Word stress in prosodic theory

Constantijn Kaland

Studies in Laboratory Phonology 15

language
science
press

Studies in Laboratory Phonology

Chief Editor: Martine Grice

Editors: Doris Mücke, Taehong Cho

In this series:

6. Feldhausen, Ingo, Jan Fliessbach & Maria del Mar Vanrell (eds.). Methods in prosody: A Romance language perspective.
7. Tilsen, Sam. Syntax with oscillators and energy levels.
8. Ben Hedia, Sonia. Gemination and degemination in English affixation: Investigating the interplay between morphology, phonology and phonetics.
9. Easterday, Shelece. Highly complex syllable structure: A typological and diachronic study.
10. Roessig, Simon. Categoriality and continuity in prosodic prominence.
11. Schmitz, Dominic. Production, perception, and comprehension of subphonemic detail: Word-Final /s/ in English.
12. Schubö, Fabian, Sabine Zerbian, Sandra Hanne & Isabell Wartenburger (eds.). Prosodic boundary phenomena.
13. Albert, Aviad. A model of sonority based on pitch intelligibility.
14. Wehrle, Simon. Conversation and intonation in autism: A multi-dimensional analysis.
15. Kaland, Constantijn. Word stress in prosodic theory.

ISSN: 2363-5576

Word stress in prosodic theory

Constantijn Kaland

language
science
press

Constantijn Kaland. 2025. *Word stress in prosodic theory* (Studies in Laboratory Phonology 15). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/472

ISBN: 978-3-96110-512-0 (Digital)
978-3-98554-141-6 (Hardcover)

ISSN: 2363-5576
DOI: 10.5281/zenodo.15437940
Source code available from www.github.com/langsci/472
Errata: paperhive.org/documents/remote?type=langsci&id=472

Cover and concept of design: Ulrike Harbort
Typesetting: Constantijn Kaland
Proofreading: Bruno Behling, Jeroen van de Weijer, Ludger Paschen, Mary Ann Walter, Raquel Benítez Burraco
Fonts: Libertinus, Arimo, DejaVu Sans Mono
Typesetting software: XƎLaTeX

Language Science Press
Scharnweberstraße 10
10247 Berlin, Germany
http://langsci-press.org
support@langsci-press.org

Storage and cataloguing done by FU Berlin

FREIE
UNIVERSITÄT
BERLIN

voor mijn ouders

Contents

Acknowledgments

This book was defended as cumulative habilitation thesis by the author in 2024. Chapter 2 was published as Kaland (2019). Chapter 3 was published as Kaland (2020). Chapter 4 was published as Kaland et al. (2021). Chapter 5 was published as Kaland (2021).

Apart from the people acknowledged in each of the published studies, this book has been shaped throughout the years by a number of people in particular, who I would like to thank here.

Martine Grice; thank you for guiding me in the habilitation process and providing me with constructive feedback in the recent years. Over a much longer period of time, basically since the start of my studies, the work coming from the Phonetics Institute in Cologne has inspired me deeply. I am particularly grateful to you for this.

Nikolaus Himmelmann; thank you for the supervision of my work over the last years. It has shaped my research in many ways and was invaluable for the outline of the thesis. Thank you also for fuelling my interest in doing typological work. Your input has sharpened my thinking and helped me to keep a critical view on my own work.

Angela Kluge; thank you for the important work you did on Papuan Malay before I started working on it. It has provided an ideal basis to work with. Thank you also for the collaboration and help, in particular with Chapter 4.

Vincent van Heuven; thank you for the critical comments on several draft versions of the studies outlined here as well as on an earlier version of this book. You inspired me for over two decades now and it is a privilege to benefit from your understanding and experience. It made me realise that there is so much more to understand and to do.

Rob Goedemans; thank you for the lively discussions on the phonetics and phonology of word stress in Indonesian and beyond. They have reassured me that the two fields are too deeply intertwined to ignore one or the other. Thank you also for your valuable and encouraging input on several studies presented in this book. You remained patient and understanding when I went nuts on acoustic evidence.

Anne Cutler; although we were in touch only briefly, she provided me with very clear and constructive feedback. I am thankful for that, but even more so for the truly inspiring research she left behind. It has definitely set the direction of the experimental approaches in this book, as well as my thinking about language production and perception. Because unfortunately I cannot thank her in person any more, I believe mentioning her here is the least I can do.

Maximilian Hörl; thanks, Max, for your very helpful statistical advice. You always returned my questions with constructive feedback and many times presented solutions in a ready-to-use fashion. Often, you managed to understand where I wanted to go before I did. I am also thankful for our friendship and the nice times we have spent together.

Maria Bardají; thank you, Maria, for all the moments we shared in the office, and the great times we had after work. You are the ideal colleague; smiling, always in a good mood and at any moment ready to help, to listen or in for a joke. I value our friendship and miss our office life.

I am deeply indebted to the following people for having collected, transcribed, or annotated the data used in some of the studies presented in this book: Apriani Arilaha, Lisa Barz, Christoph Bracks, Pascal Coenen, Lenice Harms, Jonas Heinen, Marc Heinrich, Jeanete Lekeneny, Jan-Niklas Linnemeier, Jimi Kirihio, Jonathan Reich, Sonja Riesberg, Yusuf Sawaki, Emanuel Tuturop, Volker Unterladstetter, Boas Wabia, Katherine Walker, Katja Wiesner.

Thanks to the members of the habilitation committee for making the thesis succeed: Sefan Baumann, Martine Grice, Vincent van Heuven, Nikolaus Himmelmann, Doris Mücke, Sophie Repp.

A great word of thanks also goes out to the many colleagues in Cologne, in particular in the SFB-1252, the Linguistics and Phonetics departments. It has been an honour to work with or so close to you, exchange ideas, get inspired and encouraged or simply share a nice time. Among many others, thank you Aviad, Britt, Clare, Elina, Fafa, Frank, Heiko, Job, Katherine, Kurt, Malin, Mark, Nuria, Simon, Timo, Trudel.

Special thanks for the great editorial support from the Language Science Press provided by Aviad Albert, Bruno Behling, Raquel Benitez Burraco, Amir Ghorbanpour, Sebastian Nordhoff, Ludger Paschen, Elliott Pearl, Mary Ann Walter and Jeroen van de Weijer.

1 Introduction

(1) a. You need a permit to climb Mount Everest.
    b. They don't permit sailing on this lake.

Consider the word *permit* when reading out loud 1a and 1b. It becomes clear that the two versions of that same word sound different. The sole distinction between the two is generally considered as a difference in word stress (henceforth also 'stress'). Word stress is commonly understood as a single, most prominent syllable in a word that may be acoustically marked (e.g. Hyman 2006; Ladd 2008; Gordon & Roettger 2017). In the example above either the first syllable /pɜr/ (1a) or the second syllable /mɪt/ (1b) is the most prominent one, signalling a noun or a verb respectively.

For illustrative purposes, the example above is taken from English, simplified to show the most essential aspects of word stress. These are important observations to understand the motivation for the research conducted in this book, as they reflect a number of issues regarding our current understanding of word stress.

It is safe to assume that the majority of stress research to date has been conducted on English and other (related) well-studied languages. This is problematic because it has shaped the linguistic definition and understanding of word stress in ways that are not representative of a vast number of other languages. The problem of generalizing scientific theory based on predominantly Western data, is a problem found in other scientific disciplines as well (Henrich et al. 2010). Indeed, some aspects of word stress in English appear exceptional when compared to other languages (Cutler 2012).

Scientific work on stress research can be divided according to a number of subtopics, each representing a different aspect of word stress. Phonetic research has mainly focused on revealing which acoustic features make the stressed syllable prominent (e.g. Gordon & Roettger 2017). Language description, as often found in grammars, generally analyses word stress on the basis of its distribution; i.e. the possible positions in a word to which the stressed syllable can be assigned. These descriptions aim to reveal which rules predict the location of the stressed syllable, i.e. its phonology (e.g. Hyman 2006) and its role in prosodic structure (e.g. Gordon 2014). For a small number of languages the communicative functions of word stress were investigated, i.e. how listeners perceive stress (e.g. Peperkamp & Dupoux 2002) and how it might help them to process the incoming speech signal (e.g. Cutler 2005).

Thus, languages for which stress has been studied on all of the above mentioned aspects are still few and generally biased towards West-Germanic languages such as English, Dutch or German. The aim of this book is to contribute to stress research by including a typologically more distinct and less researched language; Papuan Malay. In addition, this book provides experimental investigations of most word stress aspects that have previously only been studied in well-documented languages. In this way, this book provides a solid basis for new typological comparisons that refine our understanding of word stress.

The remainder of this introduction provides an overview of stress research and is categorised according to the core aspects of word stress (Section 1.1). A discussion of why Papuan Malay is particularly relevant to study in this context is given in Section 1.2. Thereafter, the research questions addressed in the remaining four chapters are provided (Section 1.3).

1.1 Core aspects in research on word stress

This section discusses several aspects that have all been reported as relevant for the manifestation of word stress cross-linguistically. These aspects are often studied in different research (sub-)disciplines. They are categorised depending on whether they are relevant for the acoustic realization of word stress (Section 1.1.1), determining the position of the stressed syllable (Section 1.1.2), the role of word stress in the prosodic structure (Section 1.1.3) or the communicative function of word stress (Section 1.1.4).

1.1.1 Acoustics

When a language has word stress it often has one or more acoustically measurable correlate(s) that make stressed syllables more prominent than unstressed syllables. In a survey of 75 languages, Gordon & Roettger (2017) investigated which acoustic correlates turned out to be most revealing for word stress. The results showed that duration is the most often measured as well as the aspect most often reported to correlate with word stress. That is, stressed syllables were often reported to be longer than unstressed syllables. The lengthening due to word stress mostly affected syllable rhymes or entire syllables, depending on the language.

Measures of F0 were reported as a second major correlate, followed by intensity, formants and spectral tilt. It was noted that F0 is a particularly confounding correlate as it has been shown to (also) reflect phrase level prosody in a number of languages (Gordon & Roettger 2017). Interestingly, while in most of the investigated languages F0 was higher in the stressed syllable, some languages showed lower F0 in stressed syllables than in unstressed syllables (e.g. Lahore Urdu: Hussain 1997, and Italian: Eriksson et al. 2016). As for intensity, few studies have shown overall intensity to correlate with word stress. Measures that take the intensity of low versus high frequencies in the speech spectrum into account (e.g. spectral tilt), however, were more accurate. That is, the natural reduction of intensity towards higher frequencies is steeper in unstressed syllables than in stressed syllables (e.g. Dutch; Sluijter & van Heuven 1996). In other stress languages, spectral tilt did not correlate with word stress (e.g. Peninsular Spanish: Ortega-Llebaria & Prieto 2011). Formants, in particular F1 and F2, correlate with word stress as well. Commonly observed is the effect that vowels occupy a more peripheral position in the acoustic space when stressed (Crosswhite 2004). This form of decentralization has been explained from the perspective of enhancing prominence and can be observed in languages where corner vowels (/i/, /u/ and /a/) are likely candidates for word stress (e.g. in Bulgarian). An opposing, vowel centralizing effect of word stress has been observed in some languages and has been explained in terms of contrast enhancement, where the articulatory goal is not to increase prominence, but rather to be maximally different from other vowels (e.g. in Belarusian). In the latter case, corner vowels are less likely candidates for stress (Crosswhite 2004). Crucially, none of the above mentioned acoustic correlates can serve as an exclusive diagnostic for word stress. In particular, correlates of phrase level prosody can overlap with word stress cues and many studies have not explicitly separated the word level from the phrase level in their description of stress (Roettger & Gordon 2017). This problem exists in particular for languages in which phrase level accents align with stressed syllables at the word level. In these languages, e.g. West-Germanic ones, it is particularly challenging to disentangle the two prosodic levels, although research has shown that their acoustic cues are distinct (e.g. van Heuven 2018 on 'sentence stress'). It can therefore be difficult to group languages based on which acoustic cues signal word stress. Nevertheless, some generalizations can be made.

Several studies have observed that in tone languages, F0 is not or only weakly available as a correlate of word stress (Gordon & Roettger 2017). Thus, the set of acoustic correlates in a particular language depends on other functions these correlates might have in the prosodic structure. The role of word stress in prosodic structure is further discussed in Section 1.1.3. In addition, it has been shown that stress is weakly realised acoustically in languages with a highly predictable stress pattern (Dogil & Williams 1999). In other words, the less the position of word stress varies, the less the communicative need for large acoustic differences between stressed and unstressed syllables. The next section provides a more detailed discussion of various factors affecting the position of word stress across languages.

1.1.2 Distribution

When it comes to the position of the stressed syllable in a word, a general distinction has been made between languages in which the position of the stressed syllable is fixed, e.g. on the final syllable of a word, and languages in which the position of the stressed syllable can vary. In the latter type of language, stress is labeled "variable" or "free" as its position is not defined relative to word boundaries, but rather depends on the (morpho-)phonological properties of the syllable. It is more useful to interpret the fixed versus free/variable stress languages as a continuum, rather than two distinct groups. It has been shown, for example, that languages can shift over time between one and the other (Baerman 1999). Where a language is placed on the fixed-variable continuum depends on a number of factors that determine the position of stress. These factors, as discussed in the remainder of this subsection, define the relationship between stress as a suprasegmental notion on the one hand, and the actual consonantal and vocalic makeup of the segments in the syllable structure on the other hand.

For example, the quantity of a syllable (i.e. weight: Hyman 1985) can play a role such that heavy syllables are more likely to be stressed than light syllables. Which syllable structures count as heavy or light differs per language, but the distinction always depends on their segmental makeup. To capture the fine-grained weight distinctions, phonological theories assume the mora on an intermediate level between the segments and the syllable (McCawley 1968). Furthermore, much of the stress positioning across languages can be explained by the sonority of the segments in the syllable. Vowels are assumed to form the most sonorous class of speech sounds with potentially relevant sonority differences among them as well (Parker 2002). Vowel quality as such can determine the position of stress, as has been shown for Kobon (Kenstowicz 1997). Although it is challenging to find accurate acoustic correlates of sonority (Parker 2002; Albert & Nicenboim 2022), it should be noted that sonorous segments are particularly rich in acoustic properties that could correlate with stress. For example, F0 or spectral tilt are more meaningful correlates for vowels than for consonants. Articulatorily, it is easier to increase or decrease the duration or intensity of sonorous speech sounds than of obstruents. This makes sonorous segments the preferred loci for suprasegmental contrasts such as the difference between stressed and unstressed syllables. The close relationship between segments and the position of word stress can be a challenge to a strict suprasegmental interpretation of word stress, which is discussed in more detail in the next section with examples from English and Dutch.

In addition, directionality influences the way stress is assigned to syllables. That is, the location of the stressed syllable is either counted from the left or the right edge of the word (Howard 1972). An example of the former type can be found in Latvian (Halle & Vergnaud 1987), whereas the latter, cross-linguistically more common type can be found in Armenian (Vaux 1998). It has also been shown that the position of the stressed syllable can be "bounded" in that it falls within a certain number of syllables in the word (e.g. in Kobon: Kenstowicz 1997. Other languages, however, are "unbounded" in that there is no such restriction (e.g. Hindi: Bailey 1995). Note that the predictability of the stress position differs per language and affects the perception of word stress (Peperkamp et al. 2010). That is, for some languages (morpho-) phonological rules are more successful in predicting the position of stressed syllables (e.g. Finnish), than for others (e.g. Spanish). Predictability in this sense refers to how advantageous the ruleset is for listeners to process stress patterns. This notion is crucially different from the fixed-free continuum, in that high predictability does not necessarily mean that a language has a small number of exceptions to the standard stress pattern. It has been shown that in a language with more unpredictable exceptions to the default stress placement, listeners perform better when asked to recall sequences of nonsense syllables with a given pattern of stressed and unstressed syllables (Peperkamp et al. 2010). Thus, listeners appear to memorise stress patterns better when these patterns are more challenging to capture by rules.

1.1.3 Prosodic structure

Another aspect according to which stress languages differ is the role of word stress within larger linguistic structures. We can distinguish work on the role of stress in several stages of speech production (e.g. Levelt 1989) from work on the role of stress on different phonological levels (e.g. Nespor & Vogel 2007). As for the former, the aim of investigation is where in the speech production process word stress information is available. A distinction has been made between languages in which stress is an inherent property of words in the lexicon and languages in which stress information is not stored in the lexicon. In the former language type, word stress needs to be learned by heart and plays a more important role in distinguishing between word meanings. Word stress in this type of language is termed "phonemic" when it can be the only distinction between two otherwise identical words. This is the case in English, where word class can be indicated by stress on the first syllable (noun) or on the second syllable (verb) in words such as *permit*.

Note that the primary acoustic difference between stressed and unstressed syllables in English concerns vowel quality, whereas other cues are present in a rather marginal way (Fear et al. 1995). This indicates that word stress is not strictly suprasegmental, as it can be a property of the segments. Studies have therefore referred to "strong" versus "weak" syllables instead of "stressed" versus "unstressed" syllables (see Cutler et al. 1997 for an overview). However, in a closely related language such as Dutch, vowel quality is, among other cues, a suprasegmental property of word stress (Sluijter & van Heuven 1996). To illustrate the difference between Dutch and English stress; Dutch listeners are better at recognizing stressed syllables in English than native listeners of English, because the former are sensitive to more acoustic cues than the latter (Cutler et al. 2007). In Dutch, stress has also been described as "lexical", thus referring to the fact that there are minimal stress pairs (e.g. /ˈkaːnɔn/ and /kaːˈnɔn/, translating as 'canon' in the musical and military sense, respectively). Given the differences between Dutch and English just described, languages thus differ in the extent to which word stress is phonemic and lexically stored. Although this aspect relates to the mental representation of stress, it clearly parallels the mobility distinction in terms of "free" and "fixed" (see Section 1.1.2). That is, in languages where the position of word stress is generally fixed (e.g. Polish; Jassem 1962), the need to store stress information lexically is limited.

When it comes to other levels in the prosodic structure, stress has been shown to interact with phrase-level prosody in a number of languages. This has been a recurring problem in the search for an accurate set of acoustic correlates of word stress (Roettger & Gordon 2017) and for the phonological definitions of "stress" and "accent" (Beckman 1986; Hyman 2006). Research on Germanic languages has been dominant in both revealing and dealing with this problem. This is most likely the result of a prosodic feature found in these languages: phrase-level prominences in terms of pitch accents are structurally aligned with stressed syllables at the word level (e.g. Ladd 2008). In other languages, however, the prosody of words and phrases are related in a different way or not related at all. For example, acoustic analyses of Kuot (Lindström & Remijsen 2005) have shown a strict separation of the word-and phrase-prosodic level, both in terms of acoustic cues and in terms of functionality. It appeared that in Kuot, duration is exclusively correlated with word stress, whereas pitch is used to signal phrase boundaries. Crucially, there is no overlap between these cues or functions in this language. Furthermore, research has not been able to find reliable acoustic correlates of word stress in Seoul Korean, let alone a predictable position of stressed syllables. While early work analyzed Seoul Korean as a stress language, later work showed that the phrase level rather than the word level is relevant for predicting prominent syllables (Jun 1993). Recent accounts, therefore, separate word prosody explicitly from phrase prosody in order to draw more accurate distinctions between languages (e.g. Gordon 2014). It has been proposed, for example, that word-level prominences in many languages might actually reflect phrase-level prominences. That is, languages sometimes avoid the co-occurrence of phrasal boundary tones and phrase accents on the same single syllable (i.e. tonal crowding) by means of edge repulsion. Note that other languages add vocalic material to achieve the same goal (e.g. Grice et al. 2018). In this case, pre-final instead of final syllables in the phrase are marked by F0 movements that constitute the phrase accent. Gordon (2014) observes a gap in the typology of word and phrase prosody in that there seem to be no languages with edge repulsion at the word level and not at the phrase level. Languages have edge repulsion either at both levels (e.g. Egyptian Arabic; Hellmuth 2007), only at the phrase level (e.g. Chickasaw; Gordon 2005b), or at neither level regardless of whether the language makes use of pitch accents (e.g. Hebrew; Becker 2003) or not (e.g. Wolof; Rialland & Robert 2001). Thus, there is no typological evidence for languages where stress shows edge repulsion and phrase level accents do not, as would be the case for languages with penultimate stress where phrase level accents are realised on the final syllable. This observation is taken as an indication that word stress is actually a reflection of phrase prosody in languages of the world. Note that this analysis is centered around the finding that F0 is a correlate of phrase prosody rather than word prosody (e.g. in Dutch: Sluijter & van Heuven 1996 and in Chickasaw: Gordon 2005b).

1.1.4 Communicative functions

The final aspects of word stress discussed here are its communicate functions. Most of these functions relate to the processing and recognition of words, while others concern the contribution of word stress to the rhythm of a language. This distinction thus reflects the extent to which word stress is functional only for the word domain or (also) beyond. One function that has often been attributed to word stress is its ability to discriminate between two otherwise identical words, as referred to in the discussion of phonemic stress above. However, it appears that the number of minimal stress pairs in languages is marginal (Cutler 2005).

Thus, minimal stress pairs mainly provide interesting opportunities for experimentation, instead of reflecting a core function of word stress.

Psycholinguistic studies make a distinction between word stress as a cue to word segmentation and word stress as a cue to word identification (Cutler et al. 1997). In the former, listeners use stress cues to detect word boundaries, whereas in the latter listeners use stress cues to detect the target word among possible competitors. As for word boundaries, studies on induced misperceptions have shown that listeners expect a word boundary before a stressed syllable, a perceptual strategy that explains much of the stress distributions in Dutch (Vroomen et al. 1996) and English (Cutler & Butterfield 1992). Note that languages in which the position of the stressed syllable varies less than in Dutch or English, stress cues could be assumed to be more helpful to listeners for word segmentation (i.e. fixed stress languages). However, the acoustic differences between stressed and unstressed syllables are small in languages where stress is rather fixed (see Cutler 2005: p.73 for a discussion). It therefore remains to be seen to what extent word stress cues are functional in this type of languages. The latter question is a challenge, given that listeners are less sensitive to acoustic cues, when stress patterns in their language are more predictable (Peperkamp et al. 2010). For example, ERP data showed that Turkish listeners were hardly sensitive to the default (fixed) ultimate stress, whereas deviations were readily detected (Domahs et al. 2013). A study on Polish revealed similar results in that listeners mainly showed an ERP effect associated with anomaly detection (P300) for the non-default stress pattern rather than for the default one (Domahs et al. 2012). These results corroborate the idea that the number of exceptions to the default position of word stress is of crucial importance to determine the functionality of stress cues for listeners (Peperkamp et al. 2010; Cutler 2005). When it comes to actual word identification, studies on Dutch and English have shown that word stress has a facilitating effect in tasks without time-pressure (see Cutler et al. 1997 for an overview). However, in online (time-pressured) processing, listeners have more difficulty selecting target words among competitors that only differ in the position of word stress. Strong-weak patterns were recognised consistently faster than weak-strong patterns, which seemed to suggest a stress related facilitation effect (Cutler & Clifton 1984). This could have been interpreted as an indication that word recognition starts as soon as the stressed syllable is heard; hence shorter reaction times for strong-weak patterns than for weak-strong patterns. However, the reaction time differences turned out to be artifacts of the stimulus duration. When stimulus duration was corrected for, both patterns led to similar reaction times (Cutler & Clifton 1984).

Word stress also contributes to the perceived rhythm of a language. Studies have attempted to find acoustic correlates of isochrony, the idea that speech rhythm is achieved by intervals of equal length between either stresses or syllables (Pike 1945; Abercrombie 1967). No acoustic support has been found for the dichotomous distinction between stress-timed and syllable-timed languages (e.g. Dauer 1983; Arvaniti 2012), although languages appear to differ on a continuum when other rhythm metrics are taken into account (e.g. Grabe & Low 2002; Ramus et al. 2000; Prieto et al. 2012). As the search for strict isochrony in speech production has been largely abandoned, the view that speech *perception* relies on equal intervals of some sort has become dominant (e.g. Lehiste 1977; Ramus et al. 2000). It is furthermore clear that syllable structure, vowel reduction and stress all contribute to speech rhythm (Dauer 1983). It has also been shown that phrase-level phenomena such as final lengthening and accentual lengthening have consequences for speech rhythm metrics (Prieto et al. 2012). How the contribution of word stress to perceived rhythm can best be quantified, remains an open question (see Jun 2014 for a proposal on *macro-rhythm*). In sum, most of the functions of word stress have been demonstrated for languages with variable positions of word stress (e.g. English and Dutch). Although there is evidence that the predictability of the stress pattern influences its cognitive processing (Peperkamp et al. 2010; Domahs et al. 2012; Domahs et al. 2013) and helps to segment incoming speech into words (Cutler & Butterfield 1992; Vroomen et al. 1996), more research on other languages with predictable stress patterns is needed (Cutler 2012).

1.2 Papuan Malay

Particularly interesting to our understanding of word stress are Indonesian languages. It has been a matter of debate whether Indonesian is a stress language (e.g. Odé 1994) and studies have pointed out that this question is too generic. That is, the specific language variety needs to be taken into account to answer this question (e.g. Goedemans & van Zanten 2007). Although more recent research has addressed this, there are still many aspects uncovered. Before the studies presented in this book were undertaken, no Indonesian variety had been studied systematically and comprehensively for its existence of word stress. Javanese, Toba Batak, Ambonese Malay, Manado Malay, Betawi Malay, Besemah and some other varieties all received some attention in stress studies. However, none of them have been studied for acoustic, perceptual and functional aspects together.

Before the studies in this book, Papuan Malay had been studied extensively in a grammar (Kluge 2017) and preliminary work on phrase-level prosody had been carried out (Riesberg et al. 2018). Nevertheless, word prosody had not been studied empirically, despite the claim that this language has regular penultimate stress (Kluge 2017). The language is spoken in the Indonesian provinces Papua and West Papua by approximately one million speakers. More relevant background information on this language is given in the respective studies. The reader is referred to the most recent grammar (Kluge 2017) for an extensive historical and sociolinguistic overview of Papuan Malay. The availability of native speakers makes this language suitable for extensive fieldwork and experimentation, which elicited (semi-)spontaneous speech (e.g. Riesberg & Himmelmann 2012, Kluge et al. 2014). The work conducted for this book would not have been possible without the existing documentation just mentioned. This applies in particular to the clearly defined and testable stress hypothesis, which led to the studies outlined here (Kluge 2017). Thus, Papuan Malay provides an opportunity to study word stress from different angles, combining new as well as (modern versions of) existing methodologies. The reader is referred to Kaland & Baumann (2020) and Kaland et al. (2023) for studies on phrase-level aspects, and to Kaland et al. (2018) and to Kaland & Gordon (2022) for other word-level aspects of Papuan Malay prosody.

1.3 Summary and research aims

The literature overview above is far from complete. However, the previous sections have provided a coarse categorisation of aspects that should be covered in a comprehensive study of word stress: acoustic cues (in production and perception), distributional aspects, its role in the prosodic structure and its communicative functions. While in theory all stress aspects could be covered using this categorisation, it is likely that when research advances, different perspectives ask for a different categorisation or shed a different light on existing categories. It should also be noted that there is natural overlap between these aspects. For example, the investigation of stress perception cannot go without a study of its acoustic or distributional aspects. Furthermore, the current book maintains an empirical approach to word stress, using acoustic analyses, experiments and corpus analyses. This does not entail that the approach is purely phonetic. As becomes clear from the studies discussed above, the understanding of word stress crucially relies on our understanding of the phonetics-phonology interface. This book is no exception and the (phonological) distribution of word stress patterns has therefore a central place in this work. At the same time, this book does not cover each of the aforementioned aspects in equal detail. The studies presented in the following chapters cover the acoustic realisation, the perception of stress patterns, the distribution and function of word stress in the lexicon, and whether listeners benefit from stress cues for word recognition. Thus, the interaction between word stress and other levels in the prosodic structure are covered only to smaller extent.

Research question: *Does Papuan Malay have word stress?*

Ch.2) *To what extent does the acoustic signal provide evidence for word stress?*
Ch.3) *To what extent are listeners able to perceive the stress patterns?*
Ch.4) *To what extent do lexical analyses provide evidence for word stress?*
Ch.5) *Are listeners able to use stress cues to identify words?*

The central research question (RQ) investigated in this book is formulated above. Four sub-questions are formulated that each cover a specific aspect, following the order of chapters. RQ Ch.2 is investigated by an acoustic analysis of spontaneous monologues, covering more than ten different acoustic measurements of stressed versus unstressed syllables. RQ Ch.3 is investigated by a series of three perception experiments on the relevance of acoustic cues to listeners in online and offline processing of words. RQ Ch.4 is investigated using two word-corpus-based studies on the word-disambiguation function of stress patterns and the phonological properties that underlie these patterns respectively. RQ Ch.5 is investigated using a gating experiment.

2 Acoustic correlates of word stress in Papuan Malay

2.1 Introduction

Papuan Malay is considered a Trade Malay variety, spoken in the Indonesian provinces Papua and Papua Barat. Other Trade Malay varieties include Ambonic (Ambonese, Banda), Kupang, Larantuka and Manadoic (Manado, North Moluccan), see Paauw (2009) for a comparative overview. The relatively little research that has been carried out on these languages consists mostly of descriptive grammars, with little to no attention paid to prosody. Most of the Trade Malay varieties are claimed to have word level stress. In some varieties stress occurs at variable locations and might be phonemic (i.e. the sole distinction between two otherwise identical words), while in others the stress location is rather fixed and non-phonemic. In the current study, both types are henceforth referred to as (word) stress; i.e. the single acoustically most prominent syllable in a word. The goal of the present study is to investigate whether there is acoustic evidence for word stress patterns in Papuan Malay. The communicative function of potential word stress patterns is beyond the scope of the current study, although some discussion is provided in the final section.

As the majority of the stress claims for Trade Malay varieties have remained without empirical support, it has been questioned to what extent prosodic descriptions were influenced by the (mostly Western) background of the authors (Himmelmann 2018). In fact, recent work on Ambonese Malay did not find acoustic evidence for word stress, counter to earlier claims (Maskikit-Essed & Gussenhoven 2016). Studies on varieties of Indonesian indicated that stress, if present at all, has a different status in the prosody compared to well-studied languages such as English. The current study on word stress correlates in Papuan Malay provides a necessary empirical investigation. In addition, it sheds light on an under-researched language with potential implications for the study of prosody in related languages.

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

In the next (sub)sections an overview is given of word stress claims in all Trade Malay varieties (Section 2.1.1). Thereafter, studies on word stress in Indonesian, a closely related language, are discussed (Section 2.1.2). Furthermore, an overview of common acoustic correlates found cross-linguistically is given in Section 2.1.3 and the methodological concerns in previous work on word stress are listed in Section 2.1.4. Finally, Section 2.1.5 states the research questions.

2.1.1 Stress in Trade Malay

In this section, all Trade Malay varieties distinguished in Paauw (2009) are discussed to the extent that word stress claims have been made. Phrase level prosody is discussed to a limited extent, focusing mainly on its relation to word level phenomena. Table 2.1 gives a summary of the aspects of the literature on Trade Malay, which are most relevant to the current study. These include the position of word stress, whether it is phonemic, whether acoustic measures were carried out, whether the collected samples consisted of free speech (e.g. spontaneous unscripted) or constrained speech (due to scripting or prompting) and the context in which the samples appeared (isolation, phrase-medial, phrase-final or uncontrolled).

2.1.1.1 Papuan Malay

Kluge (2017) carried out extensive fieldwork on 44 speakers, predominantly from the Sarmi region in the Papua province. Observations based on a list of 1116 words consisting of Papuan Malay roots revealed that word stress is located on the penultimate syllable in 90% of the cases. The remaining 10% show stress on the ultimate syllable. In many of the cases in which stress is located on the ultimate syllable, the penultimate syllable contains the vowel /ɛ/. It cannot be concluded that /ɛ/ rejects word stress, as 7% of the words with penultimate stress contain /ɛ/ (Kluge 2017: 96). The other reported Papuan Malay vowels are /i/, /a/, /ɔ/ and /u/. Although the observations in Kluge (2017) are mainly based on the author's auditory impressions, a preliminary acoustic analysis of spectral tilt as correlate of word stress confirmed the stress claims to a large extent (Kaland 2018). Furthermore, no minimal stress pairs are available in the word lists in Kluge (2017). It deserves to be noted that near-minimal stress pairs tend to consist of /ɛ/ in the penultimate syllable, for example: [ˈbe.bas] 'be free' and [be.ˈban] 'burden', [ˈbeŋ.kɔk] 'be crooked' and [beŋ.ˈkak] 'be swollen', [ˈe.nak] 'be pleasant' and [e.ˈnam] 'six', and [ˈme.maŋ] 'indeed' and [me.ˈnaŋ] 'to win'. Also in older and smaller descriptions of Papuan Malay, word stress has been undisputedly assumed to occur regularly on the penultimate syllable (e.g. Donohue & Sawaki 2007). As for phrase prosody, the few studies available suggest that marking of information structure using pitch movements (i.e. pitch accents) is limited to phrase boundaries in Papuan Malay (Kaland et al. 2018; Riesberg et al. 2018).

2.1.1.2 Ambonese Malay

In a description of Ambonese Malay, van Minde (1997) assumed irregular word stress. No indications are given about the (default) location of word stress and, crucially, no report on the existence of schwa is provided (Paauw 2009). The main reason for irregularity is the phonemic nature of stress; hence the minimal pair /ˈba.rat/ 'west' and /ba.ˈrat/ 'heavy'. However, a re-evaluation and acoustic measurements were carried out on read speech samples from four speakers (Maskikit-Essed & Gussenhoven 2016). The samples were comparable to the ones in van Minde (1997) and led to the conclusion that word stress does not exist in Ambonese Malay. The vowel /a/ that occurs in syllables described as unstressed by van Minde (1997) was seen as a different phoneme (a-caduc) in Maskikit-Essed & Gussenhoven (2016) on the basis of its distinct spectral characteristics. In addition to the reanalysis of the phoneme inventory, Maskikit-Essed & Gussenhoven (2016) did not find reliable acoustic evidence for stress in duration, F0 peak alignment and spectral tilt for the examples provided in van Minde (1997). Furthermore, no support was found for the use of pitch accents in this language.

2.1.1.3 Kupang Malay

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

2.1.1.4 Larantuka Malay

Word stress in Larantuka Malay was described in Kumanireng (1993) and cited in Paauw (2009). Historically, Larantuka did not lose schwa and therefore developed regular rather than phonemic word stress (Paauw 2009). The default location, like in other Trade Malay varieties, is the penultimate syllable. The literature does not report acoustic analyses of word stress in Larantuka Malay.

2.1.1.5 Manado Malay

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

2.1.1.6 Tidore (North Moluccan Malay)

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

2.1.1.7 Ternate (North Moluccan Malay)

A corpus of short stories and anecdotes was collected for Ternate and extensively described in a grammar (Litamahuputty 2012). Stress was claimed to fall on the penultimate syllable regularly. Stress occurred on the ultimate syllable when the penultimate syllable would have consisted of a schwa in other Malay varieties.

Ternate was assumed not to have schwa in its vowel inventory and therefore to have developed phonemic stress. No acoustic analyses were provided for word stress in Ternate.

2.1.1.8 Summary Trade Malay

The general claim made in descriptive work on Trade Malay varieties is that word stress occurs on the penultimate syllable unless that syllable contains a schwa or a vowel that can be historically linked to schwa. Crucially, the only acoustic analysis available showed no support for the existence of word stress or pitch accents in Ambonese Malay (Maskikit-Essed & Gussenhoven 2016). These results contrast with the analysis of Manado Malay in which stress and pitch accents could co-occur (Stoel 2007). In sum, although stress claims are highly similar for all Trade Malay varieties, at least two issues remain unresolved. First, it is by no means granted that impressionistic stress claims find acoustic support. Second, if stress indeed does exist, it cannot be assumed that all Trade Malay varieties have similar stress placement due to the vast and geographically remote areas in which they are spoken. Thus, existing stress claims need to be complemented with in-depth acoustic investigations of the Trade Malay languages. The discrepancies between the outcomes of descriptive work and those of experimental investigations are already reported in the literature on word stress in Indonesian, which has been studied more extensively. The current state of the work for Indonesian therefore illustrates the potential added value of (more) empirical investigations of word prosody in otherwise under-researched languages. For this reason, and because Indonesian has a considerable influence on Trade Malay varieties (Paauw 2009), a summary of work on word stress in Indonesian is given in the next section.

2.1.2 Stress in Indonesian

Various studies have investigated to what extent word stress exists in Indonesian. Early work claimed that stress occurs regularly on the penultimate and only moves to the ultimate syllable when the penultimate contains a schwa (Alieva et al. 1991; Teeuw 1978). In other work, the ultimate syllable has been claimed the default location for word stress (Samsuri 1971). An overview of mainly impressionistic work on stress in Indonesian is provided in Odé (1994: 41), followed by a perception experiment on phrasal prominence. The remainder of this subsection, however, is devoted to a discussion of experimental work on Indonesian word stress.

Table 2.1: Overview of stress claims for native lexical roots in studies on Trade Malay varieties. For varieties not listed no stress claims have been made. Abbreviations: **P**enult/**U**ltimate/**I**rregular word stress; **F**ree/**C**onstrained speech samples in **I**solated/**M**edial/**F**inal/**U**ncontrolled (phrase) contexts; n.a. = information not available or not reported.

Variety	Source(s)	Position	Phonemic	Acoustics	Sample		Context			
					F	C	I	M	F	U
Papuan	Donohue & Sawaki (2007)	P, alt. U	n.a.	No	x	x	x			x
	Kluge (2017)	P, alt. U	No	No	x	x	x	x	x	x
Ambonese	van Minde (1997)	I	Yes	No		x	n.a.			
	Maskikit-Essed & Gussen-hoven (2016)	-	-	Yes		x		x	x	
Kupang	Steinhauer (1983)	P, alt. U	Yes	No	n.a.		n.a.			
Larantuka	Kumanireng (1993); Paauw (2009)	P	n.a.	No	n.a.		n.a.			
Manado	Stoel (2007)	P, alt. U	Yes	No	x	x		x	x	x
Tidore	van Staden (2000)	P, alt. U	No	No		x				x
Ternate	Litamahuputty (2012)	P, alt. U	Yes	No	x					x

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

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

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

These results are in line with an earlier study on the perception of manipulated pitch movement timings by Toba Batak and Javanese speakers (van Heuven & van Zanten 1997). It was predicted that the accuracy of locating pitch movements within word boundaries would be better for Toba Batak speakers, because of word stress in this language. Locating pitch movement across word boundaries was predicted to be similarly accurate for both languages, as both show evidence for the use of phrase accents. It was found that speakers of Toba Batak were more accurate in indicating the location of pitch movements than speakers of Javanese, regardless of whether pitch movements were within or across word boundaries. This outcome was taken as evidence that the timing sensitivity of Toba Batak speakers at the syllable level (within word), shaped their sensitivity at the higher level. Javanese listeners were presumably lacking this sensitivity at the lower level and therefore it would not emerge at higher levels at all (van Heuven & van Zanten 1997).

Another acoustic study compared Toba Batak with Betawi Malay (Roosman 2007). For each variety, four speakers read carrier phrases which were analysed by several spectral and temporal measures as well as by expert listeners. Target words occurred in and out of focus as well as in phrase medial and final position.

Toba Batak was observed to have contrastive word stress, with the penultimate syllable as default location. Word stress in Toba Batak was signalled by a clear pitch movement, regardless of focus condition. In Betawi Malay, no evidence for word stress was found. This finding is in line with another study on Betawi Malay that found highly variable pitch movements on the penultimate syllable (van Heuven et al. 2008). The pitch movements were mainly found on the ultimate syllable in cases where the penultimate consisted of a schwa or when it occurred in phrase final position, indicating a privileged status of schwa and a crucial role of phrase level factors determining the F0.

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

2.1.3 Word stress correlates in a cross-linguistic context

Word stress' most common acoustic correlates include duration, pitch (F0), formant frequencies (F1 and F2), intensity and spectral tilt (see Gordon & Roettger 2017 for an overview of 110 studies on 75 languages). Although duration appeared to be the most consistent correlate of word stress across languages, not all acoustic correlates are equally reliable. For example, some of the studies on Indonesian word stress found pitch to be the most important correlate (Halim 1981; Laksman 1994; Roosman 2007). While these outcomes hold for the Indonesian varieties that were studied, work on Germanic languages generally found pitch as a correlate of phrase level prosody (i.e. Bolinger 1958; Sluijter & van Heuven 1996). Moreover, many studies on word stress have failed to keep word and phrase level phenomena apart (Roettger & Gordon 2017). In particular, words read in isolation as well as the lack of control for phrase accents in the experimental design have confounded the two levels. It could therefore be problematic to rely on pitch as a main correlate of word stress, in particular if little is known about a language's prosody at other levels.

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

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

In a limited number of studies, frequency sensitive intensity measures appear as stress correlates (i.e. spectral tilt, spectral balance, spectral emphasis, spectral slope). The correlation with word stress is defined by the intensity reduction towards higher frequencies, which is smaller in stressed syllables (shallow downward tilt) than in unstressed syllables (steep downward tilt). Although there are various ways in which this correlate has been measured (Heldner 2003), it has been taken as a reliable indicator of whether word stress exists or not, as it reflects vocal effort. This has been shown for a variety of languages of the world (Gordon

& Roettger 2017); among other languages, Dutch (Sluijter & van Heuven 1996), Catalan (Ortega-Llebaria & Prieto 2011) and Ambonese Malay (Maskikit-Essed & Gussenhoven 2016). However, in some studies, frequency sensitive intensity measures did not reliably correlate with word stress, even though languages in these studies were shown to make use of word stress. For example, in American English (Campbell & Beckman 1997) and Swedish (Heldner 2003) these measures rather correlated with focal accent.

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

2.1.4 Methodological considerations

In addition to the above mentioned issues concerning acoustic measurements, there are some methodological pitfalls frequently observed in studies on word stress. That is, little or no attention has been given to the importance of the formality of speech. On a continuum of formality, read aloud laboratory speech and unscripted interactive talking would constitute either extreme end. It was observed in Roettger & Gordon (2017) that the majority of studies on word stress make use of scripted laboratory speech. To determine the existence of stress it is preferred to investigate spontaneous and interactive speech for two reasons. First, languages of the world tend to have an oral tradition in most cases. If a writing tradition exists, it would have commonly developed from orality (Ong 1982). In other words, read-aloud speech would compromise the extent to which a given speech sample represents natural language. Although previous work made use of elicited readings (e.g. Maskikit-Essed & Gussenhoven 2016), it has to be noted that most Trade Malay varieties in particular have a limited writing tradition (Paauw 2009). Therefore, there is considerable discrepancy to expect between scripted and spontaneous speech in these languages. Second, the realisation of stress differs significantly between read and spontaneous speech samples (Howell & Kadi-Hanifi 1991), although it is possible that this finding was influenced by phrase level phenomena as well. Throughout the literature on word stress in general (Roettger & Gordon 2017) and on Trade Malay prosody in particular (Table 2.1), reading tasks have commonly been used. Here it is argued that, in the investigation of word stress cross-linguistically, unscripted spontaneous speech leads to the most representative samples and is preferred over read speech. Given the variable nature of unscripted spontaneous speech, more conservative decisions need to be made to select a corpus of samples suitable for acoustic analysis (see Section 2.2.3).

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

2.1.5 Research questions

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

To investigate these issues, the current study focuses on Papuan Malay. Word stress in Papuan Malay is claimed to occur regularly on the penultimate syllable (Kluge 2017). This claim was tested by means of acoustic analysis in the current study. Two specific questions were investigated for Papuan Malay. First, to what extent is there acoustic evidence for word stress? Second, if there is, which acoustic measures predict word stress best? Spectral, temporal and amplitudinal measures were computed, both in raw form and in derived form. The derived measures were taken to avoid acoustic effects that were unrelated to the word stress hypothesis. Linear mixed model analyses were carried out to investigate these questions (Section 2.2.4). Crucially, the investigated speech data consists of spontaneous narratives.

2.2 Methodology

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

2.2.1 Data collection procedure

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

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

2.2.2 Participants

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

2.2.3 Data processing and selection

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

A subset of the labelled syllables was selected on the basis of the following criteria. Syllables in utterances that were interrupted or cut off, were omitted. Reduplicated words (e.g. tiba-tiba) were omitted, whereas the syllables of single occurrences (e.g. tiba) were taken into account. This was done because reduplication could affect acoustic properties of the word being reduplicated and make that word less comparable to single occurrences. The syllables of words produced with hesitation or that were unidentifiable due to laughter, severe speech reduction (i.e. mumbling) or background noise were also omitted. Words containing double vowel sequences were also omitted, because they allow for two ways of syllabification (either as VV or V.V, see Kluge 2017). These sequences appeared to lead to inconsistent annotations in the current study.

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

The Papuan Malay lexicon has a considerable number of loanwords, mainly originating from Indonesian. It is known from previous work that Indonesian varieties differ with respect to their use of word stress (Section 2.1.2). In order to exclude possible influences of loanwords on the realisation of word stress, only syllables of words that were classified as native roots (Kluge 2017) were selected. Given the common occurrence of affixation in Malay varieties, the selected forms were unaffixed (roots). Furthermore, single syllable words were omitted as their acoustic characteristics cannot be compared to other syllables in the word. This comparison is required to investigate possible stress differences between syllables within the same word. In addition, the most common word length in Papuan Malay is two syllables (Kluge 2017), as confirmed by the collected data in the current study. In order to have a consistent dataset, words with a length other than two syllables were omitted.

Table 2.2: Vowel counts in words with presumed penultimate and ultimate stress in syllables labelled as stressed and unstressed in collected data. For readability purposes the position of the syllable in the word is indicated (pre-final/final).

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

After applying the selection criteria just described, a total of 3437 syllables were left for acoustic analysis (20% of the data). They were all labelled as ‘stressed’ or ‘unstressed’ by taking over the stress indications in word lists in Kluge (2017).

Table 2.3: Counts of syllables labelled as stressed and unstressed in different syllable structures. In italic: absolute standardised residuals for each of the two chi-square tests described in Section 2.2.3. For readability purposes the position of the syllable in the word is indicated (pre-final/final).

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

This resulted in 1682 syllables labelled stressed and 1755 syllables labelled as unstressed, occurring in 187 unique words. An overview of how the vowel nuclei are distributed among syllables labelled for stress, is given in Table 2.2. An overview of the syllable structures and stress positions is given in Table 2.3. The most common stress position was the penultimate syllable, and when the ultimate syllable was stressed, the unstressed (penultimate) syllable always contained /ɛ/. In a number of cases, however, /ɛ/ can occur in stressed penultimate position. This reflects the distributions reported in Kluge (2017) and, crucially, does not mean that /ɛ/ is the only factor explaining the mobility of word stress.

In order to investigate to what extent there is a relationship between stress and syllable structure, chi-square tests (R Core Team 2017) were performed on the counts in Table 2.3 for each of the stress positions separately (penultimate and ultimate). The tests revealed significant differences between the observed values and values that would be expected on the basis of a chance-level distribution, both for penultimate stress [$\chi^2(5, N = 3239) = 272.36, p < 0.001$] and ultimate stress [$\chi^2(1, N = 198) = 80.16, p < 0.001$]. The differences between observed and expected are given in Table 2.3 as absolute standardised residuals. This allows to assess the difference between observed and expected values across syllable structures, and accounts for the differences in row and column totals (Agresti 2007). For penultimate stress, the largest differences between observed and expected values were found for V and CVC syllables. For ultimate stress, the only available syllable structures for the chi-square test (CV and CVC) both show high residual values. Note that for CVC syllables, these results mainly indicate that this syllable structure is most frequent in word-final position. When stress position and syllable position are compared for CVC syllables, the distribution is not significantly different from chance level [$\chi^2(1, N = 660) = 3.05, p = 0.08$]. This sheds a different light on the potential relation between word stress and syllable structure. It appears from Table 2.3 that the most common word structure is CV.CVC. Given that penultimate stress is the most common stress pattern, it is challenging to disentangle word stress and syllable structure. However, a more general observation can be made, in that syllables with higher vocalic proportions (V, CV, VC) are more often stressed than syllables with low vocalic proportions (CVC, CCV, CCVC); [$\chi^2(1, N = 3437) = 79.17, p < 0.001$], even when CV and CVC syllables are left out [$\chi^2(1, N = 535) = 86.45, p < 0.001$]. Thus, the distributions of word stress among syllable structures indicate some sensitivity for syllable structure. It remains unclear, however, to what extent this sensitivity indicates a phonological distinction between "heavy" and "light" syllables. To date, however, there is no known source on Papuan Malay that explicitly distinguishes light and heavy syllables for word stress. The only context in which heavy syllables are mentioned in Papuan Malay is (partial) reduplication (Kluge 2017: ch.4). Furthermore, ultimate stress in the selected dataset only occurred on syllables containing /i/, /a/ or /u/. While this could be a reflection of their frequency of occurrence, it is unclear whether there is an additional phonological reason for this distribution. For example, Papuan Malay word stress could be attracted to the most extreme vowels in the acoustic space.

2.2.4 Acoustic measures

All acoustic measures were taken using Praat (Boersma & Weenink 2017). The measurements were taken automatically using scripts. Measures were categorised according to which aspect of the acoustic signal they describe; spectral, temporal and amplitudinal measures. Both raw acoustic measures and derived acoustic measures were taken. The raw (unconverted) measures were not expected to reliably correlate with word stress cross-linguistically (Gordon & Roettger 2017). None of the raw measures were therefore included in the statistical analyses (Section 2.2.5). For reference purposes and because until now it is unclear how word stress could potentially be realised in Papuan Malay, the descriptive values of the raw measures were nevertheless reported (Table 2.7). The derived measures were either a correction, conversion or alternative calculation of the raw measures, in order to improve their accuracy and to minimise the possible effects of acoustic processes that were not part of word stress (e.g. phrase prosody). That is, relative measures of F0 and intensity were calculated on the basis of the difference between syllables labelled as stressed and syllables labelled as unstressed within each word, as described in more detail in the subsequent sections. The acoustic measures were taken from different (sub)intervals relative to the annotated syllables (Figure 2.1); either the entire syllable (syllable), a voiced subinterval (svoiced) or a voiced subinterval around the intensity peak (intpeaks). The different subintervals were used to ensure accurate spectral measures (Table 2.4). The syllable level intervals (syllable) were obtained from the manual annotation procedure (Section 2.2.3). The two subintervals were extracted in an automatised way in two subsequent rounds. In the first round, syllable subintervals for which Praat (Boersma & Weenink 2017) was able to detect periodicity were taken (svoiced). Syllables for which F0 tracking errors occurred were omitted. In a second round, portions were taken from the voiced subintervals where formants showed stable values. Stable formant values guarantee accurate frequency measurements and were mostly found close to where intensity levels reached their peak. Therefore, boundaries of these stable portions were calculated relative to the intensity peak (intpeaks). This was done by taking the time stamps at which the intensity dropped 4% relative to the peak intensity on either side. The 4% margin was chosen after manual inspection of the spectrogram in order to find an optimum between stability of formant detection and interval length. That is, smaller margins resulted in intervals that were too short for (reliable) formant measures, whereas larger margins more often resulted in intervals where formants were harder to distinguish from each other due to higher amounts of variability in the detected frequency values. When the 4% drop on either side of the intensity was not found within the boundaries of the syllable, the syllable boundary was set as the boundary for the subinterval with stable formant values. This occurred in a minority of the cases, and if so, the syllable mostly started or ended with a vowel. All measures taken from this subinterval (intpeaks) accounted for gender differences in the frequency range of speech. Maximum formant frequencies were set to 5000 Hz for male speakers and to 5500 Hz for female speakers.

2.2.4.1 Spectral measures

The following spectral measures were distinguished: the F0 measures raw F0, relative F0 and F0 minima/maxima, and the formant measures F1 and F2. F0, F1 and F2 measures were expressed on logarithmic scales (F0 in semitones, F1 and F2 in Bark) to account for how pitch is perceived and, in the case of F0 movements, to abstract over gender differences (Traunmüller & Eriksson 1994). F0 minima

Table 2.4: Overview of all acoustic measures, units and intervals from which the measures were taken.

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

and F0 maxima were measured for both their level in ST and their timestamp in ms. From the minima/maxima values the measures F0 movement, rise/fall ratio and movement onsets/offsets were derived. Raw F0 was measured as the mean F0 value per syllable.

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

F0 movement was measured by subtracting the minimum F0 from the maximum F0 in the voiced subinterval in the syllable (svoiced). Measures of F0 change have been shown to be more reliable indicators of word stress than static F0 measures in Estonian, Italian and Thai (Gordon & Roettger 2017). Labels for direction of movement (rise or fall) were derived from the F0 minima and maxima timestamps. That is, whenever the F0 minimum occurred before the F0 maximum, the direction was labelled "rise" and whenever the F0 minimum occurred after the

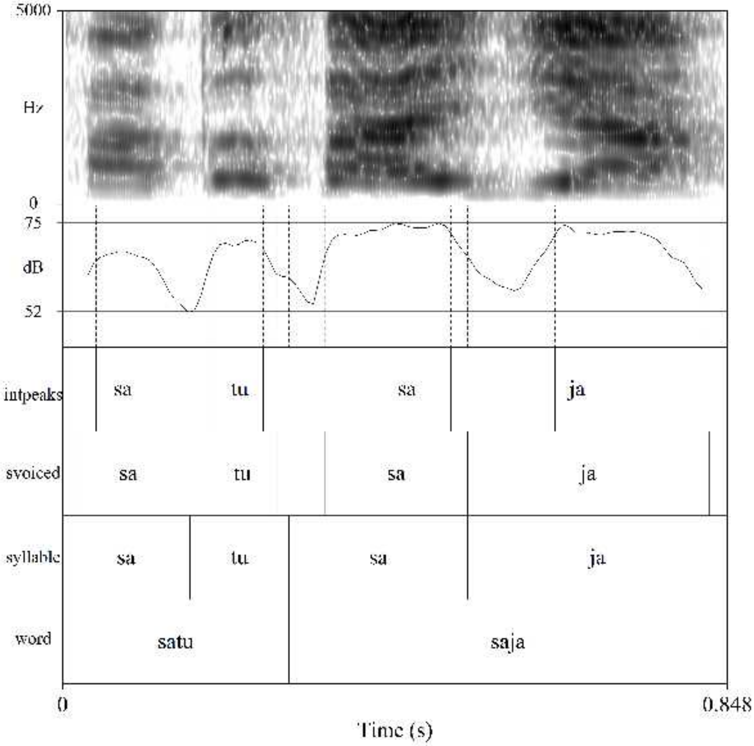

Figure 2.1: Example phrase "satu saja" ('only one') with different annotation levels (bottom), intensity curve (mid) and spectrogram (top) from which different acoustic measures were taken.

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

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

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

2.2.4.2 Temporal measures

Three temporal measures (expressed in milliseconds) were taken: raw duration, duration per phoneme, and duration deviation. Raw duration was measured by taking the absolute length of the entire syllable. It is known that word stress in some languages is attracted by phonologically heavy syllables. The weight of the syllable commonly depends on its segmental structure, which is not accurately captured by absolute syllable duration. Although Kluge (2017: ch.4) makes reference to heavy syllables in relation to reduplication processes, it is unclear to what extent there is a morpho-phonological preference for word stress placement in

Papuan Malay. In order to account for this possibility, the two other temporal measures took into account the segmental makeup of the syllable. Duration per phoneme was computed by taking the raw syllable duration and dividing it by the number of phonemes in the syllable. It has to be noted that the relationship between number of phonemes and syllable duration is not linear. That is, final lengthening generally affects codas more than onsets (e.g. Campbell & Isard 1991). Number of segments has nevertheless been a reliable and commonly used correlate in the modelling of syllable duration (Campbell & Isard 1991; Fletcher & McVeigh 1993). Furthermore, in some languages, consonant durations correlated better with word stress than vowel duration (Gordon & Roettger 2017). In order to account for the actual vocalic and consonantal makeup of the syllable, duration deviation was computed separately for each of the syllable structures in the data (CV, CVC, V, VC, CCV, CCVC). This was done by subtracting the mean duration of all syllables with a particular structure from the absolute duration of the syllable with that structure. The outcome represented the deviation in milliseconds of a particular syllable relative to the mean duration. Positive duration deviation was reflective of long syllables, whereas negative duration deviation was reflective of short syllables. Thus, both duration per phoneme and duration deviation take into account the segmental makeup of the syllable and were expected to be more accurate correlates of word stress compared to raw duration.

2.2.4.3 Amplitudinal measures

Six amplitudinal measures (expressed in decibels; dB) were taken: raw intensity, relative intensity and two measures of spectral tilt, each in uncorrected and corrected form. Spectral tilt was grouped under the amplitudinal measures as it contributes to the perceived loudness of speech and is expressed in dB in the current study. Raw intensity was measured by taking the mean intensity from the entire syllable. It has to be noted that raw intensity is not expected to be a reliable correlate of stress as the recording procedure did not control for the distance between the mouth of the speaker and the microphone (see Section 2.1.3). For this reason, relative intensity was calculated in the same way as was done for relative F0. That is, the intensity of the syllable labelled as unstressed was subtracted from the intensity of the syllable labelled as stressed. Positive relative intensity values therefore indicate that the syllable labelled as stressed had a higher intensity than the syllable labelled as unstressed. In this way, possible head movements, changes in speaking style (i.e. excited louder speech) or phrase level effects are better controlled for.

As for spectral tilt, the method of Sluijter et al. (1995) and Stevens & Hanson (1995) was adopted for half of these measures. That is, intensities of the second formant (A2) or the third formant (A3) were subtracted from the intensity of the F0 (first harmonic; H1). H1 and A2 were corrected for the effect of F1, and A3 was corrected for the effect of both F1 and F2. The overall mean F1 and overall mean F2 were used as reference values for the corrections, following the formulas in Hanson (1997: 113-115). Spectral magnitude correction (Iseli et al. 2007) has been applied in previous work to account for the increased intensity around formant frequencies, in particular F1 and F2 (Sluijter et al. 1995; Stevens & Hanson 1995). Note that other work did not apply this correction and argued that the corrected measure is not able to reliably separate word stress effects on vowel quality from those on spectral tilt, as the two are highly correlated (see Ortega-Llebaria & Prieto 2011 for Spanish). Furthermore, spectral magnitude correction has been argued to result in less accurate formant intensity measures (Caballero & Carroll 2015). In the current study, therefore, both the uncorrected (H1-A2 and H1-A3) and corrected (H1*-A2* and H1*-A3*) were taken as approximations to the intensity slope (tilt) of the spectrum. H1, A2 and A3 intensities were extracted using a Praat script based on Mayer (2014) from the voiced portions of syllables where formants showed stable values (intpeaks). Additional measures of spectral tilt in Papuan Malay are discussed in Kaland (2018).

2.2.5 Statistical analysis

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

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

With regard to F0 movement, movement onset/offset, formant displacement, duration per phoneme, duration deviation, H1-A2, H1*-A2* , H1-A3 and H1*-A3* , LMMs fit by maximum likelihood (using Satterthwaite approximations to degrees of freedom to calculate p-values) were carried out using the "lme4" package

(Bates et al. 2015). A separate analysis was done for each acoustic measure as response, with stress (two levels: stressed, unstressed) and stress position (two levels: penultimate, ultimate) as predictors and with subjects (speakers) and items (words) as random intercepts. As for movement onset/offset, also direction (two levels: rise, fall) was added as predictor to analyse timing differences between rises and falls. As for movement onset/offset and duration per phoneme, syllable structure (six levels: V, CV, VC, CVC, CCV, CCVC) as additional random intercept was added to account for timing differences due to the respective syllable structures. Rise/fall ratio was response in a generalised linear mixed effect model analysis using the "lme4" package (Bates et al. 2015). Stress (two levels: stressed, unstressed) and stress position (two levels: penultimate, ultimate) were predictors and subjects (speakers) and items (words) were included as random intercepts.

Concerning formant displacement, additional post-hoc pairwise comparisons using Tukey HSD test (Bonferroni corrected) were performed on the interactions between the fixed factors stress (two levels: stressed, unstressed) and vowel (five levels: /i/, /ɛ/, /a/, /ɔ/ and /u/) with subjects (speakers) and items (words) as random intercepts. This was done to test for displacement effects due to stress for each vowel separately (see Section 2.1.3) using the "multcomp" package (Hothorn et al. 2008). To investigate the extent to which each of the acoustic measures could predict word stress, generalised linear mixed model (GLMM) analyses were carried out with model fit comparisons "lme 4" package; Bates et al. 2015). A full model was created with stress (two levels: stressed, unstressed) as response, and with F0 movement, rise/fall ratio, movement onset, movement offset, formant displacement, duration per phoneme, duration deviation, H1-A2, H1*-A2* , H1-A3 and H1*-A3* as predictors, and with subjects (speakers) and items (words) as random intercepts. Note that the acoustic measures relative F0 and relative intensity were not included in the model as these measures expressed word stress relative to neighbouring syllable(s) and had therefore no predictive value for the dichotomous dependent variable word stress (see Section 2.2.4). In the full model, some of the variables were highly correlated, in particular the temporal measures and the measures of spectral tilt (Table 2.5). Therefore, the contribution of the independent variables was assessed in a stepwise manner using LRTs (Likelihood-Ratio-Tests) between a null model (with only the random variables subjects and items) and a model in which one independent variable was added to the null model. As an estimate of how well a model fitted the data, Akaikes An Information Criterion (AIC; Akaike 1998) was computed for each model in the comparisons. Although the absolute AIC values have no predictive value, AIC values of the models relative to each other can be used to obtain a ranking, with the lower AIC values indicating better model fit (Burnham & Anderson 2004).

Table 2.5: Moderate to (very) strong correlations (> 0.40; Evans 1996) between predictors in the full GLM model as measured by Pearson $r(N = 3395, p < 0.001)$. Other predictors did not correlate or correlated only weakly.

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

2.3 Results

This section reports the (G)LMM and LRT effects (Table 2.6) and means (Table 2.7), mean formant values and their pairwise comparisons (Table 2.8), and the results of the model fit comparisons (Table 2.9).

2.3.1 Spectral measures

Relative F0 deviated from zero for both penultimate and ultimate stress (Figure 2.2). In the case of penultimate stress, the deviation was negative, indicating overall lower F0 for syllables labelled as stressed than for syllables labelled as unstressed. In the case of ultimate stress, the deviation was positive, indicating overall higher F0 for syllables labelled as stressed than for syllables labelled as unstressed. The deviation from zero was only significant in the case of ultimate stress. For penultimate stress, a marginally significant difference from zero was found. As for position, a significant effect indicated that relative F0 was larger for ultimate stress than for penultimate stress.

F0 movements showed larger values for syllables labelled as stressed than for syllables labelled as unstressed (Figure 2.3). This difference was, however, not significant. A marginally significant effect was found for stress position, in that

Table 2.6: Overview of effects for all acoustic measures. Interactions not reported were not significant.

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

Table 2.7: Means and SDs for all acoustic measures according to stress and stress position.

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

Table 2.8: Mean F1 and F2 (Bark) and SDs per vowel for syllables labelled as stressed and unstressed, and effects of stress (Tukey HSD) on formant displacement from/towards the vowel space centre.

	F1		F2		Displacement	
Vowel	Stressed	Unstressed	Stressed	Unstressed	Re. centre	*p*
/i/	4.86 (1.06)	5.13 (1.09)	13.34 (1.08)	13.17 (1.04)	from	n.s.
/ɛ/	5.38 (0.84)	5.45 (0.93)	12.47 (0.97)	12.87 (1.09)	towards	= 0.06
/a/	6.42 (1.02)	5.94 (0.90)	11.80 (1.01)	11.94 (1.10)	from	< 0.001
/ɔ/	5.83 (0.81)	5.89 (1.24)	10.33 (0.91)	10.57 (1.36)	from	n.s.
/u/	5.16 (1.18)	5.32 (0.92)	11.31 (1.60)	11.59 (0.89)	from	< 0.001
Centre	5.72 (1.15)		12.20 (1.28)			

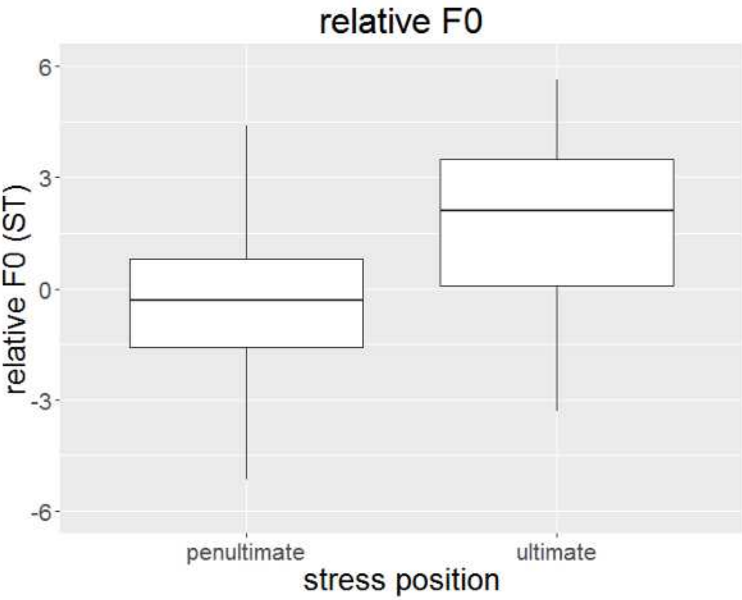

Figure 2.2: Boxplot (bar indicates median) of relative F0 (ST) as a function of stress position.

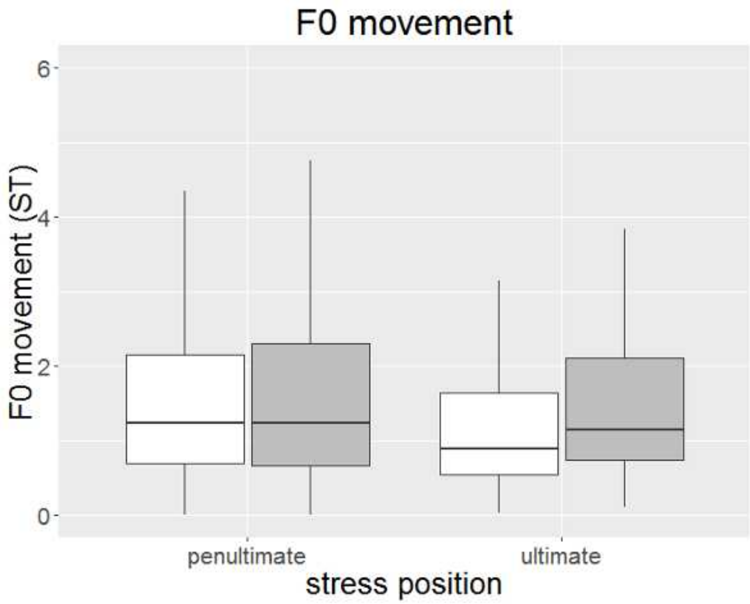

Figure 2.3: Boxplot (bar indicates median) of F0 movement (ST) as a function of stress (grey = stressed, white = unstressed) and stress position.

F0 movements were larger in the case of penultimate stress when compared to ultimate stress. A marginally significant interaction effect showed that stress differences were larger in ultimate position than in penultimate location.

Overall, rise/fall ratios indicated more falls than rises (Table 2.7). The ratios differed significantly as a function of stress, in that higher ratios (increase in rises/decrease in falls) were found for syllables labelled as stressed than for syllables labelled as unstressed (Table 2.6). No (interaction) effect involving position was found.

As for the timing of the F0 movements (Figure 2.5), onsets occurred significantly later (i.e. closer to the syllable midpoint) in syllables labelled as stressed than in syllables labelled as unstressed. The significant effect of position indicated that F0 movement onsets occurred later for ultimate stress than for penultimate stress. The effect of direction indicated that onsets occurred earlier for rises than for falls. The interaction between stress and direction indicated that the effect of stress was larger for rises than for falls.

Offsets occurred significantly later in syllables labelled as unstressed than in syllables labelled as stressed. The effect of direction indicated that offsets occurred earlier for rises than for falls. The interaction between stress and position indicated that the effect of stress was larger for ultimate stress than for penultimate stress. The interaction between stress and direction indicated that the effect of stress was larger for rises than for falls. The three-way interaction indicated that offsets were significantly affected by stress, position and direction such that the latest offsets were found for falls in syllables labelled as stressed when stress was penultimate and the earliest offsets were found for rises in syllables labelled as unstressed when stress was ultimate (see also Table 2.7). Stress differences were largely absent for fall offsets when stress was penultimate.

Formant displacement relative to the centre of the vowel space was overall significantly larger for syllables labelled as stressed than for syllables labelled as unstressed (Table 2.6). No significant difference in formant displacement was found between the respective stress positions. The interaction between stress and position was significant in that the formant displacement effect due to stress was mainly present in the case of penultimate stress. The post-hoc pairwise comparisons (Table 2.8) revealed significant displacement from the centre of the vowel space for /a/ and /u/, marginally significant displacement for /ɛ/ and no significant displacement for /i/ and /ɔ/. Crucially, the displacement of /ɛ/ was towards the centre of the vowel space, while other vowels were displaced away from the centre (Figure 2.4).

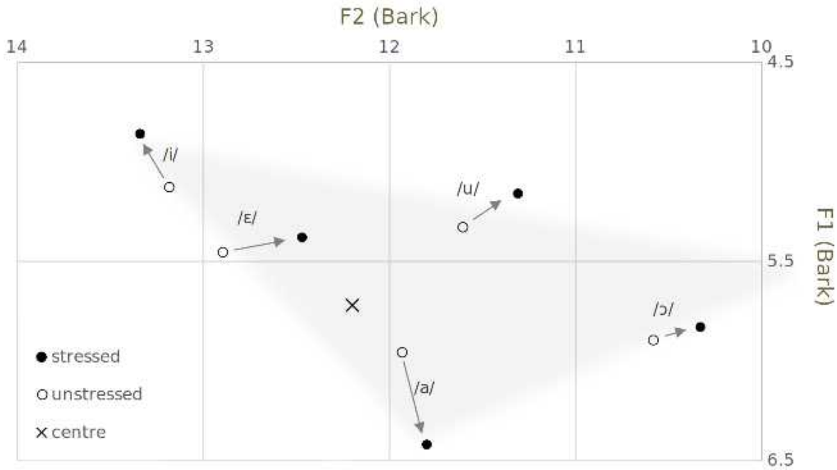

Figure 2.4: Papuan Malay vowels in syllables labelled as stressed and unstressed. The grey area represents the estimated vowel space, with the centre mark indicating the overall mean formant values (Table 2.8).

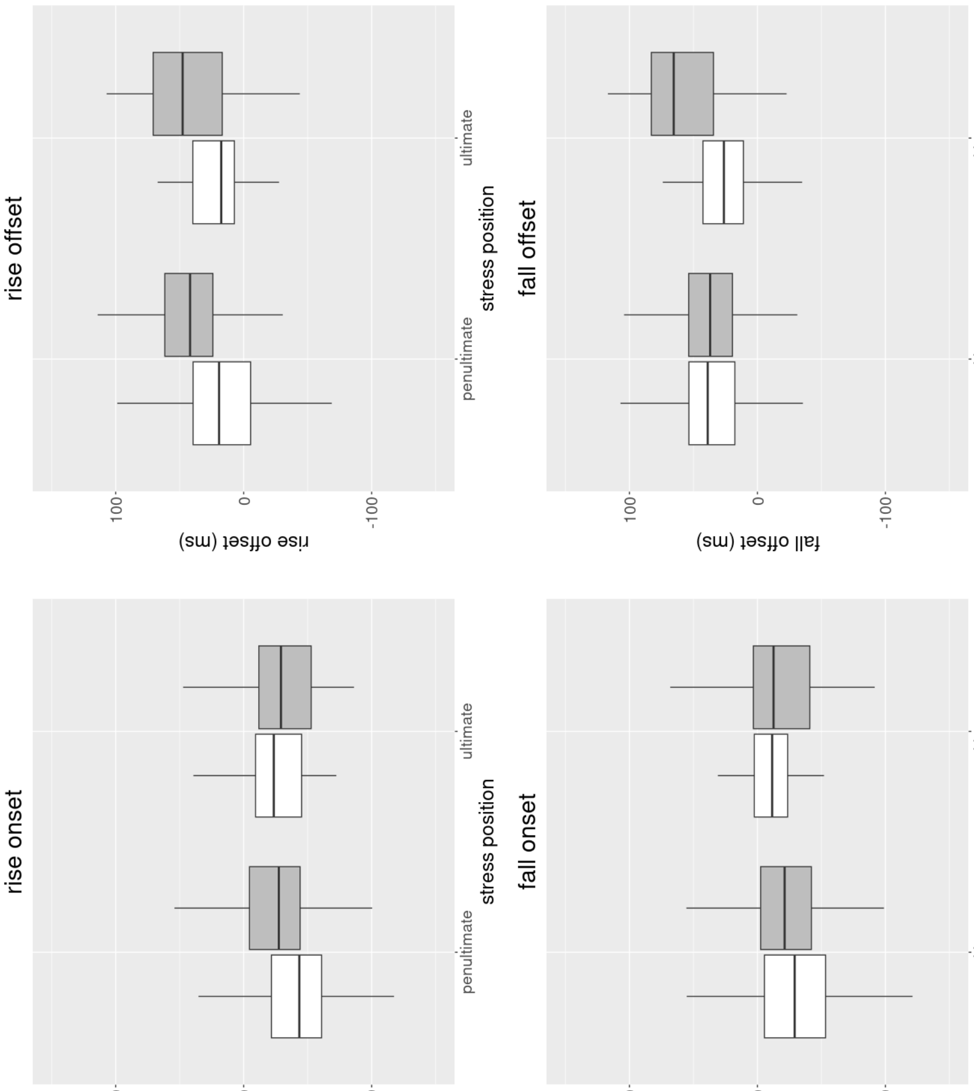

Figure 2.5: Boxplots (bar indicates median) of onsets (left) and offsets (right) of rises (top) and falls (bottom) as a function of stress (grey = stressed, white = unstressed) and stress position.

2.3.2 Temporal measures

Duration per phoneme showed larger values for syllables labelled as stressed than for syllables labelled as unstressed (Figure 2.6); a difference that appeared to be significant. The effect of position was only found significant in interaction with stress, indicating that the effect of stress (around 8 ms increase in either stress position) occurred on overall longer phoneme durations for penultimate stress than for ultimate stress.

Overall positive values of duration deviation (Figure 2.7) were found for syllables labelled as stressed and negative values were found for syllables labelled as unstressed. This effect was significant. The effect of position was only significant in interaction with stress, indicating that duration deviation was generally larger for ultimate stress than for penultimate stress.

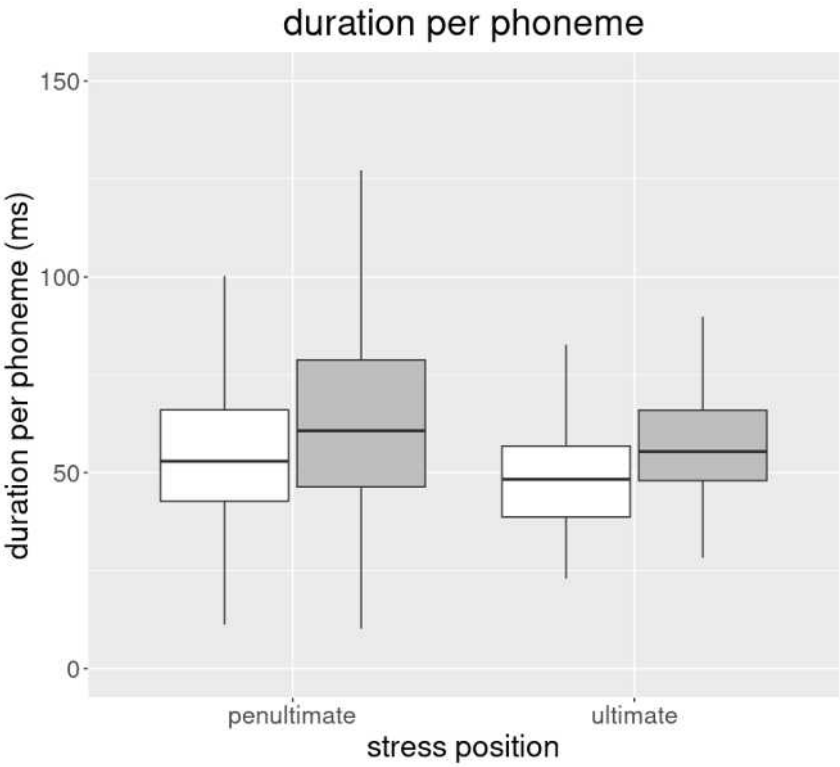

Figure 2.6: Boxplot (bar indicates median) of duration per phoneme (ms) as a function of stress (grey = stressed, white = unstressed) and stress position.

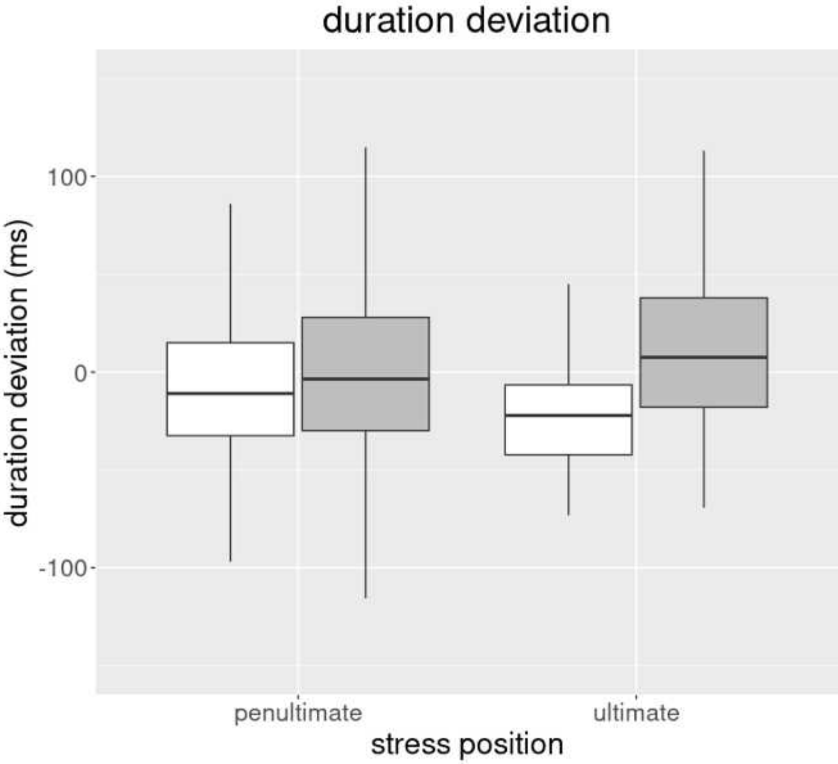

Figure 2.7: Boxplot (bar indicates median) of duration deviation (ms) as a function of stress (grey = stressed, white = unstressed) and stress position.

2.3.3 Amplitudinal measures

Relative intensity generally showed positive deviations from zero (Figure 2.8). This deviation was, however, only significant in the case of ultimate stress. The effect of position indicated that relative intensity values were larger for ultimate stress than for penultimate stress.

As for H1-A2, syllables labelled as stressed showed significantly lower values compared to syllables labelled as unstressed (Fig. 9). No effect of position was found. The same results were obtained for H1*-A2* (Figure 2.9).

H1-A3 values were significantly lower in syllables labelled as stressed than in syllables labelled as unstressed (Figure 2.9). There was no significant effect of position. H1*-A3* revealed no significant effects involving stress or position (Figure 2.9).

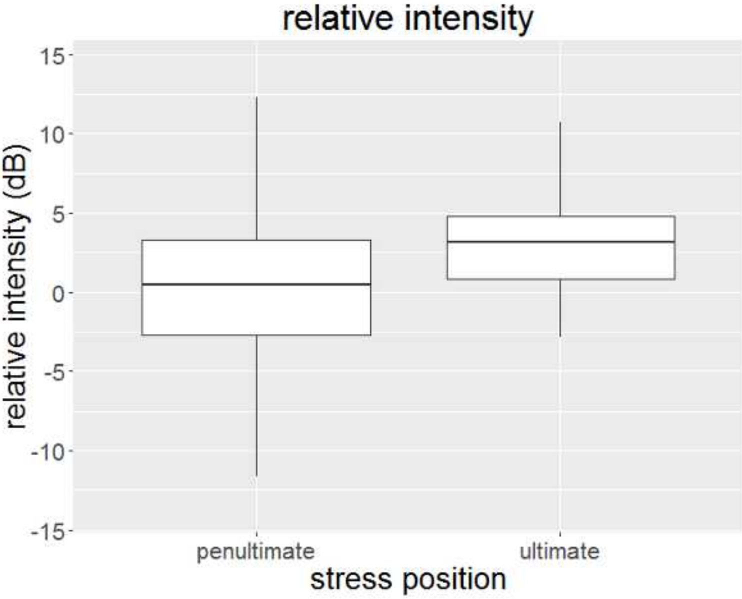

Figure 2.8: Boxplot (bar indicates median) of relative intensity (dB) as a function of and stress position.

2.3.4 Predictors of word stress

Table 2.9 reports the relevant statistical outcomes for each of the acoustic measures in the GLMM analysis. As for F0 movement, the null model comparison indicated no significant differences when adding this variable. Crucially, the AIC value indicated slightly better fit for the null model, where F0 movement was not included. Adding rise/fall ratio significantly improved the null model, with AIC values showing a small decrease. Movement onset as well as movement offset significantly improved the null model, with the offset showing the best model fit obtained in this analysis (lowest AIC) and onset showing the second best model fit. Adding formant displacement significantly improved the model fit and this variable was ranked relatively high, as indicated by the AIC scores. The model comparisons showed the third largest AIC improvement for duration per phoneme, which appeared to be significant. Duration deviation significantly improved the model fit, with a moderately ranked AIC value. H1-A2 contributed significantly to the model fit when added to the null model. The AIC values indicate that this contribution was moderate. H1*-A2* showed significant effects, with relatively low AIC rankings. As for H1-A3, a significant effect was found when adding this variable. AIC rankings showed a low contribution to the model fit for H1-

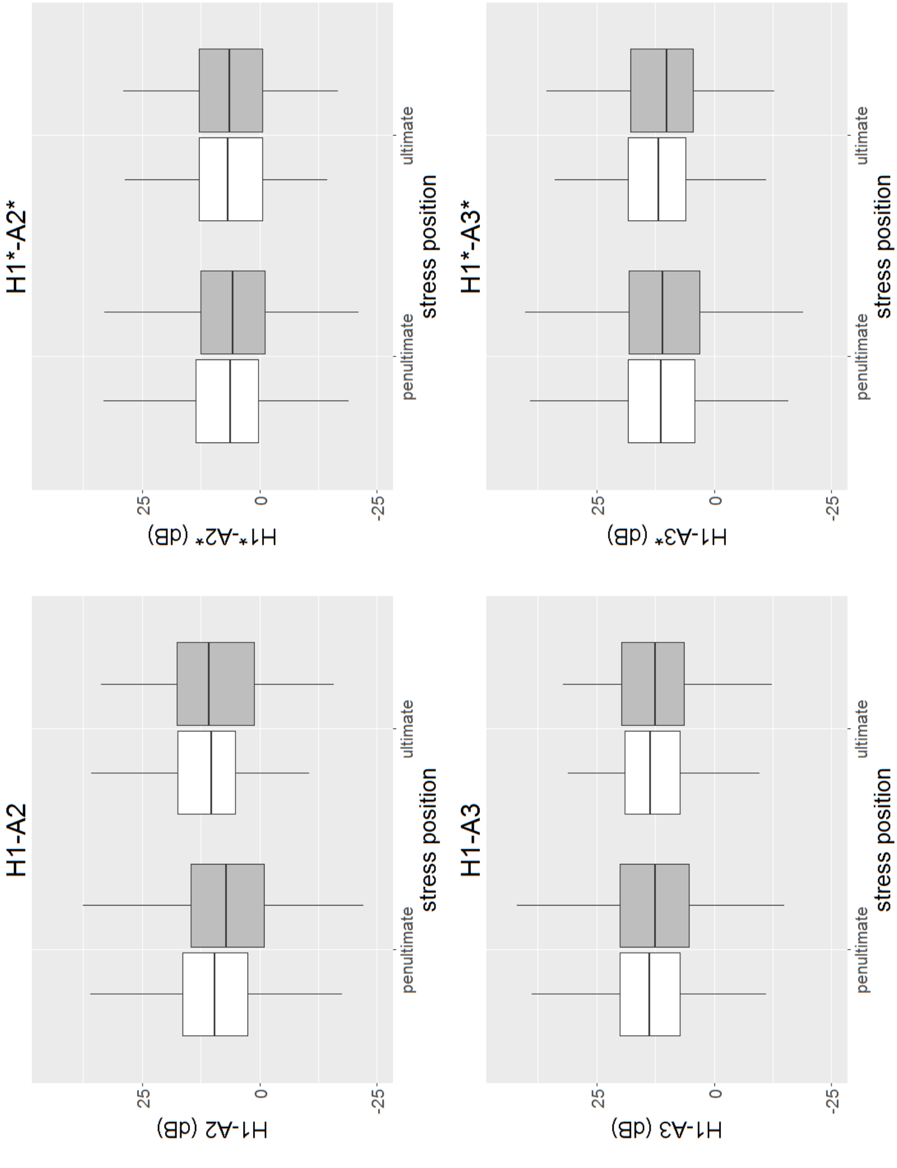

Figure 2.9: Boxplots (bar indicates median) of all spectral tilt measures (dB) as a function of stress (grey = stressed, white = unstressed) and stress position.

Table 2.9: Results of the GLMM analysis on the full model (b, SE, z) and LRT model fit comparisons with null model (AIC,χ^2, p). Ranks of null model comparisons are indicated in superscript. Null model AIC given in heading.

	GLMM (full)			Added to null model		
Acoustic measure	b	SE	z	AIC (4711.22)	χ^2	p
F0 movement	0.04	0.02	2.44	4711.31^{10}	1.92	n.s.
Rise/fall ratio	0.34	0.08	4.44	4699.41^{7}	13.82	< 0.001
Movement onset	1.50	1.38	7.59	4633.08^{2}	80.15	< 0.001
Movement offset	7.53	1.38	5.45	4609.00^{1}	104.22	< 0.001
Formant displacement	0.35	0.05	7.01	4649.25^{4}	63.97	< 0.001
Duration per phoneme	27.34	3.56	7.68	4642.75^{3}	70.47	< 0.001
Duration deviation	-0.01	0.00	-5.65	4697.34^{6}	15.89	< 0.001
H1-A2	-0.06	0.01	-6.51	4680.66^{5}	32.57	< 0.001
H1*-A2*	0.05	0.01	4.90	4706.25^{8}	6.97	< 0.01
H1-A3	0.03	0.01	3.68	4709.29^{9}	3.94	< 0.05
H1*-A3*	-0..02	0.01	-3.48	4712.03^{11}	1.19	< n.s.

A3. With regard to H1*-A3*, no significant effect was found and the AIC value worsened when including this variable in the model.

2.4 Conclusion and discussion

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

2.4.1 Spectral evidence for word stress

Significant effects involving stress were found for relative F0 (only for ultimate stress), rise/fall ratio, movement onset, movement offset and formant displacement. Support for the hypothesis that Papuan Malay has stress was found in that syllables labelled as stressed had higher F0 than the unstressed syllable in that word when stress was penultimate. As for F0 movement, the significant interaction between stress and position indicated that mainly in the case of ultimate stress, larger F0 movements were found on syllables labelled as stressed than on syllables labelled as unstressed. Furthermore, syllables labelled as stressed were more likely to show a rising F0 movement. While syllables labelled as stressed showed generally later onsets and offsets than syllables labelled as unstressed, this was mainly true for penultimate stress. For ultimate stress, onsets occurred earlier and offsets occurred later in syllables labelled as stressed than in syllables labelled as unstressed. Furthermore, rises on syllables labelled as stressed generally had a later offset compared to syllables labelled as unstressed, regardless of stress position. Falls, however, mainly showed later offsets in syllables labelled as stressed when stress was ultimate and a minimal difference when stress was penultimate. These differences are reflected in the significant three-way interaction between stress, position and direction for movement offsets.

Displacement relative to the centre of the vowel space provided evidence for stress for /a/ and /u/ (Table 2.8, Figure 2.4). The displacement of /i/ and /ɔ/ was too small to be significant. Remarkably, /ɛ/ was realised closer to the centre in syllables labelled as stressed. This result is counter to the expectation that stressed vowels are realised further away from the centre to maximise (perceptual) vowel contrasts (i.e. Liljencrants & Lindblom 1972; Flemming 2004). It can furthermore be observed that the phonetic realisation of vowels in Papuan Malay do not always match their phonemic representation as expected in the acoustic space. That is, /u/ is realised relatively central, similar to [ʉ], and /ɛ/ is realised relatively high, similar to [e] (Figure 2.4). While this could be the result of allophonic variation, none of these realisations match with the allophones reported in Kluge (2017). This is particularly noteworthy in the case of /ɛ/, of which its displacement is counter to the stress hypothesis. One could therefore question to what extent a reanalysis of the vowel inventory based on acoustic measures would still be compatible with the one in Kluge (2017). While this investigation is beyond the scope of the current study, a reanalysis of the Ambonese Malay vowel inventory shed a different light on earlier stress claims (Maskikit-Essed & Gussenhoven 2016). In sum, the formant displacement measure provided some support for word stress when individual differences between vowels are taken into account. Further discussion of the formant displacement of /ɛ/ is provided in Section 2.4.5.2.

2.4.2 Temporal evidence for word stress

Consistent support for the stress hypothesis was obtained for duration per phoneme and duration deviation. Both measures were significantly affected by word stress in that longer durations were found for syllables labelled as stressed than for syllables labelled as unstressed. Therefore, both duration per phoneme and duration deviation can be taken as correlates indicating acoustic evidence in favour of word stress in Papuan Malay. Both temporal measures also showed significant interactions between stress and position, indicating that there was more lengthening of syllables labelled as stressed when stress was ultimate than when stress was penultimate. It has to be noted that the measures did not account for the tendency of word final syllables to be longer (word final lengthening). These effects are likely, given the mean raw durations (Table 2.7). Therefore, the stress-position interactions found for both temporal measures could have been affected by word-lengthening.

2.4.3 Amplitudinal evidence for word stress

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

2.4.4 Relative contribution of the acoustic measures to word stress

From the acoustic measures that were tested for their contribution to word stress, the measures movement onset/offset and duration per phoneme were most indicative correlates of word stress in Papuan Malay. Other relatively strong correlates of word stress were formant displacement and H1-A2. Duration deviation, H1*-A2* and H1-A3 correlated with word stress in a weaker manner. F0 movement and H1*-A3* did not predict word stress at all. Overall, the results of the GLMMs and model fit comparisons provide additional insight into the relative contribution of the possible acoustic correlates of word stress. The results of the model comparisons were furthermore in line with those of the LMMs on each of the acoustic measures.

2.4.5 Discussion

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

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

2.4.5.1 Word stress position

A couple of factors need to be taken in account when interpreting the stress position difference just observed. First, due to the asymmetry in the distribution of words with presumed penultimate (95%) or ultimate stress (5%), statistical tests could falsely indicate that there was no difference in stress realisation in the respective positions (type II error). If this error indeed applies, the current data does not allow to either confirm nor to reject the mobility of word stress to the ultimate syllable. As a consequence, the acoustic analysis would then only confirm penultimate stress. It has to be noted that the penultimate/ultimate asymmetry in the current dataset reflects the presumably natural distribution of word stress in Papuan Malay (Kluge 2017). For this reason, increasing the sample size of words with presumed ultimate stress in order to reduce the type II error rate, would yield unrealistic word stress distributions. In addition, inspection of the most indicative correlates reveals that mean acoustic differences between syllables labelled as stressed and syllables labelled as unstressed were often larger for penultimate stress as opposed to ultimate stress (Table 2.7). These data suggest that the existence of ultimate stress is realistic despite the asymmetric distribution.

Second, the asymmetrical distribution of word stress in Papuan Malay hints at a marked (i.e. exceptional) status for ultimate stress. When ultimate stress is indeed the marked stress position – as supported by most acoustic measures – the question remains to what extent stress is acoustically marked in penultimate position. Although there is clear acoustic evidence for stress across the respective positions, it could be the case the penultimate stress is the default pattern which does not require large acoustic differences. In this view, the notion of "stress" in terms of markedness rather applies to ultimate positions. Such an asymmetry between the respective stress positions would be explained by the acoustic data investigated here for Papuan Malay. As a consequence, the notion of stress marking is more constrained to position in the word than the attribution of a specific amount of acoustic prominence to a single syllable in a word would imply. The literature has observed that in languages where there is little variation in the position of the stressed syllable (i.e. "fixed stress" languages), the acoustic difference between stressed and unstressed syllables is small (Cutler 2005; Dogil & Williams 1999). This can be explained when considering that speakers and listeners of this type of languages expect the most frequent pattern by default and that only deviations from this pattern need to be acoustically salient enough to warrant successful communication. Indeed, EEG studies on languages where the stressed syllable has limited mobility, have shown that listeners are particularly sensitive for non-default stress patterns (Domahs et al. 2012, Domahs et al. 2013). Given the current results, the acoustic realisation of word stress in Papuan Malay would fit the asymmetry observed for other languages. While the communicative importance of a deviant stress pattern could explain much of the acoustic asymmetry, the position of the stressed syllable in the prosodic structure of Papuan Malay needs to be considered as well. That is, ultimate syllables are particularly suitable to stand out as acoustically more prominent due to their final position. Word final lengthening, which appeared to be present in the current data (Section 2.4.2), could thus have had an additional, strengthening effect to ultimate word stress. In other words, longer syllables give more room for acoustic cues to be realised more saliently compared to shorter ones. It is likely that the ultimate stress patterns found in the current study exhibit such a combined effect of word stress and final lengthening, as hinted at by the effects of position reported in Table 2.6.

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

2.4.5.2 Formant displacement of /ɛ/

As for F1 and F2, consistent formant displacement was observed. For most vowels this meant displacement from the centre of the acoustic space. An exception was found for /ɛ/, which was centralised in syllables labelled as stressed (Figure 2.4). This is counter to the cross-linguistically common observation that vowels are centralised, or likely to be realised as schwa in unstressed positions. It has to be noted, however, that schwa can be stressed in languages where this sound is part of the phoneme inventory (e.g. Romanian; Chițoran 2002). In Papuan Malay, /ɛ/ appears more often in unstressed positions than the other Papuan Malay vowels (Table 2.2). It is therefore not immediately clear how the formant displacement of /ɛ/ should be interpreted.

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

In the second explanation, the formant displacement of /ɛ/ is interpreted in accordance with the claim that Papuan Malay has word stress (Kluge 2017). Although it is unclear in this view how exactly the realisation of stress could lead to vowel centralisation, the position of schwa in Papuan Malay might provide further insight. That is, Papuan Malay /ɛ/ often corresponds to schwa in Indonesian (Donohue 2003). Like for Romanian, it has been claimed that schwa can be stressed in Jakarta Indonesian (Laksman 1994), see Section 1.2. Although there is debate about the existence of word stress in Indonesian (e.g. Goedemans & van Zanten 2014), regular word level prosody has been observed for some of the Indonesian languages (e.g. Toba Batak; Goedemans & van Zanten 2007). Crucially, the large influence of (Jakarta) Indonesian on Papuan Malay is undisputed (Kluge 2017). Therefore, the centralised realisation of stressed /ɛ/ can be explained by assuming that /ɛ/ is underlyingly treated as schwa, however only when stress is applied (i.e. like in Jakarta Indonesian). In this way, stress on /ɛ/ is defined relative to schwa in the centre of the vowel space, not relative to unstressed /ɛ/. As a consequence, stressed /ɛ/ is realised as a "displaced schwa", and located between unstressed /ɛ/ and the position of schwa. The Papuan Malay vowel space allows for this midway location and ensures that stressed /ɛ/ remains acoustically distinct from both unstressed /ɛ/ and schwa. Such an unusual application of word stress would be a phonologically costly process, which explains why stressed /ɛ/ is relatively rare. That is, /ɛ/ occurs more often in syllables labelled as unstressed (>80%) than in syllables labelled as stressed. This distribution is different from the other Papuan Malay vowels, which occur more equally frequently among syllables labelled as stressed and syllables labelled as unstressed (Table 2.2).

Additional support to explain the formant displacement of /ɛ/ within the claim that word stress is present, is the potentially limited communicative function of word stress in Papuan Malay. This is particularly clear from the lack of minimal stress pairs, suggesting that the contribution of word prosody to distinguish word meanings is minimal. Due to this limited function, the hypothetical influence of the Indonesian stressed schwa just outlined is most likely harmless for word recognition in Papuan Malay. The adoption of stressed schwa could rather reflect the influence of Indonesian in formal settings, such as interactions with Western researchers described in Kluge (2017: 18). The recording procedure described in Section 2.1 could have caused such a level of formality for the participants in the current study. In addition to possible influences of register, it has been shown that language contact situations can be reflected in the variation of a single phoneme (Kaland, Galatà, et al. 2019). Similarly for North Moluccan and specifically Manado Malay, schwa was originally lacking, but does occur "in words which might represent code-switching to Indonesian" (Paauw 2009: 95).

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

2.4.5.3 Outlook

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

The current results deserve further caution on several aspects. First, the stress hypothesis tested in the current study was taken over from Kluge (2017). Although consistent acoustic evidence was found, a more unbiased hypothesis would have been based on the basis of purely acoustic criteria. These criteria would then have to correspond with a specific expectation of the acoustic properties of a stressed syllable. While the outcome of such an investigation would constitute stronger empirical support on word stress, the approach would face a number of difficulties. Foremost, the acoustic realisation of word stress differs cross-linguistically (Section 2.1.3), which makes it challenging to find the acoustic criteria needed to call a syllable "stressed". This is particularly complicated for languages like Papuan Malay, for which little to no acoustic research is available. The investigation outlined in this chapter furthermore suggests that acoustic realisations differ depending on where in the word stress is realised. A second word of caution concerns Papuan Malay phrase prosody, of which to date little is known. While recent work suggests that this language does not make use of pitch accents (Kaland et al. 2018; Riesberg et al. 2018), a direct assessment of the functions of F0 in Papuan Malay phrases is lacking. While the influence of phrase level phenomena was kept to a minimum in the current study, it cannot be fully excluded that some of the words in the data subset were subject to acoustic processes resulting from phrase prosody. Whether or not these processes originated from accent placement cannot be concluded at this stage. The interaction between word level and phrase level prosody is left for future research. Third, it remains to be investigated to what extent the word stress correlates found here are perceptually relevant. Minimal stress pairs are not reported for Papuan Malay (Kluge 2017) and the acoustic differences found in the current study could therefore turn out to be irrelevant for listeners. An investigation of the extent to which the acoustic cues facilitate word recognition is planned in a follow-up study.

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

3.1 Introduction

The current chapter is based on the acoustic analysis of word stress cues in Papuan Malay (Chapter 2) and on a preliminary investigation of its phonological status (Kaland, Himmelmann, et al. 2019). Several acoustic correlates were found to be strong indicators of word stress in this language, despite the previous claims that word stress is absent in (Eastern) Indonesian languages, including Papuan Malay (e.g., Goedemans & van Zanten 2014; Riesberg et al. 2018, Himmelmann 2018).

Crucially, the cross-linguistic 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: 196), "no primary word stress at all" (van der Hulst 2012: 1515), or is a language for which "the need for *w*-final accents is not hard to establish" (Gussenhoven 2004: 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 (Section 3.1.1) and the literature on word stress perception cross-linguistically (Section 3.1.2). Section 3.1.3 presents the research questions and hypotheses.

3.1.1 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 summarises the current state of research on word stress in Papuan Malay and related languages.

Papuan Malay is reported as having regular word stress on the penultimate syllable (Donohue & Sawaki 2007; Kluge 2017), except when that syllable contains /ɛ/ (which is claimed to reduce to schwa /ə/, but see Chapter 2 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., [ˈde.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, Himmelmann, et al. 2019). Both /ɛ/ and /ɔ/ are the only mid vowels 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 cross-linguistic 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; Chapter 2). For disyllabic Papuan Malay roots (loanwords excluded), consistent acoustic evidence was found to show that stressed syllables are more prominent than unstressed syllables (Chapter 2). 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 are longer than unstressed ones), formant displacement (stressed syllables are 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 realised 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 findings on 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 on the boundaries and little agreement on 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 & 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 & 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 & 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 & 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 sheds light on the strategies that listeners use to process speech, which are language-specific by nature (Cutler 2012). The next Section 3.1.2 discusses this issue in further detail.

3.1.2 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 cross-linguistically.

3.1.2.1 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 & Edwards (1994). In general, F0 has been claimed to be a better correlate of phrase level prosody than of word prosody (Sluijter & 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 & Munro 1978; Isenberg & 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 & 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 & 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 & 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 & 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).

3.1.2.2 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 cross-linguistically 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: 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 & 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. In languages with more variable stress patterns such as English and Dutch, there also appears to be a tendency for stress to be located on the word-initial syllable (English: Cutler & Carter 1987; Dutch: van Heuven & Hagman 1988). Listeners of these languages were also found to default to a word-initial stress pattern when reporting induced misperceptions (Cutler & Butterfield 1992; Vroomen et al. 1996). Given that two-syllable words are relatively common cross-linguistically (e.g., Vihman & 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.

3.1.2.3 Stress predictability and perception

Cross-linguistic 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 between regular and 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 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 & Williams 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 & 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 & 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 Section 3.1.3.

3.1.3 Research questions and hypotheses

The current study hypothesises 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 factor, it is hypothesised 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 /ɛ/ (or schwa; see Chapter 2). Given that much of the stress distribution actually depends on the presence of /ɛ/ 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 for word stress in this language.

Concerning the second factor, Papuan Malay could be classified as having rather fixed stress (Section 3.1.1), 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 (Section 3.1.2). There are, however, valid reasons to investigate the extent to which relatively 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: 282–283). Furthermore, Papuan Malay does not appear to be a fixed-stress language for which acoustic cues are only weakly realised (cf. Chapter 2 and Dogil & Williams 1999). That is, the acoustic signal provides the Papuan Malay listener with multiple consistent suprasegmental cues (Chapter 2). 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 & 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 & 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 (Experiments 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 minimise 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 3.2, 3.3, and 3.4 describe each of the experiments and Section 3.5 provides an overall conclusion and general discussion.

3.2 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 which out of two written two-syllable words 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).

3.2.1 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: 62–63 for recording details). This syllable was chosen such that subsequent acoustic manipulations would not be interrupted. That is, in voiced segments (the nasal and vowel in "ma") F0 is continuously present, unlike, for example, in voiceless stops or fricatives. Before manipulation of the word stress correlates, the syllable "ma" was filtered using a sequence of second order filters in Praat [Boersma & 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 Figure 3.1). 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 & 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 3.1.

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 3.1). This factor corresponded to the overall relative difference found between first (121 ms) and second syllables (128 ms) in the Papuan Malay words found in Chapter 2. The

Table 3.1: 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 Chapter 2.

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		-5.85 dB

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 & 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 3.1). In this way, the difference between "ta" in kata and "i" in itu was 3 semitones (ST), matching the F0 manipulations described in Section 3.2.2.

3.2.2 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 (Chapter 2), 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 Chapter 2 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 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.

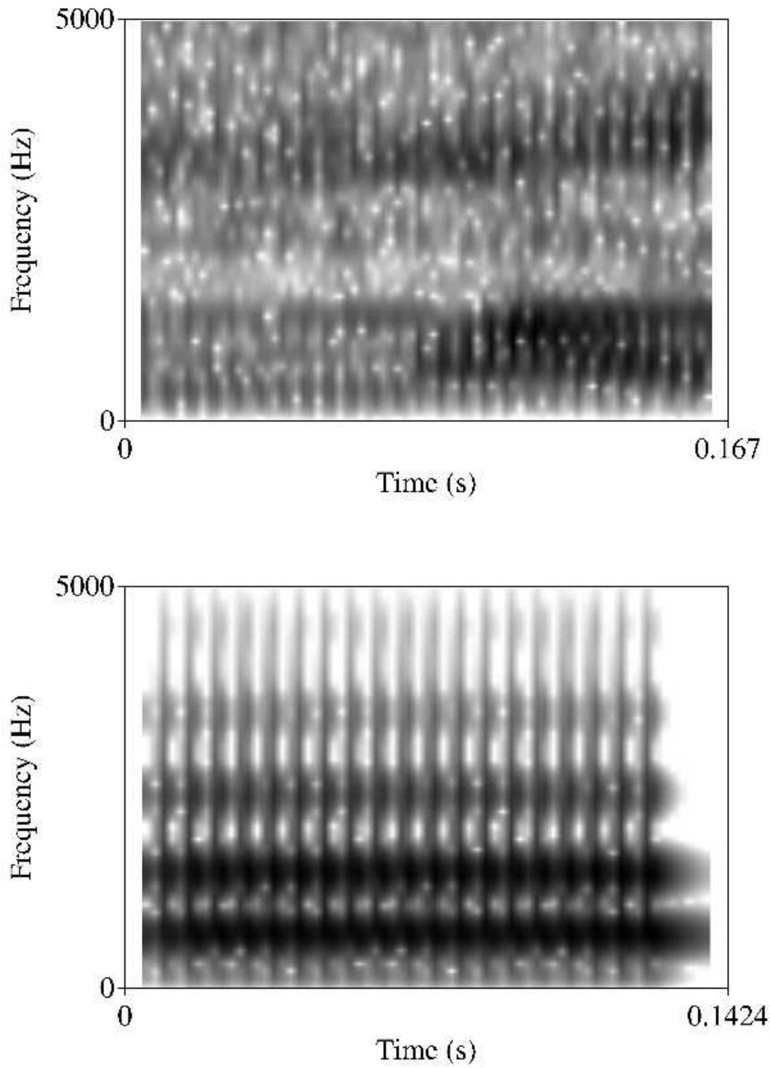

Figure 3.1: Spectrograms of the template syllable “ma” before (top) and after (bottom) formant conversion.

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 of the filtered syllable.

Although spectral tilt appeared a strong correlate of word stress in the production data (Chapter 2), 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 & 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 3.1 and Chapter 2).

3.2.3 Design

All combinations of acoustic manipulations were generated following a 2 x 2 x 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 Section 3.2.2). 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 (Chapter 2). In this way, the resulting set consisted of 22 manipulated syllable sequences (Table 3.2).

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 (Figure 3.2). 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

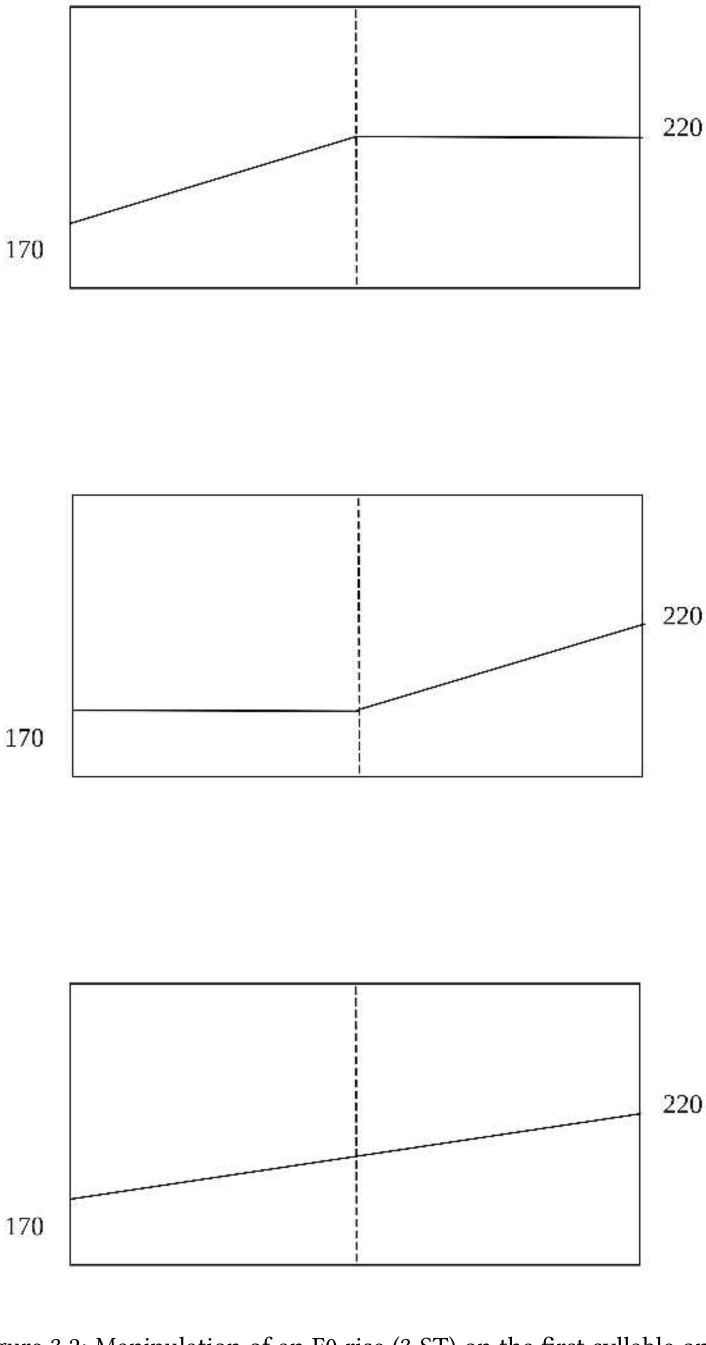

Figure 3.2: Manipulation of an F0 rise (3 ST) on the first syllable only (top), the second syllable only (mid) or both (bottom). Numbers (Hz) and solid lines represent F0; dashed lines represent the syllable boundary.

Table 3.2: 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

the F0 levels in the parts of 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 Table 3.2), 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.

3.2.4 Participants

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

3.2.5 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 & De Boer 1991) and 22 wave files (Table 3.2). Materials are available on https://osf.io/zsvd2/. 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 Table 3.3 (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 (Figure 3.3). They were instructed to pay close

Table 3.3: 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).

	Penultimate stress		Ultimate stress		Additional difference	
Pair no.	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	dengang	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	to be long (duration)	lema	be weak	s	v
19	lime	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 reach each others		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

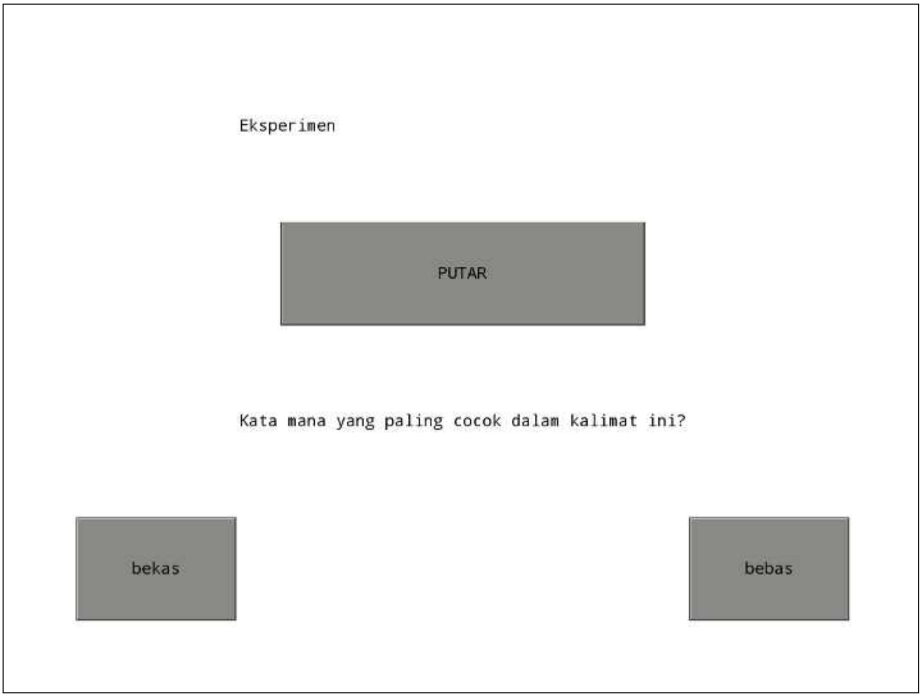

Figure 3.3: Screen capture of Experiment I displaying the play button (*putar*), the question "Which word fits best in this phrase?" and two response options representing a near-minimal stress pair.

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 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 familiarise 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 short break. The experiment lasted approximately 20 minutes. 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.

3.2.6 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, generalised 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 included 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 Section 3.2.3). Furthermore, the design of the cue manipulations (Table 3.2) 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 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.

3.2.7 Results

Overall, the mean of the response values was above chance level (M = 0.53, V = 1550700, $p < 0.01$). The Wilcoxon tests per syllable indicated chance level responses for the first syllable (M = 0.50, V = 362690, n.s.) and above chance level responses for the second syllable (M = 0.56, V = 412950, $p < 0.001$). The Wilcoxon tests per acoustic cue (Table 3.4) 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 3.5), 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 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 (b = 0.28, SE = 0.08, z = 3.41, $p < 0.001$).

Table 3.4: Results of the Wilcoxon signed rank tests (μ = 0.50) and the generalised linear mixed model analyses for the acoustic cues on each syllable in Experiment I.

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

Table 3.5: Results of the generalised linear mixed model analyses for each syllable and acoustic cue. Effects of number of cues (N cues) only reported when yielding significance.

Syllable	Cue	*b*	*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.

3.2.8 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 (μ = 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 (Chapter 2). 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 (Chapter 2).

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 a stress cue (on either the first or the second syllable), see Section 3.2.2. 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 Section 3.3, a second experiment is reported that accounted for the two issues just discussed.

3.3 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 (Section 3.2.1 and 3.2.2). The setup and procedure differed in the ways discussed in Section 3.3.2.

3.3.1 Participants

A total of 21 participants carried out Experiment II. All were students at the University of Papua; 5 male and 16 female participants (μ 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.

3.3.2 Setup and procedure

In Experiment II, one stimulus consisted of one written target word and two wave files. The audio files corresponded to two different manipulated syllable sequences in isolation taken from the same set of stimuli used in Experiment I (Table 3.2 and https://osf.io/zsvd2/). 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 Table 3.3. 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 x 11, see Table 3.2). 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 (Table 3.3) 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 (Figure 3.4). 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 let 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 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 (Section 3.2.6) and are not repeated in this section.

3.3.3 Statistical analysis

Statistical analyses were identical to the ones described for Experiment I (Section 3.2.6).

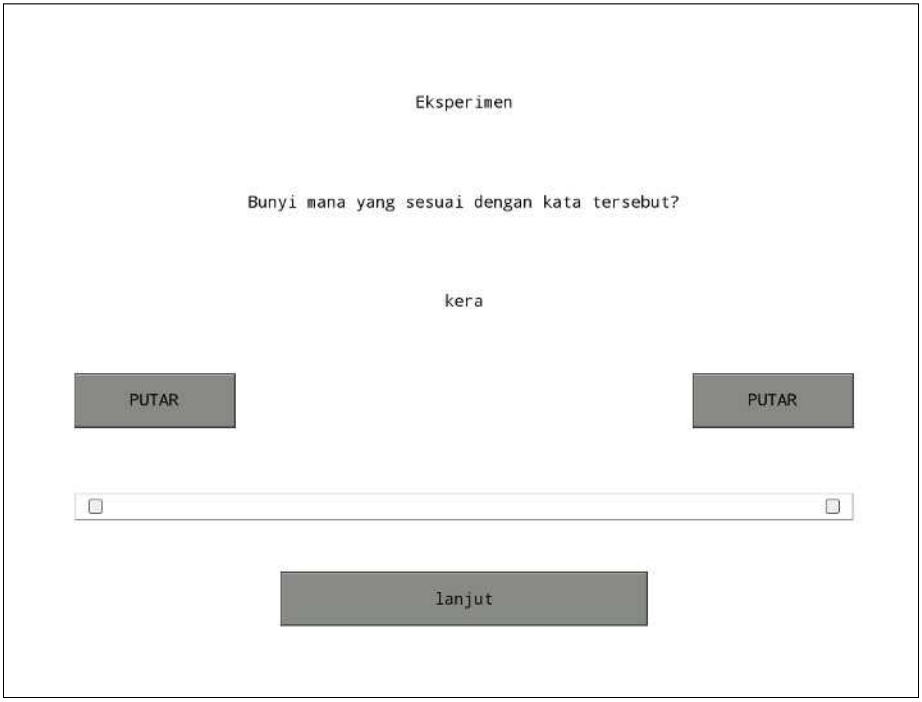

Figure 3.4: Screen capture of Experiment II displaying the question "Which sound corresponds to the word?," the target word, two play buttons (*putar*) to play the wave files in the pair with accompanying selection boxes, and a button to continue (*lanjut*).

3.3.4 Results

The results show that participants overall scored above chance level (μ = 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 (μ = 0.46, V = 346 240, p < 0.01), whereas for ultimate stress, the scores were significantly above chance level (μ = 0.59, V = 512 070, p < 0.001), see also Table 3.6.

Furthermore, the GLMM results (Table 3.7) 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 other cues. Higher correctness scores for ultimate stress were found when comparing the word-stress positions in the GLMM (b = 0.58, SE = 0.09, z = 6.33, p < 0.001).

Table 3.6: Results of the Wilcoxon signed rank tests and the GLMM analysis on the acoustic cues in each word stress position in Experiment II.

			Wilcoxon (μ = 0.50)	
Word stress	Cue	μ 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 3.7: Results of the generalised linear mixed model analyses for each syllable and acoustic cue. Effects of number of cues (N cues) only reported when yielding significance.

Syllable	Cue	b	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.

3.3.5 Discussion

The results of Experiment II show that listeners benefit from acoustic cues for 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.

3.3.6 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 (Section 3.2.8), 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 recognised. 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 resynthesised 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 & Cutler 1988) holds for Papuan Malay listeners is investigated in Experiment III, as further discussed Section 3.4.

3.4 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, see https://osf.io/zsvd2/). 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 Section 3.2.1).

(1) a. 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)'

b. 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 ______.'

Given the limited availability of recorded words with ultimate stress, a subset of the recordings was selected for use in the current experiment (with the most frequent syllable structure and stress pattern: ˈCV.CV, see Kluge 2017). 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 & 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 3.8 and generally confirm that stressed syllables stand out acoustically compared to unstressed syllables, in particular with respect to duration (Chapter 2).

3.4.1 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

Table 3.8: Mean (*SD*) differences scores (D) for F0 movement (ST), duration (ms) and intensity (dB) in the target words (N = 80) in either phrase position (medial/final).

Measure	Medial	Final
Δ F0 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)

or 1b). 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 3.9 for examples). It should be noted that apart from stress cues, the segmental makeup of the cue syllable was strictly speaking sufficient to recognise the word (cf. *lida* and *lada*, Table 3.9). 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 3.9 could be

Table 3.9: 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)

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 Section 3.4.6.

3.4.2 Setup and procedure

The task was designed and run in OpenSesame using the legacy backend (Mathôt et al. 2012). For each stimulus, the script generated a screen (Figure 3.5). 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 familiarise 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 & 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 familiarise 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 short break after completing half of the task. The experiment lasted approximately 20 min.

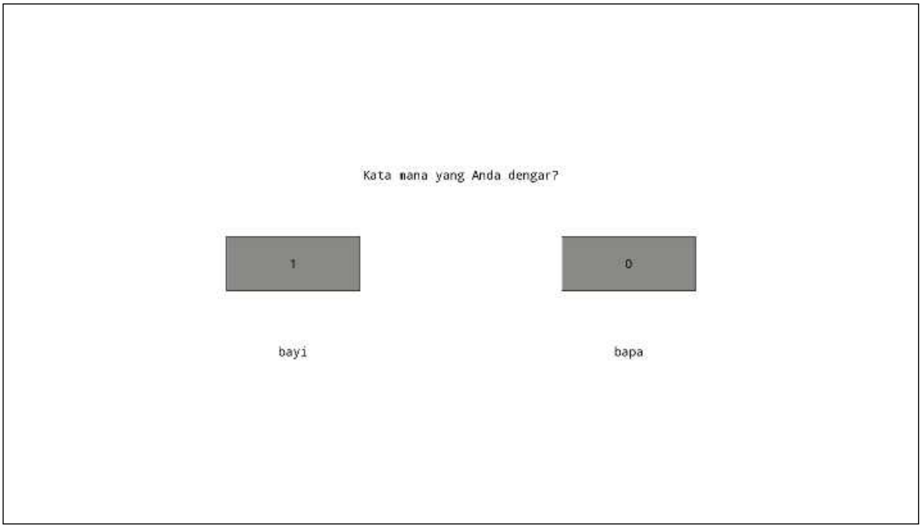

Figure 3.5: Screenshot showing an example stimulus in Experiment III.

3.4.3 Participants

All 22 participants were students at the University of Papua, Manokwari; 5 male and 17 female participants (age μ = 22.05, age 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.

3.4.4 Data processing and statistical analyses

Before statistical analysis, reaction times were removed from the data if participants had not correctly recognised 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 extreme reaction times (N = 184 in total) 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 & Milin 2010). After removal, 1576 reaction times remained for analysis.

GLMMs are proposed to be particularly suitable for reaction time measures (Lo & Andrews 2015). These measures appear to have an inverse Gaussian distribution (Baayen & Milin 2010; Lo & 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 (LMMs) 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 & Gupta 2005). The predictor target word duration indeed showed an effect for the target onset measures ($b = 0.45$, $SE = 0.16$, $t = 2.72$, $p < 0.01$) and not for the target offset measures ($b = -0.25$, $SE = 0.19$, $t = -1.31$, 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 ($N = 80$).

3.4.5 Results

Results indicate that participants were faster when the cue syllable was the first (stressed syllable) in the word (Figure 3.6, Tables 3.10 and 3.11). 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 3.12). 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.

Table 3.10: 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)

3.4.6 Discussion

Experiment III showed that Papuan Malay listeners were faster when the stressed syllable was the syllable needed to recognise 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 & Clifton 1984). This effect disappeared when stimulus length was controlled for. Stimulus length was also controlled for in the current study, making

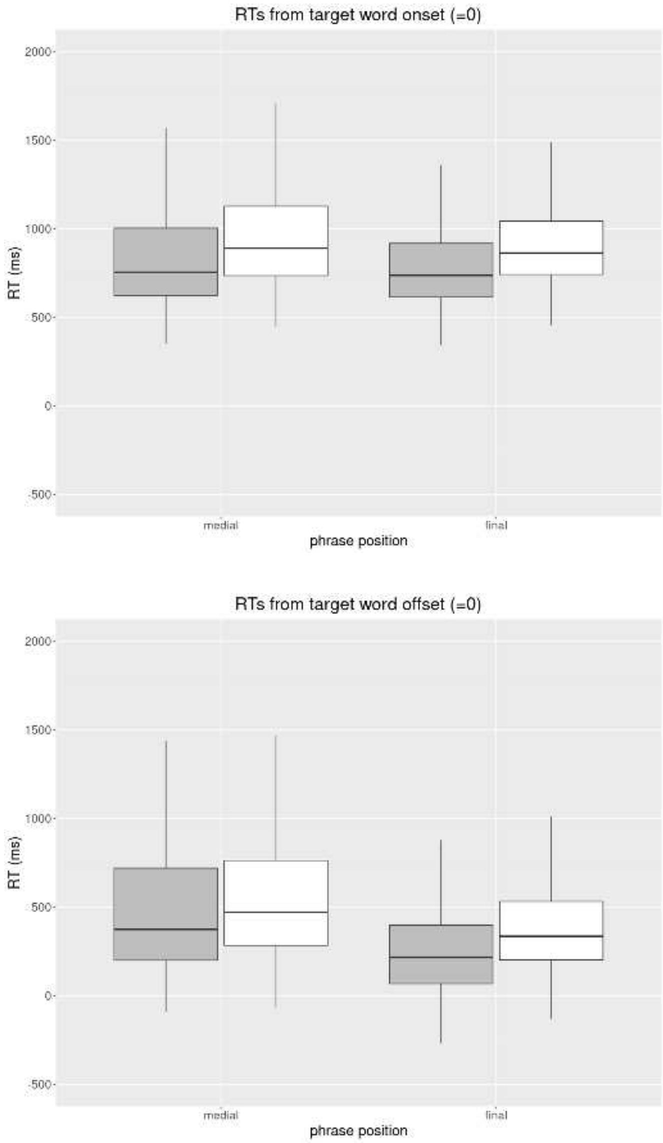

Figure 3.6: Reaction time boxplots (bar indicates median) measured from target word onset (top) and offset (bottom) in phrase-medial and phrase-final position when the cue syllable was the first (grey) or second (white).

Table 3.11: 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	*b*	*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
	Δ 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
	Δ 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.

Table 3.12: 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.)

it unlikely that the facilitative effect of the word-initial syllable was only an artefact of the stimulus material (but see Section 3.5 for more discussion on this issue). It seems more reasonable, therefore, to interpret the 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 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 (Section 3.4.4). 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 3.12). 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 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 hypothesised, 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.

3.5 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 recognise 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 (Chapter 2). 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 & Williams 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 recognise 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 (Section 3.4.6). Such an effect could be explained by the recently reported longer reaction times for later segmental uniqueness points (UPs; 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 Section 3.1.3, 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 & 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 seem to hold for Papuan Malay in the strict sense. 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 & 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 likely that which is reflected in both the production of ultimate stress (Chapter 2) 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 realised 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 & 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 & 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.

4 Lexical analyses of the function and phonology of Papuan Malay word stress

4.1 Introduction

A number of recent studies found support for the existence of word stress in Papuan Malay, a language spoken in the easternmost provinces of Indonesia (Chapter 2, Chapter 3, Kluge 2017). The findings are a challenge to integrate with the different outcomes in studies on closely related languages such as Ambonese Malay (Maskikit-Essed & Gussenhoven 2016). In addition, the word stress controversy in Trade Malay languages fits into an existing debate on word stress in Indonesian languages (e.g., van Zanten et al. 2010).

The current support for word stress in Papuan Malay comes from a grammar based on auditory impressions (Kluge 2017), acoustic analyses (Chapter 2) and perception experiments (Chapter 3). These studies agree that the penultimate syllable in a word is stressed by default, except when that syllable contains /ɛ/, in which case the ultimate syllable is stressed. This stress distribution has been reported (and disputed) for numerous Austronesian languages (van Zanten et al. 2010). However, detailed phonetic or phonological analyses of stress in these languages are few. Unlike other Malay variants that have at least one mid centralised vowel (schwa), Papuan Malay has no schwa; its five vowels are /i, ɛ, a, ɔ, u/. Papuan Malay is in this respect similar to Toba Batak, both in the lack of schwa and in stress placement (Goedemans & van Zanten 2007; van Heuven & van Zanten 1997; van Zanten & van Heuven 1984). Recent analyses have advanced the study of stress in Papuan Malay. Duration, vowel quality and spectral tilt were found to be the strongest acoustic correlates of word stress (Chapter 2). Regarding perception, Papuan Malay listeners appeared to assume the penultimate stress pattern as a default and were sensitive mainly to deviant (ultimate) stress patterns in lexical decision tasks (Chapter 3). These perception effects could be ascribed to acoustic differences in duration, (spectrally weighed) intensity and fundamental frequency (F0). However, none of these cues alone was sufficient for listeners to discriminate between penultimate and ultimate stress. Although vowel quality was not tested perceptually, it was found to be a strong acoustic correlate (Chapter 2).

The work just discussed focused on the acoustic nature of stress in production and perception. Less is known about whether and how Papuan Malay stress patterns play a functional role in the lexicon. Although word stress in Papuan Malay is unlikely to be lexically contrastive (e.g., no minimal stress pairs were reported in Kluge 2017), the distinction between penultimate and ultimate stress could still facilitate word recognition (e.g., Cutler 2005; van Zanten & van Heuven 2004). The applicability of such a function could make the case of word stress in Papuan Malay less controversial. In addition, acoustic patterns of penultimate word stress in Austronesian languages have often been explained as reflexes of phrase prosody (van Heuven et al. 2008 on Betawi Malay; van Zanten et al. 2010). To substantiate the claim that the acoustic results obtained for Papuan Malay are indeed bound to the word level, additional research is needed on the morphological and phonological factors underlying these stress patterns. This advances the existing sketch of Papuan Malay phonology (Kluge 2017). Therefore, the current study extends the existing work in two ways. First, the role of word stress in word disambiguation is investigated, comparing Papuan Malay with other languages for which similar analyses were carried out. Second, the phonological nature of Papuan Malay word stress is further explored comparing multiple (morpho-)phonological factors and their effects on stress placement. To this end, a lexical analysis of word embeddings and a random forest analysis were carried out respectively, using a list of 1127 words obtained from extensive fieldwork (Kluge 2017; Kluge et al. 2014).

4.1.1 Stress controversy in Indonesian languages

The existence of word stress patterns in Indonesian languages has been the topic of several studies, sometimes leading to contradictory findings (see Goedemans & van Zanten 2007 for an overview). Papuan Malay, the focus of this chapter, is spoken by more than one million people and serves as the common language for many different communities in the Indonesian provinces of Papua and West Papua (Kluge 2017). For most speakers, Papuan Malay is therefore their second native language, alongside their inherited language. Before illustrating the controversy in a discussion of stress in Trade Malay languages, to which Papuan Malay belongs, some general remarks should be made that are also relevant for other Indonesian languages. Part of the controversy resulted from the lack of distinguishing word stress from phrase prosody in early work (see Goedemans

& van Zanten 2007 for a discussion). In addition, attempts to generalise over language families in Indonesia have been prone to overlook language variation. Regional variation is a crucial variable in explaining the distribution of stress patterns (Goedemans & van Zanten 2007; Himmelmann & Kaufman 2020). Although generalizations over language families could provide important typological insights, they are at best based on a limited number of studies, as most languages of Indonesia are still under-researched.

Contradictory claims on word stress can be observed for at least two Trade Malay languages: Ambonese Malay and Papuan Malay. As for Ambonese Malay, word pairs with penultimate/ultimate stress as their minimal difference were reported (van Minde 1997). However, an acoustic analysis of correlates of word stress and pitch accents in this language provided a different explanation for these minimal pairs. Vowel quality differences were no longer seen as the main correlate of stress in the minimal pairs, but were rather interpreted as acoustic evidence for two different vowels. Thus, Ambonese Malay was re-analyzed as a language without stress, having six instead of five vowels (Maskikit-Essed & Gussenhoven 2016).

For Papuan Malay, a rating task on phrase-level prominences and boundaries indicated that native listeners mainly agreed on where prosodic phrase-final boundaries occurred (Riesberg et al. 2018). It was concluded that prosodic prominence is unlikely to be a relevant concept in this language and that Papuan Malay therefore does not have word stress or pitch accents, following the most recent claim for Ambonese Malay. However, the rating task concerned phrase prosody only and did not provide direct evidence about word stress patterns. An analysis of word stress correlates (Chapter 2), crucially excluding phrase-final words, showed that the distinction between penultimate and ultimate stress is signaled acoustically, confirming earlier reports (Kluge 2017). In perception studies, Papuan Malay listeners were found to be sensitive to acoustically prominent syllables at the word and phrase level. It should be noted that the perceived prominence resulted from either sequences of manipulated Papuan Malay syllables (Chapter 3) or German phrases with salient pitch accents (Riesberg et al. 2020).

4.1.2 Types of evidence for lexical stress

Given these contradictory findings, it is important to consider the type of evidence used in the studies. For Ambonese Malay, the argument in favor of word stress appeared to be mainly stem from the author's impressions based on a small number of minimal pairs (van Minde 1997). The acoustic analysis (Maskikit-Essed

& Gussenhoven 2016) was carried out on a small number of speakers (*N* = 4) and stimuli (*N* = 9), the latter being selected to test minimal pair contrasts only. In this respect, it should be noted that the literature has cast doubt on diagnosing word stress based only on the existence of minimal pairs. Psycholinguistic studies have shown that even in Germanic languages that are uncontroversially analyzed as having stress, minimal stress pairs (except for ones involving affixation) are highly infrequent in their lexicons (e.g., Cutler 2012). Thus, the lexically contrastive function does not necessarily provide a useful stress diagnostic. More evident and insightful functions of stress were revealed in studies on listener expectations and lexicon structure. For example, it has been shown that rhythmical expectations based on the (regular) placement of the stressed syllable facilitate word recognition processes (see Cutler 2005 for an overview). Furthermore, studies have shown that these processes are not identical across languages. Much depends on whether stress cues that are available in the speech signal are indeed used by listeners. It has been shown that in early stages of word recognition, stress cues do not have a word discriminatory function in English (Cutler 1986), although they do in Dutch (van Heuven 1988). How sensitive listeners are to stress cues depends to some extent on the distribution of regular versus irregular stress patterns in the lexicon (Domahs et al. 2012; Peperkamp et al. 2010). In particular, it has been shown that stress may help listeners to disambiguate words (Cutler & Pasveer 2006; Cutler et al. 2004).

4.1.3 Word disambiguation

The disambiguating role of stress can be illustrated when considering the English embedded word *bee* in either the carrier word *beanie* or the carrier word *belay*. In this example, suprasegmental stress cues could disambiguate bee and belay, as the first consonant-vowel sequence is enough to recognise it as a stressed (*bee*) or unstressed (*belay*) syllable. However, stress is not a disambiguating cue between *bee* and *beanie* as the matching first syllable is stressed in both words. In the latter case, word disambiguation is rather based on segmental differences between the embedded word and its carrier word. Studies have compared the total number of embedded words and the number of embedded words when stress was taken into account as a disambiguating cue for a small number of languages (Cutler et al. 2004). These counts revealed differences between English on the one hand and Dutch, German and Spanish on the other (Table 4.1). That is, the relative decrease in mean embeddings per word due to mismatching stress was the largest in Spanish and the smallest in English. Dutch and German occupied middle positions. For all of these languages the mean value of stress-matched embeddings was below the limit of one per carrier word, showing that stress information successfully reduces the competition between the candidates. The limit of one embedding per carrier word is crucial, as the disambiguation problem that listeners face can be successfully reduced by stress cues only if there are sufficient embeddings (Cutler & Pasveer 2006). For English, the statistics showed that even when stress information was ignored, the mean number of embeddings per carrier was below one. This would indicate that in English there is little need for word disambiguation, and therefore limited room for stress cues to play a role, even when they are present in the signal (Cooper et al. 2002). It has been argued that the differences among the lexical statistics of each language could indeed be explained by the type of stress cues listeners exploit in word recognition. In Spanish, listeners use mainly suprasegmental cues (see also Peperkamp et al. 2010). In Dutch and German, both suprasegmental and segmental cues are used, whereas in English, segmental cues are most important (Cooper et al. 2002; Yu et al. 2020). In English, vowel reduction is the most important segmental cue to stress and serves to distinguish, for example, word class in disyllabic words, e.g., *subject* is a verb when the vowel in the first syllable is reduced (ultimate stress) and a noun when the vowel in the first syllable is full (penultimate stress). The limited role of suprasegmental cues was also hypothesised to be the consequence of phoneme inventory size (Cutler et al. 2004), such that languages with large inventories (e.g., 44 in English) have more options to disambiguate words by segmental means than languages with small inventories (e.g., 25 in Spanish). This would explain the limited degree of disambiguation by means of suprasegmental stress cues found for English (Table 4.1). Although the hypothesis suggested by Cutler et al. (2004) is intuitively quite appealing, there is considerable counterevidence against it from more recent typological quantification. Maddieson (2011: and references therein) demonstrated that size of the consonant and vowel inventories, syllable complexity measures, size and complexity of tone systems and the distinguishing role of stress are all positively correlated in the UCLA Phonological Segment Inventory Database (UPSID), a genealogically representative sample of between 450 and 650 languages (depending on the linguistic property at issue).

4.1.4 Vowel quality

It remains to be seen whether the disambiguating role of stress can be used to shed light on the stress controversy in Austronesian languages. As for Papuan Malay, field elicitations reported in Kluge (2017) do not contain any minimal stress pairs. The acoustic evidence (Chapter 2) was based on unscripted story re-tellings by 19 speakers and showed structural support for penultimate word

Table 4.1: Mean number of embedded words per complex word when ignoring stress (left) and when considering stress (right). Data from Cutler et al. (2004) and Cutler & Pasveer (2006).

	All embeddings	Stress-matched	Proportion
Dutch	1.52	0.74	0.49
English	0.94	0.59	0.62
German	1.62	0.80	0.49
Spanish	2.32	0.73	0.31

stress as the default pattern and ultimate stress as the exceptional pattern (i.e., when the penultimate syllable contains /ɛ/). Vowel quality appeared as one of the stronger acoustic correlates, showing that vowel formants in stressed syllables are further displaced from the center of the acoustic space compared to those in unstressed syllables. However, this result was found for the Papuan Malay vowels /i/, /a/, /ɔ/ and /u/, but not for /ɛ/, which showed the opposite effect (i.e., further away from the center when unstressed). A speculative explanation was provided, in which /ɛ/ was produced as stressed schwa, a possibility reported for Jakarta Indonesian (Laksman 1994; see Chapter 2). However, schwa as the only cause of a stress shift has been reported as a doubtful indicator of the existence of stress, and the avoidance of schwa could actually be an effect of phrase prosody (Goedemans & van Zanten 2014: 88). It thus remains to be seen how the role of /ɛ/ should be interpreted in the light of potential other phonological factors that could cause a stress shift. The interpretation is particularly challenged by the fact that /ɛ/ appears to be the main reason for stress to shift from penultimate to ultimate (e.g., *lama*, /ˈla.ma/, ‘to be long in duration’ vs. *lema*, /lɛ.ˈma/, ‘to be weak’), and at the same time shows exceptional formant displacement when stressed (Chapter 2). The latter finding could be related to the extreme sparsity of stressed /ɛ/, a pattern mainly observed in penultimate position when the ultimate syllable also contains /ɛ/. It remains unclear, however, to what extent the vowel quality difference between centralised /ɛ/ (stressed) and decentralised /ɛ/ (unstressed) should indeed be interpreted as a stress difference.

One option would be to assume that the alleged stress differences found for /ɛ/ are actually segmental differences, i.e., two distinct but acoustically similar vowels. This type of explanation, as provided for Ambonese Malay (Maskikit-Essed & Gussenhoven 2016), could in theory also be applied to the other vowels, assuming that the Papuan Malay inventory includes twice as many vowels as reported by Kluge (2017). Such an inventory would then consist of five vowels each in a full/long version and a reduced/short version, which would imply that Papuan Malay does not have word stress. Vowel quality or length differences are indeed commonly reported as distinguishing between two subsets of vowels in an inventory (e.g., Maddieson 1984). However, given that the acoustic distinctions between the vowel categories concern several prosodic cues at the same time (Chapter 2), the stance taken in this chapter is that the acoustic results found so far need to be explained as being suprasegmental (i.e., word stress) rather than segmental properties. More research is needed to complement the existing acoustic findings. The research questions addressed in this chapter are outlined in what follows.

4.1.5 Research questions

Given the above discussion of the literature, this chapter addresses two different aspects of the lexical status of word stress in Papuan Malay using non-acoustic analyses. The first is a confirmatory analysis investigating the disambiguating function of word stress (RQ1). We predict that if stress is indeed a suprasegmental property of Papuan Malay words, listeners might benefit from how the patterns are distributed. If this holds true, stress could facilitate word disambiguation. This hypothesis is investigated in a lexical analysis similar to the ones carried out by Cutler et al. (2004) and Cutler & Pasveer (2006). The null hypothesis would be that Papuan Malay stress patterns do not facilitate word disambiguation, which could be an indication that stress needs to be interpreted as a segmental property, similar to English (Cooper et al. 2002).

RQ1: Do Papuan Malay stress patterns reduce the number of alternative word candidates?

Second, more research is needed on the phonology of the stress placement in Papuan Malay, in particular to understand the unique role of /ɛ/. This is investigated in the current chapter by means of an exploratory analysis using the random forests classification technique (Breiman 2001). This type of analysis ranks the relative importance of multiple interrelated variables. In this way, the effect of /ɛ/ can be compared to factors that could also be relevant for stress placement, such as other vowels, syllable structure or word class.

RQ2: What phonological factors determine stress placement in Papuan Malay?

These issues are investigated in this chapter by means of a lexical analysis of word embeddings (Section 4.2) and a random forest analysis of morpho-phonological predictors of stress (Section 4.3) using a corpus of Papuan Malay words.

4.2 Lexical analysis of word embeddings

The corpus consisted of Papuan Malay words, as provided by Kluge (2017) and Kluge et al. (2014). The words selected for the analyses in this study concerned native roots only (Kluge 2017: Appendix A.1). Loanwords, which occur frequently in this language, were discarded. In this way, potential influences from stress patterns that originate from other languages were avoided. The word lists consisted of a written lexeme, phonetic transcription with syllable boundaries and stress indicated, word class label, and an English gloss.

Before obtaining the number of embedded words, duplicates (e.g., homonyms such as *pasang* for 'pair' or 'to install') were purged from the word list such that only single instances of sound shapes were left. Given the paucity of four-syllable words in the list (N = 3), they were excluded from the counting procedure. Based on the syllable boundary indications in the phonetic transcriptions, the number of syllables was counted per word. In addition, the stressed syllable was indicated as a number referring to the position of that syllable in that word, based on the stress marks in the phonetic transcriptions. The final word list used for analysis contained 1106 words, 1062 of which were polysyllabic and potential candidates for carrier words. Table 4.2 provides the word counts for the analyzed word list.

In the existing studies on word embeddings (Cutler & Pasveer 2006; Cutler et al. 2004) the mean embeddings were weighted by word frequency. For Papuan Malay no corpus data is available to provide word frequencies. In the current analysis, the embedding statistics are therefore unweighted (see Section 4.2.2 for further discussion). The analysis carried out in this study is based on the word list underlying Table 4.2. Although a fair number of words might not frequently occur in spontaneous speech, the word list still provides a subset that is representative for the language. That is, the words in the list were elicited in spontaneous conversations and formed the basis for the phonological analysis (Kluge 2017).

In the absence of frequency data, the current analysis established the number of polysyllabic words which contained one or more embeddings and counted the subset of these embeddings for which stress matched between carrier word and embedded word. Following previous lexical analysis studies, syllable boundaries were taken into account. For example, *ke* 'to' was counted as an embedding in *kewa* (/ˈkɛ.wa/ 'dance party'), but not in *kembang* (/ˈkɛm.baŋ/ 'flower'). As for stress-matching embeddings, ka (/ˈka/, 'or') would count in *kali* (/ˈka.li/ 'river') or in *sikakar* (/si.ˈka.kɐr/ 'to hold onto tightfisted') but not in *muka* (/ˈmu.ka/ 'front'). The counts were done automatically using syllable-level string matching based on the phonetic transcriptions in the word list. String matching was applied after diacritics were removed from the transcriptions, as these marks indicate variation in segmental surface realizations. This was done in order to find phonemically identical (i.e., matching) syllables, following the methods of Cutler et al. (2004) and McQueen et al. (1995).

Table 4.2: Word list counts by number of syllables (σ) and word stress.

Number of σ	Penultimate stress	Ultimate stress	All stresses
1	-	-	44
2	892	103	995
3	63	4	67
Total	955	107	1106

4.2.1 Results

As reported in Table 4.3, the list consisted of a total of 159 polysyllabic carrier words, i.e., words for which embedded words could be found (column "carrier words"). The number of embedded words was 170 (column "embeddings – all"). The embedded words had a length of either one syllable (in carrier words of two or three syllables) or two syllables (in carrier words of three syllables). The total

Table 4.3: Word counts and embeddings for each word-length in syllables in the Papuan Malay word list.

Word length (σ)	Carrier words	Embeddings	
		All	Stress-matched
2	136	139	84
3	23	31	18
All lengths	159	170	102

number of embeddings was overall slightly higher than the number of carrier words (170 versus 159 respectively). This result indicated that carrier words had more than one embedding on average (μ = 1.07). The counts decreased when considering only the embedded words that matched for stress with the carrier word (column "embeddings – stress-matched"). The latter observation is an indication that when stress is taken into account, the mean number of embeddings per carrier word drops below one (μ = 0.64). Note that from the 84 stress-matched embeddings in disyllabic carrier words, 82 matched with penultimate stress and 2 with ultimate stress. From the 18 stress-matched embeddings in trisyllabic words all matched with penultimate stress.

Table 4.4: All embeddings: length and location of embedded word (E) in carrier word (C) for each carrier word length in syllables (σ).

Length (σ)		Location of onset of E in C		
Carrier	Embedding	σ1	σ2	σ3
2	1	95	44	
3	1	8	8	3
	2	4	8	0

Table 4.5: Stress-matched embeddings: length and location of embedded word (E) in carrier word (C) for each carrier word length in syllables (σ).

Length (σ)		Location of onset of E in C		
Carrier	Embedding	σ1	σ2	σ3
2	1	82	5	
3	1	0	8	0
	2	2	8	0

In addition, the location of the embedded word in the carrier word was counted, providing an insight into where in the carrier word the embedding started (see Cutler et al. 2004 for comparable tables with ratios). Tables 4.4 and 4.5 report the locations for all embeddings and for stress-matched embeddings respectively. Both tables show that the decrease in embeddings due to the consideration of word stress was particularly large for disyllabic carrier words with embeddings starting in the second syllable (e.g., *yang*, /ˈjɐŋ/, relativiser, in *goyang*, /ˈgɔ.jɐŋ/, 'to shake'), and for trisyllabic carrier words with embeddings starting in the first syllable (e.g., *sa*, /ˈsa/, 1SG, in *sarana*, /sa.ˈɾa.na/, 'facility'). In the latter case, all embeddings (N = 8) could be disambiguated on the basis of stress information.

In order to assess the potential influence of syllable structure on the degree of disambiguation by stress, the counts of embeddings were split by structure of the first syllable in the embedding (Table 4.6). These counts show that CV syllable structure is the most frequent. The degree of disambiguation does not appear to differ due to syllable structure. It should be noted, however, that the counts challenge the comparison of syllable structures due to the overall frequency differences. The largest degree of disambiguation by stress is found in monosyllabic embeddings with an onset in the second syllable of the carrier word, both for CV and for CVC syllables.

Table 4.6: Number of all/stress-matching mono- and polysyllabic embeddings with location in the carrier ($\sigma1$, $\sigma2$, $\sigma3$) as a function of syllable structure of the initial syllable of the embedding.

Structure	Monosyllabic emb.			Polysyllabic emb.		N total
	$\sigma1$	$\sigma2$	$\sigma3$	$\sigma1$	$\sigma2$	
CV	95/74	38/8	2/0	1/0	6/6	142/88
CVC	6/6	14/2	1/0	3/2	0/0	24/10
CCV	2/2	0/0	0/0	0/0	0/0	2/2
V	0/0	0/0	0/0	0/0	2/2	2/2
N total	103/82	52/10	3/0	4/2	8/8	170/102

4.2.2 Conclusion

The lexical analysis shows that many word embeddings can be successfully eliminated during word recognition on the basis of mismatching stress patterns in Papuan Malay. The reduction is most clearly found for embeddings with an onset in an unstressed syllable of the carrier word (Tables 4.4 and 4.5). This can be explained when considering that stress is highly regular in Papuan Malay (penultimate) and that most of the embeddings concern monosyllabic words. This result resembles the one reported for Spanish, a language with predominant penultimate stress, for which stress-matched embeddings were found mainly in penultimate syllables (Cutler et al. 2004: Table 4.6). It should also be noted that for both

Papuan Malay and Spanish, monosyllabic embeddings were counted as stressed and were the most frequent length of embedding. A crucial difference concerns the fact that the words under analysis were shorter in Papuan Malay than in Spanish, despite their similar phoneme inventory sizes (23 and 25 respectively). It is unlikely that the small Papuan Malay corpus did not reflect word lengths in a natural way. For instance, the predominance of disyllabic words has also been observed in spontaneous speech in Chapter 2.

In this respect it is interesting to reconsider the hypothesis that languages with small phoneme inventories resort to suprasegmental cues to signal stress to a larger extent than languages with large phoneme inventories (Cutler et al. 2004). While suprasegmental cues do indeed play a role in stress perception in Papuan Malay (Chapter 3), it is unclear how these compare to segmental cues such as vowel quality (further discussion in Section 4.4). Given the higher degrees of disambiguation in Spanish than in Papuan Malay, it seems that phoneme inventory is not the only factor predicting the role of suprasegmental cues (Maddieson 2011). The morphological composition appears equally important, as Spanish has longer words than Papuan Malay. This caused stress-matched embeddings to occur mainly in word-initial syllables in Papuan Malay (predominantly disyllabic words), whereas in Spanish stress-matched embeddings were found in penultimate syllables of longer words as well. From a processing point of view, it has been argued that both the size and the location of the embedding are crucial for successful speech processing (McQueen et al. 1995). Given the word length differences between Papuan Malay and Spanish, however, the disambiguating function simply has less opportunity to facilitate speech processing in the former language. This is plausibly reflected in the relative amount of disambiguation found in both languages.

Although word frequency could not be taken into account in the current study, the mean number of embeddings per word can now be provisionally compared with the data from the other languages (Dutch, English, German and Spanish). For a more direct comparison between the Papuan Malay values and those in Table 4.1, the proportions of stress-matched embeddings can be computed by subtracting the mean value of stress-matched embeddings per carrier word from the mean value of all embeddings per carrier word. In this way, low proportions predict large facilitation, whereas high proportions predict small facilitation. These proportions thus give an insight into the relative magnitude of the facilitatory effect of stress on word recognition and abstracts over language-specific numbers of embeddings (Table 4.7, right column). The highest proportions are found for English (0.62) and Papuan Malay (0.60), followed by German and Dutch (each

0.49), whereas Spanish shows the lowest proportion of stress-matched embeddings (0.31).

Table 4.7: Mean number of embedded words per word when ignoring stress (left), when considering stress (mid), and the proportion of the latter (right). Data from Cutler et al. (2004), Cutler & Pasveer (2006) and the current study.

	All embeddings	Stress-matched	Proportion
Dutch	1.52	0.74	0.49
English	0.94	0.59	0.62
German	1.62	0.80	0.49
Spanish	2.32	0.73	0.31
Papuan Malay	1.07	0.64	0.60

The observations of this analysis lead to the conclusion that word stress in Papuan Malay has a potential function in word recognition in that it may aid the process of rejecting alternative word candidates. Given the similar proportions of stress-matched embeddings in the current study compared to the results of Cutler et al. (2004) and Cutler & Pasveer (2006) for English, it needs to be further discussed to what extent the facilitatory effect of word stress in Papuan Malay can be found in suprasegmental or segmental cues. This question is particularly relevant given that the relatively low degree of disambiguation in English (i.e. high proportions of stress-matched embeddings) can be ascribed to the fact that stress differences in this language are mainly signaled by vowel quality differences and to a much lesser extent by suprasegmental differences. Although it was found that vowel reduction is an acoustic cue to stress in Papuan Malay (Chapter 2), it remains to be seen to what extent listeners make use of this cue, in particular given that /ɛ/ did not show the type of formant displacement found in other vowels (Section 4.1). If suprasegmental stress information is indeed less important for word recognition in Papuan Malay compared to languages such as Dutch, German or Spanish, a larger role could be reserved for vowel reduction in this language.

An important difference between Papuan Malay and English concerns the mean embeddings per carrier word when all embeddings are counted. This can be illustrated when recalling that one is the crucial limit for the (mean) number of embeddings per carrier word (Section 4.1.2). In English, the mean value is just below one (0.94), whereas in Papuan Malay this number is just above one (1.07). In English, therefore, disambiguation is less of a challenge for listeners to begin with. In Papuan Malay, however, there is more need to disambiguate than in English, predicting that the relative importance of suprasegmental cues is larger in the former language. It has to be noted that taking into account frequency data could still somewhat change this number for Papuan Malay. Despite the fact that the corpus in the current study consisted of commonly used words, no weight differences between high and low frequency words were taken into account. This could have resulted in a more coarse-grained analysis provided here, compared to the analyses in the literature (Cutler & Pasveer 2006; Cutler et al. 2004; McQueen et al. 1995). In this respect it is important to consider that the small corpus of common words in Papuan Malay (compared to the large corpora used in previous studies) could have reduced the chance that many low-frequency words were taken into account and as a consequence were given too much weight.

Given that vowel quality plays a crucial role in stress distinctions (Section 4.1.4), as established for English and as suggested for Papuan Malay, it is important to further explore the phonological factors underlying stress placement in the latter language, as further outlined in Section 4.3.

4.3 Random forest analysis of stress placement factors

Random forest analysis is a classification method based on the construction of a large number of decision trees (Breiman 2001). In order to assess which variable splits (classifies) the data best, trees are constructed on the basis of random data and variable subsets. Random forests are particularly useful to determine the predictive value of a large set of variables and a small number of observations. Compared with other statistical methods, random forests are more resistant to overfitting and collinearity between predictors. Random forest analyses have only recently been introduced into linguistics (Tagliamonte & Baayen 2012), and more specifically into phonetics and phonology (e.g., Arnold et al. 2013; Baumann & Winter 2018; Grafmiller & Shih 2011; Grice et al. 2015). The method is promising as notions such as prominence, stress or phonological weight typically correlate with a large number of acoustic and/or linguistic variables. Random forests can help reveal underlying mechanisms of linguistic structure, by providing powerful generalizations based on a relatively small set of field data. In addition, predictors are allowed to be derivatives of each other, even if there is a considerable degree of correlation among the predictors. For example, a variable with five levels corresponding to all vowels in an inventory and a variable with two levels only distinguishing low and high vowels from the same inventory can be both included in the model without losing predictive accuracy (Strobl et al.

2008). The predictive value of a certain variable in a random forest is expressed by means of its variable importance. Although the variable importance values are affected by correlation among predictors, their ranking relative to each other remains unaffected. The absolute variable importance values are irrelevant, as they are randomly generated (hence "random forest"). Thus, the interpretation of variable importance generally relies on the relative differences between the respective values (Shih 2013).

The selected subset of the Papuan Malay corpus for the random forest analysis consisted of two-syllable words only, in order to obtain a homogeneous set; words with one syllable (N = 46) or more than two syllables (N = 73) were relatively infrequent. Thus, the representativeness of the corpus was compromised to a minimal extent. An overview of the number of words per lexical category is given in Table 4.8. Note that words which translate to adjectives in English are expressed by stative verbs in Papuan Malay. For example, /bɛ.ˈsar/ ('big' - lit. 'be big') is labeled as a verb in the corpus.

Table 4.8: Distribution of word classes in the corpus.

Word class	Count
V bi(valent)	341
V mono stative	205
V mono dynamic	51
V other	7
Adverb	32
Noun	355
Function word (all)	49
Total	1040

4.3.1 Predictors

Of interest to the current analysis are phonological factors that make a syllable likely to be stressed. In this respect it is important to consider the abstract phonological notions of sonority and weight. Sonority has been reported to be an underlying principle of syllable structure, reaching its peak in the vocalic nucleus and its valley at the edge consonant(s). Perhaps the simplest phonetic correlate of sonority is intensity, although the search for such a correlate shows that the relation between the phonetics and the phonology of sonority is far more complex than single correlates can explain (e.g., Parker 2002; Albert 2023). Based on sonority scales, speech sounds can be classified following a particular hierarchy, generally with vowels at the top and obstruents at the bottom (e.g., Clements 1990). More sonorous sounds often play a more decisive role in stress placement than less sonorous sounds, with a sub-hierarchy assumed within sound classes. For example, vowels have been reported to be more sonorous when they are peripheral and when their height is low (Parker 2002: 27). While sonority is a phonological property generally attributed to individual phonemes, it affects the status of the syllable in terms of weight (Gordon 2006). In many stress languages the weight of the syllable is decisive for stress placement, since heavier syllables generally attract stress more than light ones (weight-to-stress principle; Chomsky & Halle 1968). One criterion distinguishes light and heavy syllables according to whether the syllable ends with a short vowel (light) or not (heavy), i.e., the "Latin-criterion" in Ryan (2016). There are different criteria to distinguish light from heavy syllables and it is open to discussion to what extent a strictly binary division or more gradient scales of weight are needed to account for cross-linguistic observations (Ryan 2016). Vowel nuclei (and therefore sonority) are often most determining for weight-sensitive stress placement, although in some languages the coda (and sometimes even the onset) of the syllable co-determines its weight (e.g., Goedemans 1998; Gordon 2005a; Hayes 1995; Ryan 2014). In the current study of the phonological factors underlying Papuan Malay stress, it is important, therefore, to consider the precise segmental makeup of the syllable. In particular, it is crucial to note that /ɛ/ can be realised as schwa in Papuan Malay (Kluge 2017), similar to other Trade Malay varieties (Paauw 2009). Schwa, the most central vowel, generally has lower sonority compared to the other vowels in the vowel inventory. The frequently reported role of /ɛ/ (or schwa) in Indonesian languages as a reliable indicator of the presence of stress has been questioned (Goedemans & van Zanten 2014).

For the current analysis a set of predictors was chosen that potentially affect syllable weight. Some of the predictors were derived from others. This was done to test whether individual sounds (particular vowels or consonants) or rather phonological classes of sounds predict stress placement better, as explained above. The included predictors (in italics, IPA notations following Kluge 2017) concerned *structure* in terms of consonantal and vowel segments, from which the *openness* of the syllable and the actual segments in the *onset*, *nucleus* or *coda* were derived. Papuan Malay has five vowels (/i, ɛ, a, ɔ, u/) and 17 consonants (stops: /p, b, t, d, k, g/; affricates: /ʧ, ʤ/; nasals: /m, n, ŋ/; fricatives: /s, h/; rhotic: /r/; approximants: /l, j, w/). The predictors *manner* of articulation and *height* were derived from the actual consonants and vowels respectively. Voicing of the onset and constriction type of the coda were added as these have been reported as possible contributors to syllable weight (e.g., Ryan 2014). Furthermore, recent work showed a difference in F0 excursion size between function words and content words, in that the former were smaller than the latter in phrase-final positions (Kaland & Baumann 2020). Although this difference applied to the phrase level, potential implications for word prosody cannot be excluded based on these results. In fact, in some languages word class correlates with word stress placement (e.g., in English Latinate pairs: *ˈsubject* (noun) and *to subˈject* (verb) form a minimal stress pair; in Dutch and German a difference in stress correlates with verbs with and without separable prefixes, e.g., in German *ˈübersetzen – setzte über – ˈübergesetzt* 'ferry someone across, inf., past, participle' when stressed on *ˈüber*; *überˈsetzen – überˈsetzte – überˈsetzt* 'translate(d) inf, past, participle' when stressed on the second part of the compound). Therefore, *word class* was included as a predictor.

4.3.2 Statistical analysis

The analysis was done in R (R Core Team 2019) using the package "ranger" (Wright & Ziegler 2017), which offers a computationally less intensive way to perform random forests compared with packages such as "party" (Strobl, Hothorn, et al. 2009) or "randomForest" (Liaw & Wiener 2002). The response variable in the random forest was stress location (2 levels: penultimate, ultimate). The predictors (in italics) were *syllable structure* (6 levels: CCV, CCVC, CV, CVC, V, VC), *onset* (18 levels: /b, ʧ, d, g, h, ʤ, k, l, m, n, ŋ, p, r, s, t, w, j/, no onset), *nucleus* (5 levels: /i, ɛ, a, ɔ, u/), *coda* (12 levels: /j, k, l, m, n, ŋ, p, r, s, t, w/, no coda), *openness* (2 levels: open, closed), *manner of articulation in onset and coda* (each 6 levels: plosive, fricative, nasal, rhotic, approximant, none), *height of the vowel nucleus* (3 levels: open, mid, close), *voicing of the onset* (3 levels: voiced, voiceless, no onset), *constriction type of the coda* (3 levels: sonorant, obstruent, no coda), and *word class* (7 levels: see Table 4.8). In addition, a control-predictor *gloss* (the English translation of each word) was added. Gloss is not expected to be of any predictive value and should therefore have a low variable importance. Therefore, variable importance values of other predictors that lie around or below the one of the control-predictor can be used as an additional indication of which predictors do not affect the response variable at all. Except for word class and gloss all predictors were included for both the first and second syllable in the word (total: 22 predictors, see Table 4.9).

Table 4.9: Overview of all 22 predictors (σ for respective syllables) in the random forest analysis and their values for the example word /ˈbam.bu/ ‘bamboo’.

	Predictor	Value
1	Syllable structure σ1	CVC
2	Syllable structure σ2	CV
3	Onset σ1	b
4	Onset σ2	b
5	Nucleus σ1	a
6	Nucleus σ2	u
7	Coda σ1	m
8	Coda σ2	*no coda*
9	Openness σ1	closed
10	Openness σ2	open
11	Manner onset σ1	plosive
12	Manner onset σ2	plosive
13	Manner coda σ1	nasal
14	Manner coda σ2	*none*
15	Height nucleus σ1	open
16	Height nucleus σ2	close
17	Voicing onset σ1	voiced
18	Voicing onset σ2	voiced
19	Constriction coda σ1	sonorant
20	Constriction coda σ2	*no coda*
21	Word class	noun
22	Gloss	bamboo

The number of trees in the analysis was increased in steps of 1000, starting from 1000 trees. The variable importance of the predictors reached a stable ranking around 5000 trees. To obtain a robust result, the final number of trees was set to 10000 (Shih 2013). The number of randomly preselected predictors was set to the square root of the total number of predictors in the analysis ($\sqrt{22}$), and variable importance mode was set to “permutation”. These settings are recommended for analyses with correlating predictors, following Strobl et al. (2008) and Strobl, Malley, et al. (2009).

The distribution analysis (Table 4.2) consisted of counts; 0 for each word with penultimate stress and 1 for each word with ultimate stress. The proportion of ultimate/penultimate stresses was then calculated by taking the mean of all counts. The two analyses combined appeared particularly helpful to interpret the variable importance values, as their absolute values are not indicative (Section 4.3.3).

4.3.3 Results

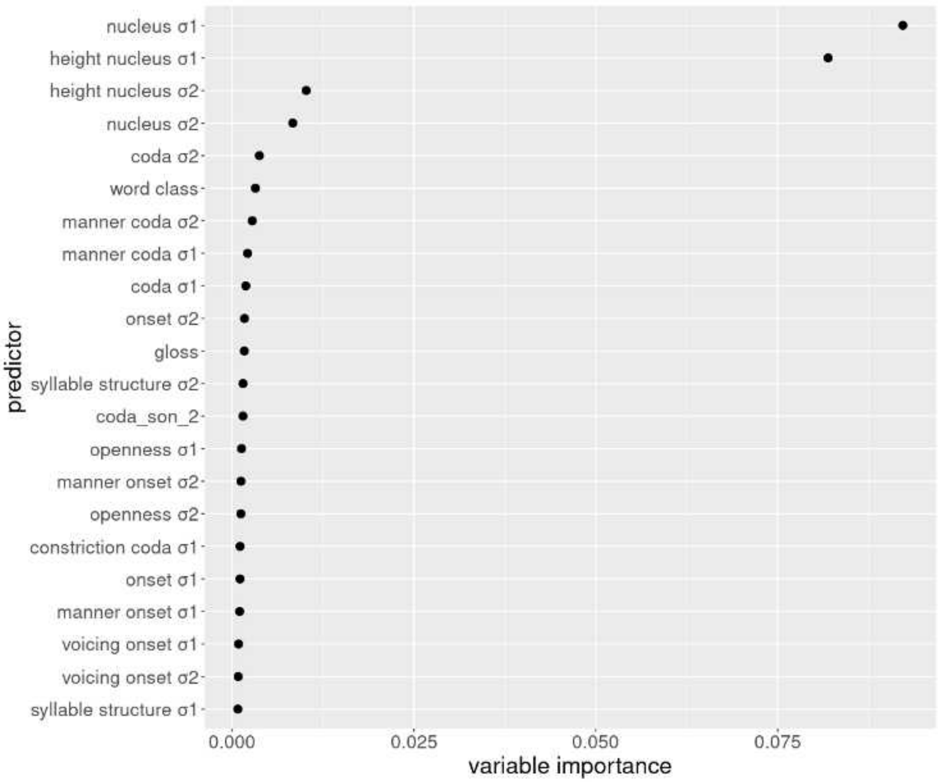

Figure 4.1: Variable importance plot with the predictors ranked from high (top) to low (bottom).

Two factors stand out as predictors for the location of stress in Papuan Malay (Figure 4.1): the nucleus in the first syllable and the height of the vowel nucleus in the first syllable. Other predictors showed considerably lower variable importance values, although the height of the vowel nucleus in the second syllable as well as the nucleus of the second syllable appeared more predictive than the lowest ranked ones. Given the hypothesised irrelevance of the control predictor *gloss* (ranked 11/22), predictors with similar or lower ranking have little to no predictive value. Indeed, from the fifth ranked predictor onwards the variable importance values hardly vary (and yield 0) compared to higher ranked ones.

Table 4.10: Ultimate/penultimate stress ratio for the four most important predictors (italics) in the random forest analysis (N = 1040).

Nucleus	Height	*Nucleus σ1*	*Ht. nucl. σ1*	*Ht. nucl. σ2*	*Nucleus σ2*
/a/	open	0	0	0.14	0.14
/ɛ/	mid	0.63	0.43	0.05	0.06
/ɔ/	mid	0			0.05
/i/	close	0.01	0.01	0.08	007
/u/	close	0.01			0.10

The highest ratio of ultimate stresses (0.63) was predicted by *nucleus σ1* (Table 4.10). This appeared to be the result of /ɛ/ causing stress to move from default penultimate to ultimate syllable (e.g., *lemba* /lɛm.ˈba/ 'valley'). *Height nucleus σ1* showed the highest ratio for mid vowels. This was again caused by only /ɛ/, as /ɔ/ was always stressed when it occurred in the first syllable (e.g., *otak* /ˈɔ.tɐk/ 'brain'). Lower predictive importance was found for *height nucleus σ2* and *nucleus σ2* (Table 4.10), indicating that only a small number of stress patterns could be predicted based on the nucleus in the second syllable. The largest ratio of ultimate stress cases was found when /a/ occurred in the second syllable, indicating that this vowel attracts stress in a small number of cases (see discussion below on the predictive strength of this factor). Note that /a/ is the only open vowel in Papuan Malay, explaining why the predictors *height nucleus σ2* and *nucleus σ2* had similar effects (Figure 4.1). The exact influence of vowel nuclei is further schematised in Table 4.11. This table shows that /ɛ/ in the first syllable causes stress to shift to the ultimate syllable. A consistent exception to this rule, however, appears to be when the ultimate syllable already consists of either /ɛ/ or /ɔ/ (e.g., *pendek* /ˈpɛn.dɛk/ 'to be short'; *belok* /ˈbɛ.lɔk/ 'to turn'). In these cases, the ultimate stress ratios are particularly low (Table 4.11). This result indicates that mid vowels in the second syllable generally prevent a stress shift.

However, none of the predictors found by the random forest predicted the stress distribution entirely. To analyze precisely how the different predictors work phonologically, three criteria were formulated (Table 4.12). First, ultimate stress is mainly found when /ɛ/ occurred in the first syllable, confirming Kluge (2017). The three exceptions to this criterion are *kitong* /ki.ˈtɔŋ/ 'we/us, including adressee', *kumur* /ku.ˈmur/ 'rinse mouth' and *kuskus* /kus.ˈkus/ 'cuscus', see also

Table 4.11: Proportion of ultimate stress cases (N = 108) as a function of the identity of the vowel in the first (V1) and second (V2) syllable.

V1	V2					
	/i/	/ɛ/	/a/	/ɔ/	/u/	Total
/i/	0	0	0	0.01	0	0.01
/ɛ/	0.12	0.04	0.60	0.02	0.19	0.97
/a/	0	0	0	0	0	0
/ɔ/	0	0	0	0	0	0
/u/	0	0	0	0.02	0	0.02
Total	0.12	0.04	0.60	0.05	0.19	1.00

Kluge (2017: 96). Note that /ki.ˈtɔŋ/ is short for /ki.ˈtɔ.raŋ/ (Kluge 2017: 326), which has penultimate stress (from /ˈki.ta/ and /ˈɔ.raŋ/, lit. 'us humans'). Re-evaluation of /kus.ˈkus/ showed that it could be analyzed as a Malay loanword (Scott 1896), indicating that its inclusion in the corpus might not have been justified.

Table 4.12: Word counts after applying the criterion that decreased the penultimate stress ratio/increased the ultimate stress ratio (Table 4.9).

Criterion	N ˈpenult.	N ˈult.	Exceptions
Total	932	108	-
Nucleus $\sigma1$ = /ɛ/	61	105	/ki.ˈtɔŋ/ /ku.ˈmur/ /kus.ˈkus/
Height nucleus $\sigma2 \neq$ mid	25	100	/tʃɛ.ˈrɛj/ /sɛ.ˈrɛj/ /dʒɛ.ˈlɛk/ /dʒɛm.ˈpɔl/ /sɛ.ˈdɔt/
Nucleus $\sigma2$ = /a/	16	65	...

Second, 61 words had /ɛ/ in the first syllable and penultimate stress. From these words, 36 had a mid vowel (/ɛ/ or /ɔ/) in the second syllable. Five exceptions to this criterion had ultimate stress, with /ɛ/ in the first syllable and a mid vowel in the second syllable; /ʧɛ.ˈrɛj/ ‘to divorce’, /sɛ.ˈrɛj/ ‘lemongrass’, /ʤɛ.ˈlɛk/ ‘be bad’, /ʤɛm.ˈpɔl/ ‘thumb’, /sɛ.ˈdɔt/ ‘to suck’. Note that [ɛj] in /ʧɛ.ˈrɛj/ and /sɛ.ˈrɛj/ is analyzed as the realisation of underlying /aj/ due to the liquid in the onset of the second syllable (Kluge 2017: 84). With /ɛ/ in the first syllable, underlying /a/ could make the second syllable the preferred location for stress. The status as native root of two more words is doubtful. That is, /ʤɛm.ˈpɔl/ is reported as a loanword from Javanese/Sundanese (Haspelmath & Tadmor 2009) and /sɛ.ˈdɔt/ is reported as a loanword from Javanese (Stevens & Schmidgall-Tellings 2010).

Third, the presence of an open vowel (/a/) in the second syllable increases the likelihood of ultimate stress. However, from the words with /ɛ/ in the first syllable and /a/ in the second syllable, only 65 had ultimate stress. Given that there were 108 ultimate stress cases in total (Table 4.12), the open vowel in the second syllable did not predict stress placement as strongly as the first two criteria. In other words, the open vowel in the second syllable was of minor importance and could only explain a small additional number of stress cases after the main criteria were applied. This result is reflected in the large variable importance difference between the first two predictors and the lower ranked ones (Figure 4.1). For this reason, no more criteria were formulated.

4.3.4 Conclusion

The results are best summarised by assuming that the default position of word stress in Papuan Malay is the penultimate syllable and that mid vowels generally reject stress. It is crucial, however, to note the difference in stress rejection between /ɛ/ and /ɔ/. When the first syllable contains /ɛ/, stress shifts to the ultimate syllable when the ultimate syllable does not contain a mid vowel. This result indicates that /ɛ/ rejects stress regardless of syllable position, although in 25 words the first syllable was stressed and contained /ɛ/ while no mid vowel was found in the second syllable. The occurrence of /ɔ/ in the first syllable did not cause a stress shift; all words with /ɔ/ in the first syllable and no mid vowel in the second syllable (N = 29) had penultimate stress. However, with /ɔ/ in the second syllable only three out of 66 words had ultimate stress (/ki.ˈtɔŋ/, /ʤɛm.ˈpɔl/, /sɛ.ˈdɔt/), which could all be explained as phonological exceptions or loanwords. Thus, /ɔ/ only rejects stress in the second syllable. This difference could be an indication of a stress hierarchy within the vowel inventory of Papuan Malay. In this hierarchy, corner vowels /i/, /a/ and /u/ can be stressed regardless of position, /ɔ/ can only be regularly stressed (i.e., penultimate), whereas /ɛ/ should be avoided as stressed regardless of position. In addition to the mid vowels, /a/ was found to attract stress to a limited extent, although it did not predict a stress shift.

The results are in line with the literature on Trade Malay with respect to the role of /ɛ/ (schwa) in stress placement (Kluge 2017; Paauw 2009). Furthermore, the role of /a/ as stress attractor is compatible with phonological accounts that distinguish open and close vowels as more and less sonorous respectively (Kenstowicz 1997; Selkirk 1984). Note, however, that the infrequently stressed mid vowels in Papuan Malay cannot be explained on the basis of openness as main correlate of vowel sonority. This would mean that /i/ and /u/ are poorer candidates for stress than /ɛ/ and /ɔ/, which is incompatible with the observations of this study. The results rather indicate a difference in phonological status between corner vowels and mid vowels in the Papuan Malay inventory.

The analysis has shown that random forests provide an insightful analysis of which phonological factors play a role in stress placement. It is worth emphasizing that without the complementary distribution analysis (Tables 4.10, 4.11, and 4.12), the role of the predictors would have been difficult to interpret. Moreover, the direction of the effect of the most powerful predictors in the random forest analysis could be understood when interpreting the stress ratios. The predictive power of the random forest model is particularly clear from the relatively small number of exceptions with ultimate stress (N = 8) after applying the first two criteria in Table 4.12. In fact, the analysis revealed that three of these words were loanwords, which should not have been included in the corpus. For another three words alternative explanations could be found, indicating that their stress pattern did not violate the phonological criteria (Section 4.3.3). As for the exceptions with penultimate stress after applying the first two criteria listed in Table 4.12 (N = 25), we cannot provide alternative explanations or additional criteria that explain why stress did not move to the ultimate syllable in these cases. Nevertheless, 25 of the 932 penultimate stress cases and 8 of the 108 ultimate stress cases constitute less than four percent of all words in the corpus. Additional phonological criteria could theoretically be derived from the remaining highest ranked predictors in the random forest analysis, although these have the risk of generating more exceptions than explained cases (Table 4.12, bottom row).

4.4 Discussion and conclusion

This study has shown that non-acoustic distribution analyses complement existing acoustic research on stress in important ways. The primary aim for Papuan Malay was to investigate whether stress patterns could facilitate word disambiguation. Results showed support for this hypothesis. It has to be noted, however, that the total number of embeddings found in this study make up less than 15% of the word list. This means that for maximally twice that percentage (roughly all embeddings plus their carrier) there is a need for stress-based disambiguation. Thus, the segmental information is sufficient to recognise the majority (>70%) of the Papuan Malay words. The relative frequencies of embeddings in the other languages discussed here are unavailable. However, for these languages too it is expected that prosodic cues are of secondary importance for word recognition compared to segmental cues.

Although it remains to be tested perceptually how vowel quality affects stress judgments in Papuan Malay (Section 4.1), the random forest analysis was able to shed new light on the role of /ɛ/ (Section 4.3). This analysis confirmed that /ɛ/ in Papuan Malay indeed determines most of the patterns by rejecting stress. A crucial additional insight concerned /ɔ/, the other mid vowel in the inventory. It was shown that this vowel rejects stress in ultimate position, which indicates that the role of /ɛ/ in Papuan Malay stress might not be so unique as acoustic results have suggested (see Section 4.1). Rather, stress placement seems to follow phonological principles such as weight sensitivity, based on the sonority hierarchy, with an important distinction between corner vowels and mid vowels. In this respect it is worth noting that the mid vowels were the least frequent vowels in the Papuan Malay syllables analyzed in Chapter 2. In addition, shifted stress (to ultimate position) only occurred for the three corner vowels (Chapter 2: Table 2a). It seems likely, therefore, that the phonological class of mid vowels are avoided as stressed nuclei, if possible. Stressed mid vowels would appear only when the segmental structure offers no alternative, i.e., when there are only mid vowels in the word. These infrequent cases might therefore explain why the acoustic realization of stress is particularly weak or atypical (Section 4.1.2). The acoustic effects of stress on mid vowels thus seem to mainly reflect their weak position in the sound inventory of Papuan Malay, with differences among them in the way they reject stress.

Concerning the controversy of word stress claims in Indonesian languages, this study has provided new data in support of the stress claim for Papuan Malay. Given the acoustic and perceptual support (Chapter 2, Chapter 3) and the current results, it seems that, with respect to the disambiguation function, word stress in Papuan Malay is at least as functional as in English, which is uncontroversially a stress-language. In addition, the outcomes strongly suggest that Papuan Malay stress placement follows phonological principles that apply to the syllable structure of words (i.e., weight). This limits the options to maintain the claim that Papuan Malay is a language without word stress, or to interpret word prosodic patterns as reflexes of phrase prosody (e.g., van Zanten et al. 2010), although it is likely that both levels interact (Kaland & Baumann 2020). In this respect, Papuan Malay appears different from Ambonese Malay, which has been claimed not to have word stress at all (Maskikit-Essed & Gussenhoven 2016). The present study therefore reconfirms that regional variation, even among closely related languages, is an important factor (Goedemans & van Zanten 2007). Note, however, that the type of analysis of Papuan Malay and Ambonese Malay fundamentally differs, plausibly affecting the conclusions drawn so far (see Section 4.1.2 for a discussion). A major difference concerns the availability of several acoustic, perceptual and (in this chapter) lexical analyses for Papuan Malay, whereas stress in Ambonese Malay has only been investigated in a small number of studies and with an acoustic analysis of data from a small number of speakers. For a more thorough comparison, stress in other Trade Malay varieties, including Ambonese Malay, should be studied using a variety of empirical methods that crucially go beyond acoustic investigations and the presence of minimal pairs, as aimed for in the current chapter.

The outcomes of this study alone do not allow an answer regarding whether Papuan Malay has word stress or not, although the results are predominantly compatible with an affirmative answer. Rather, this study primarily provided complementary insights into the lexical status of Papuan Malay word prosody. More research is needed and a plausible outcome is that the nature of word stress found in this language crucially differs from the one found in other languages, as hinted at in the comparisons between Papuan Malay, Spanish and English discussed in the previous sections. For example, it should be investigated perceptually whether native listeners can indeed use acoustic cues to make lexical decisions. Answering these types of questions is left for future work.

5 The perception of word stress cues in Papuan Malay: a typological perspective and experimental investigation

5.1 Introduction

Word stress (henceforth also *stress*) refers to the presence of a single acoustically most prominent syllable in a word (Hyman 2006). Languages vary on several aspects of word stress, such as the degree to which the location of the stressed syllable in the word can be predicted by rules, which acoustic correlates are used to make the stressed syllable prominent and the extent to which stress patterns are useful for listeners. Although the presence of word stress is well established for some languages of the world, in other languages it is controversial (e.g. Indonesian, French, Korean). Examples of the latter group of languages are found in Indonesia, an area with a considerable amount of linguistic diversity. Apart from this diversity, the limited number of well documented languages and (consequently) empirical investigations have contributed to diverging claims over the last decades. To date, there are still new studies that counter earlier work in fundamental ways. The results from a small number of perception studies plays a key role in (resolving) this controversy. To further add to this research, the goals of the current study are two-fold. First, this chapter reviews the quantitative studies on Indonesian languages in order to provide structure in the arguments in favour or against the presence of word stress, and to reveal where more research is needed. Second, a perception experiment is carried out to further complete the current research on Papuan Malay word stress, in particular concerning its potential function in word identification.

The next sections summarise three key aspects of word stress across different languages before turning to how these aspects have been covered in Indonesian languages (Section 5.2); its acoustic realisation (Section 5.1.1), its role in speech perception (Section 5.1.2) and potential communicative functions (Section 5.1.3).

Section 5.1.4 summarises the state of the art in word stress research and based on the current gaps in the literature defines the goals of the current study in more detail.

5.1.1 Acoustic realisation

When a language has word stress, prosodic parameters such as duration, (spectrally weighed) intensity, vowel quality and F0 contribute to a more or lesser degree to its acoustic realisation (e.g. van Heuven 2018 for an overview). These make stressed syllables stand out in acoustic prominence compared to unstressed syllables. It should be noted that there are two main reasons why the acoustic cues are not all equally important for stress realisations. First, some correlates have properties that make them intrinsically less suitable for signalling stress. For example, F0 has been claimed to be a primary correlate of phrase level prosody, with limited or only indirect contributions to the realisation of word stress (Gordon 2014; Gordon & Roettger 2017; but Vogel et al. 2016). Second, languages differ in how they deploy the available cues. The Functional Load Hypothesis (FLH; e.g. Hockett 1955; Berinstein 1979) explains why in some languages not all cues are available to signal word stress. That is, if one acoustic cue serves a prosodic function other than word stress, such as F0 in lexical tone (Potisuk et al. 1996; Remijsen 2002) or duration in final lengthening (McDonnell 2016), this cue is not or only limitedly available for word stress. At a more general level the FLH also holds true. It was found that in languages with a fixed position of the stressed syllable the acoustic differences between stressed and unstressed syllables are smaller (i.e. stress is weakly realised), than in languages with more positional variation of the stressed syllable (e.g. Dogil & Williams 1999). Given the higher functional load on stress in the latter type of languages, the FLH can also be taken as an explanation of the strength differences of the acoustic correlates as observed cross-linguistically.

5.1.2 Perception

The extent to which listeners perceive the differences between stressed and unstressed syllables not only depends on their acoustic realisation. It has been shown that the way word stress patterns are distributed in the lexicon determines how sensitive listeners are to stress cues (Peperkamp et al. 2010). With a high number of exceptions to the phonological rules, there is a higher need to store prosodic information of the words in the lexicon. Spanish is an example of this type of language, and its listeners are therefore highly sensitive to the acoustic realization of stress in order to successfully identify words. It differs per language to which cues listeners attend exactly, with notable differences among closely related languages. It was found that vowel quality was the strongest perceptual cue to stress in English, Mandarin and Russian (Chrabaszcz et al. 2014), while duration was the strongest in Dutch and German (Sluijter et al. 1997; Mengel 2000). When the stressed syllable in the word is highly predictable by phonological rules, listeners do not need to store the stress information in their lexicon. This is the case in French, for which listeners have been reported as being *stress-deaf* (Dupoux et al. 1997). It should be noted that *deafness* referring to the insensitivity of (French) listeners to word stress cues is a relative notion. Experiments also showed that French listeners do hear the acoustic cues, but they do not process them at an abstract phonological level as these cues do not have a function there (Dupoux et al. 1997). In Polish, a language with largely predictable stress and a small number of exceptions, listeners were found to be mainly sensitive to the exceptional stress pattern (Domahs et al. 2012). Similarly, Italian listeners recognised the most dominant stress pattern by default and showed sensitivity to particular cues associated with the non-default pattern (Sulpizio & McQueen 2012).

5.1.3 Functions

Then, what types of functions could word stress have? In this brief overview, three main functions are distinguished; lexical contrast, word segmentation and word identification (but see Cutler 2005 for a more fine-grained overview). A well-known function is that stress parameters alone can distinguish one word from the other. This is often referred to as the lexically contrastive function of word stress, as illustrated with the Dutch word pair /ˈkaː.nɔn/ and /kaː.ˈnɔn/, translating to ‘canon’ (music) and ‘cannon’ (military), respectively. The number of minimal stress pairs in the lexicon differs per language and negatively correlates with the predictability of the stress patterns. With a high number of minimal stress pairs there is a low degree of predictability and vice versa. In languages without minimal stress pairs, patterns can be mostly predicted by phonological rules. Word stress can also help listeners to detect word boundaries and thus help them to segment the incoming speech signal (Cutler 2012). This holds – unsurprisingly - for languages where stress has a largely fixed position, such as the initial syllable in Slovak (Hanulíková et al. 2010). However, in languages with more variability in the stress position, such as in English and Dutch, segmentation is facilitated as well (Cutler & Butterfield 1992; Vroomen et al. 1996). The function of word stress central to the current study concerns the facilitation of word identification. Studies have shown that listeners can correctly discriminate words based on segmentally identical first syllables, which only differed in stress such as *admiral* and *admiration* in English (Cooper et al. 2002) and /ɔr.ˈkɛst/ (‘orchestra’) and /ˈɔr.gə l/ (‘organ’) in Dutch (van Heuven 1988). The word identification function of stress has mainly been shown for English and Dutch; languages with more fixed stress positions lack experimental investigation (Cutler 2005). Studies have shown how the facilitation effect on word identification originates from how stress is distributed in the lexicon. These lexical analyses investigated the occurrence of word embeddings, such as bee, which counts as embedding in belay and beanie, due to the fact that their initial syllables match for their segmental make-up. It appeared that the number of embedded words is reduced when taking into account stress information. Thus considering stress, bee (stressed) only counts as embedding in beanie (first syllable stressed) but no longer in belay (second syllable stressed). Lexically stored stress information could therefore help listeners to reject alternative word candidates that would otherwise be activated (and compete) during processing. The amount of reduction differed per language, with Spanish showing the largest degree of reduction, followed by Dutch and German, and English showing the smallest degree of reduction (Cutler et al. 2004; Cutler & Pasveer 2006).

5.1.4 Current study

Generalizing from the brief overview of acoustic, perceptual and functional aspects of word stress cross-linguistically, one factor stands out as determining all these aspects to some extent: the degree of stress predictability by phonological rules. The lower the degree of stress predictability, the more need for clear acoustic correlates and sensitive listeners, and the more important stress is for word processing. As the overview has shown, there are several fine-grained differences between languages. It appears that neither the ability to perceive stress patterns (e.g. French, Polish) nor the presence of the lexically contrastive function alone can be sufficient to conclude whether a language has word stress or not. Studies on the perception or functions of word stress are generally limited to a small number of languages, suggesting that our knowledge of stress perception and stress processing might be far from generalizable to underdescribed languages. This stands in large contrast to the extensive list of languages for which acoustic correlates to stress were reported (see the overview in Gordon & Roettger 2017). The state of the art in word stress research therefore allows for two observations. First, more research needs to be done to complement acoustic studies, in particular with regard to perception and the functions of word stress.

This type of research, as already mentioned, lacks investigations of more diverse languages. Second and more theoretically, with a growing body of perception research on word stress more attention should be given to its interpretation relative to the existing acoustic work. A central issue addressed in the current study concerns the extent to which perception studies contribute to the question of whether a language could be analysed as a stress-language. A conservative answer would be that acoustic evidence is sufficient to show whether systematic alternations between stressed and unstressed syllables exist in the speech signal. However, perception research sheds important light on the many different types and the communicative functions of word stress attested in languages of the world, as already briefly illustrated in the overview above. The goal of the current study is therefore two-fold. The primary goal is illustrating how exactly word stress has been diagnosed in a small number of Indonesian languages (Section 5.2). The linguistic diversity in this area is large and there are diverging claims on the status of word stress in some of its languages. Perception research has played a key role in resolving some of this controversy. The secondary goal is to add more perception research using more diverse languages. This is done by an experimental investigation of the word identification function of stress in Papuan Malay (Section 5.3). Recent studies have found a number of indications for the existence of word stress in this language, crucially still lacking perceptual verification. The results of the word identification experiment are discussed by evaluating how they contribute to the diagnosis of word stress in Papuan Malay and cross-linguistically.

5.2 Experimental research on stress in Indonesian languages

The literature overview in this section focuses on languages of Indonesia for which quantitative analyses on word stress have been carried out relatively recently. Much of the quantitative research has experimentally tested impressionistic claims originating from a limited number of grammar sketches with sometimes little coverage of phonology. The latter type of work is therefore excluded from the current overview. The reader is referred to Odé (1994) for an extensive review of stress claims from mainly non-experimental work on Indonesian languages. The overview in this study is furthermore limited to nine languages in order to obtain a linguistically diverse and yet relevant impression of the research. Language diversity is revealed by the diverging results in the studies and the different areas where the languages are spoken. The relevance of the selected languages for the current study is shown by the inclusion of several (Trade) Malay languages (Paauw 2009). Papuan Malay, investigated in the current study, belongs to this language group. It is to date one of the few Indonesian languages for which word stress has been experimentally studied for its acoustic realization, perception and function. Note that Papuan Malay is a regional language spoken in the provinces Papua and West Papua, across different major urban areas (Kluge 2017: xxiv), which are the home of many smaller local languages as well. The research conducted on this language so far involved participants from the Sarmi region (Kluge 2017; Chapter 4), Manokwari (Chapter 2, Chapter 3, Kaland & Baumann 2020) and Sentani (this study). Although it should be noted that these regions exhibit dialectal differences, Papuan Malay is a distinct language due to its "structural uniqueness, limited or nonexistent inherent intelligibility, and the lack of shared ethnolinguistic identity with other Malay varieties" (Kluge 2017: 9). In addition, studies on word stress conducted so far show a considerable amount of consistency in their results (see discussion in Section 5.2).

The overview furthermore includes Javanese accented Indonesian, Toba Batak accented Indonesian, Betawi Malay, Besemah, Ambonese Malay, Manado Malay, as well as two related standard languages: Indonesian (as spoken in Jakarta) and Malay (as spoken in Malaysia). It should be noted that the standard varieties included in this study are distinguished from regional varieties, which appeared a crucial distinction in earlier stress research (e.g. Goedemans & van Zanten 2007). Thus, Javanese and Toba Batak accented Indonesian refer to Standard Indonesian as spoken by speakers with either Javanese or Toba Batak as their first language. Note that although Betawi Malay is spoken in Jakarta, it is distinguished from Standard (Jakartan) Indonesian because it is spoken by a homogeneous group (the Betawi; see also van Heuven et al. 2008). The overview furthermore includes Manado Malay, for which only minimal experimental data on stress was reported (Stoel 2005; Stoel 2007). This language is still included to complement the other Trade Malay varieties (Ambonese Malay and Papuan Malay), and because its phonological description is elaborate and has been carried out systematically.

The literature overview is structured according to the three aspects of word stress discussed in Section 5.1 (acoustic realisation, perception, function). Not all aspects are covered in the available literature, given the current lack of research. Note that word stress in this overview is separated from phrase prosodic events such as pitch accents and boundary tones. Some studies have made claims on word stress based on phrase prosodic data only (Odé 1994 for Standard Indonesian; Riesberg et al. 2018; Riesberg et al. 2020 for Papuan Malay) and are therefore excluded from this overview. The importance of separating word-level from phrase-level prosody in Malayo-Polynesian languages of South East Asia has been pointed out in recent work (Kaufman & Himmelmann 2024: 12), in particular because the absence of phrase-level pitch accents does not imply the absence of word stress (e.g. Gordon 2014; Lindström & Remijsen 2005). Table 5.1 gives a schematic overview of the three stress aspects and lists additional information about the reported stress distribution and vowel inventories for each language. The stress distributions are notated following the coding system in the StressTyp database (Goedemans et al. 2020; codes explained in Hulst et al. 2010). The distributions relevant for the current study concern penultimate (P) and ultimate (U) stress, with P being the dominant pattern on heavy syllables (P/U), or with some variability between the two positions (P;U), being either lexically contrastive (LEX) or not, or no main stress at all (NMS). Note that the StressTyp codes were derived from the available reports in the reviewed literature, sometimes lacking precise descriptions of the stress distribution. Figure 5.1 shows the geographical area for each language in the overview.

Figure 5.1: Map showing provinces and major governmental districts of Indonesia (filled dots) and Malaysia (open dots), and the geographical location (coordinates from Glottolog, Hammarström et al. 2021) of the nine languages in the overview (red triangles). Created using the `maps` package (Becker et al. 2003).

Table 5.1: Schematic overview of stress research for nine Indonesian/-Malay languages. **ISO** lists the ISO 639 three letter language codes (non-Malay L1 in italics). **Vowel** inventories are taken from the experimental studies (alternatively from the most recent literature source that provides explicit listing). **Distributions** are as reported in the reviewed literature and notated using StressTyp codes (Goedemans et al. 2020; Hulst et al. 2010), with claims unconfirmed by acoustic research between brackets. **Acoustic correlates** list d(uration), i(ntensity), v(owel quality), s(pectral tilt) and/or F0 if measured and found as correlate (> indicating strength differences). **Perception** research summarised per stress location (and acoustic cues tested, if any). **Functions** are listed with regard to the existence of minimal pairs and (potential for) word identification. Shaded cells indicate lack of research.

Language	ISO	V inventory	Distribution	Acoustic corr.	Perception	Function
Std. Indonesian	ind	/i,e,a,o,u,ə/	P(/U)	F0/i > d	pref. for P if heavy (F0)	no m. pairs, no w. ident.
Std. Malay	zsm	/i,e,a,o,u,ə/	NMS	-		
Javanese acc. Ind.	*jav*	/i,e,a,o,u,ə/	NMS (P/U)	-	no pref. for P or U (F0)	
Betawi M.	bew	/i,e,a,o,u,ə/	NMS	-		
Toba Batak acc. Ind.	*bbc*	/i,e,a,o,u/	P(;U LEX)	d/i > F0	pref. for P (d/i/F0)	m. pairs
Besemah	pse	/i,a,u/	U	F0/i > d		
Ambonese M.	abs	/i,e,a,a^c,o,u/	NMS (P;U LEX)	- > s		no m. pairs
Manado M.	xmm	/i,e,a,o,u,ə/	(P;U LEX)			m. pairs
Papuan M.	pmy	/i,e,a,o,u/	P/P	d/v/s > F0	sens. for U over P stress	no m. pairs, potential w. ident.

5.2.1 Acoustic realisation

After a series of impressionistic claims on word stress in Standard Indonesian (Odé 1994 for an overview), Laksman (1994) found F0 to be the strongest stress correlate, on the basis of data from a single speaker from Jakarta. This study concluded that stress always falls on the penultimate syllable (P) and that schwa in that position can be stressed as any other vowel. This claim does not appear to hold in basic descriptions of the IPA where schwa is part of the vowel inventory and causes stress to shift to the ultimate syllable (P/U; Soderberg & Olson 2008).

Acoustic work on Standard Malay agrees on the lack of word stress, as duration, intensity and F0 did not show effects of stress (Mohd Don et al. 2008; Wan 2012). Given the lack of stable F0 alignment with syllables, Mohd Don et al. (2008) concluded that the syllable is not a relevant unit in the prosody of Malay (see Odé 1994: 63 for a similar conclusion based on the perception of phrase prosody in Standard Indonesian). It is important to note that neither of the studies on Standard Malay systematically compared alleged stressed syllables with alleged unstressed syllables, thus lacking the test of a specific stress hypothesis.

For Javanese accented Indonesian, Goedemans & van Zanten (2007) found no stress related differences in duration and intensity that could constitute evidence for the P/U stress claim (duration results confirmed the ones in van Heuven & van Zanten 1997). F0 showed a rise on the penultimate syllable, which was claimed to be a pre-boundary phrase prosodic phenomenon.

Betawi Malay (vowel inventory in Ikranagara 1975) was investigated for F0 using words obtained in and out of focus in phrase final and phrase medial positions (van Heuven et al. 2008). No systematic F0 alignment with allegedly stressed penultimate syllables was found, but rather a large degree of variability in F0 movements. This result led to the conclusion that stress is absent in Betawi Malay. Note that correlates which have been reported traditionally as stronger indicators of word stress (duration, intensity or spectral tilt) were not investigated for this language.

Support for penultimate stress (P) was found for Toba Batak accented Indonesian in duration and intensity (Goedemans & van Zanten 2007, see also van Heuven & van Zanten 1997). F0 was also measured and correlated with focus rather than with word stress (i.e. increased F0 in focus compared to non-focus).

Besemah (also Pasemah or Central Malay) is a Malay variety with a different stress distribution compared to the other languages in this overview. The strongest correlates were F0 and intensity, supporting the claim that stress is always ultimate in this language (McDonnell 2016). Duration showed minimal effects due to its co-occurrence with final lengthening.

From the Trade Malay varieties, Ambonese Malay was shown to lack word stress (Maskikit-Essed & Gussenhoven 2016), counter to the P;U (LEX) claim in van Minde (1997). A small effect of spectral tilt could be found, but no systematic acoustic differences due to stress in duration or F0. The re-analysis of Ambonese Malay stress also led to a different claim with regard to its vowel inventory. Given the lack of acoustic effects, alleged stress differences in van Minde (1997) were re-analysed as segmental differences between /a/ and a slightly raised/centralised

/a/, which was termed ‘a-caduc’ (a^c^). Note that both /a/ variants have highly similar vowel quality.

The analysis of Manado Malay concerns an impressionistic interpretation of elicited material (Stoel 2005) and led to the claim of regular penultimate stress. An elicitation task was carried out to obtain additional impressions on variable stress; words that were sometimes produced with P stress and sometimes with U stress (i.e., P;U, Stoel 2005: 16).

A series of duration, intensity, spectral tilt, F0 and vowel quality measures taken from spontaneous Papuan Malay data revealed that duration, formant displacement (vowel quality) and spectral tilt were the strongest stress correlates (Chapter 2). F0 alignment correlated strongly with word stress, although the direction of the effects were different for P and U stress, which could be explained as originating from duration differences. Importantly, F0 excursion size was among the weakest stress correlates. Overall, these results confirmed the impressionistic claim in Kluge (2017) and showed that Papuan Malay word stress falls on the penultimate by default, shifts to ultimate when /ɛ/ is in the penultimate, and remains penultimate when the final two syllables have /ɛ/ (P/P; see also Kaland, Himmelmann, et al. 2019).

Before turning to the perception and functions of word stress, some remarks should be made on the relevance of acoustic verification of stress claims. The literature has claimed that stress distributions in Indonesian/Malay languages follow a geographical division (Prentice 1994). That is, in the Western part (Borneo, Sumatra and the Malay Peninsula) P stress can be realised on schwa, e.g. /ˈdɛ.ŋar/ (‘to hear’), whereas in the Eastern part (Java, Celebes and the islands to the east thereof) schwa in the penultimate syllable causes stress to move to the ultimate syllable (P/U distribution), /dɛ.ˈŋar/ (‘to hear’). It should be noted that the claim that schwa can be stressed in Standard Indonesian as spoken in Jakarta was supported in Laksman (1994). Geographical divisions have been widely adopted in typological accounts of prosody in Austronesian languages (e.g. Goedemans & van Zanten 2014; Kaufman & Himmelmann 2024). The relatively small overview of languages discussed here already shows that geographical generalisations on word stress do not always hold. Acoustic measures supported the P/U distribution in Toba Batak accented Indonesian (Goedemans & van Zanten 2007), even though it would classify as a Western variant according to a purely geographical division. And for Ambonese Malay, an Eastern variant, the traditional P/U claim did not find any acoustic support (Maskikit-Essed & Gussenhoven 2016). Another closely related factor that has been claimed to affect the stress distributions in Malay languages is the presence of schwa in the vowel inventory. Trade Malay languages (except Larantuka Malay spoken in East-Flores) were reported to have lost schwa, which should have led to the development of word stress (Paauw 2009). Originally, though, the lack of stress was hypothesised to be a feature of all Trade Malay varieties (Goedemans & van Zanten 2014). Again, the overview presented here shows that the schwa-claim does not hold for Ambonese Malay, which has no schwa in the inventory and is analysed as stressless (Maskikit-Essed & Gussenhoven 2016). Manado Malay, claimed to have both stress and schwa (Stoel 2005; Stoel 2007) also counters this generalisation. Whether these languages are midway in the development of acquiring word stress remains – at the moment - fruitless speculation due to the lack of empirical data. Nevertheless, it is clear that more regional variation among these languages could be expected when more detailed prosodic investigations are carried out. It seems that the grouping of languages according to geographical location or according to shared traits as currently done in typological accounts are often too course grained for this area. This observation should not come as a surprise, given the vast archipelago where these languages are spoken and given the lack of research on prosody in this area. The importance of quantitative research on stress claims will furthermore show from the next sections on the perception and functions of word stress, which shed a new light on some of the acoustic results.

5.2.2 Perception

In van van Zanten & van Heuven (2004) Indonesian recordings of three trisyllabic target words embedded in a carrier sentence were manipulated for F0. That is, the position and shape of the F0 movement on the target word (a rise-fall) was varied systematically between the three syllables, such that the different onsets of the rise and fall generate six positions and twelve shapes. Listeners were presented the manipulated stimuli and indicated which syllable they perceived as stressed. Results showed significantly more indications for the penultimate syllable (compared to other syllables) as being stressed for one of the three target words (anaknya, /a.ˈnak.ɲa/, ‘his child’). This word was the only one among the target words with a heavy (closed) penultimate syllable, plausibly attracting acoustic prominence. Overall, most stress positions were acceptable to listeners, which was taken as an indication that stress is not bound to a specific syllable (i.e. free stress) and therefore not present in Indonesian. It should be noted that other cues than F0 were not tested perceptually in this study.

Javanese and Toba Batak accented Indonesian were both tested in the same perception study (Goedemans & van Zanten 2007). Listeners chose the preferred word from a pair of acoustically manipulated words (comparison task) and rated the acceptability of these words (acceptability task). The words were embedded in a phrase. Manipulation concerned the position of the stressed syllable using the relevant acoustic correlates as found in each language, which was the alignment of the F0 fall for Javanese and F0, duration and intensity for Toba Batak. All stimuli were presented to both Javanese and Toba Batak listeners. Results were similar for the comparison task and the acceptability task. They showed that Javanese listeners accepted different locations of the stressed syllable, in particular the final two syllables. Toba Batak listeners, however, had a clear preference for stress on the penultimate syllable. The results were interpreted as an indication for the presence of word stress in Toba Batak accented Indonesian and for the absence of word stress in Javanese. The study furthermore showed the importance of taking regional variation into account.

For Papuan Malay, listeners were presented a carrier phrase in which a disyllabic target word was replaced by an acoustically manipulated sequence of hummed speech (Chapter 3). Duration, F0, intensity and spectral tilt were manipulated such that either the first or second syllable stood out as more prominent. Listeners chose one out of two words that matched the manipulated sequence best. One word had alleged P stress, the other word had alleged U stress. Results showed overall low correctness scores, although F0 predicted the outcomes best. In a subsequent perception experiment with the same stimuli presented in isolation, the effect of F0 disappeared. Although no manipulated cue alone was strong enough to affect listeners choices, the outcomes of both experiments showed overall higher correctness scores for U stress (above chance level) than for P stress. It was therefore concluded that listeners were sensitive mainly to the irregular stress pattern in Papuan Malay.

5.2.3 Function

The lexically contrastive function of word stress was reported for the Toba Batak language (Nababan 1981: 23) in minimal pairs such as /ˈi.tɔm/ (‘black dye’) and /i.ˈtɔm/ (‘your brother/sister’). These descriptions would fit a P;U (LEX) stress claim. Although minimal stress pairs have not been directly investigated in later acoustic studies on Toba Batak and Toba Batak accented Indonesian (Roosman 2006; Goedemans & van Zanten 2007), it is possible that this function exists in this variant as well, given the acoustic support for a P distribution in these studies. However, these results would need further support given that Goedemans & van Zanten (2007) investigated the production by one speaker only. The situation is different from the one for Standard Indonesian and Besemah, for which acoustic support mainly confirmed a P or U stress distribution respectively (Laksman

1994; McDonnell 2016). Thus without alternating patterns, no contrastive function can be attributed to word stress in these languages. As for the Trade Malay varieties in the overview, each of them shows a different state concerning the contrastive function. For Ambonese Malay minimal pairs were reported in van Minde (1997), but rejected in Maskikit-Essed & Gussenhoven (2016). For Manado Malay minimal pairs such as /ˈla.la/ (girl's name) and /la.ˈla/ ('tired'), as well as /ˈse.pi/ ('edge', 'brink') and /se.ˈpi/ ('silent', 'lonely') were reported without acoustic support (Stoel 2005; Prentice 1994). For Papuan Malay no minimal pairs were reported (Kluge 2017).

As for word identification, a gating task (van Zanten & van Heuven 1998) presented Indonesian listeners with (parts of) words that were identical in segmental content and supposedly different with respect to the location of the stressed syllable. That is, word triplets with alleged stress on the first, second or third syllable (e.g. /ˈa.nak/, /a.ˈnak.ɲa/, /a.nak-ˈa.nak/) were presented in such a way that either only the first syllable (gate 1) or only the first and second syllable (gate 2) were audible. The listeners' task was to identify one out of the three words after hearing each gate embedded in a carrier sentence (presented in order of increasing gate length). The hypothesis tested in this experiment was whether stress cues in Standard Indonesian help listeners to identify words. If so, they should identify the words that matched in stress location with the gates, despite the ambiguity in their segmental content. Only one of the six Indonesian listeners correctly identified the words above chance level (only for gate 2), indicating that stress had no function in word identification. An acoustic analysis of the stimulus material for F0 and duration revealed that falling F0 in either the first or second gate predicted the alleged stress location best. The same gating task was done with Dutch listeners (with basic command of Indonesian) and revealed above chance level scores for correct word identification for nearly all participants. It was concluded that despite the presence of acoustic stress correlates in the signal, they were useless for word identification by Indonesian listeners. It should be noted that the Indonesian participants in this study (the speaker of the stimulus material and the listeners in the gating task) had different first languages (Balinese, Sundanese and Javanese). Although they were reported to be proficient speakers of Standard Indonesian, it remains unclear whether language background had an effect on the materials and/or results.

In an online word processing task (Chapter 3) it was found that Papuan Malay listeners identified disyllabic words faster when the initial syllable was stressed (P stress) compared to when that syllable was unstressed (U stress). The effect could have been partially caused by a generic processing benefit for word-initial syllables, although the predominance of P stress in Papuan Malay makes it difficult to disentangle a generic effect from one that is exclusively related to word stress. Another study showed that the number of lexical embeddings (see also Section 5.1.3) is reduced when taking into account stress information in Papuan Malay (Chapter 4). This would mean that listeners can successfully reject activated word candidates during processing, if they make use of the stress cues. Although these studies show that Papuan Malay stress patterns have the potential of facilitating word identification, direct experimental evidence is still needed to corroborate these findings.

5.2.4 Current study: implications and research questions

In the introduction, three main aspects of word stress were distinguished: acoustic realisation, perception and functions. The studies discussed in this section show that perception experiments have led to a better understanding of word stress in Indonesian and Malay than could be achieved on the basis of acoustic studies alone. In fact, for Standard Indonesian and Javanese accented Indonesian the conclusion that word stress is absent in these languages was mainly based on the results of perception studies. Due to the crucial refinement (or countering) of previous claims in the literature, these perception studies have been described as 'resolving' the discussion on Indonesian stress (e.g. van Zanten et al. 2010: 101). However, on the basis of the current overview a few more remarks should be made to illustrate the extent to which perception studies complement the acoustic studies in diagnosing word stress.

There is a crucial difference between the work on Standard Indonesian and Malay spoken by the Javanese in that studies on the former mainly investigated F0 in production and perception studies (Laksman 1994; van Zanten & van Heuven 1998; but duration is measured for the single speaker in van Zanten & van Heuven 2004). Thus, it remains unclear what exactly the effect of duration, intensity or any other possible correlate is in the production and perception of Standard Indonesian word stress, in particular because F0 could be a weak correlate (Section 5.1.1). The lack of acoustic investigations also casts doubt on what perception studies have resolved precisely, as the latter considered only F0 as well. An additional complicating factor is defining what counts as Standard Indonesian given the influence of regional languages (e.g. van Zanten & van Heuven 1984; van Heuven & van Zanten 1997). For Malay spoken by the Javanese, the state of the research is crucially different, as the conclusion that it lacks word stress was based on the acoustic measures of multiple correlates (duration, intensity and F0) and all of these were taken into account for the design of the perception experiment (Goedemans & van Zanten 2007).

It should be noted that there is no reason to reject the conclusions on the absence of word stress in Standard Indonesian based on the overview presented here. However, it becomes clear that for perception studies to maximally contribute to the word stress question, these studies crucially depend on the available knowledge of the acoustic correlates. It can be questioned whether perception studies alone can be decisive to diagnose word stress. This is an important issue, in particular given the central question in many studies of whether or not the language has stress. Studies generally take acoustic analyses as decisive means to diagnose word stress (Gordon & Roettger 2017). These acoustic studies are much more frequent than perception studies and have been taken as the basis for claims on the existence of word stress for many languages, including controversial as well as uncontroversial ones (Gordon & Roettger 2017). Perception studies, on the other hand, primarily describe the (in)ability of listeners to perceive or meaningfully use the available cues. These studies reveal crucial differences in the (functional) contribution of stress distributions to speech perception across languages (e.g. Peperkamp et al. 2010), rather than a single decisive diagnostic of whether a certain stress distribution is present in the (produced) language or not.

The observations above ask for a recall of the perception results on Standard Indonesian (van Zanten & van Heuven 1998). Listeners in this study did not pick up the acoustic cues to word stress that were present in the speech signal. This is not the only study that revealed a discrepancy between the correlates in the signal and the cues listeners attend to. This outcome reflects what has been shown in comparative studies on Dutch and English. English listeners detect stressed syllables mainly on the basis of (segmental) vowel quality differences, whereas Dutch listeners use mainly suprasegmental cues (duration, spectral tilt; Sluijter et al. 1997) to do so. Interestingly, suprasegmental stress correlates are generally present in the English speech signal. As a consequence, Dutch listeners were shown to outperform English listeners in the detection of English stressed syllables (Cooper et al. 2002; Cutler et al. 2007). These studies show, therefore, that stress perception is more intricate than listeners processing whatever the speech signal has to offer. It rather seems that listeners attend to what they need to process the signal. In English, listeners attend mainly to vowel quality differences such as in the pair 'subject' (noun) and 'to subject' (verb), as these are generally sufficient to distinguish lexical meanings. Crucially, these studies also show that non-native listeners might pick up acoustic cues as stress, even though native speakers do not. This issue makes a strong case for doing empirical work on word stress (both production and perception), rather than reporting potentially misleading auditory impressions, shaped by mainly the researchers' native language (see Odé 1994 and Chapter 2, for a discussion of this issue in Indonesian and Trade Malay stress research respectively). And even though stress patterns in Indonesian and Malay might be present, they are generally highly predictable (Table 5.1) and therefore have little or no function in distinguishing lexical meanings (e.g., small number or no minimal stress pairs). The functional load of stress parameters is therefore small and listeners have little need to attend to specific acoustic cues. This is exactly the situation reported for Standard Indonesian, which was analysed as a language without stress; Dutch listeners (with command of Standard Indonesian) were able to detect stress differences in the same stimuli for which native Standard Indonesian listeners did not (van Zanten & van Heuven 1998).

A central issue in the current study is therefore whether a language could still be analysed as having stress when listeners do not use its cues. This question applies in particular to Papuan Malay. The acoustic correlates of word stress in this language (Chapter 2) do not warrant that listeners use them, in particular given their low functional load. If Papuan Malay listeners appear to not use these cues for word identification, the challenging question remains whether the acoustic patterns in Chapter 2 can still be interpreted as word stress or not (see also the discussion in Section 5.1.4). While this question cannot be answered at this stage, it is an important one that could challenge the cross-linguistic definition of word stress. The presence of correlates in Indonesian (van Zanten & van Heuven 1998) was not enough to analyse this language as having word stress. The acoustic research on Papuan Malay, however, makes a stronger case for word stress as signalled by multiple cues simultaneously and showing at least some auditory sensitivity in listeners (Chapter 3). Although these results could be taken as indication that listeners will use them (H1), the possibility that they do not should be taken into account (H0) and might apply to other (related) languages. It should also be noted that posing the word stress question as a binary one has its limitations. That is, for investigating how meaningful word stress patterns are in a particular language, perception studies are indispensable (see also Section 5.1.2 and 5.2.2), and are likely to show different degrees of listener sensitivity for stress patterns (Peperkamp et al. 2010). Lexical analyses have shown that word stress information can help to disambiguate competing lexical candidates (Chapter 4). Thus, whether listeners actually make use of the available stress cues if the segments are ambiguous remains to be tested directly. This research question is answered in the current chapter and the hypotheses are formulated in the following:

H0: Papuan Malay listeners do not use stress cues to identify words.

H1: Papuan Malay listeners use stress cues to identify words.

To investigate the research question, the current study reports an acoustic investigation, and gating task similar to the one in van Zanten & van Heuven (1998). The next section outlines the methodology of these investigations.

5.3 Methodology

In order to investigate the extent to which stress parameters contribute to word identification a forced choice gating task was carried out. In this task, participants identified one member from a pair of disyllabic words. Each word in the pair was presented in gated fashion (Cotton & Grosjean 1984) in final position in the Papuan Malay matrix phrase *Sa blum taw ko pu kata itu, kata [word]* (I don't yet know that word of yours, the word [word]). Although phrase-final words are affected by phrase prosodic phenomena such as boundary marking (Kaland & Baumann 2020), the choice for these materials was motivated in two ways. First, co-articulatory cues from neighbouring words would have affected phrase-medial words more (both word edges) than phrase-final words (left word edge only), potentially reducing the quality of the gates. Second, the availability of phrase-medial words that would fit the design of the experiment was limited. Phrase-final words were generally more clearly articulated due to their position. The next section provides further details on the stimulus material, including the presence of co-articulation cues and the design.

5.3.1 Material preparation and design

The phrase-final words were taken including their original matrix phrase from a corpus of recordings (Kluge et al. 2014). The recordings differed in recording quality, which were largely overcome in two phases of audio processing. First, noise reduction was applied per wave file in the corpus. This was done using the Noise Reduction function in Audacity (Audacity Team 2019). Using this function, a profile of the background noise in a part of the recording where there was no speech (e.g. a silent pause) was generated. Thereafter, noise reduction based on that profile was applied across the entire wave file. Second, the intensity of all noise-reduced recordings was scaled using Praat (Boersma & Weenink 2019) such that the average intensity in each recording was 70 dB SPL.

From the word list that constitutes the corpus (Kluge et al. 2014; Kluge 2017) word pairs were created such that the words in a pair had identical segmental composition either the first or the second syllable (henceforth matching syllable). Crucially, the matching syllable was the stressed one in one word and the unstressed one in the other word. For example, the first syllables match in the pair [ˈbɛ.bɛk] (duck) and [bɛ.ˈban] (burden) and the second syllables match in the pair [ˈam.pas] (waste) and [lɛ.ˈpas] (to free). In total, 32 word pairs were created, of which half had a matching first syllable and half had a matching second syllable (Table 5.2). Then, three gates per word were created (Table 5.3). The first gate consisted of the edge segment (ES) of the word; either the first segment (matching first syllable) or the last segment (matching second syllable). Note that the word pairs where chosen in such a way that the es gate always consisted of a consonant, which reveals little or no acoustic stress cues compared to vowels. The second gate consisted of the matching syllable (MS, either the first or the second). The third gate consisted of the entire word (EW).

For the ES and MS gates, the part of the word that was not present in the gate was masked with white noise, such that the position of the gate and the word duration could still be identified. The white noise was added with 30% of the RMS amplitude of the original part of the word. This value was chosen to obtain a white noise intensity that fell well below the intensity of the original speech, such that the unmasked part of the word was still audible. The segment or syllable boundaries in the gates were drawn on the basis of auditory and visual (spectral) information. These boundaries were drawn such that there was only minimal co-articulatory cues in the gate. Nevertheless, the presence of co-articulatory cues could not be entirely avoided. In this respect, it should be noted that the segment directly adjacent to the matching syllable could have been audible in the MS-gate, causing a facilitation effect. To assess this potential facilitation effect, each word pair was categorised for whether or not the segment directly adjacent to the matching syllable was identical for both words in the pair (no co-articulatory cues; e.g. [ˈbɛ.bɛk] vs. [bɛ.ˈban]) or different (possible co-articulatory cues present). Note that the extent to which co-articulatory cues could actually facilitate recognition in the MS-gate varied depending on the type of articulatory difference between the segments directly adjacent to the MS-gate. For example, it can be expected that co-articulatory cues have a smaller effect, if at all present, in e.g. [ˈbɛn.tʃɔŋ] vs. [bɛn.ˈtʊk] than in [ˈtʃɛ.bɔ] vs. [tʃɛ.ˈɾ ɛj]), see Table 5.2.

For each word that needed to be identified (target) participants chose which of the two gates (each corresponding to one member of a word pair) matched with the target. The rationale behind presenting both gates in a forced choice manner was to make the relevant acoustic differences between stressed and unstressed

Table 5.2: Overview of the word pairs, indicating the matching syllable (σ), stress location (penultimate, ultimate), English gloss and the possible presence (yes/no) of co-articulatory cues in the matching syllable due to non-identical segments directly adjacent to the matching syllable.

Pair no.	Match.σ	Penultimate stress		Ultimate stress		Co-a.cue
1	1	bebek	*duck*	bebang	*burden*	n
2	1	bencong	*transvestite*	bentuk	*form*	y
3	1	cebo	*to wash after defecating*	cerey	*divorce*	y
4	1	gedi	*aibika*	gedung	*building*	n
5	1	gepe	*to clamp*	geli	*tickle*	y
6	1	kewa	*dance party*	kebung	*garden*	y
7	1	kempes	*to be deflated*	kembar	*twin*	y
8	1	leher	*neck*	lega	*to be relieved*	y
9	1	lembek	*to be soft*	lembar	*sheet*	n
10	1	pecek	*mud*	pesang	*order*	y
11	1	pendek	*to be short*	penting	*be important*	y
12	1	seher	*piston*	segar	*be fresh*	y
13	1	seneng	*to sign*	senyum	*smile*	y
14	1	tete	*grandfather*	tebang	*fell*	y
15	1	tempe	*tempeh*	tempat	*place*	n
16	1	tepu	*to clap*	tepung	*flour*	n
1	2	ampas	*to waste*	lepas	*to free*	y
2	2	bantal	*pillow*	kental	*to be fluent*	n
3	2	nakal	*to be mischievous*	bekal	*to be equipped*	y
4	2	pinang	*betel nut*	bernang	*to swim*	y
5	2	bintang	*star*	ketang	*crab*	y
6	2	cincang	*to chop up*	kencang	*to be speedy*	n
7	2	watak	*character*	gertak	*to intimidate*	y
8	2	glombang	*wave*	kembang	*flower*	n
9	2	gugur	*to fall (prematurely)*	tegur	*to reprimand*	y
10	2	rangkap	*to be doubled*	lengkap	*to be complete*	n
11	2	sidang	*meeting*	pedang	*sword*	y
12	2	rusak	*to be damaged*	sesak	*to be crowded*	y
13	2	sabit	*sickle*	terbit	*to rise*	y
14	2	semang	*outrigger*	temang	*friend*	n
15	2	tandang	*banana plant stem*	tendang	*to kick*	n
16	2	tukang	*craftsman*	tekang	*to press*	y

Table 5.3: Examples of the three gates from two word pairs that match either for their first or second syllable.

Matching syllable	Gating direction	Gates		
		edge segment (ES)	matching syllable (MS)	entire word (EW)
First	Forward	[b]	[ˈbɛ]	[ˈbɛ.b ɛk]
		[b]	[bɛ]	[bɛ.ˈban]
Second	Backward	[s]	[pas]	[ˈam.pas]
		[s]	[ˈpas]	[lɛ.ˈpas]

syllables available for participants (in the MS- and EW-gates). Forced choice gating tasks have been applied in previous research on word recognition (e.g. Davis et al. 2002). It is expected that little to no stress information is stored lexically for languages with highly predictable stress (Peperkamp et al. 2010). Papuan Malay has a small number of exceptional stress patterns that could increase the need for lexical storage of stress information. The availability of the relevant acoustic differences in the experiment ensured that participants, if able to use the cues, could map them readily onto their lexically stored knowledge of the target word.

Due to the forced-choice nature of the gating task participants had a 50% chance of selecting the correct member. Given the segmental material present in the three gates, it is expected that the ES-gates elicit a chance level response, as little to no unique acoustic cue is provided to identify the correct word. As for the MS-gate, participants are expected to score above chance level when they successfully use the acoustic cues to stress to identify the word and around chance level when they do not. As for the EW-gate, responses are expected to yield a 100% correct score as all (supra)segmental material is present to identify the correct word.

The three successive gates were presented in triplets in the order described above (ES–MS–EW). In this way, word pairs with a matching first syllable were forward gated, whereas word pairs with a matching second syllable were backward gated. Backward gating has been applied in previous research on word recognition (Salasoo & Pisoni 1985; Wingfield et al. 1997), indicating that listeners can successfully identify words based on their final segment(s). Both gating directions were applied in the current experiment because the Papuan Malay word pairs with a matching first syllable all consisted of /ɛ/ in that syllable. Thus, to be able to investigate the stress parameters provided by the other Papuan Malay vowels, word pairs that matched for the second syllable were included (Kluge 2017). The other vowels occurring in the matching second syllables were /a/, /i/

and /u/. Although /ɔ/ also belongs to the Papuan Malay vowel inventory, it did not occur frequently enough in the word lists (Kluge 2017) to include in word pairs in the current paradigm.

5.3.2 Acoustic measurements

The matching syllables of the 64 words that were selected for the word pairs (described above) were then acoustically measured in order to confirm that they differed in the presence of stress parameters. The measures were taken from the matching syllables only, in order to assure identical segmental composition of the stressed and unstressed syllables. Four acoustic correlates were measured in the first or second syllable of each word: duration, F0, spectral tilt (H1*-A2*) and vowel quality (F1/F2). These measures were chosen as they appeared the strongest correlates of Papuan Malay word stress in Chapter 2. F0 did not correlate strongly with word stress in Chapter 2, but showed effects of alignment with the stressed syllable (also Kaland & Baumann 2020) and showed effects on stress perception in a phrase-context similar to the one in the current study (Chapter 3). Duration of the entire syllable was measured in milliseconds. Given that segmental composition was identical across stressed and unstressed syllables, no further conversion of the duration values was applied. F0 was measured in semitones as the mean per syllable. Measures of vowel formants (F1/F2) and the first harmonic (H1) were taken to compute the spectral tilt and vowel quality values respectively. These were measured in a subinterval of the syllable. This subinterval was set around the intensity peak of the syllable, where stable formant trajectories were found. The boundaries of the subinterval were set on either side at the points where the intensity level (measured in dB) had dropped 4% relative to the peak intensity. Spectral magnitude correction (Hanson 1997: 113–115; Iseli et al. 2007) was applied to the spectral tilt measure. Note that the measurement methods were identical to the ones used in Chapter 2. The vowel nucleus of the matching syllable was exclusively /ɛ/ in first syllables and almost exclusively /a/ in second syllables (/i/ and /u/ appeared once each in matching second syllables, see also Table 5.2). For this reason, only the formant measures of /ɛ/ and /a/ in their respective syllable positions are reported here (Table 5.4).

Statistical tests on the acoustic measures were not performed due to the small number of words in the dataset (N = 64). It can still be observed that syllables are longer when stressed, with the second syllable being generally longer than the first syllable (Table 5.4). Mean F0 is higher in the stressed syllable than in the unstressed syllable with overall higher values in the first syllable than in the second syllable, possibly due to the declination effect (Breckenridge 1977). The spectral

Table 5.4: Mean values for each of the acoustic measures and their standard deviations in each syllable position ($\sigma 1$, $\sigma 2$) and stress position (penultimate, ultimate). Values in bold indicate stressed syllables.

Acoustic measure	Stress position	$\sigma 1$	$\sigma 2$
Duration (ms)	Penultimate	**275.56 (38.64)**	414.48 (94.89)
	Ultimate	203.83 (50.84)	**512.08 (96.06)**
H1*-A2* (dB)	Penultimate	**3.13 (9.27)**	4.09 (8.67)
	Ultimate	6.99 (10.83)	**4.09 (6.94)**
F0 (ST)	Penultimate	**86.73 (1.19)**	83.62 (1.68)
	Ultimate	84.92 (0.84)	**85.46 (1.65)**
F1 (Bark)		/ɛ/	/a/
	Penultimate	**4.39 (0.59)**	5.63 (1.22)
	Ultimate	4.38 (1.10)	**5.98 (0.28)**
F2 (Bark)		/ɛ/	/a/
	Penultimate	**12.38 (0.98)**	10.41 (1.79)
	Ultimate	12.10 (1.22)	**9.93 (0.60)**

tilt is more shallow (less intensity roll-off towards the higher frequencies) for stressed syllables than for unstressed syllables. As for the formant measures, the average position of the vowel in the acoustic space as measured by the actual formant values did not vary much due to stress (Figure 5.2). Note, however, that /a/ is somewhat more peripheral when stressed, whereas the position of /ɛ/ remains almost identical in either stress condition. A clearer effect of stress can be observed in the standard deviations of the formant values of both vowels; indicating more target undershoot (larger SDs) in unstressed syllables (e.g. Lindblom 1963). Standard deviations of the spectral tilt measures and those of the duration measures in first syllables indicated a similar effect. Final lengthening, applied here on two levels (word and phrase), is likely to cause larger SDs for duration measures in the second syllable, irrespective of stress.

The acoustic results therefore confirm that duration and spectral tilt are indicative of Papuan Malay word stress. A direct effect of stress on vowel quality was mainly found for /a/, as the target undershoot effects can be interpreted as an indirect effect of the shorter durations of unstressed syllables. In sum, the results of the acoustic analysis show that stress parameters are present in the selected words for the stimulus set of the gating experiment, which is further described in the following.

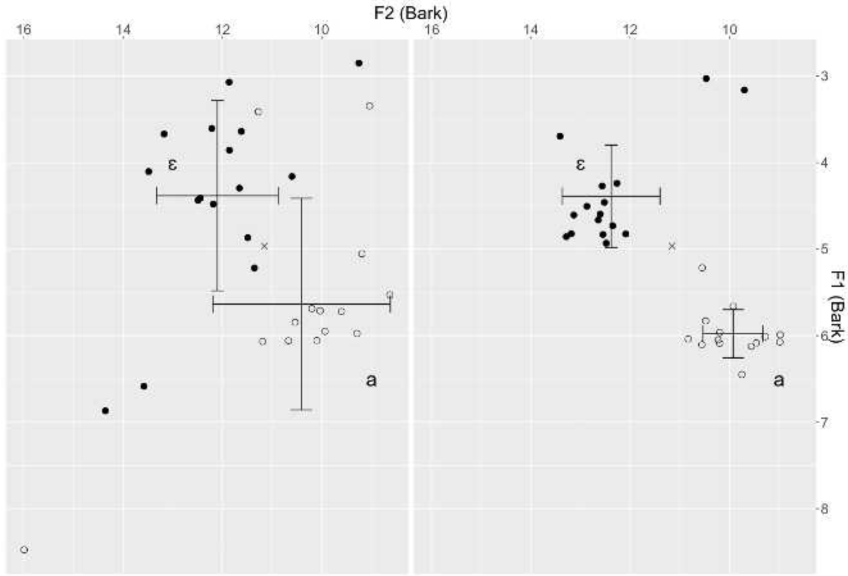

Figure 5.2: Individual values (dots) and mean F1 and F2 values (in Bark; with error bars) of unstressed (left panel) and stressed (right panel) /ɛ/ in first syllables (filled dots) and /a/ in second syllables (open dots). The cross marks the centre of the vowel space.

5.3.3 Participants

A total of 24 participants (21 female/3 male, M age: 23.40, age range: 18-35) completed the experiment. They were all native speakers of Papuan Malay without hearing problems living in the Sentani region. 11 participants reported to also speak Standard Indonesian in daily life (as native language and/or at home). All participants were remunerated for their participation.

5.3.4 Procedure

The task was designed and run online (PsyToolkit; Stoet 2010; Stoet 2017). Participants completed the task behind a laptop computer using over-ear headphones (JBL PPT-450) to listen to the stimuli. They were seated in a quiet room with limited to no background noise. For each target word the computer screen displayed a written version of the matrix phrase including that target word (Figure 5.3). The target word was chosen randomly from the word pair. In this way, correct identification of the target required participants' attention to the presence of stress parameters in the gate (matching syllable was stressed in target) for approximately half of the gates and to the absence of stress parameters in the gate (matching syllable was unstressed in target) for the other gates. Furthermore, the screen displayed two play buttons corresponding to the gates of each member in the word pair (Figure 5.3). The play buttons needed to be pressed at least once each before participants could make their choice. Each gate was presented auditorily in its original matrix phrase as recorded in Kluge et al. (2014). There was no maximum to the number of times participants could listen to a gate.

In addition, the screen displayed a visual indication of which gate in the triplet was presented (Figure 5.3). This was done by showing a rectangle directly underneath the target word. A red colored filling of the rectangle starting from either the left side (matching first syllable) or the right side (matching second syllable) indicated which gate was presented; 10% filling (ES-gate), 50% filling (MS-gate) or 100% filling (EW-gate). After completing each triplet, participants received visual feedback on the number of correctly identified words (minimally zero and maximally three). For zero and one correctly identified word a sad looking emoticon :-(was shown, for two correctly identified words a neutral looking emoticon :-| was shown and for three correctly identified words a happy looking emoticon :-) was shown. Emoticons were shown for 2.5 seconds after completing the third gate. The visual feedback (gate indicator and emoticon) were included to guide and to retain the auditory attention of the participant. A triplet of gates corresponded to one word pair, totalling 96 gates. Triplets were selected in random order and the random order was different for each participant to balance out potential learning effects. The order of the gates within each triplet was fixed as described above (ES-MS-EW). Each triplet occurred twice per experiment to balance whether participants needed to identify a target with penultimate stress or a target with ultimate stress. On the top right corner of the screen a counter indicated the percentage of the task participants had completed (Figure 5.3). The entire experiment lasted approximately 45 minutes. Responses were collected in the online experiment environment (PsyToolkit; Stoet 2010; Stoet 2017) as correctness score, counting 1 for each correctly identified target and 0 for each incorrectly identified target.

5.3.5 Statistical analysis

Generalised linear mixed model (GLMM) analysis was performed using the *lme4* package (Bates et al. 2015) in R (R Core Team 2019, R Studio Team 2019) with *correctness score* as response, with gate (three levels: ES, MS, EW) and *gating direction* (two levels: forward, backward) each in interaction with *co-articulation cue* (two levels: present, absent) as predictors, and with random intercepts for

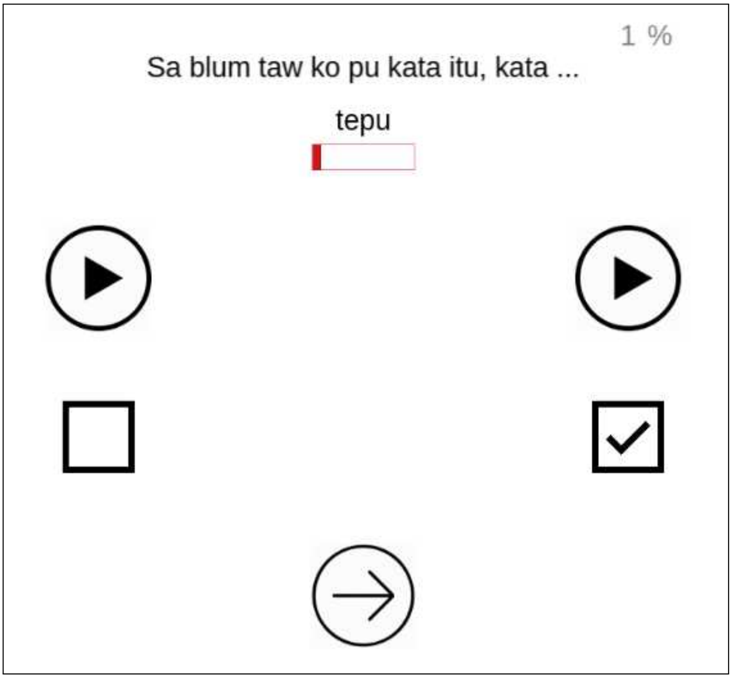

Figure 5.3: Example stimulus screen showing (from top to bottom): a percentage counter indicating the experiment progress, the matrix phrase and target word ‘tepu’, a bar indicating which part of the word was gated (here: forward ES-gate ‘t’), two play buttons and corresponding selection boxes (here: gate played with the right play button selected), and a button with an arrow to continue to the next stimulus.

participants and stimulus pair (the maximally converging model). Post-hoc pairwise comparisons using Tukey HSD test (Bonferroni corrected) were performed using the package *multcomp* (Hothorn et al. 2008) for the predictor *gate*.

Table 5.5 reports the mean correctness scores per gate, split by gating direction and the presence of co-articulation cues. Table 5.6 reports the results of the two GLMMs and pairwise comparisons. Figure 5.4 shows the mean correctness scores per gate, split by the presence of co-articulation cues in two bar charts.

Table 5.5: Mean correctness scores (standard deviations) per gate (ES, MS, EW) split by the presence of co-articulation cues and gating direction.

Co-articulation	Direction	ES	MS	EW
Present	Forward	0.50 (0.50)	0.93 (0.26)	0.99 (0.11)
	Backward	0.71 (0.45)	0.93 (0.26)	0.98 (0.14)
Absent	Forward	0.50 (0.50)	0.81 (0.39)	0.98 (0.16)
	Backward	0.64 (0.48)	0.84 (0.36)	0.97 (0.16)

Table 5.6: Results of the GLMM and pairwise comparisons on the correctness scores.

Analysis	Predictor	β	*SE*	*z*	*p*
GLMM					
	intercept	0.08	0.22	0.37	0.71
	gate:ms	1.35	0.15	9.23	< 0.001
	gate:ew	3.49	0.29	12.19	< 0.001
	co-articulation:present	0.03	0.26	0.14	n.s.
	direction:bwd	0.52	0.27	1.94	0.05
	gate:ms * co-articulation:present	0.91	0.20	4.46	< 0.001
	gate:ew * co-articulation:present	0.33	0.39	0.86	n.s.
	direction:bwd * co-articulation:present	0.14	0.34	0.42	n.s.
Pairwise comparisons					
	gate:ms-es	1.35	0.15	9.23	< 0.001
	gate:ew-es	3.50	0.29	12.19	< 0.001
	gate:ew-ms	2.15	0.29	7.29	< 0.001

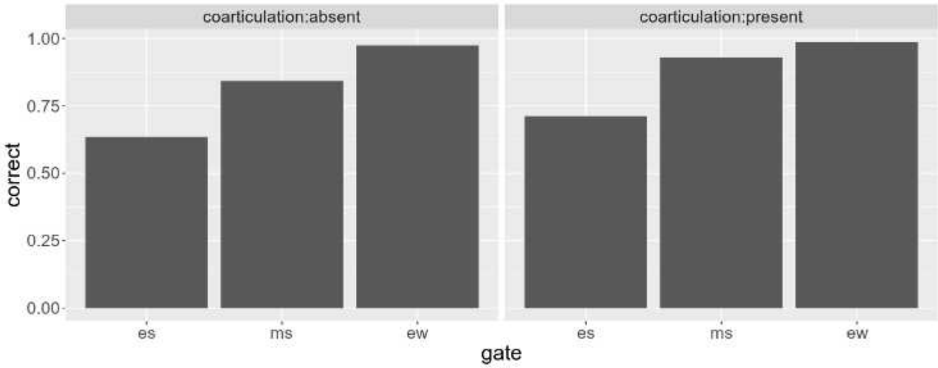

Figure 5.4: Mean correctness scores per gate (ES, MS, EW) split by the presence of co-articulation cues.

5.4 Results

The results of the GLMM shows significant effects of the predictor gate; the MS-gate elicited higher correctness scores than the ES-gate, and the EW-gate elicited higher responses than the ES-gate. The effect of gating direction showed a trend in that higher correctness scores were obtained for backward gating (i.e., when the second syllable was the matching syllable) than for forward gating. The interactions with co-articulation showed a significant effect for the MS-gate only in that higher correctness scores were obtained when co-articulation cues were present than when these were absent. The pairwise comparisons revealed that the correctness scores increased significantly with each successive gate.

5.5 Discussion and conclusion

The gating experiment shows that Papuan Malay listeners used the suprasegmental stress cues to identify words. The presence of co-articulation cues increased listeners' correctness scores for the MS-gate, indicating that listeners choices were not only affected by the stress cues. It should be noted that the correctness scores for the MS-gate was significantly lower than the correctness scores for the EW-gate (pairwise comparisons, Table 5.6). The latter scores were close to one, as expected, indicating that listeners identified the target correctly in virtually all cases in which they heard the entire word. The difference between the correctness scores of the MS-gate and the EW-gate reveals that, although stress cues facilitated word recognition, they were not sufficient to make listeners identify the target in nearly all cases (as in the EW-gate). This result can be explained by the fact that stress parameters in Papuan Malay are not lexically contrastive and have a low functional load (Section 5.2.3). Listeners are therefore used to primarily rely on segmental cues to identify words. Only for cases in which stress cues are the only ones to reveal word identity, e.g. word disambiguation such as enforced in the current task, listeners show that they are able to use them. This result crucially complements the lexical analysis in Chapter 4, in which a representative subset of the Papuan Malay lexicon showed that stress information could have a facilitating effect on word disambiguation, in addition to the segmental information. Although this facilitation is smaller in Papuan Malay than in languages with less regular stress patterns such as Spanish, the results showed that the Papuan Malay lexicon leaves room for a facilitating role of stress to similar extent as found for English (Chapter 4). The results of the current study confirm that listeners are indeed able to rely on stress cues for word disambiguation.

It is furthermore interesting to observe that the differences between the gating directions showed a trend. Table 5.5 reveals that gating direction differences were only found in the ES-gate, with higher correctness scores when the matching syllable was the second (backward gating) than when the matching syllable was the first (forward gating). This outcome indicates that listeners were better in identifying the target when listening to the final segment of that word than when listening to the first segment of that word. Identification scores were not expected to differ as a direct result of the way the gates were presented (Salasoo & Pisoni 1985; Wingfield et al. 1997), and the current design allows for alternative explanations. That is, gating direction in the current study was confound with position of the matching syllable (and therefore with stress position) and with vowel identity. That is, all forward gates concerned matching first syllables with /ɛ/ as their nucleus, whereas most backward gates concerned matching second syllables with /a/ as their nucleus (Section 5.3.1). It should be recalled that in Chapter 2 ultimate stress was realised with larger acoustic differences than penultimate stress. A possible explanation of the gating direction differences in the ES-gate could therefore lie in the stress position, which would match with the formant displacement differences observed in the stimuli (more displacement for /a/ than for /ɛ/; Table 5.4, Figure 5.2). It should also be noted that the forward ES-gate concerned a voiced consonant in 6/16 cases, whereas backward ES-gate concerned a voiced consonant in 11/16 cases. It is therefore likely that if stress realisation had a "spill-over" effect on the voiced edge segments (i.e. in the ES-gate), this effect would be larger for backward gated stimuli than for forward gated stimuli. Thus, in the current study, the acoustic cues to stress were likely to be more salient for listeners in the second (ultimate) syllable than in the first (penultimate) syllable (cf. Table 5.4).

The issue raised in Section 5.2 concerned the extent to which perception research contributes to the diagnosis of word stress in under-researched languages. The argument put forward on the basis of the literature overview is that perception studies contribute most to the word stress question when sufficient acoustic support is present (see also Table 5.1). The current study relies to a large extent on Chapter 2, reporting evidence for word stress in Papuan Malay in multiple acoustic correlates. These results were supported by the small acoustic analysis on the stimuli used in the current study. Thus, just on the basis of the speech signal, Papuan Malay could be analysed as a language with word stress. As the literature has shown for many languages, and in particular the ones spoken in Indonesia, perception studies provide a crucial insight into the functionality of the available stress parameters. On the basis of the current gating experiment it can therefore be concluded that Papuan Malay listeners are indeed able to use these cues when they don't have an alternative. Given that there is a role for Papuan Malay word stress parameters to disambiguate embedded words (Chapter 4), listeners have an incentive to use them and, given the current results, will do so. It should be noted that word disambiguation does not concern a problem Papuan Malay listeners face regularly. Embeddings are not frequent and context provides additional facilitation. As already discussed in Section 5.2.4, stress in Papuan Malay has a low functional load. This suggests that if there would be no stress cues in the Papuan Malay speech signal, it is unlikely that listeners would face perception difficulties that disrupt the communication process.

The above observations bring back the question raised in Section 5.2.4: could a language have word stress when listeners don't need to use its acoustic cues? On the basis of both Chapter 2 and the current one the answer is undoubtedly affirmative for Papuan Malay. The speech signal provides multiple stress correlates and listeners will use them to their advantage in the absence of other cues to word identification. As such, the low functional load of word stress does not justify the conclusion that this language lacks word stress. Rather, it appears to be a type of stress that requires controlled investigations to be revealed. This makes the outcomes of the current study crucially different from the ones found in a similar gating task for Standard Indonesian (van Zanten & van Heuven 1998). In that study, the stimuli also provided acoustic cues to stress, although Indonesian listeners were not able to use them functionally. The gating task in the current study was different in that participants matched one of two auditory stimuli to a single written target word, whereas in van Zanten & van Heuven (1998) participants listened to a single auditory stimulus and matched it with one of three written words. Thus, in the current study participants were presented the crucial acoustic contrast between stressed and unstressed syllables for each choice they needed to make (see also Section 5.3.4). This could have made the stress cues more salient in the current study than in van Zanten & van Heuven (1998).

Apart from methodological differences, it is important to point out the diversity among Indonesian languages (see also Section 5.2). The Trade Malay languages alone reveal two important reasons why the language diversity in Indonesia requires careful investigation and reticence in assuming overall similarities in prosodic structure. First, many empirical investigations are still lacking and do not allow firm conclusions nor generalisations on which features are unique and which are shared among all Trade Malay languages (Table 5.1). Second, the limited empirical work has already hinted at the existence of different prosodic structures among closely related languages (cf. Ambonese Malay). In this respect it is also important to reconsider that the low functional load of stress gave rise to analyses that attribute word level acoustic differences to different phonological domains. Two examples illustrate this type of analyses. For Austronesian languages in general, the alleged word stress patterns have been explained as reflexes of (intermediate) phrase prosody (Goedemans & van Zanten 2014). For Ambonese Malay in particular, re-analysis of its vowel inventory rendered the minimal differences in word pairs as segmental (Maskikit-Essed & Gussenhoven 2016) instead of suprasegmental (van Minde 1997). In impressionistic work on other Trade Malay languages (see Table 2.1 in Chapter 2 for an overview), the presence of minimal stress pairs has also been taken as the central argument in (binary) stress diagnoses. The literature has shown that the lexically contrastive function is not the only way in which word stress facilitates word identification (Cutler 2012 and Section 5.1.3). Papuan Malay appears to be a language in which stress parameters play a facilitative role for word identification despite the absence of minimal stress pairs. This observation reconfirms what perception studies have already shown: it is more fruitful to consider the presence and functionality of word stress on a scale, with the predictability of stress distributions as its core determinant (Peperkamp et al. 2010 and Section 5.1), rather than as a strictly categorical feature. In this respect, (re-)analysis of other Trade Malay and Indonesian languages is needed to understand which aspects indicate typological diversity and which aspects indicate typological homogeneity.

6 Discussion and conclusion

This book investigated whether Papuan Malay has word stress in four subsequent studies, focusing on different aspects of word stress in each of them. The short version of the answer to the research question at this stage is 'yes'. However, to understand what it means for a language to 'have stress', a brief reconsideration of the results is needed.

The first study showed that stressed syllables in Papuan Malay are acoustically more prominent than unstressed ones, in particular due to their increased duration, shallower spectral tilt and decentralised vowel quality. Furthermore, ultimate stress was found to be acoustically more marked than penultimate stress. In the second study, listeners were shown to be sensitive to this asymmetry and able to benefit from assuming the default penultimate pattern, unless otherwise signaled acoustically. The stress patterns were shown to play a word disambiguating role in the lexicon in the third study. The distinction between the stress patterns was ascribed to the avoidance of mid-vowels, rather than just /ɛ/ rejecting penultimate stress. Finally, the fourth study showed that listeners can successfully use their sensitivity to the acoustic cues to stress for word disambiguation.

While most of the evidence provided in this book speaks undoubtedly in favor of an analysis of Papuan Malay as a stress language, some issues remain unanswered. The most pregnant ones are discussed in the following and could shed a crucially different light on the results of this book when further investigated.

6.1 Acoustic signal

Chapter 2 selected disyllabic words that did not appear in phrase-final position, in order to avoid any acoustic effects related to phrase-finality in Papuan Malay, i.e., final lengthening and effects of phrase-final tonal movements such as a rise-fall. In this way, no acoustic evidence for word stress was obtained across phrase positions that differed to a larger extent. As the results stand, they mainly provide acoustic evidence for prominence differences within disyllables. As a control, antepenultimate and pre-antepenultimate syllables from longer words could have been measured. Note that no paradigmatic investigation of Papuan Malay word stress was possible due to the lack of minimal pairs (van Heuven 2018). In addition, ultimate word stress is confounded by final lengthening, which could have given speakers more time to articulate more extreme vowel positions. In other words, the word-boundary marking and potential stress marking cannot be fully disentangled. With regard to vowel quality, the acoustic measures were not corrected for intrinsic properties of vowels, which may have caused variance in the results that was not accounted for in the modeling. As for F0, the measurement did not capture fine-grained trajectories of the movement, as would a time-series measurement have done. Note that those aspects are studied in depth for phrase-final positions in Kaland et al. (2023).

6.2 Perception

The perception of the acoustic patterns reflecting word stress was tested in Chapter 3. Listeners in Experiment I and II appeared to have a bias for hearing ultimate word stress, rather than for penultimate word stress. It cannot be concluded from the results whether this bias was a consequence of the acoustic manipulation in the stimuli or the listeners' perceptual system. A control condition in which the two syllables were identical would have been needed to draw conclusions the nature of the bias. In Chapter 5, a gating task testing word disambiguation on the basis of word stress cues was carried out. The results appeared to be affected by co-articulation cues in the stimuli. One way of avoiding these effects would be to use a version of the stimuli without segmental cues (i.e., filtered or hummed speech). Although this would further reduce the naturalness of the stimuli, co-articulation would have been likely to be accounted for.

6.3 Schwa

A puzzling result in the acoustic analysis (Chapter 2) pertains to the role of schwa in the vowel inventory of Papuan Malay. Although the reports in Kluge (2017) were followed in that /ɛ/ might reduce to schwa and therefore causes a stress shift, an opposite effect was observed in the vowel quality measures: stressed syllables had a more central /ɛ/ than unstressed syllables. The number of items analysed and the resulting (in)stability of the vowel quality measures should be kept in mind. Nevertheless, schwa has played a central role in phonological theories on stress. The remaining theoretical issue here concerns the question to what extent schwa should always remain unstressed, or whether it is in fact possible to stress it as other vowels. The last possibility has been supported with acoustic data from Indonesian (Laksman 1994), but remains a difficult finding to reconcile with other stress research. The core problem lies at the central nature of schwa. Assuming that it is truly a central vowel, in what direction in the acoustic space should it go when it is stressed? For Papuan Malay, when assuming an influence of Standard Indonesian, the unexpected factors underlying the vowel quality results could be speculated on (Chapter 2). More research is needed, in particular on the extent to which schwa can or should be seen part of the vowel inventory of Papuan Malay. The importance of the status of schwa in Trade Malay has been illustrated in Paauw (2009) and one way forward would be to further investigate Papuan Malay's (historical) origin of stress loss and/or stress acquisition.

6.4 Lexical status

Traditionally, the lexical status of stress has been taken as an important diagnostic to analyse a language as a stress language. The outcomes of this book cast doubt on whether this criterion should be taken as a decisive diagnostic. It is not so much a question of whether the lexical storage of suprasegmental stress information should not be taken as an indication of word stress, but rather whether the absence of it is sufficient to conclude that a language is stressless. On the basis of the results reported in this book the answer to the latter question would be 'no'.

Chapter 4 and 5 showed that the Papuan Malay stress patterns provide cues that can be used by listeners to disambiguate words. The cues of both stressed and unstressed syllables were, however, presented to listeners (Chapter 5), eliminating the need to use lexically stored suprasegmental information. That information would have been needed in a task in which only one syllable would have been presented and in which the listeners' word choice would depend on the representation in their mental lexicon. Another way to tap into lexical storage would be by memory tasks using nonce words. The latter setup is essentially used in stress 'deafness' tasks (e.g. Dupoux et al. 1997). This was done in Kaland (2024), which compares Papuan Malay listeners' stress recall performance to those of German listeners. The results showed that Papuan Malay listeners perform worse than German ones, somewhat comparable to the French listeners in Dupoux et al. (1997). How should this result be reconciled with the results of this book, in particular with the ones in Chapter 5?

It is crucial to understand why suprasegmental information should be stored lexically. It has been shown that the need for storage largely depends on the predictability of stress patterns. Thus, it is not so much an issue of how many different stress patterns are found in a language. The crucial aspect is the extent to which listeners benefit from applying phonological rules to predict the stress patterns. For highly predictable patterns, phonological rules have a larger role to play than for unpredictable patterns. Thus, more stress information needs to be stored lexically when there is no rule that helps to predict its pattern. This is crucially a *degree*, as shown in previous research (Peperkamp & Dupoux 2002) and not a matter of finding minimal stress pairs or not (as acknowledged in Cutler 2012).

Turning back to the Papuan Malay results, it remains to be seen how a stressless analysis of this language would account for the results presented in this book. The studies so far ask for a wider definition of word stress as sometimes maintained in previous work. The proposal would be to 'diagnose' stress on a scale, similar to the 'stress deafness' one, allowing for intermediate functionality between the classic extremes of lexical stress (minimal pairs) and complete 'stress deafness'. The work conducted in Chapter 5 has shown that near-minimal pairs can be used to show listeners' ability to use suprasegmental cues to recognise words, even though the stress pattern distribution of their language is highly predictable and they are shown to be stress 'deaf' in the classic sense. The lexical analyses in Chapter 4 indicate that the lexicon is built in such a way that there is indeed a role to play for the stress patterns to disambiguate words. Given the acoustic presence (Chapter 2) and the listeners' auditory sensitivity (Chapter 3), the results of this book taken together are consistent. That is, Papuan Malay has a type of word stress that facilitates word recognition, although it does not need to be stored lexically (Kaland 2024).

6.5 Prosodic structure

Another remaining issue pertains to how word stress in Papuan Malay relates to other levels in the prosodic structure. So far, studies have mainly focused on the word level (this book and Kaland et al. 2018; Kaland & Gordon 2022) or phrase level (Riesberg et al. 2018; Riesberg et al. 2020, Kaland & Baumann 2020; Kaland et al. 2023). From these results it seems that Papuan Malay phrase prosody mainly functions as a phrase boundary marker, with little to no use of pitch accents, as described for other languages (e.g. Jun 2005; Jun 2014). The typological model by Jun (2014) relies on stressed syllables as docking sites for phrasal accents, as is reflected in the rhythmicity of head-languages. There are languages in which the word and phrase level are, however, more independent (e.g. Lindström & Remijsen 2005). Papuan Malay seems to be such a language, as stressed syllables do not appear to have a function at the phrase level in medial positions. It should be noted, though, that at phrase-final positions, stressed syllables tend to coincide with a rising F0 movement, as observed in two studies (Kaland & Baumann 2020; Kaland et al. 2023). Current research is undertaken to investigate the variation in F0 movements phrase-finally. The open question remains whether F0 movements are strictly boundary tones or whether they have a highlighting function as well, and what the difference between final and pre-final syllables is in this respect. As the phrase-final F0 movements often span (at least) two syllables, it might be a challenge for Papuan Malay to tease apart the two core prosodic functions assumed in typological models of prosody so far (Jun 2014; see Kaland & Baumann 2020).

6.6 Vowel quality and stress in English

Before wrapping up the discussion of this book, a final comment should be made on the way stress in English has been studied. English has played a central role in (shaping) our understanding of word stress, at least since Fry (1955). As shown in psycholinguistic studies, English listeners mainly attend to vowel quality differences rather than to other possible stress cues to distinguish words that are in lexical competition (e.g. Cooper et al. 2002; and discussion in Cutler 2012). Vowel quality differences in English stress research have often been referred to as 'segmental cues to stress' (e.g. Tremblay et al. 2021). As argued in the following, such a description of word stress cues is theoretically problematic.

Traditionally, word stress has been analysed as a suprasegmental property of words, i.e. concerning the prosodic cues that 'overlay' the segments. This aspect lies at the core of a widely accepted definition of word stress in prosodic theory (Chapter 1). However, allowing segmental properties to cue word stress troubles the diagnosis of stress. That is, how do we know whether we are dealing with a suprasegmental phenomenon if its main cue is categorised as segmental?

The problem lies in the use of terminology. Strictly speaking, the 'segmental cue' to English word stress, vowel quality, is not more or less segmental than other possible cues to word stress (e.g. duration). That is, these acoustic properties are physical aspects of the speech signal, regardless of how we categorise them in theories. The ambiguity with vowel quality is that it can be interpreted as the acoustic property of (phonemic) vowel identity as well as vowel reduction due to stress, where the former could be called 'segmental' and the latter 'suprasegmental'. Thus, vowel quality in English should be called a suprasegmental cue if we accept the analysis of English as a stress language. Otherwise, a re-analysis of English word stress is needed. This would mean that alleged minimal stress pairs of nouns and verbs such as (*permit* in 1a and 1b, Chapter 1) are actually segmentally different and the vowel inventory of English should be reanalysed accordingly (see Maskikit-Essed & Gussenhoven 2016 for a re-analysis along such lines for Ambonese Malay). While this re-analysis would undoubtedly be controversial for English, lexical analyses have shown that the disambiguating function of word stress in English is minimal (Cutler et al. 2004). Whether these findings indeed give rise to a rejection of word stress as a suprasegmental feature in English, therefore, remains to be a crucial topic for further investigation and discussion. For the time being, this book maintained and advocated a definition that interprets word stress as a prosodic phenomenon that is therefore signaled by acoustic cues *supra*segmentally.

6.7 Conclusion

Although future research still needs to contribute to the stress analysis in Papuan Malay, the current book has shown that, with the coverage of multiple aspects of word stress, a comprehensive understanding can be achieved. This has implications for the many languages that have only been described in grammars, often lacking empirical investigations of prosody. On the theoretical level, this book has shown that a definition of word stress as purely lexical might not apply to Papuan Malay. Thus, the results presented here ask for a view on the presence or absence of word stress that does not solely depend on its lexical status in the narrow sense. That is, this book has argued for the possibility to have predictable and functional stress patterns that do not need to be stored in the mental lexicon. Although future work should investigate this in more detail, it is possible that Papuan Malay stress is assigned post-lexically, however still at a lower level than the phonological phrase (Nespor & Vogel 2007). In traditional lexical stress languages, the assignment of stress is ‘earlier’, in the sense that it is stored with the phonemic information in the mental lexicon, whereas in the Papuan Malay type assignment is ‘later’. It is unclear whether existing phonological hierarchies are able to fully account for the potential variation among the prosodic levels at which different types of word stress could be assigned. There might be multiple underlying levels that all have a potentially similar sounding effect at the word level surface, thereby challenging the phonological analyses. A crucial question remains about which degree of memorization is needed for the Papuan Malay stress patterns, whether it is indeed possible to distinguish lexical from ‘post-lexical’ word stress, and how this can be shown in experimental (word-recognition) tasks.

References

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press. https://www.jstor.org/stable/10.3366/j.ctvxcrw9t.

Adams, Corinne & R.R. Munro. 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(3). 125–156. DOI: 10.1159/000259926. https://www.karger.com/Article/FullText/259926.

Agresti, Alan. 2007. *An introduction to categorical data analysis*. Hoboken: John Wiley & Sons. DOI: 10.1002/0470114754.

Akaike, Hirotogu. 1998. Information theory and an extension of the maximum likelihood principle. In Emanuel Parzen, Kunio Tanabe & Genshiro Kitagawa (eds.), *Selected papers of Hirotugu Akaike* (Springer Series in Statistics), 199–213. New York: Springer. DOI: 10.1007/978-1-4612-1694-0_15.

Albert, Aviad. 2023. *A model of sonority based on pitch intelligibility* (Studies in Laboratory Phonology 13). Berlin: Language Science Press. DOI: 10.5281/zenodo.7837176.

Albert, Aviad & Bruno Nicenboim. 2022. Modeling sonority in terms of pitch intelligibility with the nucleus attraction principle. *Cognitive Science* 46(7). DOI: 10.1111/cogs.13161.

Alieva, Natalia F., Vladimir D. Arakin, Alexander K. Ogloblin & Yu H. Sirk. 1991. *Bahasa Indonesia : Deskripsi dan teori* (Seri ILDEP 51). Yogyakarta: Kanisius. OCLC: 37831305.

Arnold, Denis, Petra Wagner & R. Harald Baayen. 2013. Using generalized additive models and random forests to model prosodic prominence in German. In Frédéric Bimbot, Christophe Cerisara, Cécile Fougeron, Guillaume Gravier, Lori Lamel, François Pellegrino & Pascal Perrier (eds.), *Interspeech 2013, Lyon*, 272–276. ISCA. DOI: 10.21437/Interspeech.2013-82.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40(3). 351–373. DOI: 10.1016/j.wocn.2012.02.003.

Audacity Team. 2019. *Audacity: free, open source, cross-platform audio software for multi-track recording and editing*. v.2.2.1. https://www.audacityteam.org/.

Baayen, Harald R. & Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2). 12–28. DOI: 10.21500/20112084.807.

Baerman, Matthew. 1999. *The evolution of fixed stress in Slavic* (LINCOM studies in Slavic Linguistics 15). Munich: LINCOM Europa. https://www.researchgate.net/publication/238743087.

Bailey, Todd M. 1995. *Nonmetrical constraints on stress*. University of Minnesota. (PhD thesis). Proquest/UMI: 9612955.

Balota, David A., Andrew J. Aschenbrenner & Melvin J. Yap. 2013. Additive effects of word frequency and stimulus quality: the influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(5). 1563–1571. DOI: 10.1037/a0032186.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Baumann, Stefan & Bodo Winter. 2018. What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics* 70. 20–38. DOI: 10.1016/j.wocn.2018.05.004.

Becker, Michael. 2003. Hebrew dtress: can't you hear those trochees? *University of Pennsylvania Working Papers in Linguistics* 9(1). 45–57. https://hdl.handle.net/20.500.14332/45623.

Becker, Richard A., Allan R. Wilks, Ray Brownrigg, Thomas P Minka & Alex Deckmyn. 2003. *maps: Draw geographical maps*. Version 3.4.2.1. DOI: 10.32614/CRAN.package.maps.

Beckman, Mary E. 1986. *Stress and non-stress accent* (Netherlands Phonetic Archives 7). Dordrecht: Foris. DOI: 10.1515/9783110874020.

Beckman, Mary E. & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Patricia A. Keating (ed.), *Phonological structure in phonetic form* (Papers in Laboratory Phonology III), 7–33. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511659461.002.

Berinstein, Ava E. 1979. *A cross-linguistic study on the perception and production of stress*. Los Angeles: Unviersity of California. (MA thesis). https://escholarship.org/uc/item/0t0699hc.

Boersma, Paul & David Weenink. 2017. *Praat: doing phonetics by computer*. v.6.0.28. http://praat.org/.

Boersma, Paul & David Weenink. 2019. *Praat: doing phonetics by computer*. v.6.0.56. http://praat.org/.

Bolinger, Dwight L. 1958. A theory of pitch accent in English. *Word* 14(2–3). 109–149. DOI: 10.1080/00437956.1958.11659660.

Breckenridge, Janet. 1977. The declination effect. *The Journal of the Acoustical Society of America* 61(S1). S90–S90. DOI: 10.1121/1.2015971.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32. DOI: 10.1023/A:1010933404324.

Burnham, Kenneth P. & David R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research* 33(2). 261–304. DOI: 10.1177/0049124104268644.

Caballero, Gabriela & Lucien Carroll. 2015. Tone and stress in Choguita Rarámuri (Tarahumara) word prosody. *International Journal of American Linguistics* 81(4). 457–493. DOI: 10.1086/683157.

Cambier-Langeveld, Tina. 2000. *Temporal marking of accents and boundaries.* University of Amsterdam (PhD thesis). Utrecht: LOT. https://hdl.handle.net/11245/1.172937.

Campbell, Nick & Mary E. Beckman. 1997. Stress, prominence and spectral tilt. In *Intonation: theory, models, and applications*, 67–70. ISCA. https://www.isca-archive.org/int_1997/campbell97_int.html.

Campbell, Nick & Stephen D. Isard. 1991. Segment durations in a syllable frame. *Journal of Phonetcs* 19(1). 37–47. DOI: 10.1016/S0095-4470(19)30315-8.

Chafe, Wallace L. 1980. *The pear stories: cognitive, cultural and linguistic aspects of narrative production* (Advances in Discourse Process 3). Norwood: Praeger. OCLC: 6626803.

Chafe, Wallace L. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press. OCLC: 29634150.

Chițoran, Ioana. 2002. *The phonology of Romanian: A constraint-based approach* (Studies in Generative Grammar 56). Berlin: Mouton de Gruyter. DOI: 10.1515/9783110889185.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row. https://web.mit.edu/morrishalle/pubworks/papers/1968_Chomsky_Halle_The_Sound_Pattern_of_English.pdf.

Chrabaszcz, Anna, Matthew Winn, Candise Y. Lin & William J. Idsardi. 2014. Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research* 57(4). 1468–1479. DOI: 10.1044/2014_JSLHR-L-13-0279.

Clements, G. Nick. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.), *Between the grammar and physics of speech* (Papers in Laboratory Phonology I), 283–333. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627736.017.

Connell, Katrina, Simone Hüls, Maria Teresa Martínez-García, Zhen Qin, Seulgi Shin, Hanbo Yan & Annie Tremblay. 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. *Language Learning* 68(3). 635–668. DOI: 10.1111/lang.12288.

Cooper, Nicole, Anne Cutler & Roger Wales. 2002. Constraints of lexical stress on lexical access in English: evidence from native and non-native listeners. *Language and Speech* 45(3). 207–228. DOI: 10.1177/00238309020450030101.

Cotton, Suzanne & François Grosjean. 1984. The gating paradigm: A comparison of successive and individual presentation formats. *Perception & Psychophysics* 35(1). 41–48. DOI: 10.3758/BF03205923.

Crosswhite, Katherine M. 2004. Vowel reduction. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 191–231. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486401.007.

Cutler, Anne. 1986. Forbear is a homophone: lexical prosody does not constrain lexical access. *Language and Speech* 29(3). 201–220. DOI: 10.1177/002383098602900302.

Cutler, Anne. 2005. Lexical stress. In David B. Pisoni & Robert E. Remez (eds.), *The handbook of speech perception* (Wiley Handbooks in Linguistics), 264–289. Oxford: Blackwell Publishing. DOI: 10.1002/9780470757024.ch11.

Cutler, Anne. 2012. *Native listening: language experience and the recognition of spoken words*. Cambridge: The MIT Press. DOI: 10.7551/mitpress/9012.001.0001. OCLC: 806959453.

Cutler, Anne & Sally Butterfield. 1992. Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language* 31(2). 218–236. DOI: 10.1016/0749-596X(92)90012-M.

Cutler, Anne & David M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language* 2(3-4). 133–142. DOI: 10.1016/0885-2308(87)90004-0.

Cutler, Anne & Charles Clifton. 1984. The use of prosodic information in word recognition. In Herman Bouma & Don G. Bouwhuis (eds.), *Attention and performance X: Proceedings of a symposium*, 183–196. Hillsdale: Erlbaum. https://hdl.handle.net/11858/00-001M-0000-0013-35C9-4.

Cutler, Anne, Delphine Dahan & Wilma van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40(2). 141–201. DOI: 10.1177/002383099704000203.

Cutler, Anne, Dennis Norris & Nuria Sebastián-Gallés. 2004. Phonemic repertoire and similarity within the vocabulary. In *Interspeech 2004-ICSLP, Jeju*, 65–68. ISCA. DOI: 10.21437/Interspeech.2004-61.

Cutler, Anne & Dennis Pasveer. 2006. Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In *Speech Prosody 2006, Dresden*. ISCA. DOI: 10.21437/SpeechProsody.2006-2.

Cutler, Anne, Roger Wales, Nicole Cooper & Joris H. Janssen. 2007. Dutch listeners' use of suprasegmental cues to English stress. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken*, 1913–1916. London: International Phonetic Association. http://www.icphs2007.de/conference/Papers/1108/1108.pdf.

Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11(1). 51–62. DOI: 10.1016/S0095-4470(19)30776-4.

Davis, Matthew H., William D. Marslen-Wilson & M. Gareth Gaskell. 2002. Leading up the lexical garden path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28(1). 218–244. DOI: 10.1037/0096-1523.28.1.218.

Di Cristo, Albert. 1998. Intonation in French. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation systems: A survey of twenty languages*, 195–218. Cambridge: Cambridge University Press. OCLC: 39116631.

Dogil, Grzegorz & Briony Williams. 1999. The phonetic manifestation of word stress in Lithuanian, Polish, German, and Spanish, in Harry van der Hulst (ed.), *Word prosodic systems in the languages of Europe* (Eurotyp 4), 273–311. Berlin: De Gruyter. DOI: 10.1515/9783110197082.1.273.

Domahs, Ulrike, Safiye Genc, Johannes Knaus, Richard Wiese & Barış Kabak. 2013. Processing (un-)predictable word stress: ERP evidence from Turkish. *Language and Cognitive Processes* 28(3). 335–354. DOI: 10.1080/01690965.2011.634590.

Domahs, Ulrike, Johannes Knaus, Paula Orzechowska & Richard Wiese. 2012. Stress "deafness"' in a language with fixed word stress: An ERP study on Polish. *Frontiers in Psychology* 3. 439. DOI: 10.3389/fpsyg.2012.00439.

Donohue, Mark. 2003. *Papuan Malay*. Manuscript, National University of Singapore. http://papuan.linguistics.anu.edu.au/Donohue/downloads/Donohue_2003_PapuanMalay.pdf.

Donohue, Mark & Yusuf Sawaki. 2007. Papuan Malay pronominals: forms and functions. *Oceanic Linguistics* 46(1). 253–276. DOI: 10.1353/ol.2007.0017.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian & Jacques Mehler. 1997. A destressing deafness in French? *Journal of Memory and Language* 36(3). 406–421. DOI: 10.1006/jmla.1996.2500.

Eriksson, Anders, Pier Marco Bertinetto, Mattias Heldner, Rosalba Nodari & Giovanna Lenoci. 2016. The acoustics of lexical stress in Italian as a function of stress level and speaking style. In *Interspeech 2016, San Francisco*, 1059–1063. ISCA. DOI: 10.21437/Interspeech.2016-348.

Evans, James D. 1996. *Straightforward statistics for the behavioral sciences*. Pacific Grove: Duxbury Press. OCLC: 32465263.

Fear, Beverley D., Anne Cutler & Sally Butterfield. 1995. The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America* 97(3). 1893–1904. DOI: 10.1121/1.412063.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486401.008.

Fletcher, Janet & Andrew McVeigh. 1993. Segment and syllable duration in Australian English. *Speech Communication* 13(3-4). 355–365. DOI: 10.1016/0167-6393(93)90034-I.

Fry, Dennis B. 1955. Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America* 27(4). 765–768. DOI: 10.1121/1.1908022.

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech; London* 1(2). 126–152. DOI: 10.1177/002383095800100207.

Goedemans, Rob. 1998. *Weightless segments: A phonetic and phonological study concerning the metrical irrelevance of syllable onsets*. University of Amsterdam (PhD thesis). Utrecht: LOT. https://hdl.handle.net/1887/20317.

Goedemans, Rob, Jeffrey Heinz & Harry van der Hulst. 2020. *StressTyp2*. http://st2.ullet.net/ (2 November, 2020).

Goedemans, Rob & Ellen van Zanten. 2007. Stress and accent in Indonesian. In Vincent J. van Heuven & Ellen van Zanten (eds.), *Prosody in Indonesian languages* (LOT Occasional Series 9), 35–62. Utrecht: LOT. https://hdl.handle.net/1874/296769.

Goedemans, Rob & Ellen van Zanten. 2014. No stress typology. In Johanneke Caspers, Yiya Chen, Willemijn Heeren, Jos Pacilly, Niels O. Schiller & Ellen van Zanten (eds.), *Above and beyond the segments*, 83–95. Amsterdam: John Benjamins. DOI: 10.1075/z.189.07goe.

Gordon, Matthew K. 2005a. A perceptually-driven account of onset-sensitive stress. *Natural Language & Linguistic Theory* 23(3). 595–653. DOI: 10.1007/s11049-004-8874-9.

Gordon, Matthew K. 2005b. Intonational phonology of Chickasaw. In Sun-Ah Jun (ed.), *Prosodic typology: the phonology of intonation and phrasing*, 301–330. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.003.0011.

Gordon, Matthew K. 2006. *Syllable weight: Phonetics, phonology, typology*. New York: Routledge. DOI: 10.4324/9780203944028. OCLC: 827972867.

Gordon, Matthew K. 2014. Disentangling stress and pitch-accent: A typology of prominence at different prosodic levels. In Harry van der Hulst (ed.), *Word stress*, 83–118. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139600408.005.

Gordon, Matthew K. & Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1). 20170007. DOI: 10.1515/lingvan-2017-0007.

Grabe, Esther & Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology 7* (Phonology and Phonetics 4-1). Berlin: De Gruyter. DOI: 10.1515/9783110197105.515.

Grafmiller, Jason & Stephanie Shih. 2011. New approaches to end weight. In *Variation and typology: new trends in syntactic research*, 22–23. Helsinki: Linguistics Association of Finland. https://www.academia.edu/1374901/New_Approaches_to_End_Weight.

Grice, Martine & Stefan Baumann. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191. 267–298. https://ifl.phil-fak.uni-koeln.de/sites/linguistik/Phonetik/pdf-publications/2002/Grice_Baumann-Deutsche_Intonation_und_GToBI_2002.pdf.

Grice, Martine, Michelina Savino, Alessandro O. Caffò & Timo B. Roettger. 2015. The tune drives the text: schwa in consonant-final loan words in Italian. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015), Glasgow*. Glasgow & London: University of Glasgow & International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0381.pdf.

Grice, Martine, Michelina Savino & Timo B. Roettger. 2018. Word final schwa is driven by intonation: the case of Bari Italian. *The Journal of the Acoustical Society of America* 143(4). 2474–2486. DOI: 10.1121/1.5030923.

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511616983.

Gussenhoven, Carlos, Bruno H. Repp, Toni Rietveld, Willem Rump & Jacques Terken. 1997. The perceptual prominence of fundamental frequency peaks. *The Journal of the Acoustical Society of America* 102(5). 3009–3022. DOI: 10.1121/1.420355.

Halim, Amran. 1981. *Intonation in relation to syntax in Indonesian* (Materials in Languages of Indonesia 5). Canberra: Australian National University. DOI: 10.15144/PL-D36.

Halle, Morris & Jean-Roger Vergnaud. 1987. *An essay on stress* (Current Studies in Linguistics 15). Cambridge: The MIT Press. OCLC: 48139729.

Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2021. *Glottolog 4.4.7*. Jena: Max Planck Institute for the Science of Human History. https://glottolog.org/.

Hanson, Helen M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America* 101(1). 466–481. DOI: 10.1121/1.417991.

Hanulíková, Adriana, James M. McQueen & Holger Mitterer. 2010. Possible words and fixed stress in the segmentation of Slovak speech. *Quarterly Journal of Experimental Psychology* 63(3). 555–579. DOI: 10.1080/17470210903038958.

Harrington, Jonathan. 2010. *Phonetic analysis of speech corpora.* Chichester: Blackwell. https://www.wiley.com/go/harrington. OCLC: 123375074.

Hart, Johan 't, René Collier & Antonie Cohen. 1990. *A perceptual study of intonation: an experimental-phonetic approach to speech melody* (Cambridge Studies in Speech Science and Communication). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627743.

Haspelmath, Martin & Uri Tadmor (eds.). 2009. *World loanword database (WOLD).* Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wold.clld.org/.

Hayes, Bruce. 1995. *Metrical stress theory: principles and case studies.* Chicago: The University of Chicago Press. OCLC: 28799248.

Heldner, Mattias. 2003. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* 31(1). 39–62. DOI: 10.1016/S0095-4470(02)00071-2.

Hellmuth, Sam. 2007. The relationship between prosodic structure and pitch accent distribution: evidence from Egyptian Arabic. *The Linguistic Review* 24(2-3). DOI: 10.1515/TLR.2007.011.

Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2-3). 61–83. DOI: 10.1017/S0140525X0999152X.

Himmelmann, Nikolaus P. 2018. Some preliminary observations on prosody and information structure in Austronesian languages of Indonesia and East Timor. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in austronesian languages* (Studies in Diversity Linguistics 21), 347–374. Berlin: Language Science Press. DOI: 10.5281/zenodo.1402555.

Himmelmann, Nikolaus P. & Daniel Kaufman. 2020. Austronesia. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford handbook of language* *prosody*, 369–383. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198832232.013.28.

Hockett, Charles Francis. 1955. *A manual of phonology*. 1st edn. Chicago: The University of Chicago Press. OCLC: 780743321.

Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346–363. DOI: 10.1002/bimj.200810425.

Howard, Irwin. 1972. *A directional theory of rule application in phonology*. Massachusetts Institute of Technology. (Doctoral dissertation). https://hdl.handle.net/1721.1/12982.

Howell, Peter & Karima Kadi-Hanifi. 1991. Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication* 10(2). 163–169. DOI: 10.1016/0167-6393(91)90039-V.

Hulst, Harry van der, Rob Goedemans & Ellen van Zanten (eds.). 2010. *A survey of word accentual patterns in the languages of the world*. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110198966.

Hussain, Sarmad. 1997. *Phonetic correlates of lexical stress in Urdu*. Northwestern University. (Doctoral dissertation). https://www.cle.org.pk/Publication/theses/1997/phonetic_correlates.pdf.

Hyman, Larry. 1985. *A theory of phonological weight* (Publications in Language Sciences 19). Dordrecht: Foris. DOI: 10.1515/9783110854794.

Hyman, Larry. 2006. Word-prosodic typology. *Phonology* 23(2). 225–257. DOI: 10.1017/S0952675706000893.

Ikranagara, Kay. 1975. *Melayu Betawi grammar*. University of Hawai'i at Mānoa. (PhD thesis). https://hdl.handle.net/10125/11720.

Iseli, Markus, Yen-Liang Shue & Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America* 121(4). 2283–2295. DOI: 10.1121/1.2697522.

Isenberg, David & Thomas Gay. 1978. Acoustic correlates of perceived stress in an isolated synthetic disyllable. *The Journal of the Acoustical Society of America* 64(S1). S21–S21. DOI: 10.1121/1.2004098.

Jacob, June & Charles W. Grimes. 2003. *Kamus pengantar bahasa Kupang*. 2nd edn. (Paradigma B 10). Kupang: Artha Wacana Press. OCLC: 944021922.

Jassem, Wiktor. 1962. *Akcent języka polskiego*. Wyd. 1 (Prace językoznawcze (Komitet Językoznawstwa PAN)). Wrocław: Zakład Narodowy im. Ossolińskich. OCLC: 5140392.

Jesse, Alexandra, Katja Poellmann & Ying-Yee Kong. 2017. English listeners use suprasegmental cues to lexical stress early during spoken-word recognition.

Journal of Speech, Language, and Hearing Research 60(1). 190–198. DOI: 10.1044/2016_JSLHR-H-15-0340.

Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody: Intonational phonology and prosodic structure*. Columbus: The Ohio State University. (Doctoral dissertation). http://rave.ohiolink.edu/etdc/view?acc_num=osu1220465077.

Jun, Sun-Ah (ed.). 2005. *Prosodic typology: The phonology of intonation and phrasing* (Oxford Linguistics). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199249633.001.0001.

Jun, Sun-Ah (ed.). 2014. *Prosodic typology II: The phonology of intonation and phrasing* (Oxford Linguistics). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199567300.001.0001.

Kaland, Constantijn. 2018. Spectral tilt as a correlate of Papuan Malay word stress. In *Speech Prosody 2018, Poznań*, 339–343. ISCA. DOI: 10.21437/SpeechProsody.2018-69.

Kaland, Constantijn. 2019. Acoustic correlates of word stress in Papuan Malay. *Journal of Phonetics* 74. 55–74. DOI: 10.1016/j.wocn.2019.02.003.

Kaland, Constantijn. 2020. Offline and online processing of acoustic cues to word stress in Papuan Malay. *The Journal of the Acoustical Society of America* 147(2). 731–747. DOI: 10.1121/10.0000578.

Kaland, Constantijn. 2021. The perception of word stress cues in Papuan Malay: A typological perspective and experimental investigation. *Laboratory Phonology* 12(1). 1–33. DOI: 10.16995/labphon.6447.

Kaland, Constantijn. 2024. When recall gets stressful: comparing Papuan Malay and German listeners' lexical storage of word stress. *Laboratory Phonology* 15(1). 1–27. DOI: 10.16995/labphon.11695.

Kaland, Constantijn & Stefan Baumann. 2019. Different functions of phrase-final f0 movements in spontaneous Papuan Malay. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhs), Melbourne*, 1312–1316. Canberra & London: Australasian Speech Science, Technology Association Inc. & International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_1361.pdf.

Kaland, Constantijn & Stefan Baumann. 2020. Demarcating and highlighting in Papuan Malay phrase prosody. *The Journal of the Acoustical Society of America* 147(4). 2974–2988. DOI: 10.1121/10.0001008.

Kaland, Constantijn, Christoph Bracks & Nikolaus P. Himmelmann. 2018. Repetition reduction in Papuan Malay prosody. In *Speech Prosody 2018, Poznań*, 10–14. ISCA. DOI: 10.21437/SpeechProsody.2018-2.

Kaland, Constantijn, Vincenzo Galatà, Lorenzo Spreafico & Alessandro Vietti. 2019. Which language r you speaking? /r/ as a language marker in Tyrolean and Italian bilinguals. *Language and Speech* 62(1). 137–163. DOI: 10.1177/0023830917746551.

Kaland, Constantijn & Matthew K. Gordon. 2022. The role of f0 shape and phrasal position in Papuan Malay and American English word identification. *Phonetica* 79(3). 219–245. DOI: 10.1515/phon-2022-2022.

Kaland, Constantijn, Nikolaus P. Himmelmann & Angela Kluge. 2019. Stress predictors in a Papuan Malay random forest. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhs), Melbourne)*, 2871–2875. Canberra & London: Australasian Speech Science, Technology Association & International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_2920.pdf.

Kaland, Constantijn, Angela Kluge & Vincent J. van Heuven. 2021. Lexical analyses of the function and phonology of Papuan Malay word stress. *Phonetica* 78(2). 141–168. DOI: 10.1515/phon-2021-2003.

Kaland, Constantijn, Marc Swerts & Nikolaus P. Himmelmann. 2023. Red and blue bananas: Time-series f0 analysis of contrastively focused noun phrases in Papuan Malay and Dutch. *Journal of Phonetics* 96. 101200. DOI: 10.1016/j.wocn.2022.101200.

Kaufman, Daniel & Nikolaus P. Himmelmann. 2024. Suprasegmental phonology. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the western Austronesian languages*, 703–718. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198807353.003.0042.

Kenstowicz, Michael. 1997. Quality-sensitive stress. *Rivista di linguistica* 9. 157–188. https://www.italian-journal-linguistics.com/app/uploads/2021/06/6_Kenstowicz.pdf.

Kluge, Angela. 2017. *A grammar of Papuan Malay* (Studies in Diversity Linguistics 11). Berlin: Language Science Press. DOI: 10.17169/langsci.b78.35.

Kluge, Angela, Ben E. W. Rumaropen & Lodowik Aweta. 2014. *Papuan Malay data: word list.* Dallas: SIL International. https://www.sil.org/resources/archives/59649.

Kohler, Klaus J. 2008. The perception of prominence patterns. *Phonetica* 65(4). 257–269. DOI: 10.1159/000192795.

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 arithmetic in typically developing children aged between 6 and 12 years. *Developmental Neuropsychology* 36(6). 721–740. DOI: 10.1080/87565641.2010.549882.

Kumanireng, Theresia Yosephine. 1993. *Struktur kata dan struktur frasa bahasa Melayu Larantuka*. Jakarta: Universitas Indonesia, Program Pendidikan Pasca Sarjana. (Doctoral dissertation). https://openlibrary.org/books/OL1070809M/.

Ladd, D. Robert. 2008. *Intonational phonology*. 2nd edn. (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511808814.

Laksman, Myrna. 1994. Location of stress in Indonesian words and sentences. In Cecilia Odé, Vincent J. van Heuven & Ellen van Zanten (eds.), *Experimental studies of Indonesian prosody* (Semaian 9), 108–139. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit Leiden. OCLC: 34085633.

Lehiste, Ilse. 1977. Isochrony reconsidered. *Journal of Phonetics* 5(3). 253–263. DOI: 10.1016/S0095-4470(19)31139-8.

Levelt, Willem J. M. 1989. *Speaking: from intention to articulation*. Cambridge: The MIT Press. DOI: 10.7551/mitpress/6393.001.0001.

Liaw, Andy & Matthew Wiener. 2002. Classification and regression by randomForest. *R News* 2(3). 18–22. https://journal.r-project.org/articles/RN-2002-022/.

Lieberman, Philip. 1959. Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America* 31(6). 844–844. DOI: 10.1121/1.1936148.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48(4). 839. DOI: 10.2307/411991.

Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35(11). 1773–1781. DOI: 10.1121/1.1918816.

Lindström, Eva & Bert Remijsen. 2005. Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics* 43(4). DOI: 10.1515/ling.2005.43.4.839.

Lipinski, John & Prahlad Gupta. 2005. Does neighborhood density influence repetition latency for nonwords? separating the effects of density and duration. *Journal of Memory and Language* 52(2). 171–192. DOI: 10.1016/j.jml.2004.10.004.

Litamahuputty, Betty H. J. 2012. *Ternate Malay: grammar and texts*. Leiden: Leiden University. (Doctoral thesis). https://hdl.handle.net/1887/19945.

Lo, Steson & Sally Andrews. 2015. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology* 6. 1171. DOI: 10.3389/fpsyg.2015.01171.

Maddieson, Ian. 1984. *Patterns of sounds*. 1st edn. (Cambridge Studies in Speech Science and Communication). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511753459.

Maddieson, Ian. 2011. Phonological complexity in linguistic patterning. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Conference of Phonetic Sciences (ICPhs XVII), Hong Kong*, 28–34. London: International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/PlenaryLecture/Maddieson/Maddieson.pdf.

Maskikit-Essed, Raechel & Carlos Gussenhoven. 2016. No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay. *Phonology* 33(2). 353–389. DOI: 10.1017/S0952675716000154.

Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2). 314–324. DOI: 10.3758/s13428-011-0168-7.

Mattys, Sven L., Laurence White & James F. Melhorn. 2005. Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General* 134(4). 477–500. DOI: 10.1037/0096-3445.134.4.477.

Mayer, Jörg. 2014. *Spektrales profil*. https://praatpfanne.lingphon.net/downloads/spectral_profile.praat.

McAllister, Jan. 1991. The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech* 34(1). 1–26. DOI: 10.1177/002383099103400101.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese* (Monographs on Linguistic Analysis 2). The Hague: Mouton. OCLC: 438994.

McDonnell, Bradley. 2016. Acoustic correlates of stress in Besemah. *NUSA: Linguistic studies of languages in and around Indonesia* 60. 1–28. DOI: 10.15026/87442.

McQueen, James M., Anne Cutler, Ted Briscoe & Dennis Norris. 1995. Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes* 10(3-4). 309–331. DOI: 10.1080/01690969508407098.

Mehta, Gita & Anne Cutler. 1988. Detection of target phonemes in spontaneous and read speech. *Language and Speech* 31(2). 135–156. DOI: 10.1177/002383098803100203.

Mengel, Andreas. 2000. *Deutscher Wortakzent: Symbole, Signale*. Munich: Phorm-Verlag. OCLC: 247293094.

Mo, Yoonsook, Jennifer Cole & Eun-Kyung Lee. 2008. Naïve listeners' prominence and boundary perception. In Plínio A. Barbosa, Sandra Madureira &

César Reis (eds.), *Speech Prosody 2008, Campinas*, 735–738. ISCA. DOI: 10.21437/SpeechProsody.2008-163.

Mohd Don, Zuraidah, Gerry Knowles & Janet Yong. 2008. How words can be misleading: A study of syllable timing and "stress"" in Malay. *The Linguistic Journal* 3. 66–81. https://www.linguistics-journal.com/2014/01/08/how-words-can-be-misleading-a-study-of-syllable-timing-and-stress-in-malay/.

Nababan, Partabas Wilmar Joakin. 1981. *A grammar of Toba-Batak* (Materials in Languages of Indonesia 65). Canberra: Australian National University. DOI: 10.15144/PL-D37.

Nespor, Marina & Irene Vogel. 2007. *Prosodic phonology: with a new foreword*. 2nd edn. (Studies in Generative Grammar 28). Berlin: Mouton de Gruyter. DOI: 10.1515/9783110977790.

Niebuhr, Oliver & Jana Winkler. 2017. The relative cueing power of f0 and duration in German prominence perception. In *Interspeech 2017, Stockholm*, 611–615. ISCA. DOI: 10.21437/Interspeech.2017-375.

Odé, Cecilia. 1994. On the perception of prominence in Indonesian. In Cecilia Odé, Vincent J. van Heuven & Ellen van Zanten (eds.), *Experimental studies of Indonesian prosody* (Semaian 9), 27–107. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit Leiden. OCLC: 34085633.

Ong, Walter J. 1982. *Orality and literacy: the technologizing of the word*. 1st edn. (New Accents). London: Methuen. OCLC: 8345642.

Ortega-Llebaria, Marta & Pilar Prieto. 2011. Acoustic correlates of stress in central Catalan and Castilian Spanish. *Language and Speech* 54(1). 73–97. DOI: 10.1177/0023830910388014.

Paauw, Scott H. 2009. *The Malay contact varieties of eastern Indonesia: A typological comparison*. State University of New York at Buffalo. (Doctoral dissertation). https://www.acsu.buffalo.edu/~dryer/PaauwMalayIndonesia.pdf.

Parker, Stephen G. 2002. *Quantifying the sonority hierarchy*. University of Massachusetts Amherst. (PhD thesis). DOI: 10.13140/2.1.3546.4005.

Peperkamp, Sharon & Emmanuel Dupoux. 2002. A typological study of stress deafness. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology 7* (Phonology and Phonetics 4-1), 203–240. Berlin: De Gruyter. DOI: 10.1515/9783110197105.1.203.

Peperkamp, Sharon, Inga Vendelin & Emmanuel Dupoux. 2010. Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics* 38(3). 422–430. DOI: 10.1016/j.wocn.2010.04.001.

Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press. https://archive.org/details/intonationofamer0000pike/.

Potisuk, Siripong, Jack Gandour & Mary P. Harper. 1996. Acoustic correlates of stress in Thai. *Phonetica* 53(4). 200–220. DOI: 10.1159/000262201.

Prentice, Jack. 1994. Manado Malay: product and agent of language change. In Tom Dutton & Darrell T. Tryon (eds.), *Language contact and change in the Austronesian world* (Trends in Linguistics. Studies and Monographs 77). Berlin: De Gruyter Mouton. DOI: 10.1515/9783110883091.411.

Prieto, Pilar, Maria del Mar Vanrell, Lluïsa Astruc, Elinor Payne & Brechtje Post. 2012. Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication* 54(6). 681–702. DOI: 10.1016/j.specom.2011.12.001.

R Core Team. 2017. *R: The R project for statistical computing*. v.3.4.0. https://www.r-project.org/.

R Core Team. 2019. *R: The R project for statistical computing*. v.3.5.3. https://www.r-project.org/.

R Studio Team. 2019. *RStudio: Integrated development for R*. v.1.0.143. https://www.rstudio.com/.

Ramus, Franck, Marina Nespor & Jacques Mehler. 2000. Correlates of linguistic rhythm in the speech signal. *Cognition* 75(1). AD3–AD30. DOI: 10.1016/S0010-0277(00)00101-3.

Ratcliff, Roger. 1993. Methods for dealing with reaction time outliers. *Psychological Bulletin* 114(3). 510–532. DOI: 10.1037/0033-2909.114.3.510.

Reinisch, Eva, Alexandra Jesse & James M. McQueen. 2010. Early use of phonetic information in spoken word recognition: lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology* 63(4). 772–783. DOI: 10.1080/17470210903104412.

Remijsen, Bert. 2002. Lexically contrastive stress accent and lexical tone in Ma'ya. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology 7* (Phonology and Phonetics 4-1). Berlin: De Gruyter. DOI: 10.1515/9783110197105.585.

Remijsen, Bert & Vincent J. van Heuven. 2005. Stress, tone and discourse prominence in the Curaçao dialect of Papiamentu. *Phonology* 22(2). 205–235. DOI: 10.1017/S0952675705000540.

Repp, Bruno H. 1983. Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication* 2(4). 341–361. DOI: 10.1016/0167-6393(83)90050-X.

Rialland, Annie & Stéphane Robert. 2001. The intonational system of Wolof. *Linguistics* 39(5). DOI: 10.1515/ling.2001.038.

Riesberg, Sonja & Nikolaus P. Himmelmann. 2012. *The DoBeS Summits-PAGE collection of Papuan Malay*. https://hdl.handle.net/1839/00-0000-0000-0019-FF78-5.

Riesberg, Sonja, Janina Kalbertodt, Stefan Baumann & Nikolaus P. Himmelmann. 2018. On the perception of prosodic prominences and boundaries in Papuan Malay. In Sonja Riesberg, Asako Shiohara & Atsuko Utsumi (eds.), *Perspectives on information structure in Austronesian languages* (Studies in Diversity Linguistics 21), 389–414. Berlin: Language Science Press. DOI: 10.5281/zenodo.1402558.

Riesberg, Sonja, Janina Kalbertodt, Stefan Baumann & Nikolaus P. Himmelmann. 2020. Using rapid prosody transcription to probe little-known prosodic systems: The case of Papuan Malay. *Laboratory Phonology* 11(1). 8. DOI: 10.5334/labphon.192.

Roettger, Timo & Matthew K. Gordon. 2017. Methodological issues in the study of word stress correlates. *Linguistics Vanguard* 3(1). DOI: 10.1515/lingvan-2017-0006.

Roosman, Lilie. 2006. *Phonetic experiments on the word and sentence prosody of Betawi Malay and Toba Batak*. Leiden University (PhD thesis). Utrecht: LOT. https://hdl.handle.net/1887/4371.

Roosman, Lilie. 2007. Melodic structure in Toba Batak and Betawi Malay word prosody. In Vincent J. van Heuven & Ellen van Zanten (eds.), *Prosody in Indonesian languages* (LOT Occasional Series 9), 89–115. Utrecht: LOT. https://hdl.handle.net/1874/296769.

Ryan, Kevin M. 2014. Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language* 90(2). 309–341. DOI: 10.1353/lan.2014.0029.

Ryan, Kevin M. 2016. Phonological weight. *Language and Linguistics Compass* 10(12). 720–733. DOI: 10.1111/lnc3.12229.

Salasoo, Aita & David B Pisoni. 1985. Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language* 24(2). 210–231. DOI: 10.1016/0749-596X(85)90025-7.

Samsuri, J. 1971. *Ciri-ciri prosodi dalam kalimat Bahasa Indonesia*. Flores: Nusa Indah.

Scott, Charles Payson Gurley. 1896. The Malayan words in English. *Journal of the American Oriental Society* 18. 49–124. http://archive.org/details/jstor-592501.

Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Mark Aronoff & Richard Oehrle (eds.), *Language sound structure*, 107–136. Cambridge: The MIT Press. OCLC: 9686618.

Shattuck-Hufnagel, Stefanie & Alice Turk. 1998. The domain of phrase-final lengthening in English. *The Journal of the Acoustical Society of America* 103(5). 2889–2889. DOI: 10.1121/1.421798.

Shih, Stephanie. 2013. *Random forests, for model (and predictor) selection*. Unpublished course material. UCLA 251: Variation in Phonology. https://stephsus.github.io/shih_randomforests.pdf.

Sluijter, Agaath M. C., Stefanie Shattuck-Hufnagel, Kenneth N. Stevens & Vincent J. van Heuven. 1995. Supralaryngeal resonance and glottal pulse shape as correlate of stress and accent in English. In Kjell Elenius & Peter Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences (ICPhS 95), Stockholm*, vol. 2, 630–633. London: International Phonetic Association. https://www.researchgate.net/publication/247677729.

Sluijter, Agaath M. C. & Vincent J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America* 100(4). 2471–2485. DOI: 10.1121/1.417955.

Sluijter, Agaath M. C., Vincent J. van Heuven & Jos J. A. Pacilly. 1997. Spectral balance as a cue in the perception of linguistic stress. *The Journal of the Acoustical Society of America* 101(1). 503–513. DOI: 10.1121/1.417994.

Soderberg, Craig D. & Kenneth S. Olson. 2008. Indonesian. *Journal of the International Phonetic Association* 38(2). 209–213. DOI: 10.1017/S0025100308003320.

Steinhauer, Hein. 1983. Notes on the Malay of Kupang (Timor). *NUSA Linguistic Studies of Indonesian and other Languages in Indonesia*. James T. Collins (ed.), Studies in Malay dialects II, 42–64. http://sealang.net/archives/nusa/pdf/nusa-v17-p42-64.pdf.

Stevens, Alan M. & Aluzius E. Schmidgall-Tellings. 2010. *A comprehensive Indonesian-English dictionary*. 2nd edn. Athens: Ohio University Press. OCLC: 320802745.

Stevens, Kenneth N. & Helen M. Hanson. 1995. Classification of glottal vibration from acoustic measurements. In Osamu Fujimura & Minoru Hirano (eds.), *Vocal fold physiology: Voice quality control*, 146–170. San Diego: Singular. https://www.researchgate.net/publication/312967106.

Stoel, Ruben. 2005. *Focus in Manado Malay: grammar, particles, and intonation* (CNWS Publications 134). Leiden: Leiden University. OCLC: 976569779.

Stoel, Ruben. 2007. The intonation of Manado Malay. In van Heuven, Vincent J. & Ellen van Zanten (eds.), *Prosody in Indonesian languages* (LOT Occasional Series 9), 117–150. Utrecht: LOT. https://hdl.handle.net/1874/296769.

Stoet, Gijsbert. 2010. Psytoolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42(4). 1096–1104. DOI: 10.3758/BRM.42.4.1096.

Stoet, Gijsbert. 2017. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44(1). 24–31. DOI: 10.1177/0098628316677643.

Streefkerk, Barbertje Marieke. 2002. *Prominence: acoustic and lexical/syntactic correlates*. University of Amsterdam (PhD thesis). Utrecht: LOT. https://hdl.handle.net/11245/1.198694.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(1). 307. DOI: 10.1186/1471-2105-9-307.

Strobl, Carolin, Torsten Hothorn & Achim Zeileis. 2009. Party on! *The R Journal* 1(2). 14. DOI: 10.32614/RJ-2009-013.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. DOI: 10.1037/a0016973.

Sulpizio, Simone & James M. McQueen. 2012. Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language* 66(1). 177–193. DOI: 10.1016/j.jml.2011.08.001.

Suomi, Kari, James M. Mcqueen & Anne Cutler. 1997. Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36(3). 422–444. DOI: 10.1006/jmla.1996.2495.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178. DOI: 10.1017/S0954394512000129.

Teeuw, Andries. 1978. *Leerboek bahasa Indonesia*. 3. dr. Groningen: Wolters-Nordhoff. OCLC: 246796786.

Traunmüller, Hartmut & Anders Eriksson. 1994. The frequency range of the voice fundamental in the speech of male and female adults. Manuscript. Institutionen för lingvistik, Stockholms universitet. https://www.researchgate.net/publication/240312210.

Tremblay, Annie, Mirjam Broersma, Yuyu Zeng, Hyoju Kim, Jinmyung Lee & Seulgi Shin. 2021. Dutch listeners' perception of English lexical stress: A cue-weighting approach. *The Journal of the Acoustical Society of America* 149(6). 3703–3714. DOI: 10.1121/10.0005086.

Tucker, Benjamin V., Daniel Brenner, D. Kyle Danielson, Matthew C. Kelley, Filip Nenadić & Michelle Sims. 2019. The massive auditory lexical decision (mald) database. *Behavior Research Methods* 51(3). 1187–1204. DOI: 10.3758/s13428-018-1056-1.

van Bezooijen, Renée & Charlotte Gooskens. 1999. Identification of language varieties: the contribution of different linguistic levels. *Journal of Language and Social Psychology* 18(1). 31–48. DOI: 10.1177/0261927X99018001003.

van Heuven, Vincent J. 1988. Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing. In William A. Ainsworth & John N. Holmes (eds.), *Proceedings of the 7th FASE/Speech-88 Symposium, Edinburgh*, 811–818. Edinburgh: Institute of Acoustics. https://hdl.handle.net/1887/2588.

van Heuven, Vincent J. 2018. Acoustic correlates and perceptual cues of word and sentence stress: towards a cross-linguistic perspective. In Rob Goedemans, Jeffrey Heinz & Harry van der Hulst (eds.), *The study of word stress and accent*, 15–59. Cambridge: Cambridge University Press. DOI: 10.1017/9781316683101.002.

van Heuven, Vincent J. & Peter J. Hagman. 1988. Lexical statistics and spoken word recognition in Dutch. In Peter Coopmans & Aafke Hulk (eds.), *Linguistics in the Netherlands*, 59–68. Dordrecht: Foris. https://hdl.handle.net/1887/2562.

van Heuven, Vincent J., Lilie Roosman & Ellen van Zanten. 2008. Betawi Malay word prosody. *Lingua* 118(9). 1271–1287. DOI: 10.1016/j.lingua.2007.09.005.

van Heuven, Vincent J. & Ellen van Zanten. 1997. Effects of substrate language on the localization and perceptual evaluation of pitch movements in Indonesian. In Cecilia Odé & Wim Stokhof (eds.), *Proceedings of the 7th International Conference on Austronesian Linguistics (ICAL 7), Leiden*, 63–80. Amsterdam: Rodopi. DOI: 10.1163/9789004643253_005. https://hdl.handle.net/1887/63067.

van Minde, Don. 1997. *Malayu Ambong: phonology, morphology, syntax* (CNWS Publications 117). Leiden: Leiden University. OCLC: 38075818.

van Rossum, Guido & Jelke De Boer. 1991. Interactively testing remote servers using the Python programming language. *CWI Quarterly* 4(4). 283–304. https://ir.cwi.nl/pub/18204.

van Staden, Miriam. 2000. *Tidore: A linguistic description of a language of the North Moluccas*. Leiden University. (Doctoral dissertation). OCLC: 46768773.

van Zanten, Ellen, Ruben Stoel & Bert Remijsen. 2010. Stress types in Austronesian languages. In Harry van der Hulst (ed.), *A survey of word accentual patterns in the languages of the world*, 87–112. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110198966.1.87.

van Zanten, Ellen & Vincent J. van Heuven. 1984. The Indonesian vowels as pronounced and perceived by Toba Batak, Sundanese and Javanese speakers. *Bijdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia* 140(4). 497–521. DOI: 10.1163/22134379-90003411.

van Zanten, Ellen & Vincent J. van Heuven. 1998. Word stress in Indonesian; its communicative relevance. *Bijdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia* 154(1). 129–149. DOI: 10.1163/22134379-90003908.

van Zanten, Ellen & Vincent J. van Heuven. 2004. Word stress in Indonesian: fixed or free? *NUSA Linguistic Studies of Indonesian and other Languages in Indonesia* 53. 1–20. http://sealang.net/archives/nusa/pdf/nusa-v53-p1-20.pdf.

van der Hulst, Harry. 2012. Deconstructing stress. *Lingua* 122(13). 1494–1521. DOI: 10.1016/j.lingua.2012.08.011.

Vaux, Bert. 1998. *The phonology of Armenian* (The Phonology of the World's Languages). Oxford: Clarendon Press. DOI: 10.1093/oso/9780198236610.001.0001.

Vihman, Marilyn & William Croft. 2007. Phonological development: toward a radical templatic phonology. *Linguistics* 45(4). DOI: 10.1515/LING.2007.021.

Vogel, Irene, Angeliki Athanasopoulou & Nadya Pincus. 2016. Prominence, contrast, and the functional load hypothesis: an acoustic investigation. In Jeffrey Heinz, Rob Goedemans & Harry van der Hulst (eds.), *Dimensions of phonological stress*, 123–167. Cambridge: Cambridge University Press. DOI: 10.1017/9781316212745.006.

Vroomen, Jean, Monique van Zon & Beatrice de Gelder. 1996. Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Memory & Cognition* 24(6). 744–755. DOI: 10.3758/BF03201099.

Wan, Aslynn. 2012. *Instrumental phonetic study of the rhythm of Malay*. Newcastle University. (PhD thesis). https://hdl.handle.net/10443/1682.

Wingfield, Arthur, Harold Goodglass & Kimberly C. Lindfield. 1997. Word recognition from acoustic onsets and acoustic offsets: Effects of cohort size and syllabic stress. *Applied Psycholinguistics* 18(1). 85–100. DOI: 10.1017/S0142716400009887.

Wright, Marvin N. & Andreas Ziegler. 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77(1). DOI: 10.18637/jss.v077.i01.

Yu, Jenny, Robert Mailhammer & Anne Cutler. 2020. Vocabulary structure affects word recognition: Evidence from German listeners. In *Speech Prosody 2020, Tokyo*, 474–478. ISCA. DOI: 10.21437/SpeechProsody.2020-97.

Name index

Word stress in prosodic theory

This book is composed of four studies that all investigate different aspects of word stress in Papuan Malay, an Austronesian language spoken in eastern Indonesia. These aspects, in order of presentation, include acoustic realisation, auditory perception, lexical analyses and word disambiguation. The introduction provides the theoretical background against which the studies are undertaken. All studies are empirical in nature; they either report acoustic analyses, production or perception experiments, or corpus-based analyses. Taken together, the results of all studies pose a challenge to maintain a stressless analysis of Papuan Malay. At the same time, the type of word stress that appears from the reported results is unlike its common theoretical conception and therefore requires more work to be integrated in prosodic theory. Given the controversy on word stress in Indonesian languages, the results are always discussed and carefully interpreted in a cross-linguistic context. In this way, the current thesis extends and deepens our knowledge and understanding of word stress in prosodic theory.