Betawi Malay word prosody

Vincent J. van Heuven*, Lilie Roosman1, Ellen van Zanten

LUCL, University of Leiden, The Netherlands

Received 15 September 2006; received in revised form 17 July 2007; accepted 1 September 2007

Abstract

Betawi Malay is an Austronesian language with no word-based stress, but with phrasal accent. We recorded four speakers (two male, two female), who produced target words in utterance-medial and utterance-final position, and with the target words in or out of focus. We present a detailed acoustic analysis of the pitch patterns realised on the focused target words. The results show that both the shapes and the positions of the pitch configurations are highly variable. The choice of a particular variant is governed nondeterministically (by gradient rules) by (i) the presence of schwa in the pre-final syllable and (ii) an utterance boundary following the target.

© 2008 Elsevier B.V. All rights reserved.

Keywords: Stress; Accent; Intonation; Focus; Domain-final lengthening; Gradient rule

1. Introduction

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, sentence accent typically coincides with a word stress (e.g. Jun, 2005; Ladd, 1996). Languages without word stress may also use accent to highlight words in sentences but the sentence accent is then not restricted to a particular syllable in the word. Standard Indonesian has

* Corresponding author.

E-mail addresses: v.j.j.p.van.heuven@let.leidenuniv.nl (V.J. van Heuven), lilieroosman@yahoo.co.uk (L. Roosman), e.a.van.zanten@let.leidenuniv.nl (E. van Zanten).

1 Also at Program Studi Belanda, Universitas Indonesia, Jakarta, Indonesia.

0024-3841/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2007.09.005
been claimed to be one such language (van Zanten and van Heuven, 1998; van Zanten et al., 2003).

Our study focuses on the production of word prosody in Betawi Malay (BM), an Austronesian language spoken in Jakarta, the capital of Indonesia. Just like Standard Indonesian, BM does not have word stress (Muhadjir, 1977) although it does have phrasal accent (Wallace, 1976). We will focus on the way in which accent is realised in BM.

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 accent. For a detailed analysis of the role of duration in BM we refer to Roosman (2006). We expect the exact position of the pitch change marking focus to be rather variable in BM. When a language has lexical stress, as in English or Dutch, then the accent-lending pitch change is tied rather strictly to one specific location within or associated with the stressed syllable in the word (e.g. Caspers and van Heuven, 1993; Ladd et al., 2000). We know little about the alignment of accents in languages which do not have lexical stress. However, since accents in non-stress languages are associated with words rather than syllables, we assume that there is no need for speakers of these languages to align the pitch movement as strictly as speakers of stress languages do (provided the accented word is perceived as prominent by the listeners).

Some evidence for such loose alignment was provided by van Zanten et al. (2003), who researched Indonesian, a non-stress language closely related to BM. Indonesian speakers with a Javanese background (again a non-stress language) produced (accent-lending sentence-final) pitch falls somewhere near the boundary between the pre-final and the final syllable. Javanese as well as Jakartan listeners found any accents on the right-hand side of the word equally acceptable, whether they occurred on pre-final or final syllable, or in the border region between these two syllables. Even more remarkably, van Zanten and van Heuven (2004) found that, at least sentence-finally, accent positions anywhere in the word were acceptable to Indonesian listeners. In the present study we will test the hypothesis that BM has a loose alignment of the accent-lending pitch change in words spoken in focus.

The marking of prosodic boundaries, signalling breaks between clauses and utterances, may also involve both duration and pitch. Boundaries may be signalled by boundary tones and words in sentence-final position are usually longer than words in other positions (e.g. Nooteboom, 1997). 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% vs. L%), and see how these boundary tones may interact with accent-marking tone configurations in BM.

This article is organised as follows. In section 2 we will describe the experimental design used to record materials. Section 3 specifies the acoustic and perceptual analysis of the materials, and presents the results. Section 4 links the results to the more general questions formulated in this section. However, before turning to the experimental work, we will first introduce the target language and discuss claims made in the literature with respect to its prosodic structure.

1.2. Betawi Malay

Betawi Malay, the dialect of the central part of the city of Jakarta (Dialek Kota), is used by a homogeneous ethnic group, the Betawi, and has had comparatively little influence from
other languages (Grijns, 1991). 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. These language varieties seem to resemble each other prosodically.

There has been some discussion whether Standard Indonesian has lexical stress or only phrasal accent. Gerth van Wijk (1883) claims that Indonesian has stress, but observes that it is usually very weak. Stress 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.

Others claim that main stress is generally on the word-final syllable (Samsuri, 1971) or that Standard Indonesian has no word stress (Alieva et al., 1991:34; Fokker, 1895; Halim, 1974:111–113). According to Halim, prominence depends on the position of the word in the sentence: before a sentence-internal boundary the accent falls on the final syllable of the word preceding the boundary, whereas a sentence-final accent falls on the penultimate syllable of the last word of the sentence.

Moeliono and 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 et al. (1991:34) also claim that 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 are longer. The ways in which those accented syllables are realised 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 do not signal 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 Standard Indonesian stress is either weak or non-existent. Nevertheless, rules for the placement of word stress were formulated by Cohn (1989) and Cohn and 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 and van Heuven (1998, 2004) found no preferred stress position in Standard Indonesian. Similarly, van Zanten et al. (2003) conclude on the basis of experimental evidence that Standard Indonesian does not have word-based stress, but has phrase-level accent only.

The same seems to hold for BM prosody. Wallace (1976) notices that the domain of the accent is the phrase rather than the word; there is no word stress in BM. Wallace has the impression that accent in BM is realised 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 (Wallace, 1976:56–59). The accent goes to the final position if the penultimate has schwa (1a), or if the last word of the phrase is made up of a monosyllabic stem preceded by a prefix (1b). A phrase-final monosyllabic word is always accented (1c).

(1) a. rumenye gedé [gøːdɛ] ‘the house is big’
   b. ubinnye dipél ‘the floor is mopped up’
   c. masukin di bák ‘put into the bin’
The penultimate syllable with schwa can be accented when it precedes the unaccented suffix
nye [ne], such as in itémmne [itämme] ‘the black, being black’, sambélmye [sambälme] ‘the chilli sauce’. Wallace did not consider words with schwa in both penultimate and final syllable, like deket [dakat] ‘close to’, seneng [sänæj] ‘happy’, kelelep [kälap] ‘be drowned’.

Summarising, the literature seems to indicate that BM does not have a word-based stress but rather a (postlexical) phrase-based accent. In this respect BM is not unlike languages such as French and Korean. In both these languages word prominence is not marked at the lexical level, but at the postlexical level by a rising tone at the edge of an Accentual Phrase. The difference between these languages lies in the fact that in French the last full syllable in the Accentual Phrase carries the postlexical pitch accent, whilst in Korean no syllable carries postlexical pitch accent; here prominence is imparted to a word by a boundary tone, either at the leading or the trailing edge of the Accentual Phrase (Jun, 2005:444, 447).

2. Methods

2.1. Materials

We selected eight words consisting of two or three syllables. Three words containing a schwa vowel in the pre-final syllable, pete [pate] ‘stinking bean’, deket [dakat] ‘nearby’, rejeki [rajåki] ‘fortune’, were chosen to investigate whether the schwa behaves differently from full vowels under the influence of focus and boundary marking/sentence position. A further five words containing full vowels in the last two syllables were used: kaga [kaga] ‘no, not’, kutu [kutu] ‘louse’, belaga [bålåga] ‘pretend’, pipi [pipi] ‘cheek’ and pepet [påpet] ‘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 relevant examples.

<table>
<thead>
<tr>
<th>Prominence</th>
<th>Boundary</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>+Focus +Final</td>
<td>(Die bilang ape?)</td>
<td>‘What did he say?’</td>
</tr>
<tr>
<td>-Final</td>
<td>(Die bilang [kutu] tadi)</td>
<td>‘He said [kutu]’</td>
</tr>
<tr>
<td>-Focus +Final</td>
<td>(Kapan die bilang [kutu]?)</td>
<td>‘When did he say [kutu]?’</td>
</tr>
<tr>
<td>-Final</td>
<td>(Die bilang [kutu] ape?)</td>
<td>‘What [kutu] did he say?’</td>
</tr>
</tbody>
</table>

2.2. Speakers and recording procedure

Four native speakers (two male, two female, between 30 and 55 years old) took part in the experiments. They 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 question and answer sentences were presented to the speakers in a fixed order. Another speaker of the same language read out the question sentences; the subject then responded by
reading the corresponding answers. The recordings were made in a quiet room on 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. All speech materials were then digitised (16 kHz sampling frequency, 16 bits amplitude resolution).

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.

3.1. Stylisation

For the analysis of the 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 the analysis-by-synthesis method (Nooteboom, 1997 and references therein). The point in the stylisation where a rise changes into a fall (or vice versa) is called a pivot point. The stylisation procedure is exemplified in Fig. 1. It should be noted that the overall trend of the sentence melody is not level but slopes down gently (‘downtrend’). The relevant pivot points were the following:

(2) a. p1 A 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.

b. p2 This point is defined as the F0 maximum (i.e. the highest F0) in the target word.

c. p3 A pivot point between p2 and p4 that affords the stylisation of pitch fall in terms of two straight-line segments, the first of which embodies a steep fall whilst the second part drops off at a modest rate.

d. p4 F0 minimum between p3 and the end of the utterance. When the target was utterance final, p4 is typically the terminal pitch.

Fig. 1 gives an example of a stylisation of a pitch curve produced by a male speaker in [+focus, +final] condition. In order to abstract from the overall pitch difference between male and female voices, we applied a minimal normalisation 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), as is often done when comparing vocal pitch in intonation languages across registers (Hermes and van Gestel, 1991; Ladd and Terken, 1995; Nooteboom, 1997). Pitch intervals of equal sizes when expressed in ERB should be perceptually equivalent, regardless of their absolute frequency. 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 ERB.

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.

---

2 This article reports on speech material in [+focus] condition only.
at p4. This allows straightforward comparison of pitch differences within and between utterances.

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 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 recordings. The next section describes the procedure and results of the auditory screening by a panel of expert listeners.

Preliminary auditory and visual inspection of the materials revealed that pitch movements 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 typically coincided with the final syllable; accented sentence-medial targets, however, predominantly 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 section 1.2); the distributional details were in fact predicted by Kähler (1966).

3.2. 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, a native speaker of BM, listened to the materials and decided for each of the 192 utterances (4 speakers × 2 positions × 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. Table 2 summarises 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 vs. absence of a prominence-lending pitch movement.
Table 2 shows that the male speakers quite consistently realised accent-lending pitch movements on the targets. The female speakers dropped the pitch movements in some 20–35% of sentence-medial targets; female 2 dropped her pitch movements in sentence-final position in more than 50% of the cases. We decided to exclude the 33 utterances without audible accent on the [+focus] target word from further analysis.

As a next stage in the auditory screening, the second and third authors, as well as a third expert on prosody, listened to the remaining 159 [+focus] utterances and judged syllable prominence. 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 minus 33 tokens without an audible accent) judged by three listeners yielded 477 valid judgements (231 [+final] and 246 [−final] judgements). Table 3 summarises 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. In the table we present the results broken down per speaker and aggregated over all speakers. We will first discuss the overall effects, and then consider to what extent effects may differ between speakers.

On aggregate, Table 3 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. 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.

In Table 3 the data have been aggregated over four speakers. The question may be raised to what extent the differences in the distribution of accent over the final and pre-final syllables is speaker-dependent. The first three speakers basically show the same pattern as was found in the aggregate results. These speakers stress the pre-final syllable, but are prone to shifting the accent to the final if (i) the pre-final syllable contains schwa and/or if (ii) the target word is in sentence-final position. Speaker 4, however, displays a rather different pattern. She tends to have the accent

---

3 We thank Johanneke Caspers for acting as our third expert listener.
4 For individual listener results and statistical evaluation of the between-listener agreement, see Roosman (2006).
on the final syllable unless the target word is in sentence-final position and has a full vowel in the
pre-final syllable, which is then accented instead. Speaker 4, then, concurs with the other
speakers in applying the rule that shifts the accent to the final syllable if the pre-final syllable
contains schwa. She differs from the other speakers in the effect of the boundary rule, which
seems reversed: she typically has the accent on the final syllable but shifts it to the pre-final one if
the word is in sentence-final position.

As a last check on the distribution of accents, we present Fig. 2, which is a histogram of the
timing of the F0 peaks within the target words. The alignment of the F0 peak is expressed relative
to the onset of the pre-final vowel.

We expected a uni-modal, more or less normal, distribution of the temporal locations of the F0
peaks in our materials. This would explain why our expert listeners found themselves unable to
decide whether the accent was on the final or the pre-final syllable in a fairly large number of
cases, viz. when the peak was at or near the break between the final two syllables. Counter to this
expectation, however, we find a clearly bi-modal distribution. About half of the F0 peaks are
found relatively close to the onset of the pre-final vowel. There is a second group of F0 peaks,
which are all aligned at or after the onset of the final vowel. Crucially, there is a gap in the
distribution between 133 and 167 ms. It would appear, therefore, that our BM speakers make an
effort to align the accent-lending F0 peak either on the pre-final or on the final syllable (or vowel).
It is not the case, then, that the alignment of the F0 peak is stochastically distributed over a time-
window equal to the final two syllables (this theoretical distribution is plotted in Fig. 2 as a
continuous bell-shaped curve).
We may observe that the F0 peaks are clustered more tightly when they are aligned with the pre-final vowel, and much more widely when the alignment is with the final syllable. This is quite probably due to the fact that the duration of the final syllable, but not that of the pre-final syllable, is greatly affected by final lengthening.

The following step in the auditory analysis (in section 3.3) was to establish the token frequency of specific types of accent-lending pitch movements on the prominent syllables, to allow us to determine the shapes of the various pitch configurations in acoustic terms. This will be described in section 3.4.

3.3. Token frequencies of 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 the complex rise–fall configurations, which were subdivided into early vs. 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. Table 4 presents 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 prominent syllable as determined by the listening panel.

<table>
<thead>
<tr>
<th>Shape</th>
<th>Prominence heard</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-final</td>
<td>Final</td>
</tr>
<tr>
<td>F</td>
<td>134</td>
<td>17</td>
</tr>
<tr>
<td>R</td>
<td>27</td>
<td>100</td>
</tr>
<tr>
<td>RF pre-final</td>
<td>72</td>
<td>4</td>
</tr>
<tr>
<td>RF final</td>
<td>7</td>
<td>71</td>
</tr>
<tr>
<td>Total</td>
<td>240</td>
<td>192</td>
</tr>
</tbody>
</table>

Fig. 2. Distribution of position of F0 peak (ms relative to onset of pre-final vowel) in all 159 [+focus] targets with a measurable F0 contour.
The results show first that in a large majority of the cases the panel of listeners could clearly differentiate between prominence on the final vs. pre-final syllable. Only in 9% of the cases (45/477) were the listeners undecided as to where the prominence fell. This minority of cases will not be analysed any further.

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 predominantly 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.

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.

In the following section we will present the acoustic analysis of the pitch contours on the utterances. The acoustic analysis will concentrate on the pitch configuration as found on accented targets (as judged by the second author) only. We will not consider the (basically flat) pitch pattern that was found on [–focus] targets.

3.4. Acoustic properties of accent-lending pitch configurations

ANOVAs show that the effects of word position in the sentence are highly significant on the peak timing\(^5\) \([F(1, 157) = 25.9, \ p < 0.001]\), on the height of the peak \([F(1, 157) = 41.6, \ p < 0.001]\), and on the fall excursion \([F(1, 107) = 30.4, \ p < 0.001]\). A significant effect of word position is also found on the rise onset \([F(1, 101) = 5.2, \ p = 0.025]\) and on the slope of the rise \([F(1, 101) = 4.1, \ p = 0.045]\). In sentence-final words the rises start, on average, 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 0.66 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 (0.69 ERB) than the fall excursion in sentence-final words.

Effects of vowel type are significant only on the slope of the rise \([F(1, 101) = 4.4, \ p = 0.039]\) and the slope of the fall \([F(1, 107) = 4.3, \ p = 0.040]\). 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 \([F(3, 99) = 10.2, \ p < 0.001]\), on the peak timing \([F(3, 155) = 18.7, \ p < 0.001]\) and on the rise excursion \([F(3, 99) = 5.6, \ p = 0.001]\). Furthermore, significant effects of movement shape are found on the steepness of the rise \([F(3, 99) = 2.9, \ p = 0.038]\), and on the fall excursion \([F(3, 105) = 3.0, \ p = 0.036]\).

Based on these results, we will analyse the sentence-final and the sentence-medial words separately. In each section, the F0 parameters of every shape will be analysed. 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 first present the F0 contours of target words in sentence-final position. **Fig. 3** (and **Fig. 4** below for target words in sentence-medial position) plot the stylised F0 contours in normalised ERB as a function of time, such that the timing of the movements is expressed relative to the onset of the penultimate vowel in the target word. In these figures rises, falls and rise–fall combinations are plotted separately. In the computations, rises were 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 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.

**Fig. 3** 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 characterised 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 ERB (F0 final for rise–fall) or 1.5 ERB (F0 peak for rise only) above it. Movements aligned with the pre-final syllable, however, end at approximately 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 $[F(3, 73) = 13.4, p < 0.001]$, the rise excursion $[F(3, 58) = 5.9, p = 0.001]$ and on the slope of the fall $[F(3, 44) = 5.4, p = 0.003]$. T-tests with two movement shapes, the rise–fall in pre-final syllable and that in final syllable as independent variables, show that there are significant differences between the two movements in terms of peak height $[t(31) = -2.05, p = 0.049]$, peak timing$^6$ $[t(31) = -3.47, p = 0.002]$, rise excursion $[t(31) = -4.04, p < 0.001]$, and slope of the fall $[t(31) = -2.74, p = 0.010]$. 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 0.63 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 0.88 ERB, than the pre-final syllable

$^6$ The peak timing on the final syllable is measured relative to the onset of the vowel in the 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 summarised in Table 5.

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 almost 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. The rise portion of the configuration typically occupies some 250 ms, and during this time interval the pitch rises by over 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). Apparently, BM speakers choose to truncate, rather than scale down, rise–fall configurations in final position.

Generalising 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 [rajaki], 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 schwas (in deket), there is much less space for the rise–fall configuration; as a result, the entire fall portion is dropped.

In sentence-medial position (Fig. 4) 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. Fig. 4 illustrates the four pitch movements in sentence-medial position.

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 \(F(3, 37) = 9.0, p < 0.001\) and for the peak timing \(F(3, 78) = 5.9, p = 0.001\). However, a post hoc test (Scheffé, with \(\alpha = 0.05\)) indicated that the peak timing is significantly different only between the fall and the rise–fall in the pre-final syllable. Table 6 summarises the pitch measurements of the accented words in sentence-medial position.

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 \(F(1, 54) = 22.1, p < 0.001\), with a mean difference of 0.93 ERB. Sentence-medial falls are significantly larger \(F(1, 54) = 21.9, p < 0.001\), around 0.87 ERB, and 3.9 ERB/s steeper \(F(1, 54) = 10.9, p = 0.002\), than the sentence-final falls.

Rise movements differ only in the peaks: the sentence-final rises have later peaks \((\Delta = 54 \text{ ms})\) than the sentence-medial rises \(F(1, 48) = 15.9, p < 0.001\). The sentence-medial rises also have significantly higher peaks \((\Delta = 0.63 \text{ ERB})\) than the sentence-final rises \(F(1, 48) = 18.0, p < 0.001\).

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 \(F(1, 24) = 8.6, p = 0.007\), 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 \(F(1, 24) = 4.7, p = 0.041\).

### Table 6

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Fall</th>
<th>Rise</th>
<th>RF pre-final</th>
<th>RF final</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset timing (ms)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(-171.30 (81.00))</td>
<td>(-53.20 (60.00))</td>
<td>(-187.50)</td>
<td>(-117.00)</td>
<td></td>
</tr>
<tr>
<td>Peak timing (ms)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>20.60 (14.00)</td>
<td>38.20 (28.00)</td>
<td>57.30 (57.00)</td>
<td>5.40</td>
<td>33.00</td>
</tr>
<tr>
<td>F0 peak (normalised ERB)</td>
<td>2.28 (0.70)</td>
<td>2.17 (0.58)</td>
<td>2.09 (0.78)</td>
<td>2.98</td>
<td>2.22</td>
</tr>
<tr>
<td>Rise excursion (ERB)</td>
<td>1.15 (0.49)</td>
<td>0.98 (0.63)</td>
<td>0.91</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>Slope rise (ERB/s)</td>
<td>5.91 (2.80)</td>
<td>10.46 (9.20)</td>
<td>4.70</td>
<td>8.00</td>
<td></td>
</tr>
<tr>
<td>Final F0 (normalised ERB)</td>
<td>0.46 (0.64)</td>
<td>0.52 (0.52)</td>
<td>1.31</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Fall (ERB)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.78 (0.41)</td>
<td>1.47 (0.84)</td>
<td>1.67</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>Slope fall (ERB/s)</td>
<td>9.25 (4.00)</td>
<td>8.46 (4.10)</td>
<td>13.37</td>
<td>9.10</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Relative to the vowel onset of the accented syllable.

<sup>b</sup> From the peak (p2) to the next lower point.
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.

3.5. 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 or rise–fall configurations. These findings indicate that accent position in the word is free in BM – at least within a two-syllable window at the end of the word.

The presence of a sentence boundary following the target word largely 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 to be true in sentence-final words only. Our results show that accent shifts not only due to the type of the vowel (full vs. schwa), but also due to the position of the word in the sentence.

4. Discussion and conclusion

In section 1 we formulated a hypothesis that the distribution of accents in a word in Betawi Malay would be characterised by a loose alignment. Such an alignment would show up as a unimodal distribution of accent-marking F0 peaks over the word domain. This distribution can be either uniform or normal (i.e. bell-shaped) as long as it is unimodal.

Our results show that accentuation in Betawi Malay is rather complex and variable. The basic shape of an accent seems to be a rise–fall configuration H−L, which may be centred over either the final or the pre-final syllable. When the accent 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.

The choice between the small (rise+)fall accent on the pre-final syllable and the larger rise(+fall) accent on the final syllable seems non-deterministically governed by two factors. As we saw in section 3.2, the large accent on the final syllable is most frequent with words in utterance-final position, and with words that have schwa in the pre-final syllable (see also Table 3). We would assume that accent 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% of the cases. If the word is in utterance-final position, the incidence of the large final accent
increases also by 50%. As a result, the small pre-final accent is found in 88% of sentence-medial
targets with full vowels, and the large final accents occur in 91% of sentence-final targets with
schwa in the penultimate syllable (see Table 3, not counting the cases with equal prominence).
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.

In our 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 accent 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. Notice once more that these accenetal shifts are not obligatory. The
upshot of this is that, from a surface-phonetic point of view, BM accent freely varies between the
two last syllables in the word. At a more abstract level, however, BM can best be analysed as a
language with default accent on the pre-final syllable.

The most comprehensive study on Betawi Malay was carried out by Wallace (1976). Our
results indicate that Wallace was right when he pointed out that accent shifts to the final syllable
when the penult contains schwa (see section 1). However, we have shown that the accent shift due
to schwa is optional, as has already been demonstrated for Standard Indonesian as spoken by a
Jakartan speaker (Laksman, 1994). What is new, and has gone unnoticed in the literature, is 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 accent is largely obscured.

How can we model the BM system in autosegmental-metrical terms? An AM account would
postulate a system with pre-final accent, which would account for the fact that accent is realised
on the pre-final syllable in the majority of the cases. There are two rules that shift the accent to the
final syllable in certain marked situations, viz. (i) when the pre-final syllable is light (i.e. contains
schwa) and (ii) when the word is utterance final. Process (i) is found in many languages –
including English and Dutch – and is typical of weight-sensitive stress systems. Process (ii) can
be seen as an interaction with a boundary tone. It can be modelled, and has in fact been modelled
(e.g. for Bengali, see Hayes and Lahiri, 1991), by reversing the order of the accent and the
boundary tone, in which case the accent is attracted to the right-hand word edge in pre-pausal
position. Such an account would adequately cover the facts of BM.

What seems incompatible with AM theory is that the rules that cover the marked situations in
(i) and (ii) are not deterministic. They are variable rules that roughly apply in 50% of the cases.
Variable rules were introduced into linguistic theory by Sankoff and Labov (1979) to model
variable behaviour across different language users. As we have seen in the present article, the
variable accentuation was largely speaker-independent, and there were no marked differences
among lexical items within the same category (i.e. within the category of words with full vowels
vs. schwa in the pre-final syllable). More recently, gradient rules have been proposed in
phonology. These typically operate in the postlexical stage and are often regarded as a matter of
phonetic implementation (Hargus and Kaisse, 1993). It would appear that such gradient rules
have to be invoked if we want to do justice to the phenomena found in BM as described here.
Such an account would view Betawi Malay as a language with underlying fixed accent in pre-
final position in the phrase. The two gradient rules would then generate massive variability, such
that accent would appear to be free at the observable surface.
Acknowledgements

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

References


