers in Linguistics* 22(1). 1–13.
Pierrehumbert, Janet & Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*. Cambridge, MA: MIT Press.
R Core Team. 2019. *R: The R project for statistical computing*.
Rietveld, Toni & Carlos Gussenhoven. 1987. Perceived speech rate and intonation. *Journal of Phonetics* 15(3). 273–285.
Shattuck-HufnagelStefanie & Alice Turk. 1998. The domain of phrase-final lengthening in English. *Journal of the Acoustical Society of America* 103(5). 2889.
Shields, Joyce L., Astrid McHugh & James G. Martin. 1974. Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology* 102(2). 250–255.
Silverman, Kim, Mary Beckman, John Pitrelli, Ostendorf Mori, Colin Wightman, Patti Price, Janet Pierrehumbert & Julia Hirschberg. 1992. ToBI: A standard for labeling English prosody. In *Second international conference on spoken language processing*. Banff: International Speech Communication Association.
Slote, Joseph & Julia F. Strand. 2016. Conducting spoken word recognition research online: Validation and a new timing method. *Behavior Research Methods* 48(2). 553–566.

Stoel, R. B. 2007. The intonation of Manado Malay. In Vincent J. van Heuven & Ellen van Zanten (eds.), *Prosody in Indonesian languages*. vol. 9 (LOT occasional series), 117–150. Utrecht: LOT, Netherlands Graduate School of Linguistics.

Stoet, Gijsbert. 2010. PsyToolkit: A software package for programming psychological experiments using linux. *Behavior Research Methods* 42(4). 1096–1104.

Stoet, Gijsbert. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44(1). 24–31.

Taft, Marcus & Gail Hambly. 1986. Exploring the cohort model of spoken word recognition. *Cognition* 22(3). 259–282.

Van Rossum, Guido & Jelke De Boer. 1991. Interactively testing remote servers using the Python programming language. *CWI Quarterly* 4(4). 283–304.

Watson, Charles S., Henry W. Wroton, William J. Kelly & Carole A. Benbassat. 1975. Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America* 57(5). 1175–1185.