Acoustic correlates of stress in Besemah

Bradley McDONNELL

University of Hawai'i at Mānoa

The present study investigates the location and acoustic correlates of stress in Besemah, a Malay isolect of southwest Sumatra. While a number of studies have shown that other well-known varieties of Malay do not show evidence of word-level stress, the present study is a first attempt at discovering whether Besemah shows evidence of word-level stress and if so what acoustic properties are associated with the stressed syllable. Results indicate that stress is realized on the ultima with a raised fundamental frequency, greater intensity, and longer duration. The study further investigates these results by looking at additional intonational evidence from naturalistic speech to better understand the interaction between word-level stress and phrase-level prominence. The present study adds to the current typology of prominence systems, demonstrating that Besemah is not on a par with other Malay varieties.

1. Introduction[1]

Phonetic investigations of individual languages over the past half-century have uncovered a fairly well-defined set of acoustic properties that are associated with stress. Acoustic measures, such as intensity, fundamental frequency, duration, and vowel quality, have been correlated with stress in an areally and genetically diverse group of languages, including English (Fry 1955, 1958), Tagalog (Gonzalez 1970), Mohawk and Oneida (Michelson 1988), Thai (Potisuk et al. 1996), Pirahã (Everett 1998), Chickasaw (Gordon 2004), Turkish (Levi 2005), St'át'imcets (Caldecott 2007), South Conchucos Quechua (Hintz 2006), and Turkish Kabardian (Gordon & Applebaum 2010) among others.

More recently, phonetic investigations have noted the importance of interactions between different domains of prominence at the word-level (e.g., stress) and the phrase-level (e.g., accent) when identifying acoustic correlates. Phrase-level prominence is concerned with prosodic units above the word that include larger units (e.g., Intonational Phrase) and smaller units (e.g., Accentual Phrase). In the autosegmental-metrical (AM) framework, phrase-level prominence is treated phonologically with discrete (H)igh and (L)ow tone targets that are either pitch accents (e.g., H*, L*, H*+L, L*+H) that serve a semantic or pragmatic marking function or pitch excursions at the phrase edge (or boundary tones, e.g., H%, L%, HL%, LH%) that serve a delimitive function (Ladd 2008). In a number of languages, there is a clear interaction between word- and phrase-level stress, but there are also languages that lack word-level stress altogether (Gordon 2014). Examples of such languages are found in well-known languages like French (Jun & Fougeron 2000, 2002) and Korean (Jun 1998, 1996) where prominence is realized in the domain of the Accentual Phrase and not the word. However, it is often difficult to tease apart word- and phrase-level prominence in languages lacking word-level stress. This difficulty arises because

[1] I would like to thank Matthew Gordon, Stefan Gries, Marianne Mithun, Bernard Comrie and two anonymous reviewers for their guidance and thoughtful comments and suggestions. This research was made possible by institutional support in Indonesia from the Office of Research and Technology (RISTEK), Universitas Atma Jaya, Universitas Sriwijaya, and the Max Plank Institute for Evolutionary Anthropology and by the financial support of a Fulbright Scholarship and Blakemore Freeman Fellowship. I am particularly grateful to the village of Karang Tanding for graciously taking part in this research and to Kencana Dewi and Asfan Fikri Sanaf for making fieldwork on Besemah possible. All errors are my own.

McDONNELL, Bradley, 2016. 'Acoustic correlates of stress in Besemah'. In YANTI and Timothy McKINNON, eds. *Studies in language typology and change*. NUSA 60: 1–28. [Permanent URL: http://hdl.handle.net/10108/87442]

the acoustic data are often based on recorded wordlists that are made up of words in isolation, thereby conflating word- and phrase-level phenomena. Shona, a Bantu language of Zimbabwe, exemplifies this confusion quite clearly. For example, an early report in Doke (1931) states that word-level stress is marked by a phonetically longer penult, while a later study by Fortune (1980) proposes that penultimate lengthening is most likely a phrase-level phenomenon. Recent phonological studies have cited both sources (e.g., Downing (2004, 2010) cites the former, while Beckman (1997) cites the latter). While all studies agree that prominence is realized by penultimate lengthening, there has been no in-depth phonetic study to determine whether Shona prominence operates at a word-level, a phrase-level, or an interaction between word- and phrase-level (cf. Hyman 2014).

One particularly controversial case with which there have been a number of phonetic studies is standard varieties of Malay-Indonesian. The prevailing opinion expressed in traditional descriptions of these varities, based on impressionistic judgements, states that word-level stress falls on the penultimate syllable unless the penult contains a schwa, in which case stress falls on the final syllable (Macdonald & Darjowidjojo 1967; Mintz 1994; Wolff 1986). This characterization harkens back as far as van Wijk (1883) and has been adopted in the theoretical phonological literature on stress (Cohn 1989, 1993; Cohn & McCarthy 1998). This fixed position view of word-level stress in Indonesian has been supported by acoustic phonetic studies in Laksman (1994) and Adisasmito-Smith & Cohn (1996). Adisasmito-Smith & Cohn (1996) find that word-level penultimate stress correlates with greater intensity and a higher fundamental frequency, while Laksman (1994) finds that a higher fundamental frequency is the only clear correlate of stress. Like Adisasmito-Smith & Cohn, Laksman finds that stress falls on the penultimate syllable, but, unlike Adisasmito-Smith & Cohn, she found that schwa does in fact take stress. Neither study, however, takes phrase-level prominence into account.

Others have questioned the status of word-level stress in Malay-Indonesian. Fokker (1895), for instance, states that 'Malay accent is not to be identified with that in English, German or Dutch. A peculiar feature of it is its *lack of stress*' (97). Halim (1981) presents the first acoustic analysis of Standard Indonesian prominence where words in phrase-final position behave differently than words in phrase-medial position. He finds that prominence marked by a higher fundamental frequency and longer duration falls on the final syllable in words in utterance-medial position and on the penultimate syllable in words in utterance-final position. Halim concludes that "[t]he placement of [word] accent is a function of the structure of the intonation pattern in which the accent occurs" (122). More recently, there has been a surge in experimental production and perception studies of word- and phrase-level prominence in Indonesian, all of which point to a lack of word-level stress in Standard Indonesian (van Zanten 1994; Ebing 1997; van Zanten & van Heuven 1998; van Zanten et al. 2003; van Zanten & van Heuven 2004; Goedemans & van Zanten 2007). For instance, van Zanten & van Heuven (1998), in a gating perception experiment, find that '[s]tress position . . . plays no role in word identification in Indonesian' (141), and van Zanten et al. (2003) find that Standard Indonesian listeners reported stressed syllables acceptable no matter which syllable within a word is more prominent.[2]

[2] More specifically, the gating task in van Zanten & van Heuven (1998) played portions of words with different stress patterns (e.g., *anak* 'child' with stress on the first syllable, *anaknya* 'his child' with stress on the second syllable, and *anak-anak* 'children' with stress on the third syllable) in two different passes (e.g., the first pass played the first syllable and the second pass played the first two syllables). The words were

Van Heuven & Faust (2009) even find that Standard Indonesian learners of Dutch are insensitive to narrow-focus contrasts at the sub-word level in Dutch, which again suggests that Standard Indonesian lacks word-level stress.

Only recently, however, has influence from substrate languages on Standard Indonesian been taken into account in phonetic studies of stress.[3] Preliminary evidence from Goedemans & van Zanten (2007) suggests that substrate languages have a considerable effect on the realization of word-level stress in Standard Indonesian. In the first stage of their experiment, Goedemans & van Zanten compare the pronunciation of speakers of Toba Batak, a rather different language of northern Sumatra with contrastive stress, and the pronunciation of speakers of Javanese, a language of Java with no apparent word-level stress. They find that Toba Batak speakers regularly mark word-level stress in Standard Indonesian on the penultimate syllable with higher fundamental frequency, increased loudness, and longer duration, and Javanese speakers mark prominence on either the penultimate or final syllable in Standard Indonesian. This leads Goedemans & van Zanten to conclude that Toba Batak speakers of Standard Indonesian mark word-level stress, but Javanese speakers of Standard Indonesian do not. In a subsequent pairwise-comparison perception experiment, Goedemans & van Zanten find that while Toba Batak speakers rated only the penultimate stress of other Toba Batak speakers acceptable, speakers of both Javanese and Toba Batak rated the variable stress pattern of Javanese speakers acceptable. Goedemans & van Zanten (2007) conclude 'the fact that [Toba Batak] speakers … do accept final as well as pre-final prominence, indicates to us that they enter a different "mode" when listening to a group of speakers that defines (through group size, but also greater influence in politics and the media) the most common form of Indonesian' (54). This fascinating finding raises a number of questions about the complexity of word-level stress in Standard Indonesian, one that can hardly be answered without understanding more about the interaction between Standard Indonesian and substrate languages of Indonesia.

While Standard Indonesian in all of its sociolinguistic complexity has received a lot of attention, there are a large number of Malay isolects that have received far less attention by scholars. Based on impressionistic judgements of Malay isolects, Prentice (1994) posits a broad split between western Indonesian varieties of Borneo, Sumatra, and the Malay Peninsula that mark stress on the final syllable and eastern Indonesian varieties of Java, Sulawesi, and the islands to the east that mark stress on the penultimate syllable. However, in Betawi Malay, a Malay isolect spoken in and around Jakarta on Java, the results do not follow Prentice's generalization, but pattern closely to Standard Indonesian speakers of Javanese in Goedemans & van Zanten's study. Again, their finding suggests that Betawi Malay lacks word-level stress (van Heuven et al. 2008). The question, then, becomes are Standard Indonesian and Betawi Malay, which lack word-level stress representative of other varieties of Malay? Or will other varieties of Malay not spoken in and around urban centers conform to the generalization of Prentice (1994)? To test whether

embedded in a carrier sentence. Indonesian listeners were asked to choose one of the three choices after listening to each pass (i.e., after listening to first syllable, then after listening to the first two syllables).

[3] The sociolinguistic contexts of Indonesia are varied and are generally complex. Standard Indonesian is the language of the government and education and as such is primarily learned as a second language. In major cities around Indonesia, there are large regional varieties of colloquial Indonesian (e.g., Jakarta Indonesian, Riau Indonesian) that differ significantly from Standard Indonesian and from each other (Gil 1994). For an overview of the sociolinguistic context in Indonesia see Sneddon (2003).

a lack of word-level stress can be extended to other varieties of Malay, the present study will address the following questions for Besemah, an under-described Malay isolect of southwest Sumatra.

1. Is stress, which, based on the judgements of the author, impressionistically falls on the ultimate syllable of the word, regularly associated with all or any of the traditional acoustic correlates of stress (e.g., intensity, duration, fundamental frequency)?
2. Is the domain of stress the word or some larger prosodic unit such as the Accentual Phrase or the Intonation Phrase?

The paper is split into two parts. The first part investigates elicited lexical items to discover whether the location of stress is associated with one or more acoustic correlates. The second part of the study, then, further investigates the acoustic results of the first part with additional intonational evidence from naturalistic speech to understand the interaction between word- and phrase-level prominence. By addressing these questions, the present study will add to the current typology of prominence systems, ultimately showing that Besemah is not on a par with Standard Indonesian and Betawi Malay, which lack word-level stress. Rather, the results support the conclusion that Besemah evinces word-final stress, which correlates with a higher fundamental frequency and increased intensity. Finally, an equally important component of this study is to provide a description of a heretofore undocumented prosodic system among a plethora of Malay isolects that have received very little linguistic documentation.

2. Background

Besemah is a Malay isolect spoken by some 400,000 people in the highlands of southwest Sumatra (Gordon 2005).[4] While Besemah demonstrates many of the features typical of Malay isolects, it differs from the more well-known Standard Indonesian in many respects. First, Besemah has a very different sociolinguistic status than Standard Indonesian. As was mentioned in the previous section, Standard Indonesian is typically learned as a second language by a diverse group of speakers of various local languages throughout Indonesia. Thus, Standard Indonesian phonology is affected by the substrate language (van Zanten & van Heuven 1984; van Zanten 1986; Goedemans & van Zanten 2007; Prentice 1990). Besemah, on the other hand, is a basilectal variety, acquired as a first language by a rather homogenous group of speakers. In classifying these differences along sociolinguistic lines, Adelaar & Prentice refer to the prior as ‘Literary Malay’ (i.e., ‘products of the literary Malay tradition’ (1996: 202)) and the latter as a ‘Vernacular Malay’, which they define as a variety of Malay ‘spoken in traditionally Malayic speech communities and appear to be regularly inherited from Proto-Malayic’ (203). As is the case with many of the Malay isolects of Sumatra, Besemah has received little scholarly attention. The earliest study of Besemah phonology is a short description of the phonology in a Besemah-Dutch dictionary from the Dutch Colonial period by Helfrich (1904). Additionally, the author has published a treatment of the historically conservative vowel

[4] Besemah is a part of a complex dialect cluster that I refer to as South Barisan Malay. South Barisan Malay is not mutually intelligible with Standard Indonesian or other Malay isolects of Indonesia (McDowell & Anderbeck 2008). These isolects have been recently conflated in the most recent version of the Ethnologue, under the name Central Malay (Lewis 2009).

system from an Austronesian perspective (McDonnell 2008) and contributed to a phonetic treatment of Besemah central vowels as a part of a larger study on the sonority of central vowels (Gordon et al. 2012).

While there are no previous studies devoted to Besemah prosody, Helfrich, in his section on Besemah phonology, describes stress in the following way, '[t]he accent is, in contrast to the one in Malay, quite strong and falls on the penultimate syllable, even in derived words.... However, if the penultimate syllable has the atonic "ə" and if it is not closed then the accent falls on the ultimate syllable' (1904: 200).[5] Helfrich's description of Besemah stress does not follow the impressionistic judgements of the author according to which stress falls on the ultima.[6] With little in the way of phonetic and phonological description, the following section presents a basic description of the segmental inventory, phonotactics, and segments that may affect stress assignment.

2.1 Consonants

The consonant phoneme inventory of Besemah, in many respects, represents a rather typical inventory for a Malay isolect. The only difference between Besemah and other varieties of Malay is the additional contrast between the voiced velar fricative /ɣ/ and the alveolar trill /r/ (e.g., the minimal pair /ɣagi/ 'yeast' and /ragi/ 'color') (cf. Adelaar 1992; McDonnell 2008). Besemah has 21 consonant phonemes shown in Table 1.

Table 1. Besemah consonants

	Bilabial	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Stop	p b	t d			k g	ʔ
Nasal	m	n		ɲ	ŋ	
Tap/Trill		ɾ/r				
Fricative		s			ɣ	h
Affricate			ʧ ʤ			
Approximant	w			j		
Lateral		l				

2.2 Vowels

The vowel phoneme inventory of Besemah is historically rather conservative, with only three vowel phonemes as opposed to the five in Standard Indonesian.[7] Allophonic variation in Besemah vowels largely depends on two factors, (1) on the presence or absence of a coda consonant and (2) the position within in the word. The following generalizations account for allophonic variation in Besemah. The high front vowel /i/ lowers to [ɪ] in the ultimate syllable when it is closed by any coda consonant. On the other hand, the high back vowel /u/ lowers to [ʊ] in the ultimate syllable when followed by a supralaryngeal coda consonant and to [o] when followed by a laryngeal consonant. This sub-

[5] Thank you to Bram Vertommen for providing the translation.

[6] This discrepancy could also be attributed to a change in Besemah prosody over the last century.

[7] The three vowel phoneme inventory reflects that of Proto-Malayic, which has been reconstructed with the same three vowel phonemes (Adelaar 1992).

and supralaryngeal lowering effect does not generalize to all speakers. Of the four male and four female speakers considered here, all of the male speakers show evidence of two vowel qualities, [ʊ] and [o], but only one female speaker shows evidence of two different vowel qualities. Additionally, when /a/ occurs word- or root-finally, it raises to [ɨ]. Finally, there is an epenthetic vowel [ə], referred to as the *pepet* vowel by many Austronesianists that is restricted to the penultimate syllable and only occurs between two consonants (e.g., [kəbʊn] 'field'). The vowel phonemes and their respective allophones in Besemah are shown in Table 2 below.

Table 2. Besemah vowels[8]

	Front	Central	Back
High	/i/ [ɪ]	[ɨ]	/u/ [ʊ]
Mid		(ə)	([o])
Low		/a/	

Figure 1 plots the vowel allophones of Besemah for the same four male and four female speakers used in the study of Besemah stress. The measurements include the words used in the stress study (see Appendix A) in addition to examples of words with an epenthetic vowel and words with a glottal coda that were a part of the original wordlist (see Section 3). Due to the allophonic variation that results from an asymmetry in the penultimate and ultimate syllable, vowel quality (e.g., F1 and F2) was not considered as an acoustic correlate of stress in the present study. The epenthetic (or *pepet*) vowel is also excluded from the present study as it behaves quite differently from the other vowels in its morphophonemic behavior, predictability, and phonetic properties (McDonnell 2008). For example, despite being in complementary distribution with [ɨ], the *pepet* clearly differs in both vowel quality (crucially from [ɨ] and /a/ in Figure 1) and duration. Figure 2 demonstrates that the *pepet* is significantly shorter than all other vowels in the penultimate syllable and, as a point of comparison, the [ɨ] does not significantly differ from other final vowels in open syllables.

2.3 Phonotactics and minimal words

Besemah phonotactics are typical of many Austronesian languages. The maximal syllable is CVC. Any consonant except /h/ freely serves as the onset of the syllable.[9] Coda consonants in the word medial position are restricted to nasals, the alveolar trill /r/, and in rare cases the alveolar lateral approximant /l/ or the alveolar fricative /s/. Nasals in the word-medial coda position are only followed by homoganic stops. Affricates and voiced stops never occur in coda position. The velar fricative /ɣ/ never occurs in word-medial coda position, while the palatal nasal /ɲ/ never occurs in word-final coda position. The only restriction on vowels is that the *pepet* only occurs between two consonants in pre-final syllables.

8 Vowels within slashes are considered phonemes, while vowels within brackets are considered allophones. The parentheses represent that the allophone is present for some speakers.

9 The phoneme /h/ is primarily restricted to word-final position, but does occur word-medially in a small number of loan words (e.g., [mahap] 'sorry').

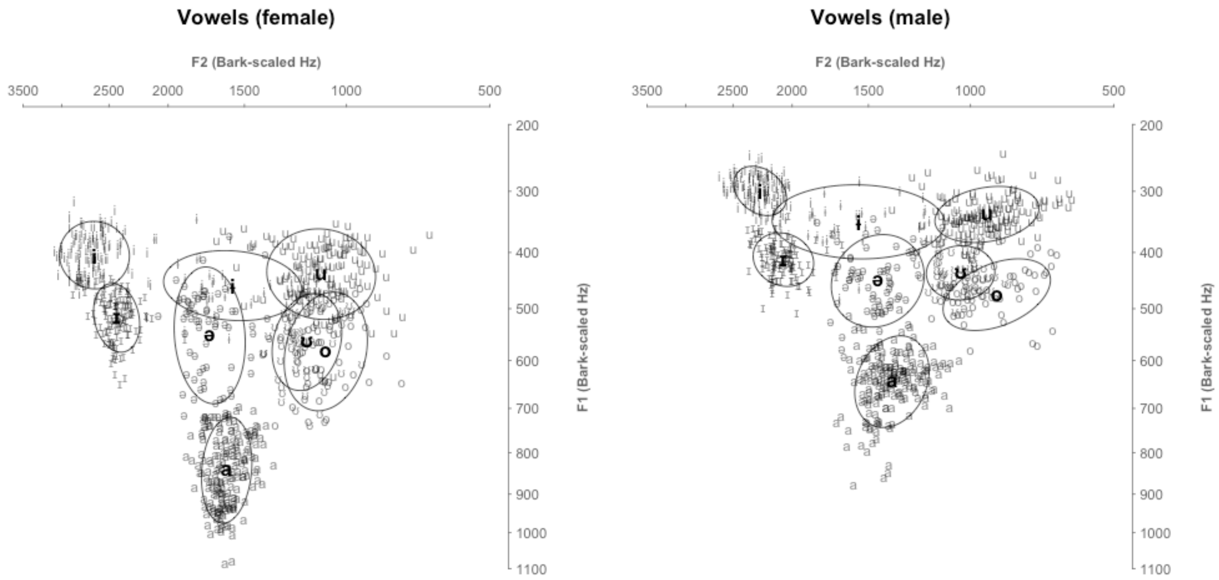

Figure 1. Besemah vowel allophones

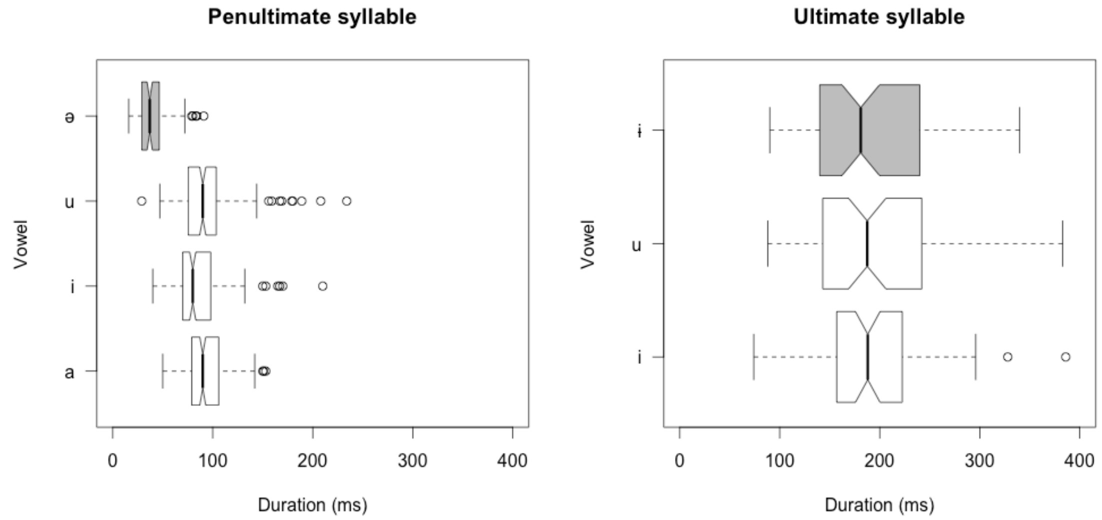

Figure 2. Vowel duration in penultimate and ultimate syllables

Words are most typically disyllabic with only a handful of words with a CVC shape (e.g., [tʃɪt] ‘paint’, [laŋ] ‘eagle’). In the case that words are more than two syllables, prepenultimate syllables are reduced to schwa (e.g., [tʃənilah] ‘sandal’), a process that occurred in many other Malay languages (Adelaar 1992).[10] The words utilized in the present study are all disyllabic.

3. Methodology

The data used in the present study originally were collected in order to investigate allophonic variation in Besemah. The constructed wordlist consists of 341 monomorphemic

[10] Unless the antepenult is a high vowel in a vowel cluster (e.g., /muanaj/ ‘older sister’).

Besemah words with each vowel occurring in various contexts (i.e., in open and closed syllables and in the penultimate and ultimate syllable of word). A total of eight Besemah speakers, four males and four females, were recorded over a period of two weeks in June 2010 in the village of Karang Tanding, a small Besemah village that is located approximately ten kilometers northwest of the small city of Pagaralam in South Sumatra, Indonesia. The participants range in age from 25 to 60+ and either reside in or near the village of Karang Tanding. Furthermore, all participants are native speakers of Besemah and have different levels of fluency in Standard Indonesian. During the recording sessions, participants were asked to say each word twice in isolation and then a third time in a carrier phrase [ŋitʃeʔkah __ agi] 'say __ again'. All recordings were made on a solid-state Marantz PMD670 recorder with an Audio-Technica AT825 stereo microphone. The recordings were made with a sampling frequency of 48 kHz and were stored as .wav files.

While the wordlist was not originally constructed or collected to investigate stress, the recordings provide a robust enough sample to be repurposed for the present study. Of the 341 words recorded, 54 words were selected based on three factors: (1) syllable weight, (2) the vowel, and (3) the surrounding consonants. Six words for each vowel phoneme (i.e., /a, i, u/) were chosen for each of the above factors (6*3*3 = 54). It is noteworthy, however, that syllable weight has not been suggested as a significant factor for stress in any Malay language (cf. Cohn 1989, 1993). The purpose in choosing words with similar syllable weight is to keep the environments of the penult and ultima as similar as possible. To militate against microprosodic effects of surrounding consonants, examples where the target vowel was surrounded by the same obstruent consonants were sought out first (e.g., [kuku] 'fingernail' and [tuntʊn] 'watch'). This, however, was not always possible. All else being equal, words with voiceless obstruents or nasal consonants were preferred, and only in one case does the wordlist depart from this preference. Finally, in a number of cases, it was not possible to find words where the vowels in the penultimate and ultimate syllable were either identical or had the appropriate surrounding consonants. In these cases, only one of the syllables was included in the corpus; these words are in gray in the corpus, which is presented in Tables 5-7 in Appendix A. Measurements and acoustic analysis of F0, duration, and intensity, were carried out using the speech analysis software Praat (Boersma & Weenink 2010). Following Turk et al. (2006), vowels in each of the three tokens were segmented at the onset and offset of a continuous second formant. All measurements were conducted using a script in Praat. F0 was measured separately for male and female participants as the average over a 30 ms window in the center of the vowel. Intensity was averaged over the entire duration of the vowel. The automated measurements resulted in a small number of outlier values, which were corrected by remeasuring the vowels manually.

The data were analyzed with a mixed-effects logistic regression.[11] The dependent variable SYLLABLE has two levels *penult* and *ultima*. The predictor variables in this study include the acoustic measurements discussed in the previous section (F0, DURATION, and INTENSITY) as well as the following control variables: SPEAKER, ITEM, VOWEL (the tar-

[11] The open-source programming language *R* (3.0.1 patched, R Core Team, 2013) was used for the statistical analyses. The *lme4* package was used for the mixed-effects modeling (Version 0.999999-2, Bates et al., 2013), and the *effects* package was used for creating plots (Version 2.2-4, Fox et al. 2013).

geted vowel), PRESENTATION ORDER (the order in which a new word was presented to the speaker), REPETITION ORDER (the order in which the token of a word is repeated by the speakers), and CONDITION (whether the token was uttered in isolation or a frame). The independent variable VOWEL has three levels representing each vowel phoneme considered /a, i, u/. To test whether the acoustic measurements (F0, DURATION, and INTENSITY) are correlated with a final stressed syllable, a mixed-effects logistic regression was fitted given the predictors above. If stress, associated with the acoustic measurements, falls on the final syllable, the level of the dependent variable *ultima* should be predicted (with a higher predicted probability) as each of the acoustic measures increases. However, if stress is not cued by the predictor variables and/or does not fall on the ultima, then the model would predict *penult* or both levels of the dependent variable equally.

The decision to consider phonemes instead of allophones was necessary for the statistical analysis because some allophones (i.e., [ɪ, ʊ, ɨ]) only occur in the ultimate syllable and are thus necessarily predictive of the ultimate syllable position. To control for individual differences between participants and words in the corpus, the variables SPEAKER and ITEM, respectively, were included as random intercepts. A likelihood ratio test showed that both random intercepts were fully justified. Originally, a single model was fitted for both male and female participants with the predictor variable GENDER, both as a main effect and interacting with all other predictors, using a backwards model selection process, excluding all non-significant predictors in a step-wise fashion. Aside from the main predictor variables, two-way interactions between the measurement variables (F0, DURATION, INTENSITY) and the control variables CONDITION and VOWEL, respectively were included in the model to control for differences that result from vowel quality and/or contextual phrasal factors. While the main effect GENDER was highly significant, it had a predicted probability of 100% for the *ultima* for male participants and a predicted probability of 0% for the *ultima* for female participants. In the end, a single model failed to adequately take into account the effect of GENDER, thus two separate mixed-effects regression models, one for male participants and one for female participants, were fitted with all same main predictors and interactions, except for GENDER to avoid any of the issue discussed above. The final models were chosen in the same way, using a backwards model selection process, excluding all non-significant predictors in a step-wise fashion. The next section reports the final results of the two models.

4. Results

The final models for both male and female participants show a highly significant correlation between the predictor variables and the dependent variable SYLLABLE. Both models are consistent with an analysis of Besemah stress in which stress falls on the final syllable of the word. The significant predictors for the model of male participants are the main effects F0 and INTENSITY as well as the interaction DURATION:CONDITION. In addition, the control variable PRESENTATION ORDER is significant. The final overall model for male participants is summarized in Table 3. The estimate for CONDITION*frame* is in relation to CONDITION*isolation*, which is the intercept. In addition, the final model includes two effects with random intercepts, SPEAKER (*Variance* = 0.639, *sd* = 0.799) and ITEM (*Variance* = 2.489, *sd* = 1.577).

Table 3. Fixed-effect coefficients, standard error, z-value and p-values in a generalized mixed-effects model for four male participants fitted to SYLLABLE, *penult* or *ultima* (Log likelihood ratio= -210.3, Dxy = 0.917, C = 0.958)

	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	-42.062104	6.716011	-6.263	$3.78e-10$	***
F0	0.116925	0.017555	6.660	$2.73e-11$	***
DURATION	82.232889	9.953877	8.261	$< 2e-16$	***
INTENSITY	0.226373	0.080633	2.807	0.004994	**
CONDITION*frame*	5.549308	1.337956	4.148	$3.36e-05$	***
PRESENTATION ORDER	0.014870	0.004035	3.686	0.000228	***
DURATION:CONDITION*frame*	-28.594746	14.421921	-1.983	0.047398	*

The significant predictors for the model of female participants are the interactions DURATION:CONDITION, DURATION:VOWEL, INTENSITY:CONDITION, and INTENSITY:VOWEL, and F0:CONDITION. The final overall model for female participants is summarized in Table 4. The estimate for CONDITION*frame* as well as VOWEL*i* and VOWEL*u* are in relation to CONDITION*isolation* and VOWEL*a*, respectively, both of which are the intercept. In addition, the final model includes two effects with random intercepts, SPEAKER (Variance = 9.913, SD = 3.148) and ITEM (*Variance* = 23.108, *sd* = 4.807). The final model for female participants is presented in Table 4.

Table 4. Fixed-effect coefficients, standard error, z-value and p-values in a generalized mixed-effects model for four female participants fitted to SYLLABLE, *penult* or *ultima* (Log likelihood ratio= -96.8, Dxy = 0.993, C = 0.996)

	Estimate	Std. Error	z value	Pr(> $\|z\|$)	
(Intercept)	−108.58650	27.75099	-3.913	$9.12e-05$	***
F0	0.11617	0.02364	4.914	$8.93e-07$	***
DURATION	190.38479	30.80847	6.180	$6.43e-10$	***
INTENSITY	0.73458	0.33332	2.204	0.027538	*
VOWEL*i*	17.86984	31.16176	3.202	0.566337	
VOWEL*u*	93.80502	29.29802	2.853	0.001366	**
CONDITION*frame*	−187.60561	50.86739	−3.688	0.000226	***
DURATION:VOWEL*i*	−22.16382	33.30613	−0.665	0.505758	
DURATION:VOWEL*u*	−80.02277	25.33431	−3.159	0.001585	**
INTENSITY:VOWEL*i*	−0.17307	0.40923	−0.423	0.672351	
INTENSITY:VOWEL*u*	−1.06439	0.37631	−2.828	0.004677	**
F0:CONDITION*frame*	0.08814	0.04899	1.799	0.071990	.
DURATION:CONDITION*frame*	−47.34075	19.10402	−2.478	0.013210	*
INTENSITY:CONDITION*frame*	2.34703	0.57351	4.092	$4.27e-05$	***

Each of the significant main effects and the interactions into which each enter are discussed separately according to the measurement variable in each of the next subsections. For each significant effect or interaction, the χ^2, degrees of freedom, and p-values are reported from an ANOVA table that compares the model with and without the significant predictor using the χ^2 test. The mean and standard variation for each of the main predictors are presented in Appendix B.

4.1 Duration

DURATION is a robust effect for predicting the *ultima*, however, as is well documented in the literature, it is confounded by word-final lengthening (Wightman et al. 1992; Turk & Shattuck-Hufnagel 2000). Thus, it is impossible to know if the longer duration is a result of word-level stress, word-final lengthening, or a combination of both. For completeness, the results of DURATION are reported, but because stress is posited to be word-final, it cannot in principle be used as evidence for (or against) word-level stress. Not surprisingly, then, the main effect DURATION is highly significant for both male participants ($\chi^2 = 77.7185$, $df = 1$, $p < 2.2e-16$) and female participants ($\chi^2 = 32.6552$, $df = 1$, $p = 1.100e-08$). What is of interest is the fact that there is a significant interaction DURATION:CONDITION for both male ($\chi^2 = 4.0657$, $df = 1$, $p = 0.04376$) and female participants ($\chi^2 = 5.7574$, $df = 1$, $p = 0.01642$). The interaction DURATION:CONDITION for males and females is shown in Figures 3 and 4, respectively.

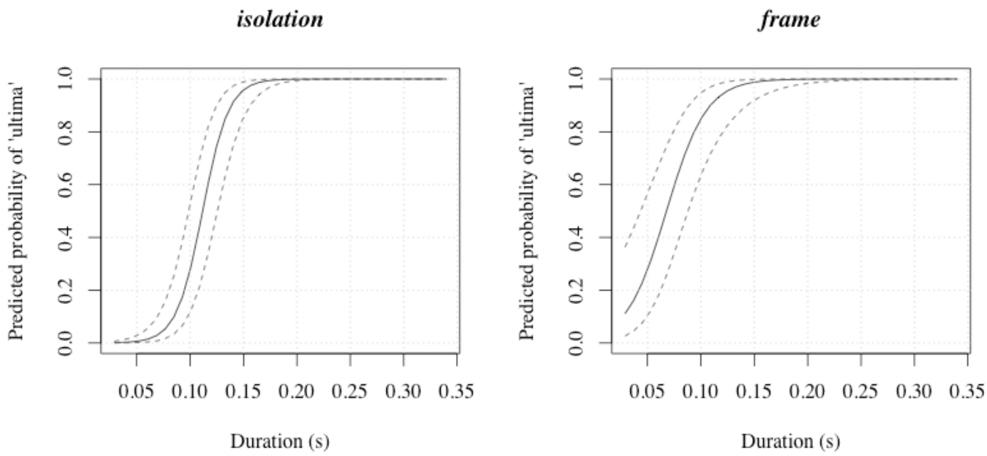

Figure 3. The interaction DURATION:CONDITION for male participants

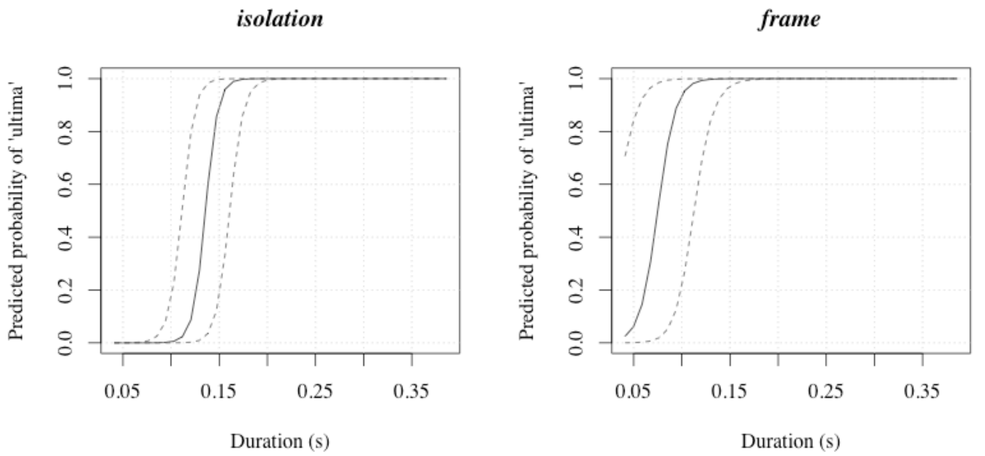

Figure 4. The interaction DURATION:CONDITION for female participants

The plots show the predicted probability of *ultima* on the y-axis and the duration on the x-axis. It is clear in each case that as duration increases, the predicted probability of *ultima* also increases. However, this differs between ISOLATION in the left panel and *frame* in the right panel. To illustrate this difference, for male participants the predicted probability crosses the 50% threshold at 0.11s in isolation and 0.06s in a frame. Likewise, for female participants the predicted probability crosses the 50% threshold at 0.13s in isolation and 0.07s in a frame. The plots reveal that while male and female speakers differ slightly, the same pattern holds whereby the predicted probability of *ultima* in a frame increases more quickly (e.g., with shorter durations) as compared to words uttered in isolation. The significant interaction may suggest the presence of phrase-final lengthening in isolated forms that is not present in words uttered in a frame.[12]

Female participants additionally showed a significant interaction DURATION:VOWEL ($\chi^2 = 14.681$, $df = 2$, $p = 0.0006486$). Surprisingly, this interaction does not appear to be reflective of the intrinsic durational properties of different vowel qualities (Beckman 1986), where low vowels (e.g., /a/) are expected to be longer than high vowels (e.g., /i,u/). Conversely, in this case, /u/ is significantly different than /a,i/. Figure 5 shows that the predicted probability for /u/ begins to rise slightly earlier than /a,i/ and does not have as steep of a slope as /a,i/. While it is not entirely clear why the interaction DURATION:VOWEL is significant, it appears that one explanation could lie in the larger number of outliers for /u/ than for /a,i/. As duration is confounded by word-final lengthening, this issue is left aside to focus on other predictors F0 and INTENSITY.

4.2 Fundamental frequency

Fundamental frequency is a good indicator of stress for both male and female participants. The main effect F0 is highly significant for male participants ($\chi^2 = 57.073$, $df = 1$, $p = 4.199e - 14$). Figure 6 presents the main effect F0. As is the case with duration, the predicted probability of *ultima* increases as F0 increases, meaning the ultima is predicted to have a higher F0 than the penult. To illustrate this point, the predicted probability of *ultima* crosses the 50% mark at approximately 118 Hz. The fact that F0 does not interact with CONDITION shows that there is not a significant difference between words in isolation and a frame. This finding is pertinent to the discussion of word- and phrase-level prominence in Section 5, as it points to the fact that words in isolation are not significantly affected by phrase-level prominence.

This is not necessarily the case for female speakers for whom the interaction F0:CONDITION is significant ($\chi^2 = 4.9508$, $df = 1$, $p = 0.02608$). Figure 7 presents the interaction F0:CONDITION for female participants. The plots in Figure 7 demonstrate that as pitch increases so does the predicted probability of *ultima* in both conditions. However, for words in isolation in the left panel the predicted probability of the *ultima* increases with a much higher F0 than for words uttered in a frame in the right panel. As a frame of reference, the

---

[12] An anonymous reviewer points out that the difference of 0.5s or 0.6s is rather minor. The reviewer further suggests that there may be other possible explanations for this interaction, including general compression of a word in a frame context. This is entirely possible. A follow-up study involving words in two different frames where the word occurs in phrase-medial and phrase-final contexts would be necessary to tease this apart. It is important to keep in mind, however, that duration cannot in principle be considered evidence for word-level stress.

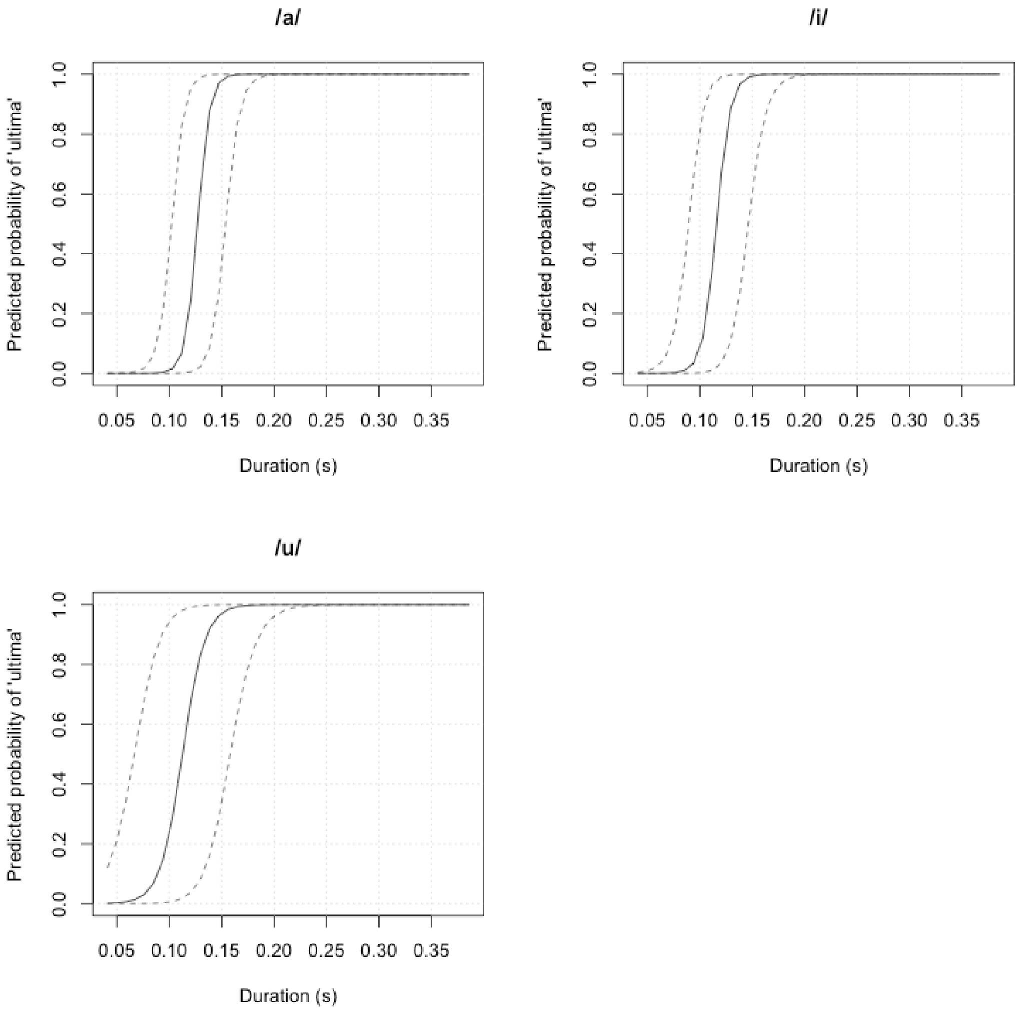

Figure 5. The interaction DURATION:CONDITION for female participants

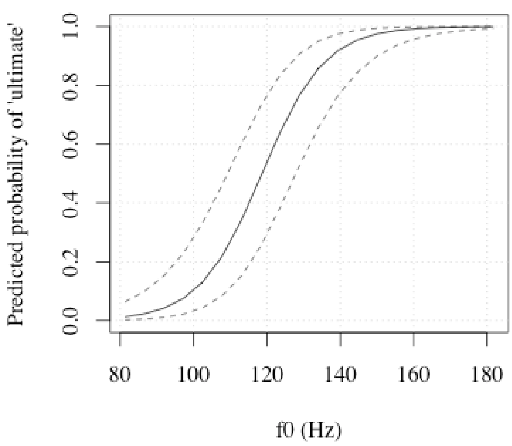

Figure 6. The main effect F0 for male participants

predicted probability of *ultima* in *isolation* crosses the 50% mark at approximately 224 Hz, and in a *frame* at approximately 188 Hz.

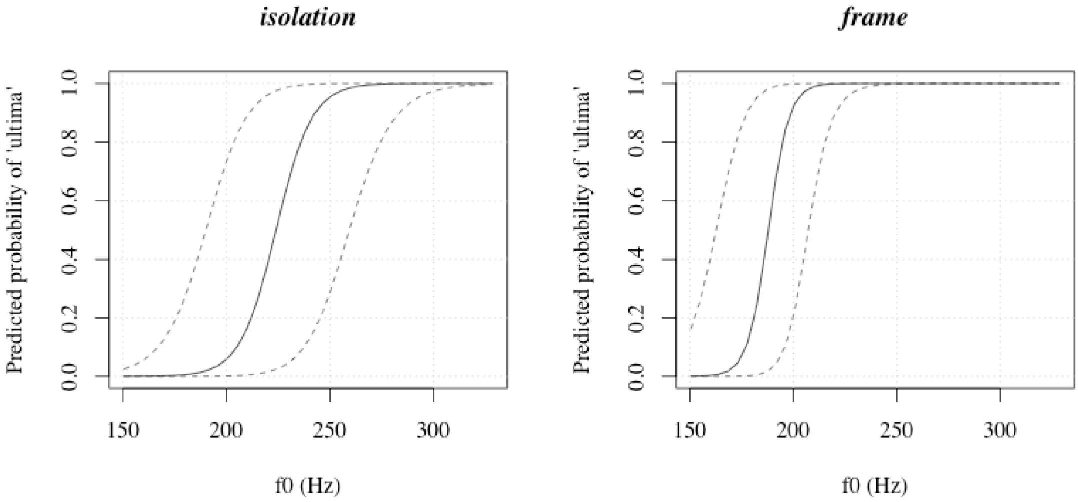

Figure 7. The interaction F0:CONDITION for female participants

The fairly large difference between F0 in a frame and in isolation appears to result from inter-speaker variation in realization of F0 in isolation, but not in a frame. In words uttered in isolation, one speaker (fs1) demonstrates a pattern whereby F0 is higher on the penult, a pattern that is not found in any of the other speakers. However, in words uttered in a frame context, all female speakers show the same pattern whereby F0 is higher on the final syllable. These patterns are shown in Figure 8.

There were two clear intonation patterns that the female speakers used with words in isolation. The female speaker (fs1) that differed from all others employed a 'final' intonation contour that is characterized as a falling F0 on the ultima. The other three female speakers used a 'continuing' intonation contour where that is characterized by a rising F0 or even a level F0 on the ultima. All speakers showed the same pattern in words uttered in a frame. When the word was uttered in a frame, the phrasal prosody does not clearly affect the target word, but rather the final word of the frame *agi* 'again'. Representative examples of the patterns discussed above and their corresponding forms in a frame are exemplified in Section 5.1 in which the interaction between word- and phrase-level prominence is discussed.

4.3 Intensity

The final predictor, intensity, is probably the best indicator of word-level stress for two reasons. First, intensity is not confounded with other factors as was the case with duration. Second, intensity is not as integrated as pitch with phrasal prominence. As is the case with F0, INTENSITY is straightforward for male participants, but not so straightforward for female participants. For male participants the main effect INTENSITY is significant ($\chi^2 = 8.325$, $df = 1$, $p = 0.00391$). Figure 9 shows that as the intensity increases, so does the predicted probability of *ultima*. As a frame of reference, the predicted probability reaches the 50% mark at approximately 77.5 dB.

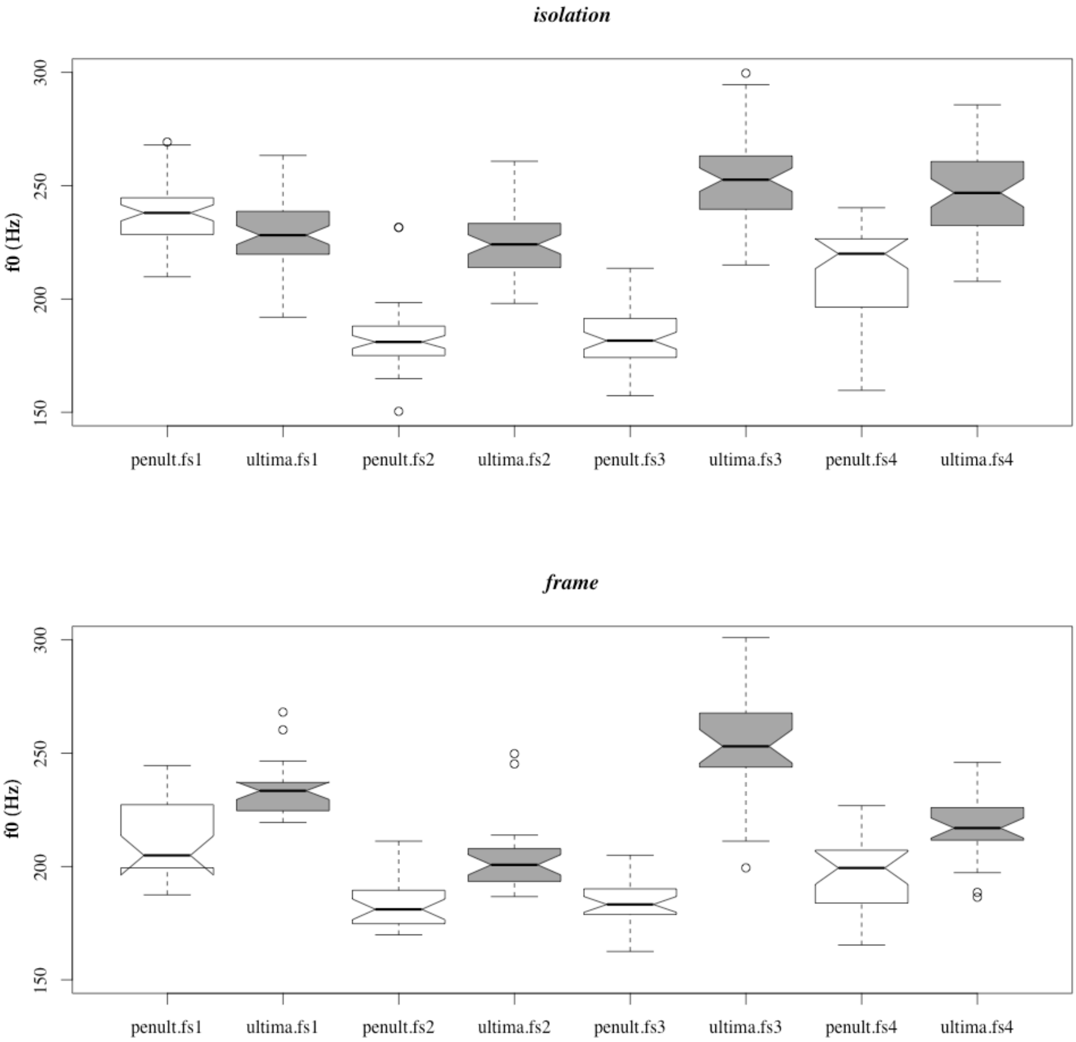

Figure 8. F0 values in the ultima and penult for each female participant in isolation and in a frame

Female participants had a highly significant interaction INTENSITY:CONDITION ($\chi^2 = 32.406$, $df = 1$, $p = 1.251e-08$) and a significant interaction INTENSITY:VOWEL ($\chi^2 = 11.11$, $df = 2$, $p = 0.003868$). The interaction INTENSITY:CONDITION in Figure 10 shows that in both cases intensity increases as the predicted probability of *ultima* increases. However, the 95% confidence intervals for words in isolation in the left panel are large, which calls the reliability of the result for words in isolation into question. For the words uttered in a frame in the right panel, the results show a very different pattern. There is a clear increase in the predicted probability of *ultima* with much smaller confidence intervals. For words in a frame, the predicted probability reaches the 50% mark at approximately 75 dB.

While it is not entirely clear why there are intensity differences between words uttered in isolation and those uttered in a frame, it is significant that words uttered in a frame show a clear indication of increased intensity on the final syllable. One possible explanation that could account for the large confidence intervals for words uttered in isolation is a wide

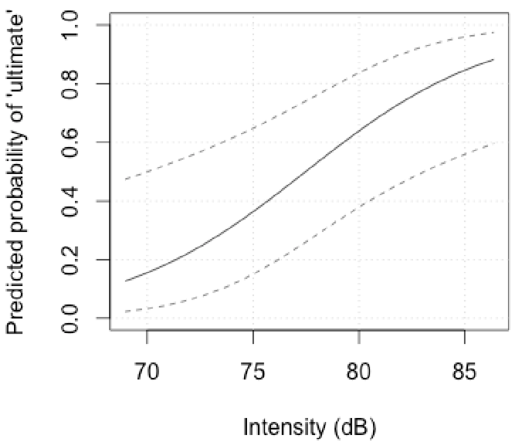

Figure 9. The main effect INTENSITY for male participants

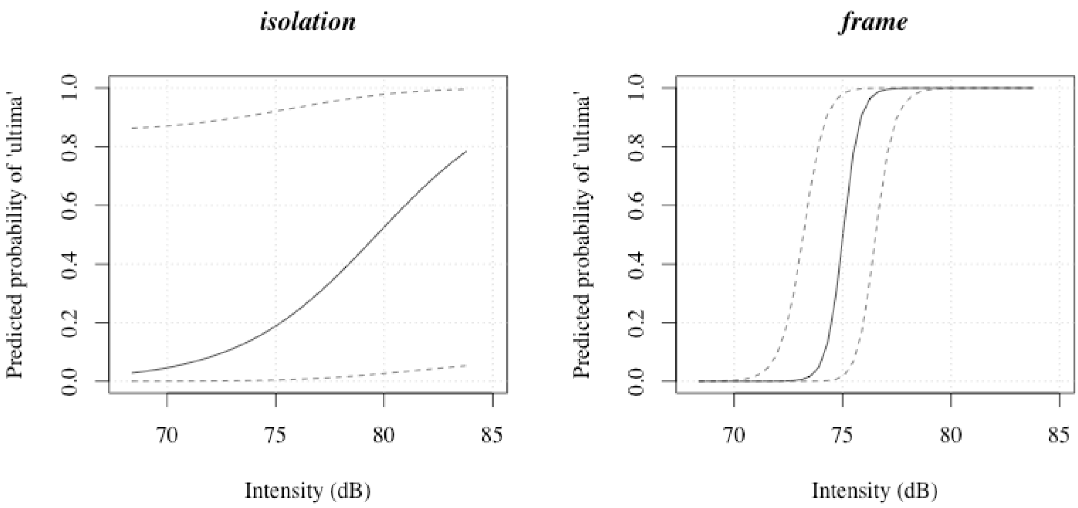

Figure 10. The interaction INTENSITY:CONDITION for female participants

range of intensity values in initial syllables. This, in turn, may result from utterance initial fortition effects (Cho & Keating 2009). That is, speakers were asked to repeat the word twice in isolation and once in a frame. While the control variable REPETITION ORDER was not significant in the model, it appears that the first repetition was quite different from the second and third repetitions shown in Figure 11. These box plots show that in the first repetition (1) the overall INTENSITY in both the penult and ultima is much higher than second and third repetitions and (2) the median intensity values were roughly the same in the penult and the final. In the second and third repetitions, on the other hand, the ultima has a higher median without any overlapping notches. All of this points to an utterance initial fortition effect that apparently skews the results of the interaction INTENSITY:CONDITION.

Finally, there is also a significant interaction INTENSITY:VOWEL ($\chi^2 = 11.11$, $df = 2$, $p = 0.003868$). Figure 12 shows that /a,i/ are significantly different from /u/; /u/ has a

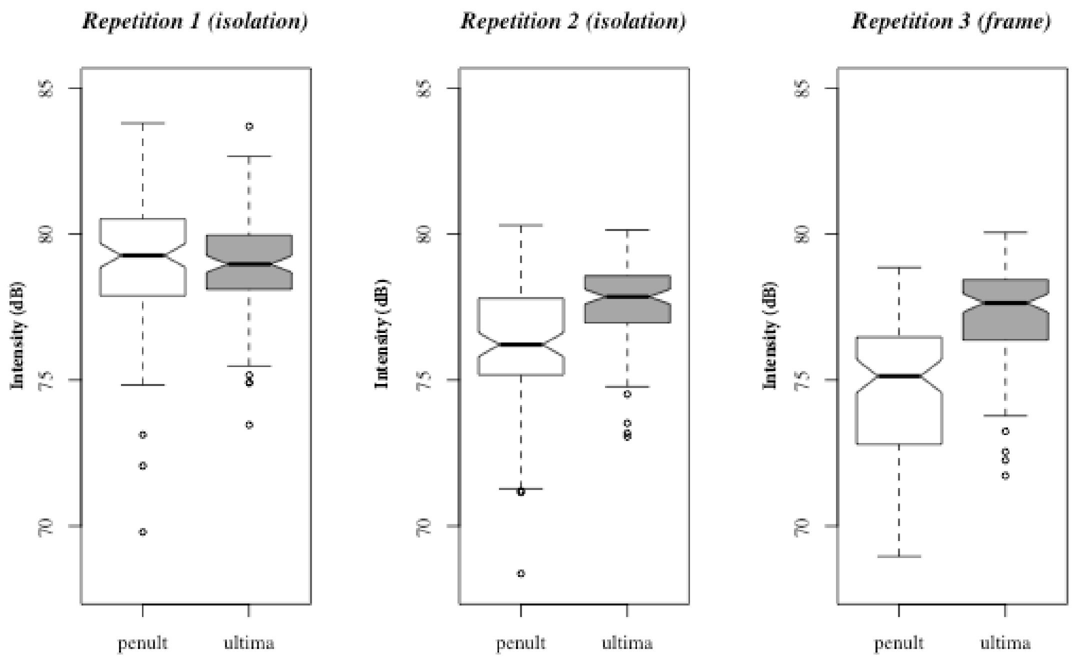

Figure 11. Three different repetitions of INTENSITY by vowel for female participants

more gradual slope than /a,i/. As was the case of the interaction DURATION:VOWEL, it is unclear why /u/ patterns differently than /a,i/, as this is not the expected result for intensity intrinsic properties of different vowel qualities.

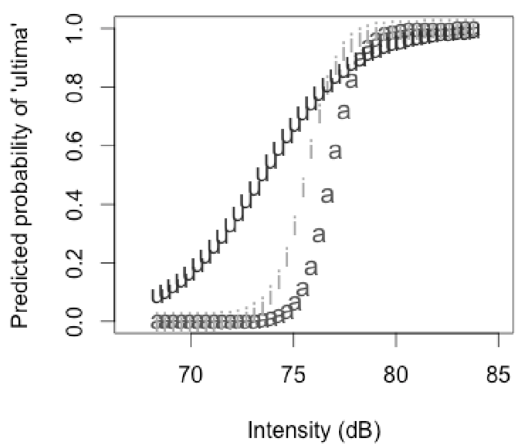

Figure 12. The interaction INTENSITY:VOWEL for female participants

4.4 Results summary

The results above support the proposal that Besemah has word-level stress that falls on the ultimate syllable, which is associated with raised intensity, higher fundamental frequency, and longer duration (although the inference from duration is confounded by word-final lengthening). The results for male speakers are straightforward, where all three acoustic measures were significant predictors and only DURATION participated in an interaction with CONDITION. The results for female speakers are less straightforward. All three measurement predictors participate in a significant interaction with the control variable CONDITION. Further, DURATION and INTENSITY, but not F0 interact with the control variable VOWEL as well. Of the significant interactions, it is noteworthy that while the acoustic measures are not always robust in isolation, they are significant in a frame. This result is consistent with an analysis of word-level stress and is especially significant for INTENSITY.

5. Discussion

In Section 1, two competing views of stress in Malay languages were presented. The first, the 'traditional' view, holds that Malay languages (especially Indonesian) have word-level stress that falls on the penultimate syllable unless the penult contains a schwa, in which case stress falls on the final syllable. The second view, the 'free' stress (or no stress) view, has called into question the 'traditional' analysis by showing that prominence is essentially free at the word-level (or more correctly, determined phrasally) in a number of production and perception studies (van Zanten 1994; Ebing 1997; van Zanten & van Heuven 1998; van Zanten et al. 2003; van Zanten & van Heuven 2004; Goedemans & van Zanten 2007). These studies find that what has been labeled as word-level stress results from higher phrase-level factors and word-level stress may be variable across speakers with different substrate influences (i.e., Javanese speakers of Indonesian do not mark word-level stress, while Toba Batak speakers of Indonesian do). This view of 'free' stress is not limited to Javanese speakers of Indonesian, but has been shown to occur in Betawi Malay, another Malay variety spoken in and around Jakarta (van Heuven et al. 2008). The present study set out to determine if the 'free' stress analysis could be extended to other Malay varieties, in particular Besemah, a historically conservative Malay language of southwest Sumatra.

Based on a statistical analysis of three acoustic properties (DURATION, F0, and INTENSITY), this study provides support for the view that stress falls on the ultima and is associated with a higher fundamental frequency, increased intensity, and longer duration (although duration is a confound with word-final lengthening). Therefore, Besemah is not classified as a language that lacks word-level stress like French and Korean. Drawing on the results of the statistical analysis as well as examples from naturalistic speech, the next section addresses the relationship between word- and phrase-level prominence in Besemah. However, without a fuller description of the intonational system, the following section draws only some preliminary conclusions about the interaction between word- and phrase-level prominence in the language.

5.1 Word-level and phrase-level prominence

While the only significant interaction between acoustic measures and the phrasal context in which the word was uttered (e.g., isolation vs. frame) in the model for male speakers is DURATION:CONDITION, the model for female speakers had interactions between all three measurement variables and the phrasal context (i.e., F0:CONDITION, INTENSITY:CONDITION, and DURATION:CONDITION). The interaction DURATION:CONDITION can be accounted for by phrase-final lengthening in both models (see Section 4.1), and the interactionINTENSITY:CONDITION in the model for female participants appears to result from utterance initial fortition effects (see Section 4.3). With the importance of F0 for phrasal prominence (Pierrehumbert 1980; Ladd 2008), one wonders why the interaction F0:CONDITION arises in the model for female speakers, but not in the model for male speakers. The data reveal that the answer to this question is most likely found in the interaction between word-level stress and phrase-level prominence. Section 4.2 demonstrates that one female speaker (fs1) appears to use a 'final' intonation pattern characterized by a falling F0 contour on the final syllable in words in isolation. All other female speakers appear to use a 'continuing' intonation pattern characterized by a level or rising F0 contour on the final syllable in words in isolation.

Utilizing notions from the AM framework, the 'final' and 'continuing' contours are best analyzed as a combination of pitch accents and boundary tones (Ladd 2008). The 'final' intonation contour is then the result of a high pitch accent (H*) and a low boundary tone (L%), and the 'continuing' intonation contour is the result of a high pitch accent (H*) and a high boundary tone (H%). Figures 13–15 demonstrate these intonational patterns from three different female speakers. Figure 13 presents a representative example from fs1 who consistently used the 'final' intonation pattern with a falling F0 on the ultima (i.e., a H* followed by L%). Figures 14 and 15 demonstrate the two patterns found under what might be considered a 'continuing' intonation contour with a higher F0 on the ultima that either rises or remains level. In the case of the latter, it is unclear whether or not the high boundary tone is present. It may be the case that F0 remains level from the H* to the H% or that the H% is absent in these cases.

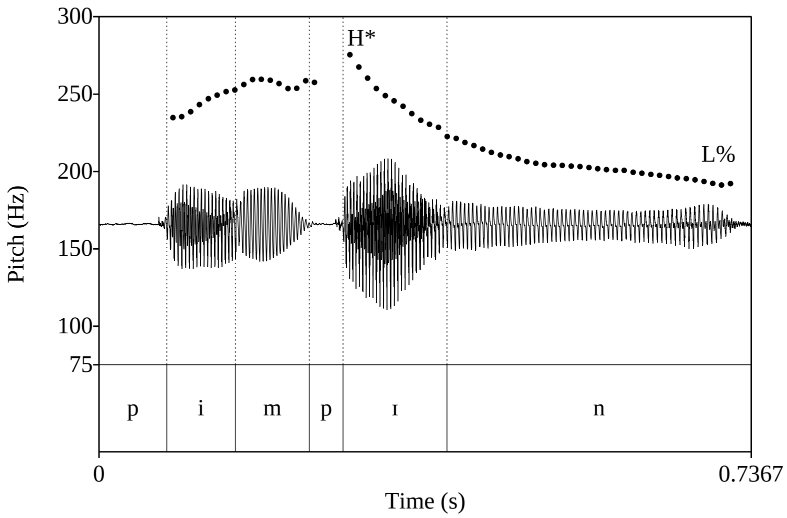

Figure 13. F0 trace for [pimpɪn] 'lead' and in a frame for fs1 using 'final' intonation

What is more, these intonation patterns are not present (crucially for fs1) when uttered in a frame as is demonstrated in Figure 16. It is especially interesting to note the sharp declination of F0 on the ultima of the final word of the frame [agi] 'again'. A comparison between the variable patterns in isolation and stable patterns in a frame point to the fact that the interaction between word- and phrase-level prominence does not appear to affect words uttered in a frame as much as words in isolation. Why then is there not a signifi-

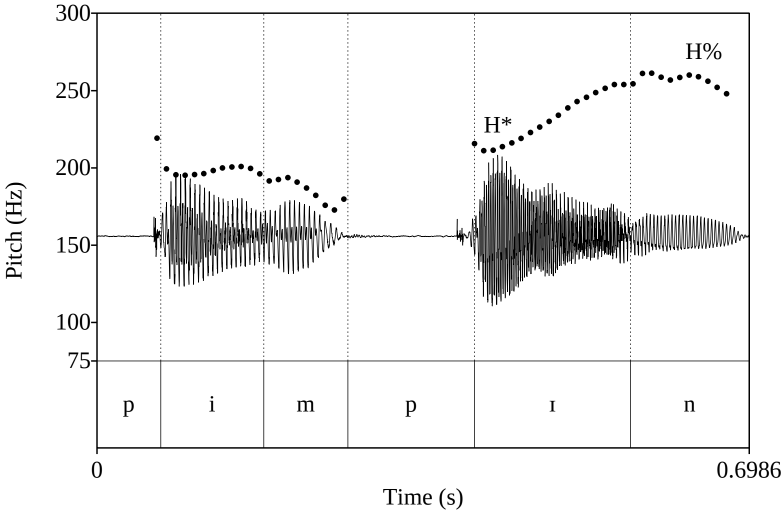

Figure 14. F0 trace for [pimpɪn] 'lead' and in a frame for fs2 using 'continuing' intonation

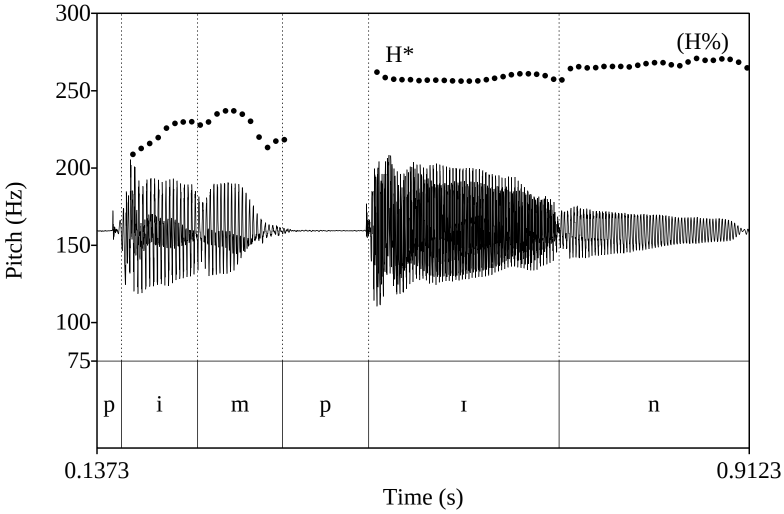

Figure 15. F0 trace for [pimpɪn] 'lead' and in a frame for fs4 using 'continuing' intonation

cant interaction between F0 and CONDITION in the model for male speakers? Based on the analysis above, and an inspection of the data, male speakers more consistently produced H% in words in isolation. In summary, words in isolation appear to be affected by boundary tones, in which L% conflicting with the H* results in a falling F0 contour on the ultima, while a H% results in a rising F0 contour on the ultima or possibly results in maintaining a level F0 contour. One way to interpret this variation in the realization of boundary tones in words in isolation is that the presence of L% or H% is up to the speaker's interpretation (and not the pragmatic conditions) of the single word phrase as 'continuing' or 'final'. This unpredictable result is most likely due to the fact that the words are uttered in an atypical communicative context. The only clearly 'final' intonation pattern comes at the end of the frame. (Remember that speakers were asked to repeat the word twice in isolation and once in a frame, then asked to do the same for a new word.)

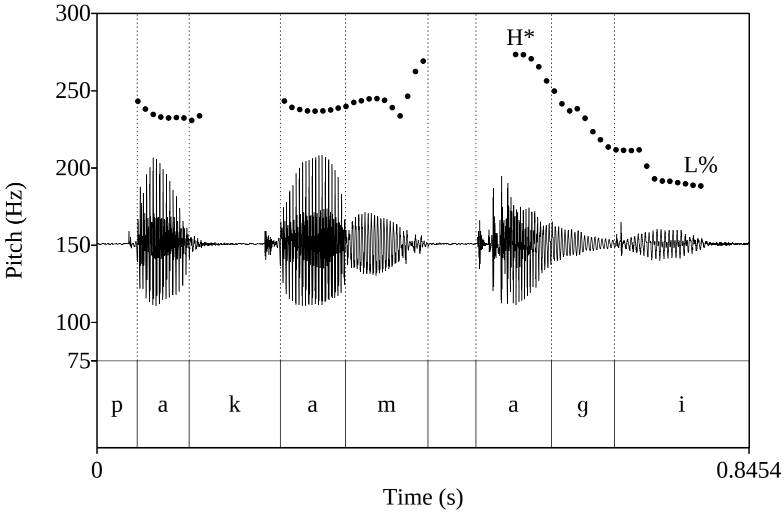

Figure 16. F0 trace for [pakam] 'sturdy' followed by the second half of the frame [agi] 'again' for fs1

One explanation for the intonational patterns presented above is tonal crowding avoidance, whereby the H* pitch accent timing shifts leftward, moving the pitch peak earlier in the ultima or even onto the penult (see below), when a L% is on the ultima. According to Gordon (2008), this timing change pattern is one of a number of strategies employed in the languages of the world to minimize tonal crowding. Other strategies include tonal undershoot for Modern Greek (Arvaniti & Baltazani 2005), segmental lengthening for Korean (Jun 2005) (see Gordon (2008) for more discussion and a detailed analysis of timing changes in Chickasaw). The conflict between high pitch accent and low boundary tone is also clear in the examples from naturalistic speech in Figure 17, where L% is realized on the ultima and H* is realized on the penult. In these data, the changes in the timing of H* is especially clear when there is a mismatch in the pitch accent and the boundary tone.

Finally, further support for the tonal crowding analysis of Besemah comes from words in phrase-final position where vowels are extra long as in Figure 18.[13] Because the vowels are long in these examples, there is no conflict between the pitch accent and the boundary tone, as both the H* and L% can be realized on the same syllable. Gordon finds similar effects with a weight asymmetry in Chickasaw, where a pitch accent and a boundary tone cannot dock on a syllable with a short vowel, but can dock on a syllable with a long vowel. The preliminary conclusion here is word-level stress and phrase-level accent in Besemah operate on orthogonal principles and are highly affected by boundary tones. This interpretation of the data provides a promising hypothesis that needs confirmation in future studies of Besemah prosody.

[13] Besemah does not have phonemic vowel length nor does it have adjacent identical vowels. The long vowel in this example is attributed to phrase-final lengthening.

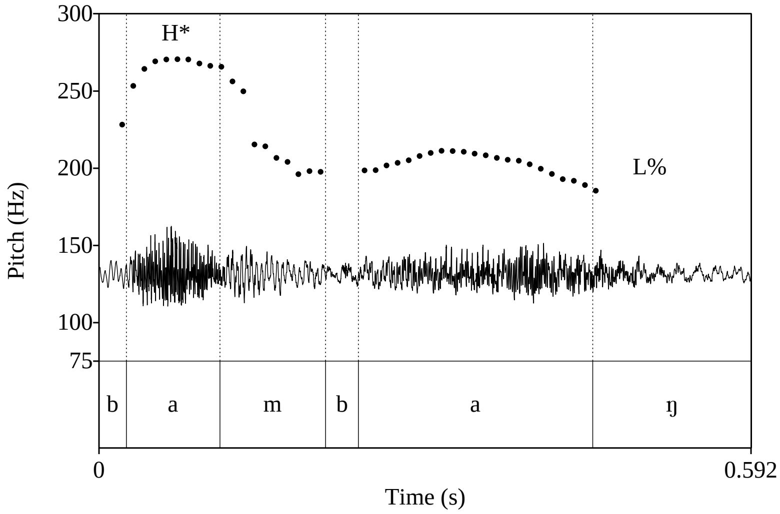

Figure 17. F0 trace of [bambaŋ] 'seedling' from a female speaker in naturalistic speech

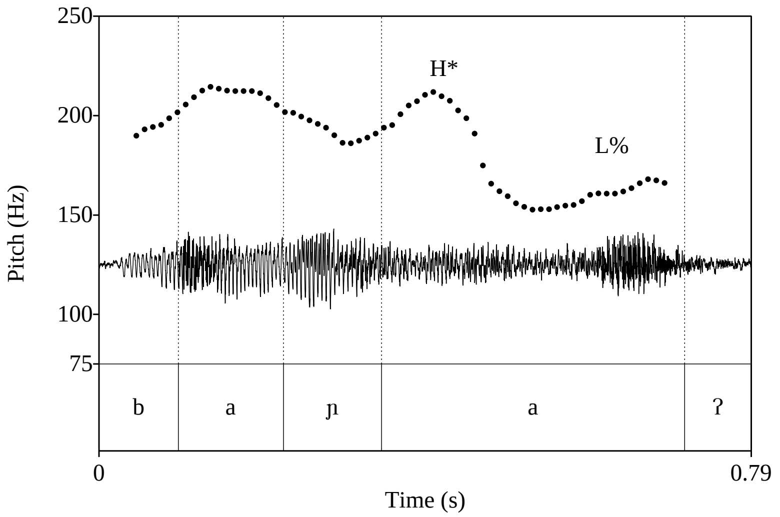

Figure 18. F0 trace of [baɲaːʔ] 'many' from a female speaker in naturalistic speech

6. Conclusion

In contrast to 'traditional' descriptions of Standard Indonesian and even Besemah as well as recent findings for Malay isolects like Standard Indonesian and Betawi Malay, this study has shown that there is clear evidence that Besemah stress falls on the final syllable of the word, associated with a higher fundamental frequency and increased intensity. While a higher fundamental frequency seems to be affected by phrase-final tonal crowding when words are uttered in isolation, increased intensity is a clearer indicator of word-level stress. This study has several implications for future studies of stress in Malay languages. Most importantly, it demonstrates that Malay languages do differ in their realization of stress, and that the conclusion drawn for varieties spoken in and around urban centers may not in fact be representative of other Malay varieties spoken throughout the Indonesian archipelago.

Appendix A. Corpus

Table 5. Tokens with two open syllables

	CV.CV		CV.**CV**	
/a/	[katɨ]	'word'	[katɨ]	'word'
	[sapɨ]	'who'	[sapɨ]	'who'
	[mati]	'die'	[kitɨ]	'we (inclusive)'
/i/	[kitɨ]	'we (inclusive)'	[mati]	'die'
	[titu]	'that'	[laki]	'husband'
/u/	[kutu]	'louse'	[kutu]	'louse'
	[kuku]	'fingernail'	[kuku]	'fingernail'
	[butɨ]	'blind'	[titu]	'that'

Table 6. Tokens with a final closed syllable

	CV.CVC		CV.**CVC**	
/a/	[tatap]	'touch'	[tatap]	'touch'
	[sakat]	'bother'	[sakat]	'bother'
	[pakam]	'sturdy'	[pakam]	'sturdy'
/i/	[kikɪs]	'scrape'	[kikɪs]	'scrape'
	[pipɪs]	'pulverize'	[pipɪs]	'pulverize'
	[tikʊs]	'mouse'	[musɪm]	'season'
/u/	[tutʊs]	'pound'	[tutʊs]	'pound'
	[musɪm]	'season'	[tikʊs]	'mouse'
	[sutar]	'k.o. game'	[takʊt]	'scared'

Table 7. Tokens with two closed syllables

	CVC.CVC		CVC.**CVC**	
/a/	[tandaŋ]	'sleep over'	[tandaŋ]	'sleep over'
	[kantɪn]	'friend'	[timpas]	'swim'
	[pantʊk]	'strike'	[antan]	'pestle'
/i/	[pimpɪn]	'lead'	[pimpɪn]	'lead'
	[timpas]	'swim'	[kantɪn]	'friend'
	[pintaʔ]	'request'	[intɪp]	'spy'
/u/	[tuntʊn]	'watch'	[tuntʊn]	'watch'
	[pumpʊŋ]	'float'	[pumpʊŋ]	'float'
	[tumpah]	'pour'	[bantʊt]	'cut'

Appendix B. Mean and standard deviation of main predictors

Table 8. Mean (*M*) and standard deviation (*SD*) of main predictors for female speakers ($n = 4$)

	Duration (s)		Intensity (dB)		F0 (Hz)	
	penult	ultima	penult	ultima	penult	ultima
M	0.09846	0.1633	76.6398	77.966	200.1942	234.9699
SD	0.02609	0.0624	2.889	1.7347	24.6273	24.2708

Table 9. Mean (*M*) and standard deviation (*SD*) of main predictors for male speakers ($n = 4$)

	Duration (s)		Intensity (dB)		F0 (Hz)	
	penult	ultima	penult	ultima	penult	ultima
M	0.0817	0.1238	77.5577	78.522	113.2832	126.2574
SD	0.0196	0.0499	2.5925	1.6111	11.9777	15.1723

References

Adelaar, K. Alexander. 1992. *Proto-Malayic: A reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.

Adelaar, K. Alexander & David J. Prentice. 1996. Malay: Its history, role and spread. In Stephen Adolphe Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, 6736–6793. Berlin: Mouton de Gruyter.

Adisasmito-Smith, Niken & Abigal C. Cohn. 1996. Phonetic correlates of primary and secondary stress in Indonesian: A preliminary study. *Working Papers of the Cornell Phonetics Laboratory* 11. 1–15.

Arvaniti, Amalia & Mary Baltazani. 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In Sun-Ah Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 84–117. Oxford: Oxford University Press.

Bates, Douglas, Martin Maechler & Ben Bolker. 2013. *lm4e: Linear mixed-effects models using s4 classes version 0.999999-2*.

Beckman, Jill N. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14(1). 1–46.

Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.

Boersma, Paul & David Weenink. 2010. Praat: doing phonetics by computer (version 5.1).

Caldecott, Marion C. 2007. Acoustic correlates of St'át'imcets stress. In *Proceedings of the 22nd northwest linguistics conference*, 1–9. Burnaby: Simon Frasier University.

Cho, Taehong & Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37(4). 466–485.

Cohn, Abigal C. 1989. Stress in Indonesian and bracketing paradoxes. *Natural Language and Linguistic Theory* 7(2). 167–216.

Cohn, Abigal C. 1993. The initial dactyl effect in Indonesian. *Linguistic Inquiry* 24(2). 372–381.

Cohn, Abigal C. & John J. McCarthy. 1998. Alignment and parallelism in Indonesian phonology. *Working Papers of the Cornell Phonetics Laboratory* 12. 53–137.

Doke, Clement M. 1931. *A comparative study of Shona phonetics*. Johannesburg: University of Witwatersrand Press.

Downing, Laura J. 2004. What African languages tell us about accent typology. In Susanne Fuchs & Silke Hamann (eds.), *ZAS papers in linguistics*, vol. 37, 101–136. Berlin: ZAS.

Downing, Laura J. 2010. Accent in African languages. In Harry van der Hulst, R. Goedemans & E. van Zanten (eds.), *A survey of word accentual patterns in the languages of the world*, 381–427. Berlin: Mouton de Gruyter.

Ebing, Ewald F. 1997. *Form and function of pitch movements in Indonesian*. Leiden: Research School CNWS.

Everett, Karen M. 1998. The acoustic correlates of stress in Pirahã. *The Journal of Amazonian Languages* 1(2). 104–162.

Fokker, Abraham Anthony. 1895. *Malay phonetics*. Leiden: E. J. Brill.

Fortune, George. 1980. *Shona grammatical constructions*, vol. 1. Harare: Mercury Press.

Fox, John, Weisberg Sanford, Hong Jangman, Robert Andersen, David Firth, Michael Friendly & Steve Taylor. 2013. *effects: Effect displays for linear, generalized linear, multinomial-logit, proportional-odds logit models and mixed-effects models version 2.2-4*.

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

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech* 1(2). 126–152.

Gil, David. 1994. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17(2). 179–200.

Goedemans, Robertus Wilhelmus Nicolaas & Ellen van Zanten. 2007. Stress and accent in Indonesian. In Vincent J van Heuven & Ellen van Zanten (eds.), *Prosody in Indonesian languages*, 35–62. Utrecht: LOT.

Gonzalez, Andrew. 1970. Acoustic correlates of accent, rhythm, and intonation in Tagalog. *Phonetica* 22(1). 11–44.

Gordon, Matthew K. 2004. A phonological and phonetic study of word-level stress in Chickasaw. *International Journal of American Linguistics* 70(1). 1–32.

Gordon, Matthew K. 2008. Pitch accent timing and scaling in chickasaw. *Journal of Phonetics* 36(3). 521–535.

Gordon, Matthew K. 2014. Disentangling stress and pitch accent: Toward a typology of prominence at different prosodic levels. In Harry van der Hulst (ed.), *Word stress: Theoretical and typological issues*, Cambridge: Cambridge University Press.

Gordon, Matthew K. & Ayla Applebaum. 2010. Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association* 40(01). 35–58.

Gordon, Matthew K., Edita Ghushchyan, Bradley McDonnell, Daisy Rosenblum & Patricia A Shaw. 2012. Sonority and central vowels: A cross-linguistic phonetic study. In Steve Parker (ed.), *The sonority controversy*, 219–256. Berlin: De Gruyter Mouton.

Gordon, Raymond G. (ed.). 2005. *Ethnologue: Languages of the world*. Dallas, Texas: SIL International 15th edn.

Halim, Amran. 1981. *Intonation in relation to syntax in Indonesian*. Canberra, Australia: Deptartment of Linguistics, Research School of Pacific Studies, the Australian National University.

Helfrich, Oscar Louis. 1904. *Bijdragen tot de kennis van het Midden Maleisch: (Běsěmahsch en Sěrawajsch dialect)*, vol. 53 Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen. Batavia: Landsdrukkerij.

van Heuven, Vincent J. & Vera Faust. 2009. Are Indonesians sensitive to contrastive accentuation below the word level? *Wacana, Jurnal Ilmu Pengetahuan Budaya* 11. 226–240.

van Heuven, Vincent J., Lilie Roosman & Ellen van Zanten. 2008. Betawi Malay word prosody. *Lingua* 118. 1271–1287.

Hintz, Daniel M. 2006. Stress in South Conchucos Quechua: A phonetic and phonological study. *International journal of American linguistics* 72(4). 477–521.

Hyman, Larry M. 2014. Do all languages have word accent? In Harry van der Hulst (ed.), *Word stress: Theoretical and typological issues*, Cambridge: Cambridge University Press.

Jun, Sun-Ah. 1996. *The phonetics and phonology of Korean prosody: Intonational phonology and prosodic structure*. London: Routledge.

Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15(2). 189–226.

Jun, Sun-Ah. 2005. Korean intonational phonology and prosodic transcription. In Sun-Ah Jun (ed.), *Prosodic typology : The phonology of intonation and phrasing*, 201–229. Oxford: Oxford University Press.

Jun, Sun-Ah & Cécile Fougeron. 2000. A phonological model of French intonation. *Intonation: Analysis, modeling and technology* 209–242.

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

Ladd, D. Robert. 2008. *Intonational phonology* Cambridge Studies in Linguistics. Cambridge: Cambridge University Press 2nd edn.

Laksman, Myrna. 1994. Location of stress in Indonesian words and sentences. In Cecilia Odé & Vincent J. van Heuven (eds.), *Experimental studies of Indonesian prosody*, Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.

Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35(1). 73–97.

Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the world*. Dallas, Texas: SIL International 16th edn.

Macdonald, Ross R. & Soenjonjo Darjowidjojo. 1967. *A student's reference grammar of modern formal Indonesian*. Washington, D.C.: Georgetown University Press.

McDonnell, Bradley. 2008. A conservative vowel phoneme inventory of Sumatra: The case of Besemah. *Oceanic Linguistics* 47(2). 409–432.

McDowell, Jonathan & Karl R. Anderbeck. 2008. Unity in diversity: Malayic varieties of Southern Sumatra. SIL, Jakarta.

Michelson, Karin E. 1988. *A comparative study of Lake-Iroquoian accent*. Dordrecht: Kluwer Academic Publishers.

Mintz, Malcolm W. 1994. *A student's grammar of Malay and Indonesian*. Singapore: EPB Publishers.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of english intonation*: Massachusetts Institute of Technology dissertation.

Potisuk, S., J. Gandour & M. P. Harper. 1996. Acoustic correlates of stress in Thai. *Phonetica* 53(4). 200–220.

Prentice, David J. 1990. Malay (Indonesian and Malaysian). In Bernard Comrie (ed.), *The world's major languages*, 913–935. London: Routledge.

Prentice, David J. 1994. Manado Malay: Product and agent of language change. In Tom Dutton & Darrell T. Tryon (eds.), *Language contact and change in the Austronesian world*, 411–441. Berlin: Mouton de Gruyter.

R Core Team. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria.

Sneddon, James N. 2003. Diglossia in Indonesian. *Bijdragen tot de Taal-, Land-en Volkenkunde* 159(4). 519–549.

Turk, Alice, Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: A practical guide. In Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter & Johannes Schließer (eds.), *Methods in empirical prosody research*, 1–28. Berlin: Mouton de Gruyter.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28(4). 397–440.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91. 1707–1717.

van Wijk, D. Gerth. 1883. *Spraakleer der Maleische taal*. Batavia: G. Kolff.

Wolff, John U. 1986. *Formal Indonesian*. Ithaca: Cornell University Press.

van Zanten, Ellen. 1986. Allophonic variation in the production of Indonesian vowels. *Bijdragen tot de Taal-, Land-en Volkenkunde* 142(4). 427–446.

van Zanten, Ellen. 1994. The effect of sentence position and accent on the duration of Indonesian words: A pilot study. In Cecilia Odé & Vincent J. van Heuven (eds.), *Experimental studies of indonesian prosody*, 140–80. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.

van Zanten, Ellen, Robertus Wilhelmus Nicolaas Goedemans & Jos J.A. Pacilly. 2003. The status of word stress in Indonesian. In Jeroen van de Weijer, Vincent J. van Heuven & Harry van der Hulst (eds.), *The phonological spectrum: Suprasegmental structure*, 151–175. Amsterdam: John Benjamins.

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* 140(4). 497–521.

van Zanten, Ellen & Vincent J. van Heuven. 1998. Word stress in Indonesian: Its communicative relevance. *Bijdragen tot de Taal-, Land-en Volkenkunde* 154(1). 129–144.

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.