



Universiteit
Leiden
The Netherlands

Word stress in Indonesian; its communicative relevance

Zanten-Wervelman, E.A. van; Heuven, V.J. van

Citation

Zanten-Wervelman, E. A. van, & Heuven, V. J. van. (1998). Word stress in Indonesian; its communicative relevance. *Journal Of The Humanities And Social Sciences Of Southeast Asia And Oceania*, 154, 129-149. Retrieved from <https://hdl.handle.net/1887/1068>

Version: Not Applicable (or Unknown)
License: [Leiden University Non-exclusive license](#)
Downloaded from: <https://hdl.handle.net/1887/1068>

Note: To cite this publication please use the final published version (if applicable).

ELLEN VAN ZANTEN and VINCENT J. VAN HEUVEN

Word Stress in Indonesian Its Communicative Relevance

1. Introduction

In lexical stress languages such as English and Dutch, stress patterns are known to listeners and stress is used in auditory word recognition. In English, words like *FOREbear* and *forBEAR* or *TRUSty* and *trustEE* (stressed syllables are capitalized) are mutually distinguishable by their stress patterns. The opposition between the stressed and unstressed syllables can be expressed by a range of prosodic means. Stress may be expressed by a change in pitch; the stressed elements *FORE* and *BEAR* will also be longer and/or louder than unstressed *for* and *bear*. At the same time, stressed syllables in English have full vowels, whereas in unstressed syllables vowels are often reduced to schwa (e.g., *TElephone* ['teləfəʊn] versus *teLEphony* [tə'lefəni]). In many Western languages stress information is important in speech perception. 'Given perceptually ambiguous information, lexical stress information can be used to resolve the ambiguity in favor of a word' (Connine, Clifton and Cutler 1987:145; see also Van Heuven 1988). In Indonesian, as opposed to many Western languages, stress is not distinctive: there are no words containing the same sequence of vowels and consonants that differ in their stress patterns (and consequently in their meanings).

In this article we will draw a systematic distinction between (word) stress and (sentence) accent. Stress is an abstract linguistic property of a word or morpheme marking the syllable within the unit that is felt by the

ELLEN VAN ZANTEN, a research associate in the Department of Linguistics at the University of Leiden, took her PhD at that University. With phonetics/phonology of Indonesian and other languages in Indonesia as her field of specialization, she is the author of *The Indonesian vowels: Acoustic and perceptual explorations*, PhD thesis, Leiden University, 1989, and 'Perception of Indonesian vowels spoken in context', *Phonetica* 45 (1988):43-55.

VINCENT J. VAN HEUVEN is Associate Professor of Phonetics at Leiden University and director of the Phonetics Laboratory at the Holland Institute of Generative Linguistics (HIL). He obtained his PhD from the University of Utrecht and has specialized in experimental linguistics and phonetics, orthography and reading, and speech technology. His publications include: (with L.C.W. Pols (eds)), *Analysis and synthesis of speech; Strategic research towards high-quality text-to-speech generation*, Berlin: Mouton de Gruyter, 1993, and (with L. Menert) 'Why stress position bias?', *Journal of the Acoustical Society of America* 100 (1996):2439-51.

Both authors may be contacted at the Phonetics Laboratory, Leiden University, P.O. Box 9515, 2300 RA Leiden, The Netherlands.

ELLEN VAN ZANTEN and VINCENT J. VAN HEUVEN

Word Stress in Indonesian Its Communicative Relevance

1. Introduction

In lexical stress languages such as English and Dutch, stress patterns are known to listeners and stress is used in auditory word recognition. In English, words like *FORE*bear and *for*BEAR or *TRUSTy* and *trustTEE* (stressed syllables are capitalized) are mutually distinguishable by their stress patterns. The opposition between the stressed and unstressed syllables can be expressed by a range of prosodic means. Stress may be expressed by a change in pitch; the stressed elements *FORE* and *BEAR* will also be longer and/or louder than unstressed *for* and *bear*. At the same time, stressed syllables in English have full vowels, whereas in unstressed syllables vowels are often reduced to schwa (e.g., *Telephone* ['teləfəʊn] versus *teLEphony* [tə'lefəni]). In many Western languages stress information is important in speech perception. 'Given perceptually ambiguous information, lexical stress information can be used to resolve the ambiguity in favor of a word' (Connine, Clifton and Cutler 1987:145; see also Van Heuven 1988). In Indonesian, as opposed to many Western languages, stress is not distinctive: there are no words containing the same sequence of vowels and consonants that differ in their stress patterns (and consequently in their meanings).

In this article we will draw a systematic distinction between (word) stress and (sentence) accent. Stress is an abstract linguistic property of a word or morpheme marking the syllable within the unit that is felt by the

ELLEN VAN ZANTEN, a research associate in the Department of Linguistics at the University of Leiden, took her PhD at that University. With phonetics/phonology of Indonesian and other languages in Indonesia as her field of specialization, she is the author of *The Indonesian vowels; Acoustic and perceptual explorations*, PhD thesis, Leiden University, 1989, and 'Perception of Indonesian vowels spoken in context', *Phonetica* 45 (1988):43-55.

VINCENT J. VAN HEUVEN is Associate Professor of Phonetics at Leiden University and director of the Phonetics Laboratory at the Holland Institute of Generative Linguistics (HIL). He obtained his PhD from the University of Utrecht and has specialized in experimental linguistics and phonetics, orthography and reading, and speech technology. His publications include: (with L.C.W. Pols (eds)), *Analysis and synthesis of speech; Strategic research towards high-quality text-to-speech generation*, Berlin: Mouton de Gruyter, 1993, and (with L. Menert) 'Why stress position bias?', *Journal of the Acoustical Society of America* 100 (1996):2439-51.

Both authors may be contacted at the Phonetics Laboratory, Leiden University, P.O. Box 9515, 2300 RA Leiden, The Netherlands.

Ebing (1997) reports phrase-final pitch rises of 2 to 4 ST. In our own material, accent-lending rises of approximately 2.5 ST were quite frequently observed in sentence-final position. Such values barely exceed the minimum for accent-lending pitch movements reported by Van Heuven (1994) for Dutch speakers, namely 3 ST. In addition, the difference in duration between stressed and unstressed syllables is comparatively small in Indonesian (Van Zanten and Van Heuven 1997). An earlier perception experiment (Van Zanten and Van Heuven submitted) provided evidence that Indonesians are relatively tolerant as regards stress and its position. Neither the form nor the position of the accent-lending pitch movement associated with the stressed syllable seems to be of crucial importance to Indonesian listeners. This would suggest that, indeed, stress is free in Indonesian.

In the present study we investigate whether stress plays a role in the auditory recognition of words in Indonesian, and thus may be communicatively relevant. To find out what is the contribution of stress perception to word identification in Indonesian we chose the so-called gating² paradigm (Grosjean 1980) for our experimental method. In gating, a spoken language stimulus is presented repeatedly, with a larger portion of the material being made audible on each consecutive pass. The listeners' task is to identify (often guess) the word presented after each pass. The gating paradigm was used previously by Van Heuven (1988) to collect evidence on the role of lexical stress in word recognition in Dutch. Van Heuven selected pairs of Dutch words of which both members had the same onset CV combination. Crucially, one member of each pair had a stressed and the other member an unstressed first syllable (e.g., *KAvia* 'guinea pig, cavy' versus *kaNarie* 'canary'). Three gates of gradually increasing length were presented to the listeners. For the first gate the stimuli were truncated just after the first syllable, so that only stressed *KA* or unstressed *ka* was audible. Of the correct responses, the great majority (76%) reflected the stress position intended by the speaker, thus indicating that in Dutch, stress information plays a role in word recognition.

Here we intend to establish in a similar way whether Indonesian listeners use stress information in word recognition. To this end, listeners are requested to identify (parts of) words which are phonemically identical but have a different stress position according to the accepted rule (stress on the penultimate syllable). If Indonesian listeners identify the stimuli correctly, we will assume that what has enabled them to achieve this is stress information (or other prosodic information, such as syllable

² The term 'gate' originally refers to the closing of an electronic circuit by which a signal (electrical current, e.g., speech) is passed on from a source (e.g., a tape recorder or computer) to a destination (e.g., an amplifier or loudspeaker). By extension, 'gate' has now come to refer to the time *interval* during which a speech stimulus is made audible.

shortening in longer words) and that stress may therefore be linguistically relevant in Indonesian. Conversely, if the subjects are not able to identify the stimuli correctly, we will conclude that stress plays no role in word identification, suggesting that stress is essentially free in Indonesian.

Some languages in the Indonesian area have distinctive stress. Nababan (1981) shows, for instance, that in Toba Batak, in contradistinction to Indonesian, stress functions contrastively, e.g. *TIBbo* 'height' ~ *tibBO* 'high', *Itom* 'black dye' ~ *iTOM* 'your sibling'. We argue that listeners who are speakers of a distinctive-stress language will be quicker to retrieve stress information than listeners who are speakers of a language with non-distinctive stress. Indeed, as Van Heuven and Van Zanten (1997) have found earlier, Indonesian listeners with the Toba Batak substrate (distinctive stress) were more accurate in locating pitch accents in a given utterance than Indonesian listeners of non-distinctive-stress backgrounds (where the same utterance was used in the case of both groups). In the current experiment we will compare the performance of Indonesian listeners with that of native speakers of Dutch, which is a distinctive-stress language. It is our prediction that Dutch listeners will be more sensitive to stress cues contained in the stimuli than Indonesian listeners.

2. Method

2.1 Gating paradigm

To assess the contribution of stress to word perception, we used the gating paradigm (see above). For our purposes two passes (using two 'gates', namely (C)V- and (C)VCV-) will be sufficient. Crucially, our stimuli differ in stress position.

2.2 Stimulus material

Ten sets of three target words each were constructed (see Appendix 1). Within each set, the beginning (or onset) of the target words contained the same sequence of phonemes, but the words differed in length. According to the accepted rule (stress on the prefinal syllable), the words thus differed in stress position; for example (supposedly stressed prefinal syllable is capitalized):

<i>Anak</i>	'child'
<i>aNAKnya</i>	'his/her child'
<i>anak-ANak</i>	'children'

To help listeners determine the stressed/unstressed character of the target onsets, the material was embedded in a prosodic context consisting of a fixed carrier sentence, as follows: *Dia mengucapkan kata (anak)*, 'He pronounces the word (anak)'. Being the only variable element in the sentence,

and in sentence-final position, target words were expected to be marked prosodically with a pitch accent on the stressed syllable (Samsuri 1978; Van Heuven 1994:15). In view of the uncertain position of the Indonesian schwa as regards its stressability (see Laksman 1994), the research was restricted to peripheral (non-central) monophthongs.

Of the ten sets of three target words, eight sets consisted of a two-, a three-, and a four-syllable word each. One set included a five-syllable (*imbang-imbangan*) instead of a four-syllable word, and one set consisted of a three-, a four-, and a five-syllable word (*mengangkat, mengangkat* and *mengangkat-angkat*).³

The thirty target words in their fixed sentence frame were randomized, typed on sheets and read aloud twice by an Indonesian speaker of Balinese descent. The speaker was instructed to speak fluently. The recordings were made in a sound-proofed booth with a Sennheiser unidirectional condenser microphone (MKH 416) onto a DAT recorder (48.1 kHz, 16 bits). They were then transferred to a Silicon Graphics computer and downsampled to 16 kHz.

The purpose of the present experiment is to find out what is the contribution of stress perception to word identification. To this end we presented the listeners with the segmentally identical beginnings of the target words, which differ, however, in canonical stress position (e.g., *Anak* vs. *aNAK-* vs. *anak-*). Stimuli were selected from the first round of speech, unless a particular utterance sounded unnatural, for instance because of hesitation, in which case it was replaced by the corresponding utterance from the second round of speech. In this way thirty stimuli, that is, pairs of gates of increasing length, were created.

For each particular target word, the first gate corresponded to the carrier sentence up to and including the first syllable of the target word⁴ (that is, the non-shaded part of Figure 1), with the final syllable(s) deleted. The truncation in each case was made just before the first segment of the second syllable, which was the same for the three target words in each set. The second (and following) syllables were completely inaudible (see Figure 1, shaded areas).

1st gate		deleted
<i>Dia mengucapkan kata a</i>		<i>nak-Anak</i>

If lexical stress contributes to word perception, we would expect listeners to be able to discriminate at this stage between two-syllable words on the one

³ For convenience's sake these stimuli will also be referred to in this report as two-, three-, and four-syllable words respectively.

⁴ The second syllable in the case of the longer target words of the *mengangkat, mengangkat* and *mengangkat-angkat* set.

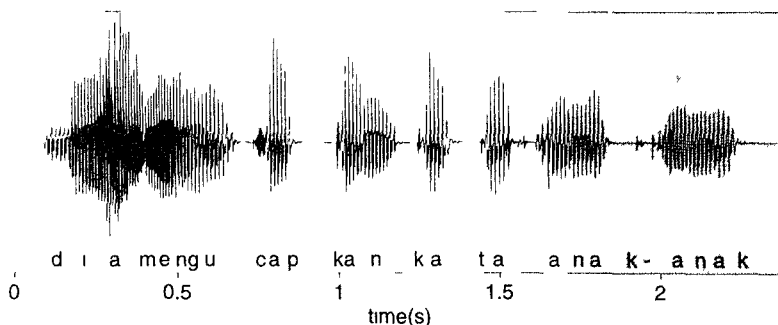


Figure 1. Wave form of sample utterance *Dia mengucapkan kata anak-anak* ('He pronounces the word "children"'). The white area corresponds to gate 1; the portion shaded in light-gray was added at gate 2; the portion shaded in dark-gray was never made audible.

hand (e.g., *A[nak]*; stressed syllable audible), and three-syllable (e.g., *a[NAKnya]*) and four-syllable (e.g., *a[nak-Anak]*) words on the other (stressed syllable deleted).

For the second gate the sentence was truncated immediately after the second vowel⁵ of the target word. We also deleted the offset part of the relatively long final vowels of *GANTI* and *Ada*, which otherwise might help identify this target as a disyllable (see Nooteboom and Doodeman 1980). The second /a/ in *aDA[lah]* was pronounced relatively long, but due to an experimental error its offset was not deleted.

2nd gate	deleted
<i>Dia mengucapkan kata ana</i>	<i>k-Anak</i>

The three target words in each set – the disyllabic targets (e.g., *Anak*) with stress (supposedly) on the first syllable vs. the three-syllable targets (e.g., *aNAK[nya]*) with stress on the second syllable, and finally, the four-syllable target words (e.g., *anak[-Anak]*) with the stressed prefinal syllable deleted – should now be identifiable if lexical stress information can resolve the ambiguity.

Stimuli (that is, pairs of gates of lesser and greater length) were copied twice on DAT tape in the same random order, with 5-second interstimulus intervals (offset to onset), both within and between pairs of gates. After every ten pairs of gates a short beep was sounded to help the listeners keep track of the stimuli on their answer sheets.

⁵ The third vowel in the case of *mengangkat*, *mengangkatkan* and *mengangkat-angkat*.

2.3 Listeners

Six Indonesian listeners took part in the experiment: two listeners of Balinese descent, including the speaker of the stimulus sentences, three Sundanese listeners, and one listener of Javanese origin. Most of them were recent arrivals from Indonesia, and all used Indonesian very frequently. In addition, six Dutch listeners participated in the experiment. These included a PhD student specializing in Indonesian intonation, an intonologist with a good knowledge of Indonesian, and a phonetician (the second author of the present article) whose knowledge of Indonesian is restricted to pronunciation rules, including the traditional stress rule (stress on the penultimate syllable). Apart from these three phoneticians, three phonetically naive Dutch subjects who frequently use Indonesian took part in the experiment.

2.4 Procedure

The tape was played to listeners individually over good-quality earphones. The instructions included a list of the three possible responses to each stimulus (namely the three target words in each set). Listeners were instructed to identify each stimulus word after each pass, immediately after hearing it, as (the beginning of) one of the three alternatives listed on the answer sheets. To this end, they were asked to indicate the most appropriate response and not to skip this for any stimulus (forced choice). Listeners were also asked to cross out the least likely candidate for each stimulus. It was pointed out to them, however, that indicating the most likely alternative was of crucial importance and had priority over crossing out the least likely alternative. The experiment was preceded by a trial run of three practice items (two gates each). After this the tape was stopped to answer any questions raised by the listeners. All instructions were given in Indonesian.

3. Results and discussion

Table 1a summarizes the results (viz. 'best fitting' alternatives⁶) for the Indonesian listeners. The columns in the left half present the results for the first presentation (that is, first gate: one syllable of target audible), and the columns in the right half those for the second presentation (second gate: two syllables of target audible). Perfect identification would result in 100% ratings on the diagonals (top left to bottom right – the figures in bold print). Listeners had to choose out of three alternatives each time. Consequently, if listeners make their choices at random, correct ratings will be 33%. For the Indonesian listeners the mean percentage of correct

⁶ 'Least likely alternative' ratings approximately mirror the 'most likely alternative' ratings for the two groups of listeners and do not provide any additional information.

stimulus identification was 34 on first presentation; these ratings improved slightly to 36% on second presentation (not significant). The 63% correct identification of disyllables on first presentation is partly due to a strong bias towards disyllables in the choices. Statistically there is no significant difference between this percentage and the (incorrect) ratings obtained for the three-syllable and the four-syllable stimuli in this column (58%). It appears, then, that Indonesian listeners are not at all helped by stress information in the identification of words.

Table 1a. Stimuli as labelled by the 6 Indonesian listeners; 10 stimuli per type, 2 repetitions per stimulus per listener.

1st gate: one syllable ((C)V-) of target audible

2nd gate: two syllables ((C)VCV-) of target audible

stim	1st gate				2nd gate			
	listeners' choice (%)				listeners' choice (%)			
	2 syll.	3 syll.	4 syll.	none	2 syll.	3 syll.	4 syll.	none
2 syll.	63	27	8	2	37	38	25	0
3 syll.	58	32	9	1	27	43	30	0
4 syll.	58	33	7	2	28	44	28	0
mean	59	31	8	2	31	41	28	0

The Dutch listeners (Table 1b) scored 37% correct on first presentation (not significant), but 61% correct on second presentation; this is highly significant: $\chi^2 = 145.2$; $p < .0001$. On second presentation, Dutch listeners seemed often able to identify the stimuli correctly.

Table 1b. Stimuli as labelled by the 6 Dutch listeners; 10 stimuli per type, 2 repetitions per stimulus per listener.

1st gate: one syllable ((C)V-) of target audible

2nd gate: two syllables ((C)VCV-) of target audible

stim	1st gate				2nd gate			
	listeners' choice (%)				listeners' choice (%)			
	2 syll.	3 syll.	4 syll.	none	2 syll.	3 syll.	4 syll.	none
2 syll.	55	27	17	1	80	14	5	1
3 syll.	45	33	22	0	17	51	32	0
4 syll.	41	35	23	1	22	26	52	0
mean	47	32	20	1	40	30	29	1

If stress contributes to word perception, listeners should be able to identify, even on first presentation, disyllabic stimuli on the one hand (stressed syllable audible), and three or four-syllable stimuli on the other (stressed

syllable deleted). It would seem impossible, in the set-up of our experiment, however, to distinguish between three- and four-syllable stimuli on first presentation on the basis of stress information, as the supposedly stressed syllable was deleted in both cases. It should not surprise us, then, that identification ratings on first presentation do not deviate significantly from chance. On second presentation, stress position, if functionally relevant, should enable listeners to distinguish between all three stimuli in each set (e.g., *Anak* vs. *aNAK[nya]* vs. *anak[-Anak]*). Our results indicate that stress is functionally irrelevant for the Indonesian listeners, while the stimuli do contain some stress (or other suprasegmental) information which was used by the Dutch listeners to identify them correctly. Before analysing this prosodic information in general, let us first consider the results for the individual listeners.

As regards the Indonesian listeners (see Appendix 2a), statistical significance was attained in only one instance, namely that of listener 5 on second presentation, where $\chi^2 = 21.6$; $p < 0.001$. This listener was also the speaker of the stimulus material. We conclude that he is able to differentiate between the stimuli spoken by himself when presented with the first two syllables (55% correct), but not on hearing only the first syllable. None of the other five Indonesian listeners identified the stimuli correctly, either on first or on second presentation.

Two of the Dutch listeners (see Appendix 2b) attained a significant level of correct stimulus identification on first presentation, namely listeners 2 (45% correct: $\chi^2 = 10.8$; $p < 0.03$) and 10 (55% correct: $\chi^2 = 18.8$; $p < 0.001$). Both are trained phoneticians. The best result was obtained by listener 10, the second author of this article, who does not speak Indonesian but is very familiar with (stress) perception experiments. Apparently, the initial syllable contained sufficient prosodic information to enable trained phoneticians to identify the stimuli correctly more often than chance would predict. It should be noted, however, that in both cases the level of correct identification was highest for the (shortest) disyllabic words and lowest for the (longest) four-syllable stimuli. After the second pass, five out of the six Dutch listeners identified the stimuli correctly ($p < 0.03$ for one listener and $p < 0.001$ for four listeners). On second presentation the stimuli contained sufficient prosodic information for naive Dutch listeners to identify them correctly.

Five out of our six Dutch listeners, as well as one Indonesian listener, were able to retrieve prosodic information from the stimuli that enabled correct identification at a frequency greater than chance after the second pass. In our experimental set-up we did not, however, distinguish between different kinds of prosodic information. Consequently, we cannot be certain whether these subjects obtained the results in question on account of their knowledge of the stress patterns of Indonesian or by using other prosodic information, such as syllable shortening in longer words or slowing down at

the end of a sentence (see Nooteboom and Doodeman 1980; Van Zanten 1994; and Van Zanten and Van Heuven 1997). The high percentage of correct identification on second presentation is due to a fair extent to correct identification of the disyllabic stimuli (80%); it is possible, therefore, that especially in these cases the end-of-sentence marking as well as stress information was detected by listeners.

In the present study our main interest is in the Indonesian subjects. The Indonesian listeners as a group were not able to identify target words at a rate greater than chance. In only one instance (namely that of listener 5; 2nd gate) did we find a significant association between stimuli and responses. As the other listener of Balinese descent (no. 9) was not able to identify the stimuli, we are inclined not to attribute this result to the substrate language; it seems more likely that listener 5 recognized his own prosodic patterns and that this enabled him to identify the stimuli. The prosodic information – including any stress cues – which apparently is available and is used by the majority of Dutch subjects, generally speaking is not used by Indonesians.

Stimulus analysis

To get an idea of the effects of the temporal and melodic information⁷ contained in the stimuli, we measured the pitch movements and durational factors that might have helped the listeners identify the stimuli (see Figure 2). Auditory pitch in spoken utterances depends on the rate of vocal cord vibration, which corresponds to the fundamental frequency (F_0 , expressed in hertz (Hz), that is, number of repetitions per second) in the acoustic signal.⁸ On the basis of the raw F_0 measurements, we defined several prosodic variables which potentially allowed listeners to identify the stimulus types at the first gate, namely RISE1, FALL1 and DURATION1.

The measurements for the entire stimulus set are listed in Appendix 1. Linear Discriminant Analysis, or LDA (Klecka 1980), was used as a heuristic tool to determine the relative contribution of each of the above prosodic variables to the automatic identification of the three stimulus types. Given three stimulus categories to be distinguished, LDA yields two discriminant functions, each of which is a different linear combination of the various (weighted and standardized) prosodic variables. The weight

⁷ These are the two strongest perceptual cues for stress and accent. Perceptually weaker cues, such as intensity and vowel quality, were not included in the stimulus analysis. The perceptual strength of spectral balance, as an acoustic operationalization of loudness, is comparable with that of duration, though only in low-quality, reverberant speech; in high-fidelity speech such as ours, spectral balance should be classed among the weaker cues for stress and accent (Sluijter, Van Heuven and Pacilly 1997).

⁸ A range of computer-implemented algorithms is available for the determination of F_0 in speech utterances. We used the method of F_0 extraction by subharmonic summation developed by Hermes (1988).

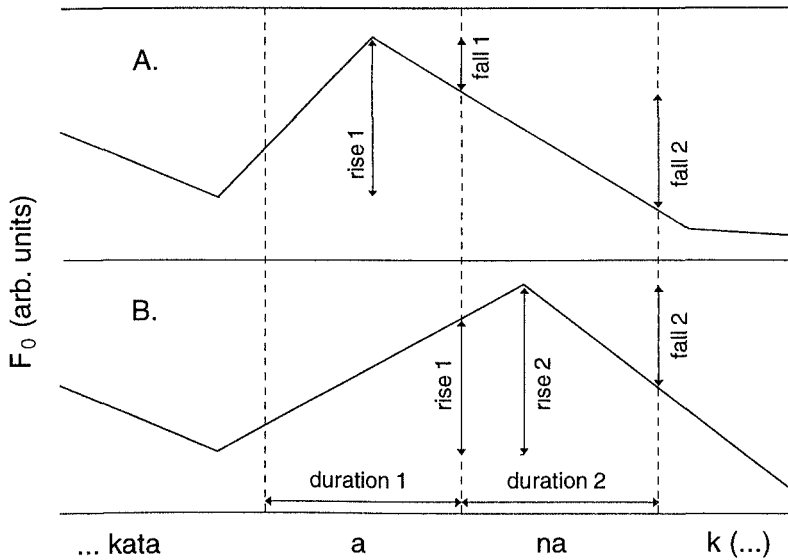


Figure 2. Schematic representation of possible F_0 contours and relevant F_0 measurement points. Panel A shows the situation where the accent-lending rise-fall contour reaches its F_0 peak in gate 1; panel B shows a similar rise-fall with the F_0 maximum in gate 2.

- RISE1 is the pitch interval (rescaled from Hz into ERB)⁹ between the highest F_0 anywhere in the first gate and the F_0 minimum anywhere prior to the onset of this gate. RISE1 was negative on one occasion, when there was no F_0 rise in the first gate.
- FALL1 is the pitch interval (in ERB) between the F_0 maximum in gate 1 and the F_0 maximum anywhere in gate 2; FALL1 is axiomatically given the value '0' if the pitch does not drop in gate 1.
- DURATION1 is the time span (in milliseconds (ms)) between the end of the carrier sentence and the end of gate 1.

At the end of the second gate, the potentially relevant prosodic variables are all those defined for gate 1 plus:

- RISE2, defined as the pitch interval (in ERB) between the highest F_0 anywhere in the second gate and the F_0 minimum anywhere prior to the onset of this gate.
- FALL2, defined as the interval (in ERB) between the highest and lowest F_0 anywhere within gate 2; FALL2 is axiomatically given the value '0' if the pitch does not drop in gate 2.
- DURATION2, that is, the duration (in ms) added to the stimulus by gate 2.

⁹ The Equivalent Rectangular Bandwidth (ERB) scale expresses fundamental frequency (F_0) in intonation languages in a perceptually optimally relevant fashion (Hermes and Van Gestel 1991; Ladd and Terken 1995).

of each variable in the identification process can be estimated by inspecting the contribution made by each prosodic factor to the two discriminant functions. Of course, this attempt makes sense only insofar as the LDA correctly postdicts the stimulus types.

At the end of gate 1, LDA yielded a 52% correct stimulus postdiction (see Table 2, left half). This means that RISE1, FALL1 and DURATION1 together distinguish the three stimulus types well above chance level at the end of gate 1. The scores were highest for the two-syllable stimuli (72%), but below chance level (30%) for the four-syllable stimuli. At the end of gate 1, 79.6% of the variance is explained by Discriminant Function 1, which largely coincides with FALL1 (correlation $r = 0.88$). Function 1 primarily distinguishes the two-syllable stimuli from three-syllable ones. Function 2 (20.4% of variance) is partly determined by RISE1 ($r = 0.72$) and by DURATION1 ($r = 0.51$). Function 2 distinguishes the four-syllable stimuli from the two- and – to a lesser extent – the three-syllable stimuli. To summarize, FALL1 turned out to be the most important postdictor, especially in discriminating the two-syllable from the three-syllable stimuli. RISE1 had some discriminating effect and DURATION1 had the smallest effect.¹⁰

Table 2. Predicted group membership of the stimuli on the basis of six prosodic variables. Ten stimuli per type.

1st gate: one syllable ((C)V-) of target audible; variables RISE1, FALL1, DURATION1.

2nd gate: two syllables ((C)VCV-) of target audible; variables RISE1, RISE2, FALL1, FALL2, DURATION1 and DURATION2.

stim	end of 1st gate			end of 2nd gate		
	predicted word length class (%)			predicted word length class (%)		
	2 syll.	3 syll.	4 syll.	2 syll.	3 syll.	4 syll.
2 syll.	72	14	14	100	0	0
3 syll.	10	60	30	0	90	10
4 syll.	30	40	30	0	0	100

At the end of gate 2, correct postdiction on the basis of all six variables averaged 96%: it was 100% for the two- and four-syllable stimuli, and 90% for the three-syllable stimuli. This is in fact much better than the 61% correct identification by the Dutch listeners as a group. Individual Dutch listeners, in particular the trained phoneticians, scored as high as 77% (listener 2) or even 82% correct (listener 10).

¹⁰ Although Dutch listeners as a group were unable to identify the stimuli above chance level, 55% correct identification was achieved for gate 1 by Dutch listener 10. We assume that this trained phonetician used the same three prosodic variables. As in the LDA, the (short) two-syllable stimuli were identified best by listener 10, which, again, points to the importance of FALL1 as a differentiating factor.

Here the variance is explained mainly by Discriminant Function 1 (87.8%). This function correlates moderately with FALL2 ($r = 0.44$), FALL1 ($r = 0.41$) and RISE2 ($r = 0.37$). Function 1 was most successful in discriminating between the two- and three-syllable stimuli. Function 2 (12.2% of variance) correlates with FALL2 ($r = 0.65$), but also – in descending order of correlation – with RISE2 ($r = 0.64$), DURATION2 ($r = 0.45$), RISE1 ($r = 0.44$), and DURATION1 ($r = 0.27$). This function distinguishes the four-syllable stimuli from the two- and three-syllable stimuli. At the end of gate 2, therefore, the most influential variables are FALL1 and FALL2.

Note that neither DURATION1 nor DURATION2 is important in the LDA. This indicates that durational patterning is not important in discriminating between the three groups of stimuli. This can be seen as a reflection of the relatively small effect of stress on syllable duration (Van Zanten and Van Heuven 1997).

At both gates, discrimination is best between two-syllable stimuli on the one hand and three-syllable stimuli on the other. The most important discriminating factors are FALL1 and FALL2. We suggest that a drop in pitch signalling the end of the sentence may be of overriding importance, rather than any accent-lending pitch movement or lengthening effect, which, as mentioned in the Introduction, are fairly modest in Indonesian.

Our provisional conclusion is, therefore, that sentence-final intonation was often picked up by the Dutch listeners and used to deduce the remaining length of the (sentence-final) target word. Stress position as such, then, plays no role in word identification in Indonesian.

4. Conclusion

In Indonesian, word stress has traditionally been described as being fixed on the penultimate syllable. If this rule holds good, stress should help listeners identify words. The results of the present experiment indicate that this is not the case, however. Segmentally identical but prosodically different stimuli failed to be identified by five out of our six Indonesian listeners. Nevertheless, as Van Heuven and Van Zanten (1997) have shown, Indonesian listeners are indeed able to locate accents within a sentence. We must conclude, then, that Indonesians can distinguish prosodic cues but do not use them to identify (parts of) words. The fact that Indonesian is spoken on the basis of a large variety of substrate languages – which may themselves be free-stress languages – may cause Indonesians not to pay attention to stress location.

It is usually assumed that non-native listeners are at a disadvantage in cross-linguistic language processing (Cutler and Otake 1996). Quite surprisingly, we found in our study that Dutch listeners performed better than their Indonesian counterparts in guessing the remaining length of

truncated words. We assume that Dutch (distinctive-stress) listeners are keener at retrieving stress and other prosodic information, such as a sentence-final drop in pitch (which is not specific to Indonesian), so that these listeners were at an advantage in our experiment.

Van Zanten and Van Heuven (submitted) have found a preference in Indonesian for stress on phonologically heavy prefinal syllables. In the present study, however, sets of target words which included stimuli with heavy prefinal syllables were not identified any better than other sets. In fact, no significant level of correct stimulus identification by the Indonesian listeners was found for any single set, including the *Ada*, *aDaLah*, *adaKaLa* set, with a long /a/ in the second syllable of *aDaLah*.¹¹ We conclude, therefore, that although Indonesian listeners may prefer stress on the (heavy) prefinal syllable, such a preference has no relevance in speech communication.

Stress is phonetically weaker in Indonesian than in, for instance, Dutch. Its effect on syllable duration is slight, and accent-lending pitch movements are less pronounced than in distinctive-stress languages. Our Balinese speaker, who has worked as a news reader and as a teacher, is possibly relatively consistent in his prosodic patterning. This consistency may have helped him as well as the Dutch listeners in identifying the stimuli; it was not, however, detected or made use of by the other Indonesian listeners. Different Indonesian speakers might have produced different prosodic patterns, which might, in turn, have led to poorer identification by the Dutch listeners. We would not expect better results from Indonesian listeners in a similar listening test with different speakers, however.

Word stress information was not used by our Indonesian listeners to differentiate between words. Our results indicate that stress is communicatively irrelevant and essentially free in Indonesian. They support the view advanced by Halim (1974:111-3) and Zubkova (1966, cited in Odé 1994) that there is no word stress in Indonesian, free stress being tantamount to no stress.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the cooperation of the informants, especially I.B.M. Palguna. Technical assistance was provided by Jos Pacilly. Comments by two anonymous referees contributed substantially to the readability of this article.

¹¹ One referee of this article pointed out that *-lah* (and *-kah*) are generally analysed as clitics rather than suffixes, and are therefore stress-neutral. Our speaker was not instructed to stress any particular syllable in any word; our listeners did not react any differently to the word containing *-lah* than to the other stimuli.

REFERENCES

- Alieva, N.F., V.D. Arakin, A.K. Ogloblin, and Yu.H. Sirk, 1991, *Bahasa Indonesia; Deskripsi dan teori*, Yogyakarta: Kanisius.
- Bolinger, D.L., 1958, 'A theory of pitch accent in English', *Word* 14:109-49.
- Cohn, A.C., 1989, 'Stress in Indonesian and bracketing paradoxes', *Natural Language and Linguistic Theory* 7:167-216.
- Connine, C.M., C. Clifton, and A. Cutler, 1987, 'Effects of lexical stress on phonetic categorization', *Phonetica* 44:133-46.
- Cutler, A., and T. Otake (eds), 1996, *Phonological structure and language processing; Cross-linguistic studies*, Berlin / New York: Mouton de Gruyter.
- Dardjowidjojo, S., 1978, *Sentence patterns of Indonesian*, Honolulu: University of Hawaii Press.
- Ebing, E.F., 1997, *Form and function of pitch movements in Indonesia*, Leiden: Research School CNWS, School of Asian, African and Amerindian Studies. [CNWS Publications 55. PhD thesis.]
- Grosjean, F., 1980, 'Spoken word recognition processes and the gating paradigm', *Perception and Psychophysics* 28:267-83.
- Halim, A., 1974, *Intonation in relation to syntax in Bahasa Indonesia*, Jakarta: Djambatan.
- Hermes, D.J., 1988, 'Measurement of pitch by subharmonic summation', *Journal of the Acoustical Society of America* 83:257-64.
- Hermes, D., and 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, 1988, 'Effects of stress and accent on the human recognition of word fragments in spoken context; Gating and shadowing', in: W.A. Ainsworth (ed.), *Proceedings of the 7th Fase/Speech-88 Symposium*, 811-8, Edinburgh: Acoustical Institute.
- , 1994, 'Introducing prosodic phonetics', in: C. Odé and V.J. van Heuven (eds), *Experimental studies of Indonesian prosody*, pp. 1-26, Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University. [Semaian 9.]
- Heuven, V.J. van, and A.M.C. Sluijter, 1996, 'Notes on the phonetics of word prosody', in: R. Goedemans, H. van der Hulst and E. Visch (eds), *Stress patterns of the world; Part 1: Background*, pp. 233-69, Leiden: Holland Institute of Generative Linguistics / The Hague: Holland Academic Graphics. [HIL Publications 2.]
- Heuven, V.J. van, and E. van Zanten, 1997, 'Effects of substrate language on the localization and perceptual evaluation of pitch movements in Indonesian', in: C. Odé and W.A.L. Stokhof (eds), *Proceedings of the 7th International Conference on Austronesian Linguistics*, pp. 63-80, Amsterdam/Atlanta: Rodopi.
- Klecka, W.R., 1980, *Discriminant analysis*, Beverly Hills, London: Sage.
- Ladd, D.R., and J.M.B. Terken, 1995, 'Modelling intra- and inter-speaker pitch range variation', in: K. Elenius and P. Branderud (eds), *Proceedings 13th International Congress of Phonetic Sciences 2*, pp. 386-9, Stockholm: Royal Institute of Technology / Stockholm University.
- Laksman, M., 1994, 'Location of stress in Indonesian words and sentences', in: C. Odé and V.J. van Heuven (eds), *Experimental studies of Indonesian prosody*, pp. 108-39, Leiden: Department of Languages and Cultures of South-East Asia and Oceania,

- Leiden University. [Semaian 9.]
- Nababan, P.W.J., 1981, *A grammar of Toba Batak*, Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Nooteboom, S.G., and G.J.N. Doodeman, 1980, 'Production and perception of vowel length in spoken sentences', *Journal of the Acoustical Society of America* 67:276-87.
- Odé, C., 1994, 'On the perception of prominence in Indonesian', in: C. Odé and V.J. van Heuven (eds), *Experimental studies of Indonesian prosody*, pp. 27-107, Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University. [Semaian 9.]
- Samsuri, 1978, 'Fokus dan alat-alat pembentukannya dalam Bahasa Indonesia'. [Unpublished conference paper.]
- Sluijter, A.M.C., V.J. van Heuven, and J.J.A. Pacilly, 1997, 'Spectral balance as a cue in the perception of linguistic stress', *Journal of the Acoustical Society of America* 101:312-22.
- Teeuw, A., 1978, *Leerboek Bahasa Indonesia*, Groningen: Wolters-Noordhoff.
- Zanten, E. van, 1994, 'The effect of sentence position and accent on the duration of Indonesian words; A pilot study', in: C. Odé and V.J. van Heuven (eds), *Experimental studies of Indonesian prosody*, pp. 140-80, Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University. [Semaian 9.]
- Zanten, E. van, and V.J. van Heuven, 1997, 'Effects of word length and substrate language on the temporal organisation of words in Indonesian', in: C. Odé and W.A.L. Stokhof (eds), *Proceedings of the 7th International Conference on Austronesian Linguistics*, pp. 201-16, Amsterdam/Atlanta: Rodopi.
- , submitted, 'Word stress in Indonesian: Fixed or free?', *NUSA*.

Stimulus word	Gloss	Rise1	Fall1	Dur1	Rise2	Fall2	Dur2
mengANG ka t	lift up	.41	.41	539	.00	.33	194
mengang KA Tkan	lift up for someone	.29	.00	517	.43	.11	218
mengang ka t-ANGkat	praise	.13	.00	455	.27	.24	198
PA sa ng	pair	.74	.39	112	.35	.38	216
pa SA ngan	couple	.44	.00	76	.44	.24	223
pa sa ngGRAhan	resthouse	.50	.07	74	.43	.24	184
TANG gu ng	guarantee; mediocre	.31	.24	266	.07	.17	94
tang GU ngan	obligation	.14	.00	224	.24	.00	111
tang gu ngJAwab	responsibility	.39	.36	194	.04	.07	90

Appendix 2a. Stimuli as labelled by the Indonesian listeners; 10 stimuli per type, 2 repetitions per stimulus per listener.

stimulus	responses at first gate				responses at second gate			
	2-syll.	3-syll.	4-syll.	none	2-syll.	3-syll.	4-syll.	none
listener 4			(n.s.)			(n.s.)		
2-syll.	20	0	0	0	20	0	0	0
3-syll.	19	0	0	1	20	0	0	0
4-syll.	20	0	0	0	20	0	0	0
listener 5			(n.s.)			(p < 0.001)		
2-syll.	6	7	7	0	12	3	5	0
3-syll.	5	9	6	0	1	11	8	0
4-syll.	4	11	5	0	2	7	11	0
listener 6			(n.s.)			(n.s.)		
2-syll.	3	13	2	2	0	5	14	1
3-syll.	5	12	3	0	1	3	16	0
4-syll.	8	9	2	1	0	7	13	0
listener 7			(n.s.)			(n.s.)		
2-syll.	11	8	1	0	0	17	3	0
3-syll.	10	10	0	0	0	15	5	0
4-syll.	10	10	0	0	0	17	3	0
listener 8			(n.s.)			(n.s.)		
2-syll.	15	5	0	0	12	6	2	0
3-syll.	16	3	1	0	9	8	3	0
4-syll.	13	4	1	2	11	8	1	0
listener 9			(n.s.)			(n.s.)		
2-syll.	20	0	0	0	0	14	6	0
3-syll.	15	4	1	0	1	14	5	0
4-syll.	14	5	1	0	1	13	6	0
all listeners	214	110	30	6	110	148	101	1

Appendix 2b. Stimuli as labelled by the Dutch listeners; 10 stimuli per type, 2 repetitions per stimulus per listener.

stimulus	responses at first gate				responses at second gate			
	2-syll.	3-syll.	4-syll.	none	2-syll.	3-syll.	4-syll.	none
listener 1		(n.s.)				(p < 0.001)		
2-syll.	9	7	4	0	17	3	0	0
3-syll.	10	6	4	0	1	11	8	0
4-syll.	10	6	4	0	4	9	7	0
listener 2		(p = 0.028)				(p < 0.001)		
2-syll.	13	4	3	0	18	1	1	0
3-syll.	5	8	7	0	0	9	11	0
4-syll.	4	10	6	0	0	1	19	0
listener 3		(n.s.)				(n.s.)		
2-syll.	16	4	0	0	17	3	0	0
3-syll.	16	4	0	0	15	5	0	0
4-syll.	17	3	0	0	16	4	0	0
listener 10		(p = 0.001)				(p < 0.001)		
2-syll.	14	2	4	0	19	0	1	0
3-syll.	4	12	4	0	1	15	4	0
4-syll.	3	10	7	0	1	4	15	0
listener 11		(n.s.)				(p < 0.001)		
2-syll.	7	6	6	1	15	1	3	1
3-syll.	9	5	5	1	1	14	5	0
4-syll.	5	5	9	1	1	5	14	0
listener 12		(n.s.)				(p = 0.021)		
2-syll.	7	10	3	0	10	9	1	0
3-syll.	10	4	6	0	3	7	10	0
4-syll.	10	8	2	0	5	8	7	0
all listeners	169	114	74	3	144	109	106	1