Chapter five

Melodic structure in Toba Batak and Betawi Malay word prosody [32]

Lilie Roosman

Program Studi Belanda, Universitas Indonesia, Jakarta
Leiden University Centre for Linguistics

5.1 Introduction

5.1.1 Background

In languages with word stress, one (and only one) syllable is perceived by native listeners of the language as stronger than the other syllables in the same word. On the higher levels, in phrases or sentences, accent is used to make particular words more prominent than other words. In stress languages, the sentence accent typically coincides with the word stress. Languages without word stress may also use accent to highlight words in sentences but then the sentence accent is not restricted to a particular syllable in the word. Indonesian has been claimed to be one such language (van Heuven & van Zanten 1997; van Zanten, Goedemans & Pacilly 2003).

Our study focuses on the production of word prosody in Toba Batak (TB), spoken on the Island of Sumatra, and Betawi Malay (BM), spoken on the Isle of Java. TB is a language with word stress (van der Tuuk 1971, Nababan 1981). BM, on the other hand, is a language – just like Standard Indonesian – that does not have word stress (Muhadjir 1977) although it does have phrasal accent (Wallace 1976). We will compare these two different Austronesian languages in their realization of accent.

When a language does not use lexical tone, accent is primarily marked by pitch. A communicatively important (‘in focus’) word will bear a perceptually prominent change in pitch which is typically omitted on non-prominent (out of focus) words. Such accented words also have longer duration than their unaccented counterparts. In this article we will concentrate on the use of speech melody, rather than on duration, as a correlate of stress and accent. For a detailed analysis of the

[32] This research was supported by the Royal Netherlands Academy of Arts and Sciences (KNAW) under project 95-CS-05, by the Netherlands Organisation for Research (NWO) under project 370-75-001, by the Dutch Language Union (NTU) and by the Leiden University Fund (LUF).

role of duration in the two target languages we refer to Roosman (2006). If it is indeed true that BM has no word stress, we expect the exact position of the pitch change marking focus to be rather variable. When a language has lexical stress, as in TB, then the accent-lending pitch change will be tied rather strictly to one specific syllable in the word. In the present study we will test the hypothesis that BM and TB differ in loose versus strict alignment of the accent-lending pitch change in words spoken in focus.

Next to accent, we will investigate boundary marking in TB and BM. The marking of prosodic boundaries, signaling breaks between clauses and utterances, may also involve both duration and pitch. Boundaries may be signaled by boundary tones and words in sentence-final position are usually longer than words in other positions (van Heuven 1994 and references therein). In most languages clause boundaries are marked by a high tone (H%), indicating that the utterance is not finished, whilst completed utterances end on a low terminal tone (L%). In our study we will concentrate on the choice of boundary tone (H% versus L%) and see how these boundary tones may interact with accent-marking tone configurations in BM and TB.

This article is organized as follows. In section 5.2 we will describe the experimental design used to record materials in TB and BM. Section 5.3 specifies the acoustic and perceptual analysis of the materials, and presents the results. Section 5.4 links the results to the more general questions formulated in the introduction above. However, before turning to the experimental work, we will first introduce the two target languages and discuss claims made in the literature with respect to their prosodic structure,

5.1.2 Betawi Malay

Betawi Malay (BM), the dialect of the central part of the city of Jakarta (Dialek Kota), is used by a homogeneous ethnic group, the Betawi, and it has had comparatively little influence from other languages. BM belongs to the Malayic subgroup of the Western Malayo-Polynesian branch of the Austronesian language family (Adelaar 2005). BM is genealogically close to Standard Indonesian (SI). These language varieties certainly seem to resemble each other prosodically. For both SI and BM there is discussion whether they have lexical stress or phrasal accent only. On the strength of the claim that the prosodic systems of BM and SI are essentially the same we will draw on publications on either language variety for a short overview of claims regarding stress and accent in either language.

Gerth van Wijk (1985, first published in 1883) claimed that Indonesian has stress but observed that it is usually very weak. All syllables are pronounced with approximately the same emphasis. Stress generally falls on the pre-final syllable of a root. If the pre-final syllable is an open syllable and contains a schwa, the stress falls on the final syllable, unless the onset of the final syllable is *ng* [ŋ], in which case stress falls on the pre-final syllable with schwa. Words with schwa in the pre-final syllable are thus pronounced as follows: *déndam, sémpit; terús, besár; déngan, béngis* (Gerth van Wijk 1985: 45–46).

Fokker (1895) claimed that – phonologically – there is no word stress in Malay. Phonetically, in two-syllable stems, both syllables have virtually the same amount of stress. However, Malay does have accent, which is signaled by duration. Accent is on the penultimate syllable, except if this syllable contains a schwa. Importantly, melodic variations are not analyzed by Fokker as a reflection of prominence whether at the word or at the sentence level.

Samsuri (1971) did research on the prosody of SI spoken by speakers from different language backgrounds. He also claims that SI has no distinctive stress; whatever the position of the prominent syllable in the word, the meaning of the word is the same. However, he found that the last syllable in a word or phrase is the most prominent one. On the other hand, in two- or three-syllable words without schwa the penultimate syllable is generally higher in pitch than the other syllables (i.e. *náma* 'name', *méja* [meja] 'table', *móbil* 'car', *usía* 'age', *seléra* [səlera] 'appetite'.[33] When the penultimate syllable contains a schwa and the final syllable does not, the last syllable is higher in two-syllable words (*senáng* [sənaŋ] 'happy', *jemú* [jəmu] 'bore'). But in three-syllable words, the first syllable can also be higher. Besides *karená* [karəna] 'because', *majemúk* [majəmuk] 'plural', also *sútera* [sutəra] 'silk' and *pútera* [putəra] 'son' occur.

According to Halim (1974: 111–113), prominence depends on the position of the word in the sentence: before a sentence-internal boundary the stress falls on the final syllable of the word preceding the boundary, whereas sentence-final stresses fall on the penultimate syllable of the last word of the sentence.

Moeliono & Dardjowidjojo (1988) state there is always one word in an utterance that is accented. That word is then highlighted by loudness, duration and pitch movement. Alieva, Arakin, Ogloblin & Sirk (1991: 34) also claim that there is no phonological word stress in SI. However, there are always syllables in sentences that are highlighted or pronounced with higher intensity and thus are louder and clearer than the other syllables in the sentence, or that have a particular melody and a higher pitch, or that are longer. The ways in which those accented syllables are realized depend on the intonation pattern of the sentences. Zubkova, (1971, in Alieva et al. 1991: 62) observes the way in which syllables are highlighted in disyllabic words. She concludes that pitch and vowel intensity are not important for word stress. Also, differences in duration between both vowels are small and inconsistent. A production experiment by Pavlenko (1969, in Alieva et al. 1991: 62–63) shows that intensity is not important.

Most authors thus claim that stress in SI is either weak or non-existent. Nevertheless, there is a group of authors who formulated rules for the placement of word stress in (Standard) Indonesian. These rules have, in fact, recently been reiterated by Cohn (1989) and Cohn & McCarthy (1994), working in a metrical framework: stress is on the penultimate syllable, unless this syllable contains a schwa, regardless of the morphological structure of the word. However, experimental work by Laksman (1994) provides evidence that schwa can be stressed. Experiments by van Zanten & van Heuven (1998, 2004) found no preferred stress position in SI. Similarly, van Zanten, Goedemans & Pacilly (2003) conclude

[33] In all examples quoted from Samsuri (1971) the acute accent denotes 'high pitch'. Most likely, high pitch should also be taken as stressed.

on the basis of experimental evidence that SI does not have word-based stress, but has phrase-level accent only.

The following description, specifically of BM prosody, is based on Wallace (1976), who notices that the domain of the accent is the phrase rather than the word; his impression is that there is no word stress in BM. Wallace has the impression that accent in BM is realized with a rising pitch; longer duration and an increased loudness are secondary cues. The accent is usually on the penultimate syllable of the last word in a phrase in BM (Wallace 1976: 56–59).

tu buku mére[34]	*buku báru*
'that book is red'	'new book'

The accent goes to the final position if the penultimate has schwa (a), or if the last word of the phrase is made up of a monosyllabic stem preceded by a prefix (b). A monosyllabic word is always accented (c).

(a) *rumenye gedé* [gədɛ]	(b) *ubinnye dipél*	(c) *masukin di bák*
'the house is big'	'the floor is mopped up'	'put into the bin'

Again, Wallace underlines that schwa is unstressed in the examples *kecepetán* /kəcəpətan/ 'to be fast' and *itemín* /itəmɪn/ 'to make black'. In one case he finds that schwa can be accented, namely when it precedes the unaccented suffix *nye* [ɲe], such as in *itémnye* [itəmɲe] 'the black, being black', *sambélnye* [sambəlɲe] 'the chili sauce'. That the accent shifts to the penultimate syllable in these instances (*ítem* → *itémnye*, and *sámbel* → *sambélnye*) is in line with the general rule that accent is penultimate, but it is at odds with the rule that accent goes to the final position when the penultimate contains a schwa. Wallace did not consider words with schwa in both penultimate and final syllable, like *deket* [dəkət] 'close to', s*eneng* [sənəŋ] 'happy', *kelelep* [kələləp] 'be drowned'.

Summarizing, the literature seems to indicate that BM does not have a word-based but rather a phrase-based accent.

5.1.3 Toba Batak

Batak belongs to the West Malayo-Polynesian languages (van der Tuuk, 1971 [1864]). The Batak dialects are divided into the northern dialects (Karo, Dairi), and the southern dialects (Toba, Angkola, Mandailing, Simalungun, cf. Adelaar 1981, 2005, Woollams, 2005, Sibeth 1991). Toba Batak (TB) is the most common dialect among the Batak dialects, spoken by about two million people living on Samosir Island and to the east, south and south west of Lake Toba in North Sumatra.

Contrary to BM, all authors agree that TB is a (distinctive) stress language. TB has lexical stress (van der Tuuk 1971, Nababan 1981). Stress in TB is penultimate

[34] Wallace's (1976) example is *tu buku mérah*. This must be a mistake. Similarly, in the next example, Wallace has *Rumahnye gedé*, instead of the correct BM *Rumenye gede*.

for nouns and verbs containing two or more syllables and final for predicatively used adjectives (Nababan 1981, Emmorey 1984). There is a clear difference, for instance, between the noun *tíbo* ‘height’ and the adjective *tibó* ‘high’ (stressed syllable indicated by acute accent mark). However, attributive adjectives, and adjectives following *na* (relative pronoun), do not have final stress; they have penultimate stress, e.g. *na tíbo* ‘which is high’.

Emmorey (1984) investigated the TB intonation system. Her research is limited to basic sentence types and a few constructions from one native speaker. Sentences were presented in isolation and in context. She found that in declarative sentences, the nuclear pitch accent is aligned with the stressed syllable of the last word of the phrase. Emphatic stress has a higher nuclear pitch accent than non-emphatic stress.

An experimental study was done by Chen (1984), who also claimed that TB is a stress language. Chen (1984) shows that in TB stress is realized by a rising fundamental frequency. The difference in fundamental frequency between stressed and unstressed syllables is less obvious in connected speech than in isolated words, while the difference in duration between stressed and unstressed syllables is more obvious in connected speech than in isolated words. If target words are not at the intonation peaks (i.e. out of focus and therefore not accented), stress is signaled by longer duration. In contrast to the above, Podesva & Adisasmito-Smith (1999) found no duration-stress relationship for TB vowels. They did, however, find a relation between pitch (but not intensity) and stress.

5.2 Methods

5.2.1 Materials

For TB we selected eight words with penultimate stress: *dakka* ['dak:a] ‘branch’, *pittu* ['pit:u] ‘door’, *jabukku* [ja'buk:u] ‘my house’ and *pagatti* [pa'gat:i] ‘be exchanged by mistake’, *pitu* ['pitu] ‘seven’, *suga* ['suga] ‘thorn’, *jabu* ['jabu] ‘house’, and *kareta* [ka'reta] ‘carriage’.

For BM we also selected eight words consisting of two or three syllables. Three words containing a schwa vowel in the pre-final syllable *pete* [pəte] ‘stinking bean’, *deket* [dəkət] ‘nearby’, *rejeki* [rəjəki] ’fortune’, were chosen to investigate whether the schwa behaves differently than full vowels under the influence of focus and boundary marking. A further five BM words containing full vowels in the last two syllables were used: *kaga* [kaga] ‘no, not’, *kutu* [kutu] ‘louse’, *belaga* [bəlaga] ‘pretend’, *pipi* [pipi] ‘cheek’ *and pepet* [pɛpɛt] ‘overtake rashly’.

The target words were embedded in fixed carrier sentences, in order to create four focus and boundary conditions. Four question sentences were devised to elicit these four sentence types. Table 1 lists examples of TB and BM materials.

Table 1: Examples of Toba-Batak and Betawi-Malay speech material in four sentence conditions (question sentences in parentheses).

		Boundary	Language	Example
Prominence	+focus	+final	TB	(Aha didokkon ibana?) 'What did he say?' Didokkon ibana [dakka] 'He said [dakka]'
			BM	(Die bilang ape?) 'What did he say?' Die bilang [kutu] 'He said [kutu]'
		–final	TB	(Aha didokkon ibana nattoari?) 'What did he say yesterday?' Didokkon ibana [dakka] nattoari 'He said [dakka] yesterday'
			BM	(Die bilang ape tadi?) 'What did he say just now?' Die bilang [kutu] tadi 'He said [kutu] just now'
	–focus	+final	TB	(Nandigan didokkon ibana [dakka]?) 'When did he say [dakka]?' Nattoari didokkon ibana [dakka] 'Yesterday he said [dakka]'
			BM	(Kapan die bilang [kutu]?) 'When did he say [kutu]?' Tadi pagi die bilang [kutu] 'This morning he said [kutu]'
		–final	TB	(Aha [dakka] didokkon ibana?) 'What [dakka] did he say?' Didokkon ibana [dakka] na togu. 'He said [dakka] which is straight'
			BM	(Die bilang [kutu] ape?) 'What [kutu] did he say?' Die bilang [kutu] buku. 'He said [kutu] of books'

5.2.2 Speakers and recording procedure

Four native TB speakers (two male, two female) and four native BM speakers (two male, two female) took part in the experiments. At the time of recording the four TB speakers (aged between 30 and 50 years old) were living in Jakarta. They had come to Jakarta from North Tapanuli (a TB region) after the age of puberty and lived among the TB community in Jakarta, so that they still used TB in their daily life. The BM speakers (between 30 and 55 years old) were living in Sawah Besar, Central Jakarta. These speakers belong to a homogenous ethnic group (*anak betawi*) and use the variety of BM spoken in the central part of Jakarta (dialek kota) in their daily life.

All questions and answer sentences were presented to the speakers in a fixed order. Another speaker of the same language read out the question sentences, and the subject then responded by reading the corresponding answers. The recordings were made in a quiet room onto a Sony TC-D5 PRO II tape recorder through head-worn Shure SM-10A microphones. Every speaker spoke all the materials three times. The total number of utterances was 384 per language. All speech materials were then digitized (16 kHz sampling frequency, 16 bits amplitude resolution).

5.3 Acoustic analysis

Each utterance was subjected to a pitch extraction algorithm (autocorrelation method as implemented in the Praat software, Boersma and Weenink, 1996). Upper and lower frequency bounds were set manually for each speaker. Raw pitch curves were visually inspected and corrected by hand whenever the algorithm had erred.

5.3.1 Toba Batak

5.3.1.1 Stylization

For the analysis of the TB materials four pitch points in each target word were located by eye, and their time/frequency coordinates were stored in a database. The pitch points were found as the result of a data reduction technique that was developed at the Institute for Perception Research. In this so-called analysis-by-synthesis method (Cohen & ’t Hart 1967, ’t Hart, Collier & Cohen 1990) the researcher replaces the original raw pitch curve of the target utterance by a straight-line stylization (fundamental frequency expressed in semitones or ERB – see below – as a function of linear time) such that perceptual equivalence is obtained between the original and the stylization using the smallest possible number of straight-line segments. The comparison between original and stylization is done by virtue of the PSOLA (Pitch Synchronous Overlay and Add, see e.g. Moulines & Verhelst 1995) signal processing technique, which affords the interactive manipulation of the fundamental frequency of an utterance (and even complete replacement or exchange of melodies between utterances) while good to excellent sound quality is maintained in the resynthesis. The result of the stylization is the reduction of the original, capricious pitch curve to a sequence of straight-line rises and falls. The point in the stylization where a rise changes into a fall (or vice versa) is called a pivot point, or just pivot. The stylization procedure is exemplified in Figure 1 below. It should be noted that the overall trend of the sentence melody is not level but slopes down gently. This so-called downtrend is indicated in Figure 1 by a dotted line fitted by hand through the lower pivot points in the stylization (i.e. where a fall ends and/or a rise begins). Downtrend is a universal characteristic of human speech. It is most likely caused by the gradual reduction of subglottal air pressure over the course of an utterance (e.g. ’t Hart et al. 1990) even though the speaker has a choice to reinforce or to counteract the effect through laryngeal muscle activity (e.g. Strik 1994). The downtrend line as drawn in Figure 1 acts as a baseline. Note that the sentence-terminal pitch, especially in statements and commands, tends to go below the baseline. As a result of this, sentence-final falls are often larger than earlier falls; if the vocal pitch reaches the baseline before the last syllable, there will be a noticeable drop in pitch during the last syllable. These effects are generally subsumed under the term ‘final lowering’ (see e.g. Ladd 1996).

The relevant pivot points for TB were the following:

- p1 a low pitch at the beginning of a rise located in the penultimate syllable in the target word (i.e. the first syllable of a disyllabic target) or in the ante-penultimate syllable of the target word (i.e. the first syllable in a tri-syllabic item). P1 is defined as the F0 minimum (i.e. lowest F0) from the beginning of the utterance onwards preceding the pitch peak on the penultimate syllable;
- p2 the peak F0 located in the penultimate syllable of the target word;
- p3 a pivot point between p2 and p4 that affords the stylization of pitch fall in terms of two straight-line segments, the first of which drops off at a modest rate whilst the second part embodies a steep fall. In a fair number of cases, and in fact in all non-final targets without focus, such a point could not be found; p3 was then left undefined;
- p4 end of the fall or F0 minimum between p2 and the end of the utterance. When the target was utterance final, p4 is typically the terminal pitch; in non-final targets without focus p4 could easily be located as the pivot point between the fall after p2 and the large rise marking the focused constituent following the non-focused target.

Figure 1 gives an example of an original F0 curve (capricious lines) and a close-copy stylization of an utterance in TB. The dotted line represents the baseline (downtrend, see above).

Given that two speakers were male and two female, some basic form of speaker normalization was unavoidable. As a first approximation we applied a minimal normalization procedure to the raw pitch data (the four pivot points). The raw pitch data in hertz were first rescaled to Equivalent Rectangular Bandwidths (ERB units, cf. Hermes & van Gestel 1991, Nooteboom 1997, Ladd & Terken 1995), which is currently held to be the psychophysically most valid scale for comparing vocal pitch in intonation languages across registers. Pitch intervals of equal sizes when expressed in ERB should be perceptually equivalent regardless of their absolute frequency in hertz. As a rough indication, the typical male vocal pitch range in speech is between 3 and 5 ERB, and that of women between 5 and 7.

Inspection of raw pitch measurements revealed that the lowest recurrent pitch that could be found in the materials, was pivot point p4 in sentence-final position in [–focus] constituents. All pitches were therefore rescaled to ERB and then expressed relative to the reference pitch at p4. This allows straightforward comparison of pitch differences within and between utterances.

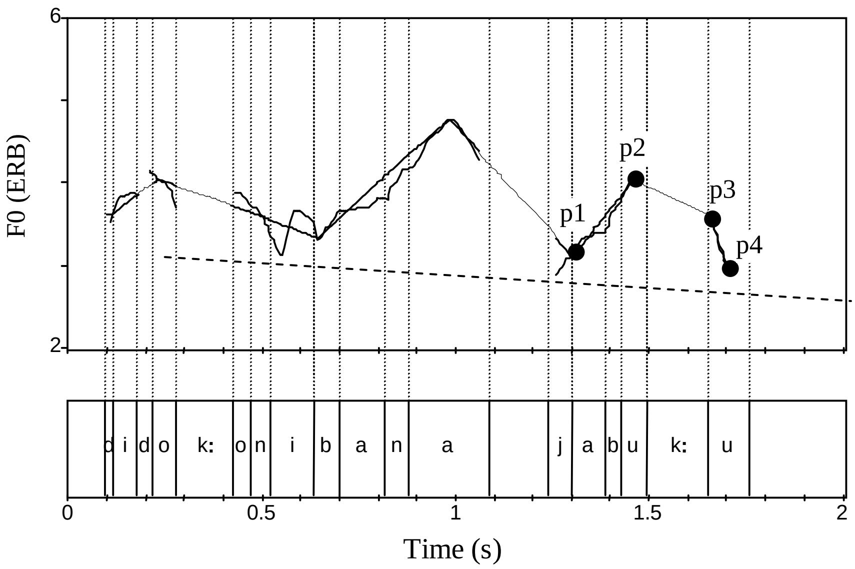

Figure 1: Original F0 curve (capricious lines) and close-copy stylization (solid straight lines) of the TB target word *jabukku* 'my house' in [+focus, +final] condition, spoken by a male TB speaker. The dotted line represents the downtrend or declination (see text).

5.3.1.2 Results

In the selected TB target words the accent-lending pitch movement occurs on the penultimate syllable, i.e. the stressed syllable. In non-focused words pitch movements occur on the penultimate syllables as well, but there the excursions are rather small. Stress in unaccented words is realized with a smaller pitch obtrusion than in accented words. This agrees with the findings of Chen (1984) and Podesva & Adisasmito-Smith (1999), who both mention a relation between pitch and stress in TB. Pitch movements in the TB targets are generally realized with a rise-fall movement.

Figure 2 illustrates the pitch contour of all TB words in normalized F0 (ERB), broken down by sentence type. The x-axis shows the time scale relative to the onset of the penultimate vowel.

One-way ANOVAs with sentence condition as a four-level fixed factor indicated significant effects for several acoustic parameters. For instance, the timing of the peak [$F(3, 163) = 8.7$, $p < .001$], the height of the peak [$F(3, 163) = 23.5$, $p < .001$], the size of the rise $F(3, 163) = 12.2$, $p < .001$] and of the fall [$F(3, 163) = 62.0$, $p < .001$] are significant. No significant effect of sentence condition was found for the beginning of the rise [$F(3, 163) = 1.9$, $p = .136$].

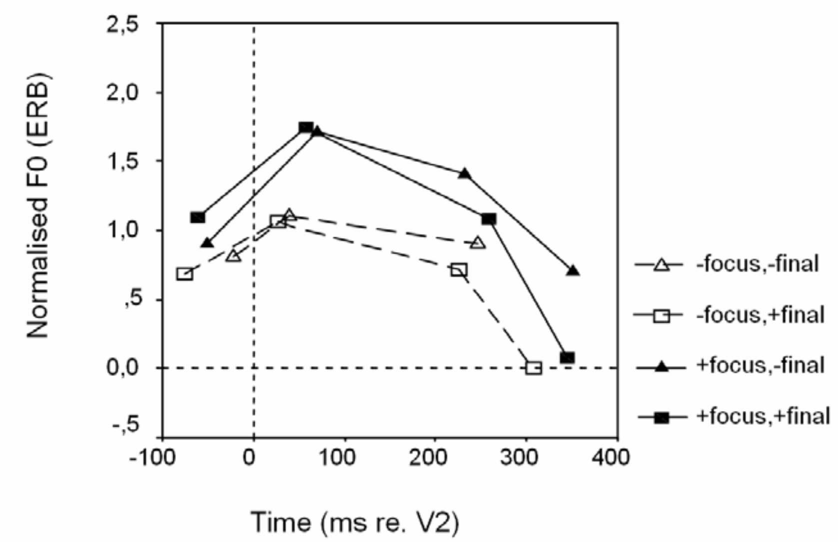

Figure 2: Pitch contour of all target words in normalized F0 (ERB), with the time scale relative to the onset of penultimate V, broken down by sentence type.

Two-way analyses of variance with focus and final position as fixed two-level factors show that the effects of focus are often significant but that the effects of finality are not. The interactions between focus and finality are hardly significant. Focus affects many aspects of the melody, except for the F0 minimum at the end of the target word, which effect is insignificant [$F(1,163) = 3.6$, $p = .059$], and a small effect on the onset of the rise [$F(1,163) = 4.1$, $p = .044$]. The effect of focus is highly significant for the timing of the peak [$F(1,163) = 21.3$], the height of the peak [$F(1,163) = 66.0$], the size of the rise [$F(1,163) = 29.0$], the size of the fall [$F(1,163) = 91.8$], and the slope of the fall [$F(1, 163) = 52.9$], all with $p < .001$.

Most pitch movements are not much affected by finality. The effect on the rising movements is insignificant, for the onset of the rise [$F(1,163) < 1$], for the peak timing, [$F(1,163) = 1.7$, $p = .194$], for the peak height [$F(1,163) < 1$], for the size of the rise [$F(1,163) = 3.6$, $p = .060$], and for the slope of the rise [$F(1,163) = 3.7$, $p = .057$]. However, there are highly significant effects on the fall movements for the size of the fall [$F(1,163) = 94.8$], for the slope of the fall [$F(1,163) = 53.0$] and for the F0 minimum at the end of the target word [$F(1,163) = 156.8$], all with $p < .001$. The interaction between the two effects is significant only for the final F0 with [$F(1,163) = 10.0$, $p = .002$].

In Table 2 the results of the various measurements (means and standard deviations) are given for each of the four sentence conditions.

Table 2: Mean of pitch accent measurements in eight TB words per sentence condition with standard deviation (in parentheses), and the mean across all conditions.

Measurements	+focus, +final		+focus, –final		–focus, +final		–focus, –final		Mean
Onset timing (ms)*	–61.50	(52.00)	–51.20	(48.00)	–74.70	(35.00)	–21.60	(31.00)	–61.30
Peak timing (ms)*	66.60	(36.00)	83.10	(60.00)	36.90	(29.00)	40.70	(21.00)	64.80
F0 peak (norm. ERB)	1.76	(0.51)	1.73	(0.47)	1.04	(0.28)	1.11	(0.34)	1.57
Rise exc. (ERB)	0.66	(0.36)	0.82	(0.40)	0.36	(0.26)	0.44	(0.18)	0.65
Slope rise (ERB/s)	5.48	(3.10)	7.02	(4.70)	3.59	(2.60)	4.57	(2.80)	5.60
Final F0 (norm. ERB)	0.09	(0.31)	0.70	(0.48)	0.00	(0.17)	1.03	(0.31)	0.37
Fall (ERB)**	1.67	(0.54)	1.02	(0.44)	1.04	(0.34)	0.08	(0.15)	1.20
Slope fall (ERB/s)	–6.52	(2.70)	–4.19	(2.30)	–4.19	(1.90)	–0.42	(0.81)	–4.77

*) Relative to the onset of the penultimate vowel
**) From the peak (p2) to the F0 terminal (p4)

The results show that focus affects the pitch movements. In sentence-final position the rise starts later in the [+focus] words, and reaches the highest pitch later as well, than in the [–focus] words. Sentence-medially, the rise starts earlier in focused words, and reaches the highest pitch later than in the [–focus] words. Accented words have significantly higher pitch peaks than unaccented words, the difference amounting to some 0.6 ERB. The excursion sizes of the [+focus] rises are about twice as large (in ERB) as those of their [–focus] counterparts. Also, the slopes of the [+focus] rises are considerably steeper. There is no systematic difference between plus and minus-focus falls in terms of final pitch, but the [+focus] falls generally have larger excursions and steeper slopes.

Accent-lending rises start on average 57 ms before the onset of the pre-final vowel, with a slope of around 6 ERB/s. The peak is reached 74 ms after the onset of the vowel in the pre-final syllable, with the F0 maximum at 1.74 ERB. After the peak, pitch goes down gradually to the final-syllable and then drops again to the end of the word.

Boundary marking affects the fall movements. In sentence-final position, the fall excursions are significantly larger than in their [–final] counterparts. The falls in sentence-final words are also steeper than the falls in sentence-medial words. The pitch movements in the penultimate syllables, which are stressed, depend on the focus condition of the words. Irrespective of focus the presence of a final boundary determines the fall movements in the final syllable. At the end of the utterance, the fall is larger and reaches the baseline.

5.3.2 Betawi Malay

Preliminary auditory and visual inspection of the BM materials revealed that pitch movements were in general completely absent when the target word was not in focus. Pitch movements were observed on [+focus] targets but these could occur in either the final or the pre-final syllable of the target word, depending primarily on the target's position in the sentence. When the focused target occurred in sentence-final position, the pitch movement seemed to coincide with the final syllable; accented sentence-medial targets, however, typically carried the pitch movement on the pre-final syllable. The shape of the accent-lending movement could be a rise, a fall or a rise-fall combination. Simple rises always occurred on final syllables, simple falls on pre-final syllables; rise-fall combinations were found in both final and pre-final positions (depending on the position of the target in the sentence). These findings seem to be in line with the view that BM has no word stress but phrasal accent only (see introduction); the distributional details were in fact predicted by Kähler (1966).

5.3.2.1 Stylization

In view of the variability in the occurrence and shapes of the pitch movements, some refinements of our stylization point p1 and p2 (as defined above for TB contours) were in order. The definitions for pitch points p3 and p4 were left unchanged.

p1 As before, this is the low pitch at the beginning of a rise associated with an accented target word. It is defined as the latest F0 minimum (i.e. lowest F0) preceding the pitch peak on the target. However, when no pitch rise could be seen on the target word (as happened in [–focus], i.e. unaccented words), p1 was considered to be absent.

p2 This point is defined as the F0 maximum (i.e. the highest F0) in the target word. However, in unaccented BM targets (without any pitch rise) p2 was located at the onset of the target.

Figure 3 gives an example of a stylization of a BM speech spoken by a BM male speaker in [+focus, +final] condition.

The raw pitch measurements were again normalized (in ERB re. the terminal F0 in [–focus, +final] targets) and aligned relative to the onset of the target in exactly the same way as was done for TB.

However, before we turn to a presentation and discussion of the acoustic analysis, we will first present the results of an auditory screening of the BM materials, so that the analysis can be done separately for the three types of pitch movement (rise, fall, rise-fall) that we found in the BM recordings (see above). The next section describes the procedure and results of the auditory screening by a panel of expert listeners.

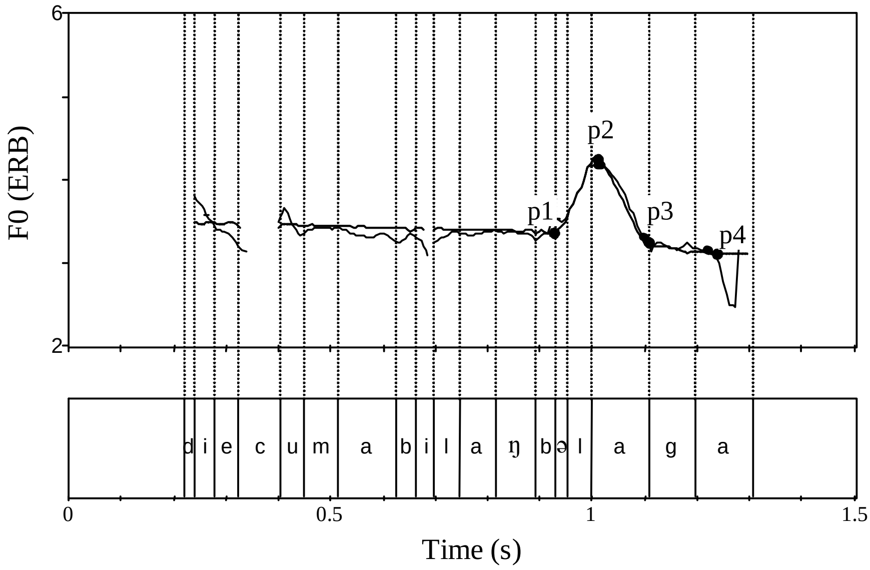

Figure 3: Original F0 curve (capricious lines) and close-copy stylization (solid straight lines) of the BM target word *belaga* 'pretend' in [+focus, + final] condition, spoken by a male speaker.

5.3.2.2 Auditory analysis

Preliminary inspection of the [+focus] utterances revealed considerable variation in the position of accent-lending pitch movements. These occurred either on the pre-final or on the final syllable but it was not always straightforward which of these two syllables was accented, nor which part of the pitch movement should be seen as accent lending.

In order to make reasoned decisions in this matter, a formal listening test was conducted. In a preliminary screening of the recordings, the second author, who is a native speaker of BM, listened to the materials and decided for each of the 192 utterances (4 speakers × 2 conditions × 8 words × 3 repetitions) whether the speaker had produced an accent on the [+focus] target word. In all, 33 out of the total of 192 utterances contained target words on which a pitch movement was clearly absent. Absent pitch movements were randomly scattered over the various conditions and repetitions, such that, in fact, every condition in the design was still represented with at least two exemplars by each speaker. Table 3 summarizes the result of this preliminary screening for the valid cases of the [+focus] words. The table specifies the number of [+focus] target words in sentence-final and medial position for each of the four speakers separately, broken down by perceived presence versus absence of a prominence-lending pitch movement.

Table 3 shows that the male BM speakers quite consistently realized accent-lending pitch movements on the targets. The female speakers dropped the pitch movements in some 20 to 35 percent of sentence-medial targets; female 2 dropped her pitch movements in sentence-final position in more than 50 percent of the cases. We decided to exclude the 33 utterances without audible accent on the [+ focus] target word from further analysis.

Table 3: Number of [+focus] Betawi Malay target words realized with and without an audible accent, broken down by position in the sentence (final, medial).

sentence condition	speaker	N perceived as	
		Accented	Non-accented
+focus, +final	male1	22	2
	male2	22	2
	female1	22	2
	female2	11	13
+focus, −final	male1	23	1
	male2	24	0
	female1	16	8
	female2	19	5

As a next stage in the auditory screening, the second and third author as well as a third expert on prosody, listened to the remaining 159 [+focus] utterances and judged syllable prominence.[35] They indicated, independently of each other, for each correctly spoken utterance (i.e. with target words bearing an audible accent) whether they found the pre-final or the final syllable of the target more prominent or, as a third option, whether they considered both syllables equally prominent. The 159 correctly pronounced utterances (192 – 33 tokens without an audible accent) judged by three listeners yielded 477 valid judgments (231 [+final] and 246 [–final] judgments). Table 4 summarizes the results of the prominence test for the [+focus] targets, for full-vowel words and schwa words, broken down by position of target in the sentence.[36]

Table 4 shows that the perception of prominence in sentence-final words with full vowels is distributed equally over the pre-final and final syllables. Words with schwa in the pre-final syllable, however, are more prone to have accent on the final syllable, regardless of whether this final syllable does or does not contain a schwa. Sentence-medially, pre-final syllables with full vowels are generally accented; however, pre-final syllables with schwa in sentence-medial words are accented in only half of the cases.

[35] I thank Drs. Ellen van Zanten and Johanneke Caspers for acting as expert listeners.
[36] For individual listener results and statistical evaluation of the between-listener agreement, see Roosman (2006).

Table 4. Relative frequency (%) of perceived prominence on pre-final versus final syllable of accented targets for BM target words in sentence final and medial position.

Target position	Vowel type	Prominence perception (%)		
		pre-final syll.	final syll.	equal
sentence-final	V – V	48	45	7
	ə – V/ə	9	86	5
sentence-medial	V – V	80	11	9
	ə – V/ə	42	41	17

The following step in the auditory analysis (in section 5.3.2.3) was to establish the token frequency of specific types of accent-lending pitch movements on the prominent syllables. We will then be in a position to determine the shapes of the various pitch configurations in acoustical terms. This will be described in section 5.3.2.4.

5.3.2.3 Token frequencies of BM accent-lending pitch movement types

For the next part of the data analysis we made a distinction between four types of pitch movement on the targets using visual criteria. These are the simple rise and simple fall, and complex rise-fall configurations which were subdivided into early versus late alignment. For early alignment the pitch peak (pivot point p2) should be located within the confines of the penultimate syllable, for late alignment the peak finds itself in the final syllable. Tables present the four shapes of the pitch contour over the final two syllables of the [+focus] targets (collapsed over sentence-medial and final positions as well as over all stimulus words and speakers) cross-tabulated against the position of the syllable that bears the accent-lending pitch movement (Table 5a) and against the position of prominent syllable (Table 5b), as determined by the listening panel.

The results show, first of all (Table 5a), that in 28 cases the panel of listeners could not detect a pitch movement on the target word – even though the second author had judged earlier that the target did bear an accent. These 28 cases probably make up a separate category of audible accents that are not marked melodically but, for instance, temporally; these cases will not be analyzed as part of the present study. Next, there is a relatively small group of tokens that were judged to bear equal prominence on the final and pre-final syllables (less than 10% of the prominence judgments are in this category); these, too, will not be analyzed any further.

Table 5: Number of accent-lending pitch configurations (a) and perceived prominences (b), heard by three experts (see text) in final and pre-final syllables in Betawi Malay [+focus] target words broken down by type of movement (F: simple fall, R: simple rise, RF pre-final: rise-fall with peak in pre-final syllable, RF final: rise-fall with peak in final syllable).

a. Perception of pitch-accented syllable.

Shape	Pitch movement heard				Total
	None	pre-final	Final	both/equal	
F	13	**131**	9	15	168
R	10	17	**103**	20	150
RF pre-final	2	**68**	4	4	78
RF final	3	5	**64**	9	81
Total	28	221	180	48	477

b. Perception of prominent syllable.

Shape	Prominence heard			Total
	pre-final	Final	both/equal	
F	**134**	17	17	168
R	27	**100**	23	150
RF pre-final	**72**	4	2	78
RF final	7	**71**	3	81
Total	240	192	45	477

For the remaining cases, there is a very strong association between the type of pitch movement and the position of the prominent syllable. If the movement is a simple fall, the prominence is on the pre-final syllable, if it is a simple rise, then the prominence is typically on the final syllable. When the pitch movement is a complex rise-fall configuration, about half of the tokens are perceived with prominence on the pre-final syllable, and the other half with final prominence. From the stylization it was found that the rise-fall could indeed occur in the pre-final syllable (N = 26) and final syllable (N = 27). In sentence-final position there are more rise-fall configurations found in the final syllable (N = 26) than in the pre-final syllable (N = 7). However, in sentence-medial position rise-fall occurs more often in the pre-final syllable (N = 19) than in the final syllable (N =1).[37]

The position of the prominent syllable depends not only on the position of the target in the sentence, but also on the type of vowel in the target word, and certainly on the shape of the curve. Final prominence, of course, is predicted when the pre-final syllable contains schwa; pre-final prominence is what we typically find when the pre-final syllable contains a full vowel. As a consequence of this, simple falls typically occur on pre-final full vowels, and simple rises are found on final vowels after schwa.

[37] See Roosman (2006: appendix 1a) for the table of shapes.

In the following section we will present the acoustical analysis of the pitch contours on the BM utterances. The acoustical analysis will first concentrate on the pitch configuration as found on accented targets (as judged by the second author) only. In the final subsection we will also consider the (basically flat) pitch pattern that was found on [–focus] targets.

5.3.2.4 Acoustic properties of BM accent-lending pitch configurations

ANOVAs show that the effects of word position in the sentence are highly significant on the peak timing[38] [$F(1,157) = 25.9$, $p < .001$], on the height of the peak [$F(1,157) = 41.6$, $p < .001$], and on the fall excursion [$F(1,107) = 30.4$, $p < .001$]. A significant effect of word position is also found on the rise onset [$F(1,101) = 5.2$, $p = .025$] and on the slope of the rise [$F(1,101) = 4.1$, $p = .045$]. In sentence-final words the rises start 162 ms before the onset of the vowel; sentence-medially rises start later, 117 ms before the onset of the vowel. In sentence-final words, the F0 peak is reached 69 ms after the vowel onset and in sentence-medial words earlier, 33 ms after the vowel onset. However, the peak in sentence-medial words is .66 ERB higher than that in sentence-final words. The slope of the rise is thus significantly steeper in sentence-medial words, with a difference of about 2 ERB/s. Also, the fall excursion in sentence-medial words is larger (.69 ERB) than the fall excursion in sentence-final words.

Effects of vowel type are significant only on the slope of the rise [$F(1,101) = 4.4$, $p = .039$] and the slope of the fall [$F(1,107) = 4.3$, $p = .040$]. Words with full vowels have on average steeper rises (2.1 ERB/s) and steeper falls (1.9 ERB/s) than words with schwa.

The effects of movement shape are highly significant on the rise onset [$F(3,99) = 10.2$, $p < .001$], on the peak timing [$F(3,155) = 18.7$, $p < .001$], and on the rise excursion [$F(3,99) = 5.6$, $p = .001$]. Furthermore, significant effects of movement shape are found on the steepness of the rise [$F(3,99) = 2.9$, $p = .038$], and on the fall excursion [$F(3,105) = 3.0$, $p = .036$].

Based on these results we will analyze the sentence-final and the sentence-medial words separately. In each section, the F0 parameters of every shape will be analyzed. We will not separate the target words with a schwa vowel from the target words with only full vowels since significant effects of vowel type occur only on two parameters and the significance levels are not high.

We will first present the F0 contours of target words in sentence-final position. Figure 4 (and Figure 5 for target words in sentence-medial position) plots the stylized F0 contour in normalized ERB as a function of time, such that the timing of the movements is expressed relative to the onset of the final vowel in the target word. In these figures rises, falls and rise-fall combinations are plotted separately. Rises are always aligned with final syllables, falls with pre-final syllables. Rise-fall combinations were separated into two alignment groups: those that were heard as imparting prominence to pre-final syllables and those that were heard with

[38] Relative to the onset of the vowel in the accented syllable.

prominence on the final syllable. As a consequence of this redefinition of movement types, subsequent data analysis, again using ANOVA, will involve a four-level factor for movement type.

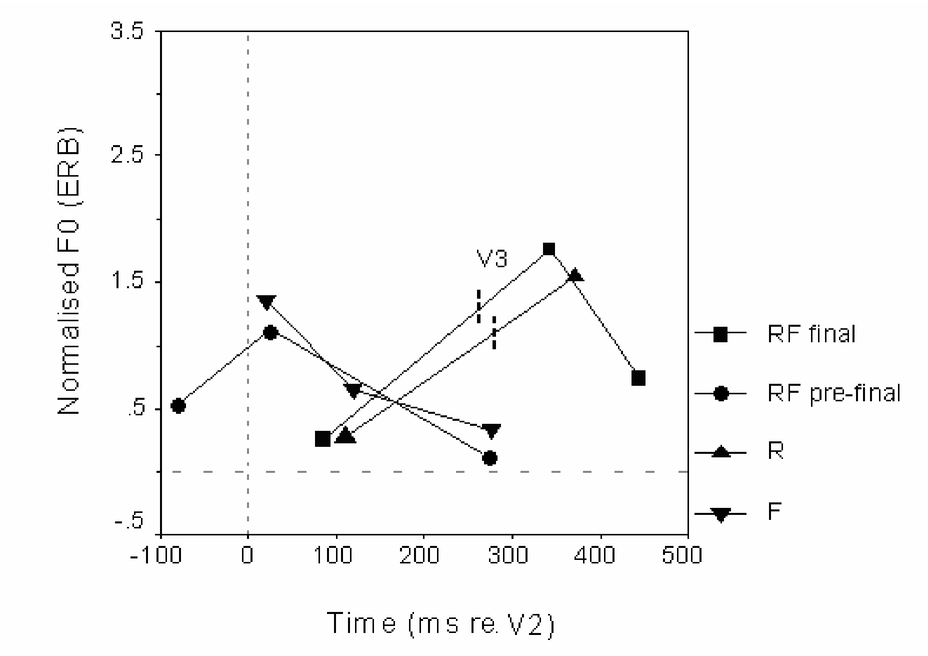

Figure 4: F0 contours of BM [+focus] words in sentence-final position in normalised ERB with on the *x*-axis a time scale relative to the onset of the penultimate vowel. The vertical dashes on the RF-final and the R contours mark the onset of the final vowel.

Figure 4 shows that the excursion sizes of the shapes are to some extent different from each other. Typically, a movement – whether fall or rise-fall – on the pre-final syllable has a small excursion size and does not reach a high peak frequency. In contrast to this, movements in final syllables, whether rise-fall or just a rise, are characterized by a very large excursion size leading to a high F0 peak. Notice that in the case of the rise, and even in the case of a rise-fall on the final syllable, the pitch does not drop down to the baseline but remains 0.7 (F0 final for rise-fall) or 1.5 (F0 peak for rise only) ERB above it. Movements aligned with the pre-final syllable, however, end at baseline level. ANOVAs for the sentence-final words, with shape of movement as a (four-level) fixed factor showed significant effects on the peak timing [$F(3, 73) = 13.4$, p < 001], the rise excursion [$F(3, 58) = 5.9$, p = 001] and on the slope of the fall [$F(3, 44) = 5.4$, p = .003]. T-tests with two movement shapes, the rise-fall in pre-final and that in final syllable as independent variables, show that there are significant differences between the two movements in terms of peak height

[$t(31) = –2.05$, p = .049], peak timing[39] [$t(31) = –3.47$, p = .002], rise excursion [$t(31) = –4.04$, p < .001], and slope of the fall [$t(31) = –2.74$, p = .010]. In the pre-final syllable the F0 peak is reached 25 ms after the onset of the vowel. In the final syllable the peak is reached later, at 84 ms after vowel onset. Also, the peak is .63 ERB higher in the final syllable than when it occurs on the pre-final syllable. The rise excursion of the final-syllable rise-fall is thus larger, by .88 ERB, than the pre-final syllable rise-fall. However, the fall of the rise-fall movement in the final syllable is 5.7 ERB/s steeper than the fall in the pre-final syllable. In the final syllable the pitch goes down very quickly from the highest point to the next point before the utterance is ended. The complete results of the measurements are summarized in Table 6.

Table 6: Means of pitch accent measurements in BM words per movement shape, in sentence-final position, with standard deviation (in parentheses), and the means across all movements.

Measurements	Fall		Rise		RF pre-final		RF final		Mean
Onset timing (ms)*			–168.10	(115.00)	–81.10	(24.00)	–176.40	(90.00)	–161.70
Peak timing (ms)*	1910	(12.00)	92.60	(54.00)	24.70	(21.00)	84.10	(43.00)	69.20
F0 peak (norm. ERB)	135	(0.50)	1.54	(0.53)	1.13	(0.58)	1.76	(0.75)	1.54
Rise exc. (ERB)			1.26	(0.46)	0.62	(0.27)	1.49	(0.55)	1.29
Slope rise (ERB/s)			5.71	(3.70)	5.83	(2.00)	6.27	(2.60)	6.00
Final F0 (norm. ERB)	0.33	(0.61)			0.11	(0.14)	0.73	(0.97)	0.52
Fall (ERB) **	0.91	0.41)			1.00	(0.59)	1.02	(0.60)	0.98
Slope fall (ERB/s)	5.33	(3.70)			4.77	(2.80)	10.54	(5.30)	8.10

*) Relative to the vowel onset of the accented syllable.
**) From the peak (p2) to the next lower point.

It seems that the canonical shape of the accent-lending pitch configuration in BM is a rise-fall combination, which can occur either on the pre-final or on the final syllable of the [+focus] target word. When the rise-fall is on the pre-final syllable it imparts prominence to that syllable. The rise portion may be absent from the contour (i.e. if the preceding context ends in a high pitch) but the temporal alignment of the fall is not affected by the presence or absence of the rise. Importantly, the fall is always complete and reaches the baseline pitch around the onset of the final syllable. Due to the severe time constraints on the rise and fall of the configuration on pre-final syllables, the excursion size of the movements is small: the configuration is scaled down. When the rise-fall is executed on the final syllable – which is then perceived as prominent – there seems to be no time constraint. The final syllable, also as a consequence of pre-boundary lengthening, provides ample space for large movements. Typically the rise portion of the configuration takes up some 200 ms, and during this time interval the pitch rises by roughly a full ERB. The rise is often, but by no means always, followed by a fall, which, however, never reaches the lower declination line (and final lowering seems to be conspicuously absent).

[39] The peak timing on the final syllable is measured relative to the onset of the vowel in the final syllable.

Apparently, BM speakers choose to truncate, rather than scale down, rise-fall configurations in final position.

Generalizing further, the small scaled-down rise-fall configurations occur on pre-final syllables that have mostly full vowels. Pre-final syllables with schwa do not carry prominence; prominence is then pushed onto the final syllable. The only exception is *rejeki* [rəjəki], where the pre-final syllable could carry the prominence due to a clear rise-fall pitch movement. If the final syllable contains a full vowel, it usually provides ample space for a large rise-fall configuration, which however, is truncated halfway during the fall portion. When both the pre-final and the final syllables contain schwa (in *deket* 'nearby'), there is much less space for the rise-fall configuration; as a result the entire fall portion is dropped.

In sentence-medial position (Figure 5) there is only one case of a rise-fall configuration in the final syllable (in *rejeki*), which is different from the pre-final rise-fall. Figure 5 illustrates the four pitch movements in sentence-medial position.

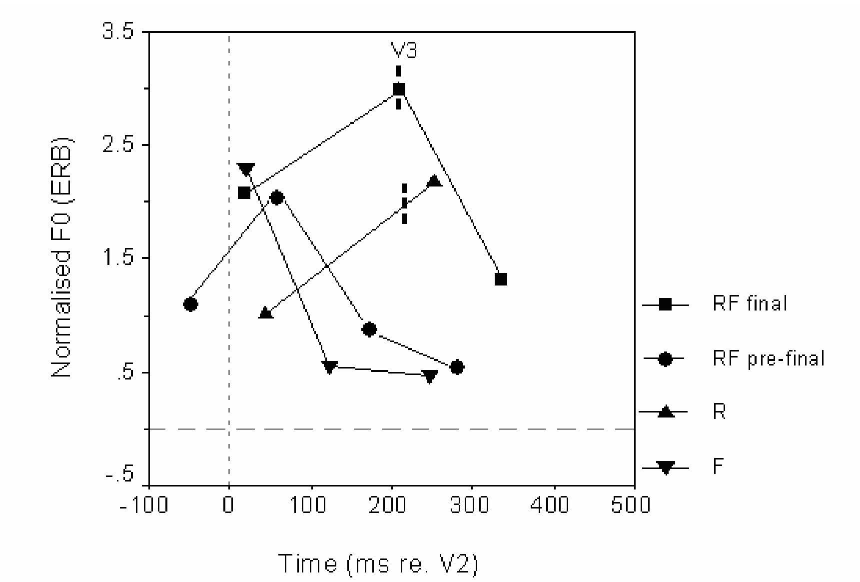

Figure 5: F0 contours of BM [+focus] words in sentence-medial position in normalized ERB with on the *x*-axis a time scale relative to the onset of the penultimate vowel. The vertical dashes on the RF-final and the R contours mark the onset of the final vowel

No statistical comparisons could be made for the rise-fall in the final syllable (one case only). One-way ANOVAs with the shape of the movement (excluding the rise-fall in the final syllable) as a fixed factor showed that effects of shape of movement are significant only for the rise-onset timing [$F(3,37) = 9.0$, $p < .001$] and for the peak timing [$F(3,78) = 5.9$, $p = .001$]. However, a post-hoc test (Scheffé, with $\alpha = .05$) indicated that the peak timing is significantly different only between the fall and the rise-fall in the pre-final syllable. Table 7 summarizes the pitch measurements of the accented words in sentence-medial position.

Table 7: Means of pitch accent measurements of BM words per movement shape, in sentence-medial position, with standard deviation (in parentheses), and the means across all movements.

Measurements	Fall		Rise		RF pre-final		RF final		Mean
Onset timing (ms)*			–171.30	(81.00)	–53.20	(60.00)	–187.50	-	–117.00
Peak timing (ms)*	20.60	(14.00)	38.20	(28.00)	57.30	(57.00)	5.40	-	33.00
F0 peak (norm. ERB)	2.28	(0.70)	2.17	(0.58)	2.09	(0.78)	2.98	-	2.22
Rise exc. (ERB)			1.15	(0.49)	0.98	(0.63)	0.91	-	1.06
Slope rise (ERB/s)			5.91	(2.80)	10.46	(9.20)	4.70	-	8.00
Final F0 (norm. ERB)	0.46	(0.64)			0.52	(0.52)	1.31	-	0.50
Fall (ERB) **	1.78	(0.41)			1.47	(0.84)	1.67	-	1.68
Slope fall (ERB/s)	9.25	(4.00)			8.46	(4.10)	13.37	-	9.10

*) Relative to the vowel onset of the accented syllable.
**) From the peak (p2) to the next lower point

The pitch movements in sentence-medial words differ to some extent from the pitch movements in sentence-final position. The falls in sentence-medial words start from significantly higher pitches than those in sentence final words [$F(1,54) = 22.1$, $p < .001$], with a mean difference of .93 ERB. Sentence-medial falls are significantly larger [$F(1,54) = 21.9$, $p < .001$], around .87 ERB, and 3.9 ERB/s steeper [$F(1,54) = 10.9$, $p = .002$], than the sentence-final falls.
Rise movements differ only in the peaks: the sentence-final rises have later peaks ($\Delta = 54$ ms) than the sentence-medial rises [$F(1,48) = 15.9$, $p < .001$]. Also, the sentence-medial rises have significantly higher peaks ($\Delta = .63$ ERB) than the sentence-final rises [$F(1,48) = 18.0$, $p < .001$].

Pre-final rise-fall movements are affected by boundary only in terms of peak height and slope of the fall part. Peaks in sentence-final words are significantly lower than in sentence-medial words [$F(1,24) = 8.6$, $p = .007$], with a difference of roughly a full ERB. The slope of the fall portion of the pre-final rise-fall is 3.7 ERB/s steeper in sentence-medial words than in sentence-final words [$F(1,24) = 4.7$, $p = .041$].

Accent-lending pitch movements in sentence-medial words seem larger and steeper than those in sentence-final position. The peaks of the accent-lending pitch movements in sentence-medial words are higher than in sentence-final words. The declination effect explains the lower pitches in pre-boundary position. On the other hand, sentence-final rises need more time to reach the centre of the accented (final) syllable, which are longer than in sentence-medial words as a consequence of pre-boundary lengthening. Rise movements in sentence-final words are thus less steep than in sentence-medial words.

5.3.2.5 Pitch in [–focus] BM targets

Listening tests on [–focus] targets were not deemed necessary as preliminary auditory and visual inspection indicated that there were no pitch movements in the [–focus] targets. The pitch contours of [–focus] words show only slight falls. Figure 6 illustrates the pitch movements in [–focus] targets compared with the [+focus] pitch movements; in the figure the horizontal axis has been time-normalized by plotting equidistant steps between successive pitch pivot points.

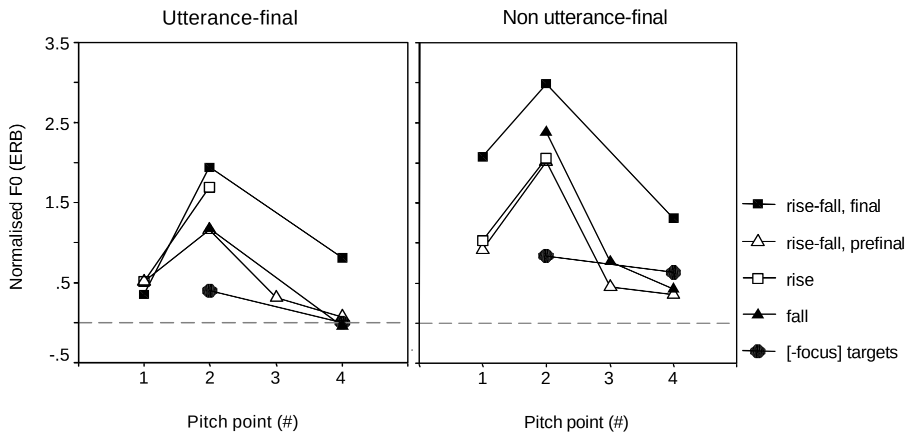

Figure 6. Pitch curves of [–focus] (black circles) and [+focus] targets (other marks) in [+final] position (left-hand panel) and in [–final] position (right-hand panel), with pitch points as a time-normalised *x*-axis.

Figure 6 shows clearly that there is no accent-lending pitch movement in [–focus] targets. In [–focus] words pitch point p2, which is the highest F0 in the target word, is lower than in [+focus] targets. To show this, we compared the pitch points p2 and p4 of the non-focus targets with those of the lowest pitch curve of the focus targets, i.e. the pre-final rise-fall. Independent t-test were therefore run separately for each finality condition, with the F0 of p2, F0 of p4 and the fall excursion as test variables and the two curves (non-focus fall and pre-final rise-fall in focus targets) as the grouping variables. The results show that the values of the F0 and the fall excursions are significantly different from each other. In sentence-final position the difference between F0 point p2 in [–focus] and the lowest peak in [+focus] targets, the peak of the pre-final rise-fall, is .66 ERB [$t(88) = 3.65$, $p < .001$]. Sentence medially this difference is even larger, 1.25 ERB [$t(104) = 8.55$, $p <.001$]. The F0 terminal (pitch point p4) in [+focus] words with falling pitch or with a pre-final rise-fall is equal to the F0 terminal of the [–focus] words [$t(88) = .46$, $p = .64$ in sentence final position; $t(104) = -.84$, $p = .40$ in sentence-medial position]. The fall excursion in the non-focused words is significantly smaller than that in the focused words, with a difference .55 ERB sentence finally [$t(88) = 2.58$, $p = .012$], and 1.37 ERB sentence medially [$t(104) = 9.44$, $p < .001$].

The melodic structure of unaccented words is also affected by boundary marking. Figure 7 plots the pitch configuration of the [–focus] words in sentence-final and sentence-medial position.

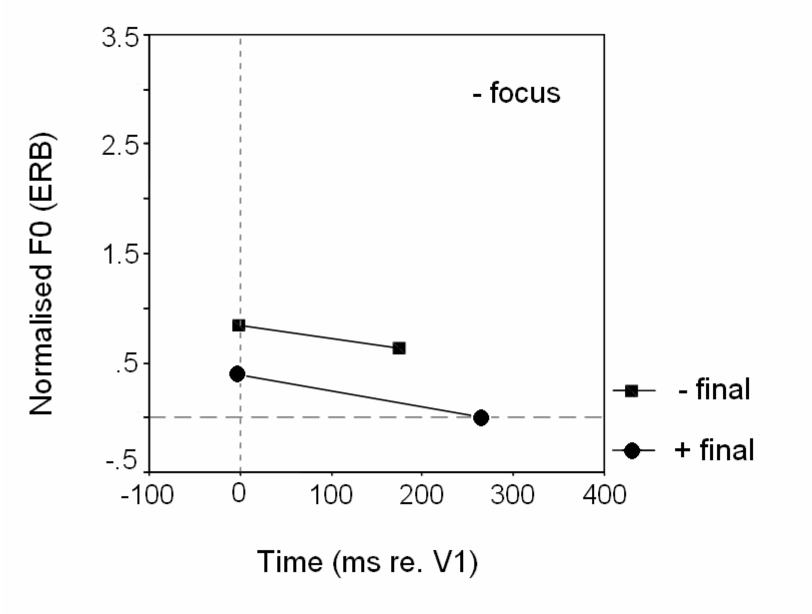

Figure 7: Pitch contours of unaccented words in sentence-final and sentence-medial position. The *x*-axis refers to the time (in ms) relative to the onset of penultimate vowel.

Figure 7 shows that unaccented words in sentence-final position are lower by about .5 ERB than in sentence-medial position. The excursion size of the fall is very small and suggests that the fall is entirely due to declination. There is no difference between medial and final targets in the steepness of the fall. The larger excursion on final falls is caused by the fact that final falls are longer than medial falls. This state of affairs is born out by a series of one-way ANOVAs which show significant effects of finality on the fall excursion [$F(1,166) = 5.1$, $p = .025$]. The fall excursion is about .5 ERB for the [+final] words, and .2 ERB for the [–final] words. However, finality does not affect the slope of the fall [$F(1,166) = 1.2$, $p = .269$].

5.3.2.6 Tonal accent in Betawi Malay

The results of the pitch analyses of the [+focus] targets show clearly that accents in BM may fall on pre-final or final syllables. Pitch movements are either rises, falls, of rise-fall configurations. These findings indicate that stress position in the word is free in BM – at least within a two-syllable window at the end of the word.

The melodic structure of BM words is influenced by both focus condition and pre-boundary position. Focus yields accent-lending pitch movements; non-focus suppresses pitch movements. The presence of a sentence boundary following the target word determines the position of the accent-lending pitch movements within the target: it attracts the accent to the ultimate position; without a following sentence boundary the accent remains on the penult. Wallace's claim that accent is on the pre-final syllable is in line with our findings for sentence-medial words only, but clearly clashes with our results for sentence-final words. Moreover, Wallace's claim that words with schwa in the pre-final syllable are generally accented on the final syllable is found true in sentence-final words only. Our results show that accent shifts not only due to the type of the vowel (full versus schwa), but also due to the position of the word in the sentence.

Pre-boundary position affects the pitch of the words significantly. Target words in sentence-medial position are generally realized at higher pitches than target words in pre-boundary position, regardless of the focus condition of the targets. The F0 terminal of target words in pre-boundary position typically reaches the base line.

5.4 Discussion and conclusion

The word prosody of Toba Batak appears to be a rather simple system. Within the restricted data set of the present study, TB has just one accent-lending pitch configuration, a rise on the stressed syllable followed by a fall on the next syllable. Such a rise-fall movement can be adequately represented as H*L in autosegmental terms, i.e. an instruction to reach a high target in the stressed syllable and to quickly go from there to a low target. Interestingly, when a TB word is not in focus, there is still a rather large rise-fall contour on its stressed syllable. The [+focus] version, however, has a longer rise, which reaches a higher peak pitch somewhat later in the syllable, followed by a larger and steeper fall. The presence of a rise-fall configuration on a [−focus] target would be unusual in many languages, including those of the Germanic family. These languages typically omit any pitch change from non-focused words.

It would seem, therefore, that both word stress and phrasal accent are signaled in TB by an H* target. When a syllable bears both word stress and phrasal accent the two H* targets are stacked on top of each other, which would be quite compatible with the phonetic implementation sketched here.

The effect of boundary in TB is quite straightforward. When the word is at the end of the utterance, the pitch goes down to the speaker's baseline, i.e. reflects final lowering. This could be modeled by associating a low boundary tone L% with the end an utterance. When the target word is followed by a phrase boundary which does not coincide with the end of the utterance, the pitch remains relatively high: it does not go down at all after an H* in a [−focus] word and it goes down only slightly after the HH* configuration in [+focus] targets. The non-final phrase boundary is most adequately represented as a high boundary tone H%. Obviously, in a sequence of H%L%, which is obtained when a phrase boundary coincides with an utterance boundary, the H% has to be deleted.

The situation is more complex, and certainly more variable, in Betawi Malay (and presumably also in Standard Indonesian). Again, the basic shape would seem to be a rise-fall configuration H*L, which may be centered over either the final or the pre-final syllable. When the F0 peak is on the pre-final syllable, the peak is relatively low; when it is on the final syllable the peak is higher (and the pitch interval larger). The rise-fall pattern has simpler variants; the rise part may be omitted when the peak is on the pre-final syllable, and the fall portion may be deleted when the peak is on the final syllable. Deletion of the rise portion may be contextually triggered: if the preceding context ends in a high pitch, then the low between it and the accent on the target word may be deleted. However, no contextual effect seems to be involved in the deletion of the fall portion in accents on the final syllable. Rather, it seems that the speakers have insufficient time to fully realize the fall; therefore they either delete the entire fall (simplifying H*L to H*), or execute only a partial fall which never goes down to the baseline level. This would then be a matter of phonetic truncation of the fall part in order to cope with the limited time frame of the final syllable.

The choice between the small (rise +) fall accent on the pre-final syllable and the larger rise (+ fall) accent on the final syllable is non-deterministically governed by two factors. As we have seen in section 5.3.2.2, the large accent on the final syllable is preferred for words in utterance-final position, and in words that have schwa in the pre-final syllable (see also Table 4). We would assume that stress is on the pre-final syllable by default. When the default syllable contains schwa, the speaker chooses the larger accent on the final syllable in some 50 percent of the cases. If the word is in utterance-final position the preference for the large final accent increases by yet another 50%. As a result the small pre-final accent is found in 88% for sentence-medial targets with full vowels, and the large final accents occur in 91% on sentence-final targets with schwa in the penultimate syllable (see Table 4). The distribution of the two accent types is roughly even for the two other combinations of vowel type and sentence position. It is unclear at this stage whether the small accents on the pre-final and the larger accents on the final syllable can be used interchangeably, i.e. are truly free variants, in spite of the statistical preferences based on vowel type and boundary position, or whether there is some semantic difference between the two. Further research is required here.

The rise-fall pattern is simply absent when the target is out of focus. In this respect BM behaves like most languages we have studied (see above), which delete pitch movements from non-focused items, and differs principally from TB.

We do not argue, of course, that the absence of a pitch mark on [−focus] words in TB reflects the absence of word stress in the language. The same deletion of pitch marks occurs in Germanic languages, which clearly have word stress and may use word stress contrastively.

Toba Batak uses stress contrastively (see introduction). As a consequence stress is functionally more important in Toba Batak than in Betawi Malay. The difference in functional importance of stress is seen in the fact that stress is always realized with a clear pitch movement in TB, even when the word bearing the stress is not in focus. Also, a TB speaker has no freedom to move the accent between two syllables. The position of the accent is tied to exactly one syllable, and we find no exception to the default alignment. In Betawi Malay, however, the speaker seems to have a great deal of freedom between either accent on the pre-final or on the final syllable of [+focus] target words. In our BM materials the distribution of final and pre-final accents is roughly equal. In the informal account of the system we presented in the preceding section we assumed that the default stress is on the pre-final syllable. The accent tends to be moved to the final syllable in exceptional circumstances, i.e., when the pre-final syllable contains a schwa or when the target word appears before an utterance boundary. We point out, once more, that these accentual shifts are not obligatory. The upshot of this is that, from a surface phonetic point of view, BM stress freely varies between the two last syllables in the word. At a more abstract level, however, BM can best be analyzed as a language with fixed stress on the pre-final syllable.

The most comprehensive study on Betawi Malay was done by Wallace (1976). Our results indicate that Wallace was right when he pointed out that stress shifts to the final syllable when the penult contains schwa, as has also been claimed form Standard Indonesian (Cohn 1989, Cohn & McCarthy 1994; see introduction). However, we have shown that the accent shift due to schwa is optional, as has been demonstrated earlier for SI (Laksman 1994). What is new, and has gone unnoticed in the literature, that there is a second optional process that shifts the accent to the final syllable when the word occurs at the end of an utterance. As a result of the interaction between the two optional accent-shift processes, the underlying default position of BM stress is completely obscured.

References

Adelaar, A. (1981) Reconstruction of Proto-Batak phonology. *NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia* 10, 1−20.

Adelaar, A. (2005) The Austronesian languages of Asia and Madagascar: a historical perspective. In A. Adelaar & N.P. Himmelmann (eds) *The Austronesian languages of Asia and Madagascar*. London/New York: Routledge, 1−42.

Alieva N.F., V.D. Arakin, A.K. Ogloblin & Yu.H. Sirk (1991) *Bahasa Indonesia: deskripsi dan teori [Indonesian: description and theory]*. Seri ILDEP. Yogyakarta: Kanisius.

Boersma, P. & D. Weenink (1996) Praat: A system for doing phonetics by computer. *Report of the Institute of Phonetic Sciences, University of Amsterdam* 132.

Chen, P. (1984) Stress in Toba Batak. In P. Schachter (ed) *Studies in the structures of Toba Batak*. UCLA Occasional Papers in Linguistics 5, 1−8.

Cohen, A. & J. 't Hart (1967) On the anatomy of intonation. *Lingua* 19, 177−192.

Cohn, A. (1989) Stress in Indonesian and bracketing paradoxes. *Natural language & linguistic theory* 7, 167−216.

Cohn, A. & J. McCarthy (1994) Alignment and parallelism in Indonesian. *Working papers of the Cornell phonetics laboratory* 12, 53−137.

Emmorey, K. (1984) The intonation system of Toba Batak. In P. Schachter (ed) *Studies in the structures of Toba Batak*. UCLA Occasional Papers in Linguistics 5, 37−58.

Fokker, A.A. (1895) *Malay Phonetics*. Leiden: Brill.

Gerth van Wijk, D. (1985, first published in 1883) *Spraakleer der Maleische Taal [Malaysian grammar]*. Batavia: G. Kolff & Co.

Halim, A. (1974) *Intonation in relation to syntax in Bahasa Indonesia*. Jakarta: Djambatan.

Hart, J. 't, R. Collier & A. Cohen (1990) *A perceptual study of intonation*. Cambridge: Cambridge University Press.

Hermes, D.J. & J.C. van Gestel (1991) The frequency scale of speech intonation. *Journal of the Acoustical Society of America* 90, 97−102.

Heuven, V.J. van (1994) Introducing prosodic phonetics. In C. Odé & V.J. van Heuven (eds)

Experimental studies of Indonesian prosody. Semaian 9. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University, 1−26.

Heuven, V.J. van & E. van Zanten (1997) Effects of substrate language on the localization and perceptual evaluation of pitch movements in Indonesian. In C. Odé & W.A.L. Stokhof (eds) *Proceedings of the 7th International Conference on Austronesian Linguistics*. Amsterdam/Atlanta: Rodopi, 63−80.

Kähler, H. (1966) *Wörterverzeichnis des Omong Djakarta*. Berlin: Dietrich Reimer.

Ladd, D.R. (1996) *Intonational phonology.* Cambridge Studies in Linguistics 79. Cambridge: Cambridge University Press.

Ladd, D.R. & J. Terken (1995) Modelling inter- and intra-speaker pitch range variation. In K. Elenius & P. Branderud (eds) *Proceedings of the 13th International Congress of Phonetic Sciences*, 386−389.

Laksman, M. (1994) Location of stress in Indonesian words and sentences. In C. Odé & V.J. van Heuven (eds) *Experimental studies of Indonesian prosody*. Semaian 9. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University, 108–139.

Moeliono, A.M. & S. Dardjowidjojo (1988) *Tata bahasa baku Bahasa Indonesia [Indonesian reference grammar]*. Jakarta: Balai Pustaka.

Moulines, E. & W. Verhelst (1995) Time domain and frequency-domain techniques for prosodic modification of speech. In W.B. Kleijn & K.K. Paliwal (eds) *Speech coding and synthesis*. Amsterdam: Elsevier Sciences, 519–555.

Muhadjir (1977) *Morfologi dialek Jakarta: Afiksasi dan reduplikasi [Jakartan dialect morphology: Affixation and reduplication]*. Seri ILDEP. Jakarta: Djambatan.

Nababan, P.W.J. (1981) *A grammar of Toba Batak*. Pacific Linguistic Series D-37.

Nooteboom, S. (1997) The prosody of speech: Melody and rhythm. In W.J. Hardcastle & J. Lavers (eds) *The handbook of phonetic sciences*. Oxford: Blackwell, 640–673.

Podesva, R.J. & N. Adisasmito-Smith (1999) Acoustic investigation of the vowel systems of Buginese and Toba Batak. *Proceedings of the 14th International Congress of Phonetic Sciences*, 535–538.

Roosman, L.M. (2006) *Phonetic experiments on the word and sentence prosody of Betawi Malay and Toba Batak*. LOT Dissertation Series 129. Utrecht: LOT.

Samsuri (1971) *Tjiri-tjiri prosodi kalimat bahasa Indonesia [Prosodic features of Indonesian sentences]*. Malang: Tim Publikasi Ilmiah, Fakultas Keguruan Sastra dan Seni, IKIP Malang.

Sibeth, A. (1991) *The Batak peoples of the island of Sumatra*. London: Thames and Hudson.

Strik, H. (1994) *Physiological control and behaviour of the voice source in the production of prosody.* PhD dissertation, Catholic University Nijmegen, The Netherlands.

Tuuk, H.N. van der (1971) *A grammar of Toba Batak.* Den Haag: Martinus Nijhoff. [first published in 1864].

Wallace, S. (1976) *Linguistic and social dimensions of phonological variation in Jakarta Malay.* PhD Dissertation, Cornell University.

Woollams, G. (2005) Karo Batak. In A. Adelaar & N.P. Himmelmann (eds) *The Austronesian languages of Asia and Madagascar*. London/New York: Routledge, 534–561.

Zanten, E.A. van, R.W.N. Goedemans & J.J.A. Pacilly (2003) The status of word stress in Indonesian. In J.M. van de Weijer, V.J. van Heuven & H.G. van der Hulst (eds) *The phonological spectrum. Volume II: Suprasegmental structure*. Amsterdam/Philadelphia: John Benjamins, 151–175.

Zanten, E.A. van & V.J. van Heuven (1998) Word stress in Indonesian; its communicative relevance. *Journal of the Humanities and Social Sciences of Southeast Asia and Oceania [Bijdragen tot de Taal-, Land- and Volkenkunde]* 154, 129–149.

Zanten, E.A. van & V.J. van Heuven (2004) Word stress in Indonesian: Fixed or free? *NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia* 53, 1–20.