

Phonetic Correlates of Primary and Secondary Stress in Indonesian: A Preliminary Study*

Niken Adisasmito-Smith and Abigail C. Cohn

We report here on a preliminary instrumental analysis of word stress in Indonesian, in order to broaden our cross-linguistic understanding of the phonetic properties of stress and to assess observations in the phonological literature about Indonesian stress based on impressionistic listening. In this study, reiterant speech of two speakers was used to investigate the contributions of F0, duration, and intensity as acoustic correlates of both primary and secondary word stress in Indonesian. Once the technique of reiterant speech was shown to be reliable for these particular speakers, acoustic correlates of both primary and secondary were studied, by analyzing the reiterant speech patterns of two- to five-syllable words. It was found in all cases that the penult showed the peak F0 and greatest intensity. Final syllables exhibited consistently low F0 and intensity, but were similar in duration to penults, showing final lengthening. This study highlights the importance of expanding the range of languages investigated to reach a full understanding of the phonetic realization of both primary and secondary stress, while providing further evidence of the reliability of the technique of reiterant speech.

1 Introduction

Our understanding of the phonetic properties of stress is informed largely by studies of English and a few other Western European languages and yet the degree to which the principal acoustic correlates of stress—F0, duration, and intensity—contribute to the percept of stress is known to be language-specific (Berenstein 1979). Thus instrumental investigations of less commonly studied languages are needed to broaden our cross-linguistic understanding of such phenomena. We also know very little about the phonetic properties of primary vs. secondary stress: do languages use the same acoustic correlates but to differing degrees, or do they choose different correlates or combinations of correlates? Thus phonetic correlates of stress in languages with regular patterns of both primary and secondary stress need to be investigated. The present study is an instrumental analysis, using reiterant speech, of word stress in Indonesian, described to have systematic patterns of both primary and secondary stress. This work is part of a larger project now underway in order to assess observations in the phonological literature about Indonesian stress based on impressionistic listening.

The basic pattern of word stress in Indonesian is presented in (1) (as described by MacDonald 1976, Halim 1981, Lapoliwa 1981 for primary stress, and Cohn 1989 for secondary stress).

* This is a written version of a talk presented at the 70th meeting of the LSA, San Diego, January 1996. Thanks to the LSA audience and members of the Cornell Phonetics Lab for insightful and helpful comments and especially to Joan Sereno and Ayako Tsuchida for comments on this draft and Eric Evans for advice on the statistics. Any errors are of course ours alone. This research is supported by NSF grant #SBR-9511185 to Abigail C. Cohn.

(1)	ó	cát	'paint'
	ó o	cári	'search for'
	o ó o	bicára	'speak'
	ò o ó o	bìjaksána	'wise'
	ò o o ó o	àntisipási	'anticipation'
	ò o ò o ó o	òtobìogràfi	'autobiography'
	ò o o ò o ó o	àmerikànisási	'Americanization'

We observe that a primary stress falls on the penult (except in monosyllabic words); a secondary stress falls on the initial syllable, as long as it is not next to the main stress and additional secondaries surface in six- and seven-syllable words. There are additional complications with words with schwas and morphologically complex forms, which we will not address here (see Cohn 1989 for discussion of these points and fuller discussion of the basic patterns).

There has been very limited instrumental work on the prosodic structure of Indonesian. The first systematic instrumental study was carried out by Halim (1974, later published as Halim 1981). The primary focus of the study was intonation, but word accent (primary stress) was also investigated. Pitch was found to be the most systematic correlate, with duration also playing a role. An important addition to the study of prosody in Indonesian is a volume of several studies—Odé & van Heuven (1994) *Experimental Studies of Indonesian Prosody*. While the volume focuses primarily on intonation, Laksman (1994) is of direct relevance to the question of word stress. Investigating primary stress in disyllabic words, she shows F0 to be a consistent correlate of word stress and she also observes duration and amplitude differences, although these were due to a number of factors. Her study is limited to one speaker and does not consider secondary stress. In an additional study, van Zanten & van Heuven (1994) investigate the effect of word length on duration of words in Indonesian, studying several speakers of Indonesian with distinct native language backgrounds. They study forms of one- to seven-syllables, mixing monomorphemic and morphologically complex forms. They find that stressed syllables in monosyllables are longer than in longer words, but are quite comparable in two- to seven-syllable words. They also find that syllable duration of stressed and unstressed syllables are quite similar and fairly constant across forms of varying numbers of syllables, however no data are presented by syllable for those syllables not receiving main stress. We return in the discussion to the findings of these studies.

Clearly further systematic study of the acoustic correlates of word stress in Indonesian is warranted, particularly to investigate the relative contributions of primary vs. secondary stress. We report here on two experiments. In Experiment 1, we assessed the reliability of reiterant speech as a means of studying word stress in Indonesian and in Experiment 2, we investigated the phonetic correlates of primary and secondary stress in monomorphemic words in Indonesian. The structure of the paper is as follows: in §2 the technique of reiterant speech is discussed and the general methodology is presented; in §§3 & 4 the methodologies and results for Experiments 1 & 2, respectively, are presented; and conclusions, discussion, and directions for further research are presented in §5.

2 Reiterant Speech & General Methodology

2.1 Reiterant Speech

As has been observed by many researchers, since F0, duration, and intensity are all affected by a wide range of aspects of the phonological and linguistic system more generally, it is difficult to tease apart the contribution of stress alone, as so many factors might be contributing to the observed patterns. To address this problem in the study of English and Swedish, some researchers have used the technique of reiterant speech, whereby a subject mimics a real speech utterance with a sequence of segmentally identical syllables, e.g. *fundamental* → mamamama (Lieberman & Streeter 1978, Nakatani et al. 1981). This technique has been used successfully to avoid the problems of teasing apart multiple influences on F0, duration, and intensity. Its reliability has been tested extensively by Larkey (1983). As pointed out by Nakatani et al. (1981, p. 85), use of reiterant speech allows one to focus on "macrostructure"—overall prosodic structure—rather than "microstructure"—differences in duration due to intrinsic differences in segment duration and contextual and syllable structure effects.

2.2 General Methodology

The basic methodology follows that of Liberman & Streeter (1978) and Larkey (1983). Two speakers of "standard" Indonesian,¹ one male (W) and one female (H), were tested for their ability to produce reiterant speech and trained on the items to be analyzed. Both served as subjects in the impressionistic studies by Cohn and McCarthy (1994). The subjects were provided with instructions in Indonesian. They were then recorded in a sound-treated booth on a Carver, model TD-1700 cassette recorder, reading words from a word list written in standard Indonesian orthography. The forms investigated were real

¹ "Standard" Indonesian is defined here as the speech of educated non-Javanese speakers living in Java.

words of two- to five-syllables, all containing full vowels. We report here on results with monomorphemic forms. (See the Appendix for a list of the forms analyzed in the present experiments.) The speakers uttered the "normal" speech (Nor) first, followed by reiterant speech (Re), using the syllable /ma/, repeating both twice.

The data were digitized at 11,025 Hz and acoustic analysis was done, using WAVES+/ESPS on a SUN-SPARC station. Using this acoustic package, both waveforms and spectrograms were generated, as well as pitch tracks and intensity measurements. Duration, fundamental frequency, and amplitude were studied for both the normal and reiterant speech, for each vowel or syllable. The measurements made are summarized in (2).

(2) Duration (in ms.)	
Vowel	
Syllable	(Reiterant only)
Word	
F0 (in Hz)	Highest F0 value during the vowel itself
Amplitude (in dB)	RMS integrated across the vowel (Reiterant only)

The segmentation criteria used were as follows. The onset of a vowel following an obstruent was measured starting from the onset of periodicity, as identified in the waveform; while following a sonorant, the onset was measured from the onset of a distinct F2, as identified in the spectrogram. The vowel offset was measured from the end of vowel F2, as identified in the spectrogram. In all cases, onset and offset of consonants could be clearly distinguished from the surrounding sounds, with the exception of a few tokens of unreleased stops in final position. In these cases the average duration of released stops in the same position was added to the total word duration of these few forms to allow comparison with the other forms.

We turn now to the specific experiments.

3 Experiment 1: Reliability of reiterant speech

As discussed by Larkey (1983), there are several issues concerning the reliability of reiterant speech. In Experiment 1, the reliability of reiterant speech for these particular speakers was assessed, by comparing the normal and corresponding reiterant patterns. The basic question is whether the macrostructure, that is, overall prosodic structure, is maintained and the microstructure—prosodic differences due to segmental context, etc.—is neutralized.

The first question is whether the speakers were able to carry out the task. Both speakers were quite successful at the task in monomorphemic forms, but found it more difficult once five or more syllables were involved. In a pilot recording of one speaker, all items were put in a frame sentence, but this made the task much more difficult and resulted in unreliable patterns of reiterant speech (often not showing the correct number of syllables.) Thus in the present study, words were uttered in isolation.

Next we judged the speech impressionistically. All normal and reiterant pairs were listened to carefully by both authors. Any pair where the reiterant speech did not match, either due to number of syllables or a different intonation or stress pattern, was rejected. For monomorphemic forms up to four syllables the results were very consistent and all but one token were used (see below), but results for five syllable words were mixed. Of the eight tokens (two forms x two repetitions x two speakers), five were used.

The speakers were also recorded producing morphologically complex forms, but the prosodic structure of these reiterant forms did not match the normal speech and thus analysis of morphologically complex forms is not included in the present paper. Either more training would be needed, or a modified technique, where for example, real speech affixes were used, might work; this is an area for further research.

Finally we turned to the question of maintenance of macrostructure and neutralization of microstructure. Macrostructure can best be assessed by comparing F0 patterns, while the microstructure is reflected in duration. Intensity cannot be reliably compared between the normal and reiterant speech, since the normal speech includes a range of vowels with different inherent intensities. This was assessed by looking at two- and three-syllable words, with different combinations of high and low vowels, voiced and voiceless consonants, and open or closed final syllables. (See the Appendix for the word list.)

Based on careful listening and comparison of normal and reiterant pitch tracks, all but one reiterant token (Speaker W, *pidana*, second repetition) were judged to have natural prosodic structure. The reiterant patterns both sounded natural and closely paralleled the corresponding normal speech form. The close similarity of Normal and Reiterant forms is exemplified in the trisyllabic form *batara* in Figure 1. The intonational contours, Mid-High-Low in both cases, are very similar, characteristic of the corpus more generally.

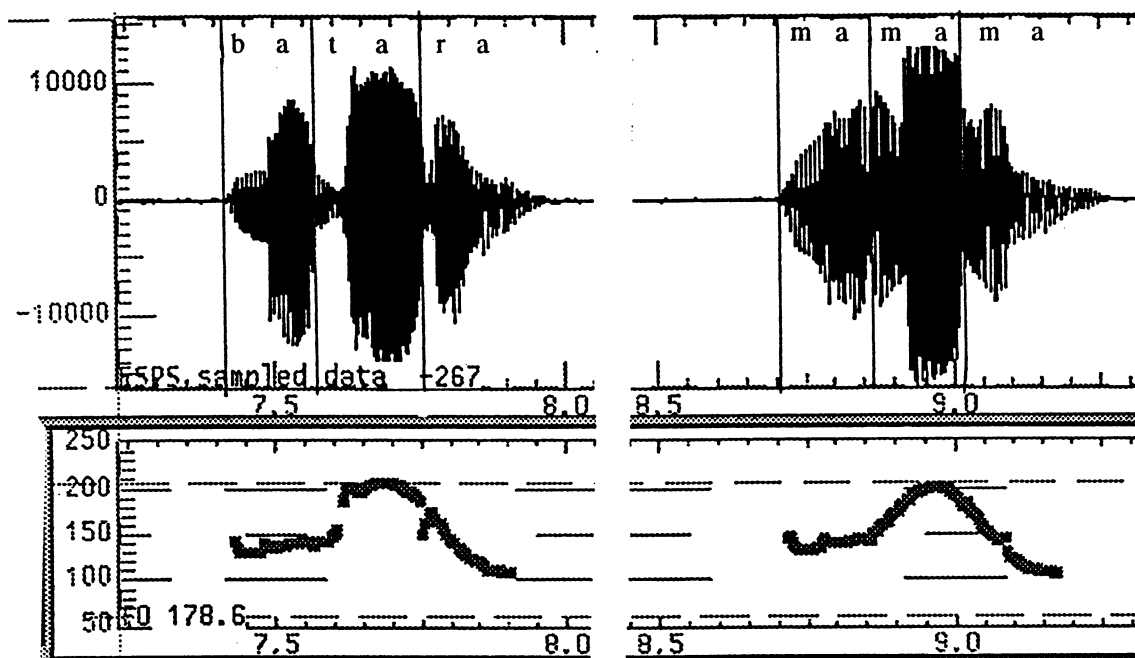


Figure 1. Representative F0 contours, with waveforms, normal [bata] and corresponding reiterant [mamama] for SpeakerW.

To assess whether there was neutralization of microstructure, we compared both word and vowel durations of normal and reiterant speech pairs. Word duration was included as this was one of the standard measures in earlier work using reiterant speech. We found that these segmental duration effects were greatly reduced or completely neutralized. This can be seen, for example, with word durations for two syllable words, as shown in Figure 2 below.

In Figure 2, we observe the range of durations for two-syllable words, in normal speech and reiterant speech for each speaker. To test the difference between ranges, we used the Siegel-Tukey method. The value of the Siegel-Tukey statistic for SpeakerW was 108 and its associated P-value = .0084; while for SpeakerH was 120 with an associated P-value = .0446. Thus for both speakers, the ranges of duration in the reiterant speech are significantly smaller. We also note that the reiterant speech was faster, not surprising since this followed the normal speech, giving the effect of a repeated utterance. In addition, it should be noted that the variation that does exist in the reiterant durations does not reflect systematic differences in the normal speech.

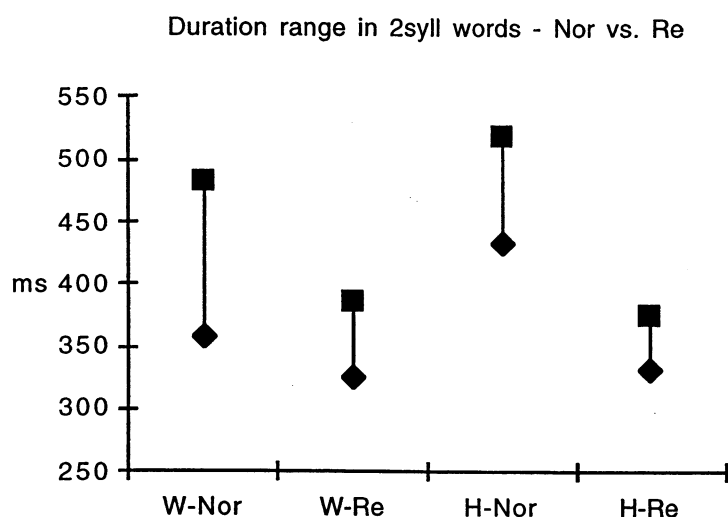


Figure 2. Duration range in two syllable words for "normal" (Nor) and reiterant (Re) forms for SpeakerW and SpeakerH.

In conclusion, reiterant speech was found to be reliable for monomorphemic words for both speakers and therefore was used as basis for Experiment 2, in which the acoustic correlates of primary and secondary stress were investigated.

4 Experiment 2: Acoustic correlates of primary and secondary stress

In Experiment 2, F0, amplitude, and duration were analyzed for reiterant speech patterns of two- to five-syllable words for the same two speakers studied in Experiment 1, although, as noted above, with a limited number of five-syllable tokens. (See the Appendix for the word list.) These forms allowed us to study the phonetic realization of penultimate stress in a range of cases, as well as initial secondary stresses in words of four- and five-syllable words. We discuss F0, amplitude, and duration in turn.

Average F0 values by syllable for two- to five-syllable words for each speaker are presented in Figure 3. In this and subsequent figures, the average value for each syllable is plotted from right to left, thus all forms have a final and a penultimate syllable, with increasing numbers of pretonic syllables in longer forms.

As observed in Figure 3, both speakers show a similar pattern, with the penult having the highest pitch and a clear drop on the final syllable. T-tests were done, by speaker and by number of syllables, to see if the observed differences were statistically significant ($P\text{value} < .05$). Comparisons were made between the penult and both the final and prepenultimate syllables, as well as between initial and other prepenultimate syllables in four- and five-syllable forms.

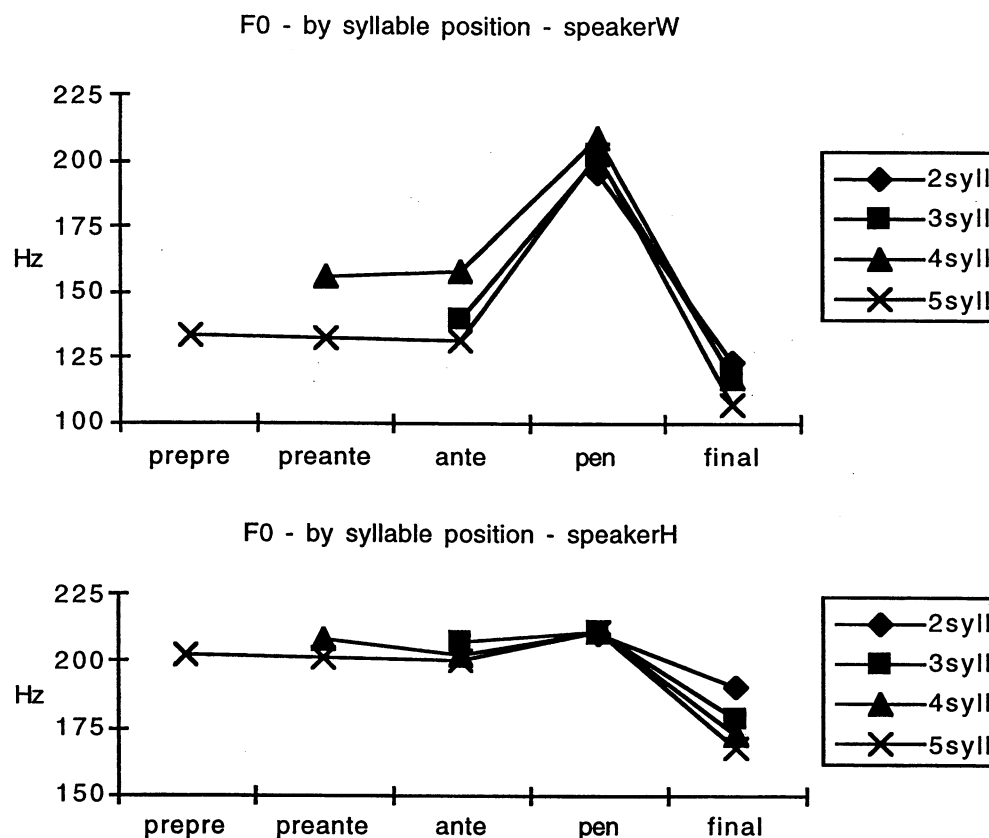


Figure 3. Average F0 values by syllable for two- to five-syllable words by speaker.

In the comparison of the penult to all other syllables, the differences in F0 were highly significant in all cases for both speakers, except in the cases of five-syllable form where the number of tokens was so small, but nevertheless the results approached significance. For both speakers, the F0 contour is quite flat for the pretonic syllables: SpeakerW seems to show a Mid-High pattern, while SpeakerH exhibits a fairly high F0 straight across. Comparing the F0 of the initial syllable to other pretonic syllables in four- and five-syllable forms, the differences were not significant for either speaker. SpeakerW used a wider pitch range, with a more marked peak on the penult and generally lower pitch, while SpeakerH showed a fairly monotone pattern, however the significance of the observed patterns did not differ for the two speakers.

The representative nature of these average patterns can be seen by looking at samples of individual F0 contours. Examples of contours in three-, four-, and five-syllable forms for SpeakerW are presented in Figure 4.

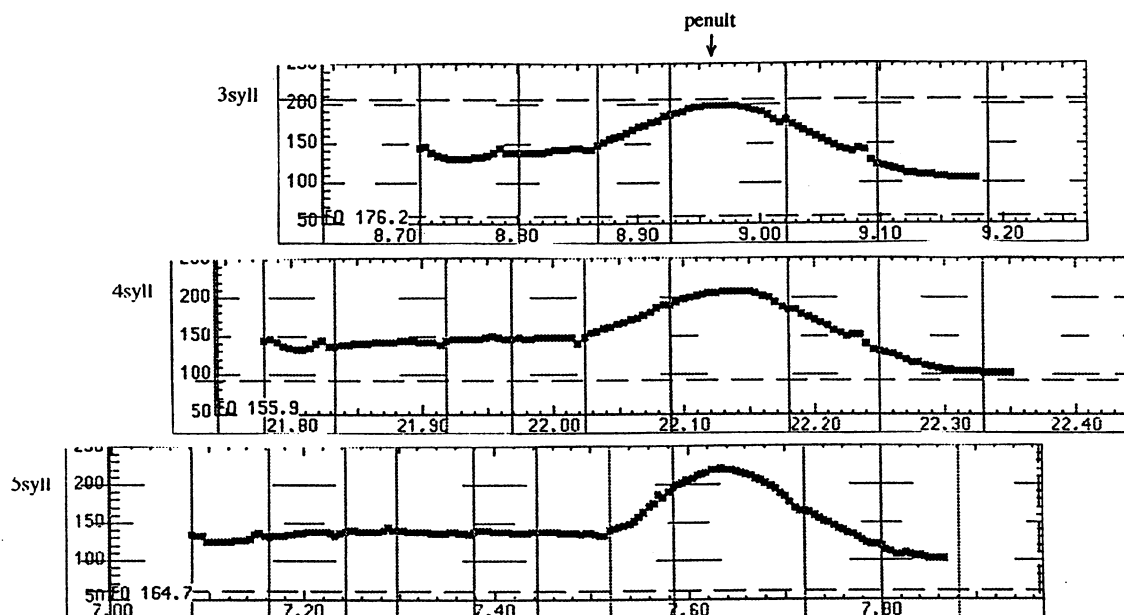


Figure 4. Sample F0 values in three-, four-, and five-syllable words for Speaker W.

Turning now to amplitude, the patterns exhibited are similar across the two speakers for the penult and final syllables, but different for the pretonic syllables. Average amplitude values by syllable for two- to five-syllable words by speaker are presented in Figure 5.

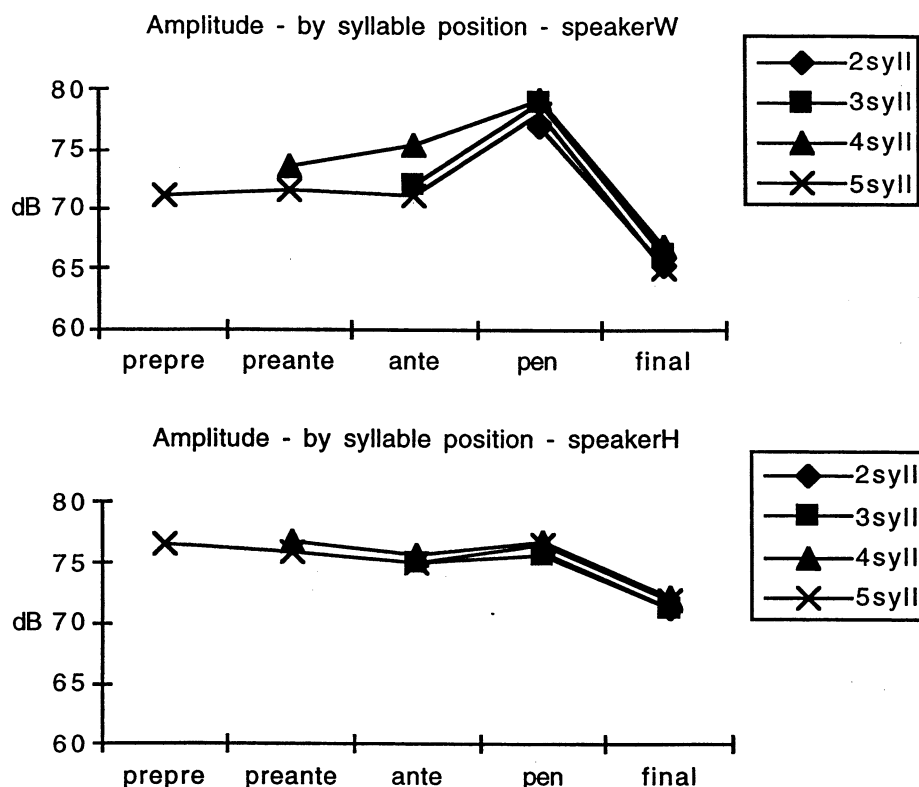


Figure 5. Average amplitude values by syllable for two- to five-syllable words by speaker.

We observe the highest amplitude on the penult, with a sharp drop on the final syllable for both speakers. Again t-tests were done, by speaker and by number of syllables to see if the observed differences were statistically significant (P value $< .05$). Comparing the amplitude of the penults to final syllables, the amplitude was significantly greater for all cases. However comparing pretonic syllables to the penults, the two speakers exhibited distinct patterns. For SpeakerW, the amplitude of all pretonic syllables was lower than that of the penult, significantly so in all cases, except the five-syllable cases, where again the number of tokens was very small. The amplitude of the initial syllable was not higher than the other pretonic syllables. (Indeed in the case of the four-syllable forms it was significantly lower). Thus for SpeakerW the amplitude patterns closely parallel the observed patterns for F0. For SpeakerH, the pretonic syllables had similar amplitude values to those of the penults and the values were not significantly different. However the initial syllables had higher values than other pretonic syllables. This increase falls on those syllables which are impressionistically observed to have secondary stress. This difference approached significance in the case of the four-syllable forms ($P = .065$), but was not significant for

the five-syllable cases. More tokens of four- and five-syllable forms would be needed to confirm this trend. Thus for both speakers, the amplitude pattern observed for penult and final syllables closely parallel those of F0; in all cases the penult shows the highest F0 and greatest intensity, while the final syllable shows a sharp drop. However, the amplitude patterns of the pretonic syllables differ for the two speakers.

We turn now to the duration patterns. Measurements were made for both syllable and vowel duration and the results were very similar. We report here on vowel durations. Overall the two speakers showed quite different patterns and within each speaker the results show more variation than those for F0 or amplitude. We discuss the two speakers in turn, starting with SpeakerW. Average duration values by syllable for two- to five-syllable words for SpeakerW are presented in Figure 6.

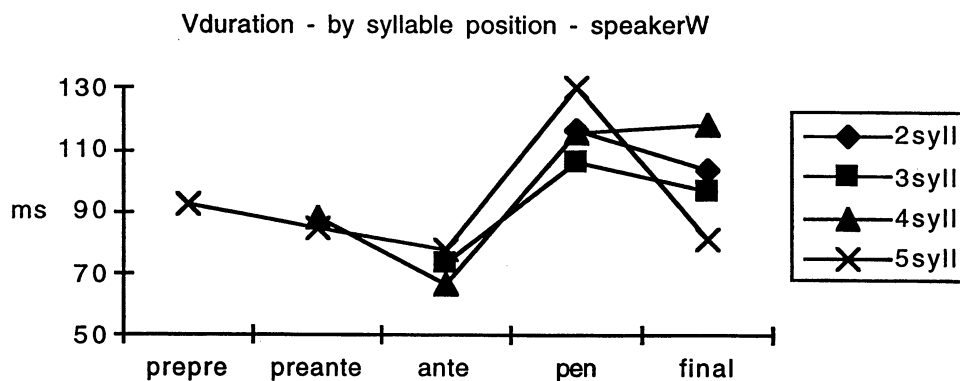


Figure 6. Average duration values by syllable for two- to five-syllable words for SpeakerW.

As observed in Figure 6, for SpeakerW the penult showed the greatest duration when compared to any preceding syllable, consistent with the patterns for F0 and amplitude. Again t-tests were done. The duration differences between the penult and pretonic syllables were significant in all cases, except the five-syllable cases where they approached significance. Final syllables were similar in duration to penults and there were no significant differences. At first blush, this might seem surprising, since the stressed penult would be expected to be the longest, however, this appears to be the result of final lengthening. Final lengthening, characteristically observed at the end of an intonational phrase, is consistent with the intonation patterns, where words appeared to be treated as whole intonational phrases. Initial syllables in four- and five-syllable words, impressionistically observed to have secondary stress, were consistently longer in duration than other pretonic syllables. The difference was highly significant in the four-syllable

cases, but not in the five-syllable cases. (Here too more tokens would be needed to confirm this trend.)

Average duration values by syllable for two- to five-syllable words for SpeakerH are presented below in Figure 7.

The duration patterns observed for SpeakerH are strikingly different from those for SpeakerW. Penults were slightly longer than the final syllables; this difference was significant in the two- and three-syllable cases, but not the four- and five-syllable ones. But in all cases of three syllables or more, the pretonic syllables were significantly longer than the penult; yet it was not the initials that were the longest, rather in the four-syllable cases it was the antepenult which was the longest (in the five-syllable cases the differences were not significant).

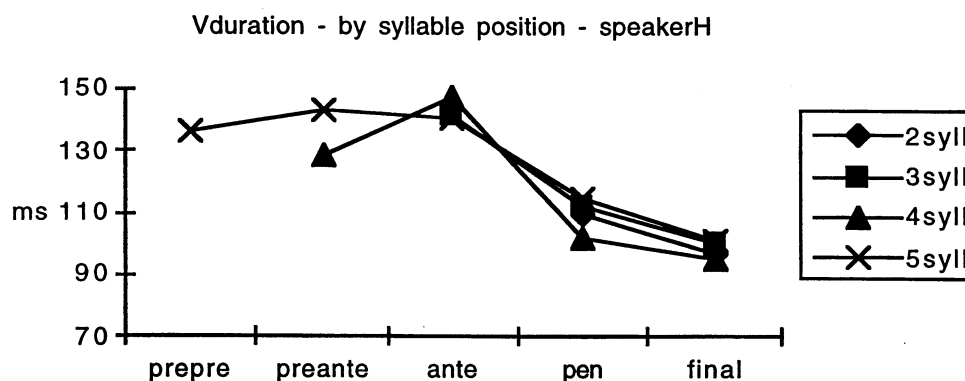


Figure 7. Average duration values by syllable for two- to five-syllable words for SpeakerH.

We are at a loss to explain these duration patterns. Perhaps the lengthening of the antepenult is, in effect, displaced one syllable to the left from the penult. Perceptual studies would be required to determine if this is a tenable account. Unlike for SpeakerW, where the duration patterns closely parallel the results for F0 and amplitude, for SpeakerH, the duration results are very different.

5 Conclusions & Discussion

In conclusion, the prominence of primary stress correlated with F0, and intensity, for both speakers. It also correlated with duration (with the confounding factor of final lengthening) for SpeakerW, but not for SpeakerH. Different correlates of secondary stress were observed for the two speakers: for SpeakerH, initial syllables in four- and five-syllable forms showed increased amplitude and for SpeakerW these syllables showed increased duration. Clearly more work is required to see if these are generally different

strategies available to speakers. SpeakerH showed rather monotone intonational contours, in contrast to SpeakerW's more animated ones. It is possible that SpeakerH was not as comfortable with the task of reiterant speech. The difference in initial-syllable prominence between the two speakers, as well as the unexpected duration patterns of SpeakerH, call for investigating additional speakers.

We now briefly compare our results to previous studies. Halim (1981), investigating the acoustic correlates of primary stress, observed that pitch is systematically highest on the stressed syllable. He observed a general Mid-High-Low pattern. Both stressed syllables and final syllables were observed to show increased duration, while no systematic pattern was observed for intensity. (It is not clear if vowel quality was controlled for in the consideration of intensity.) Our results for F0 for both speakers and duration for SpeakerW are consistent with the observations made by Halim, however we observed systematic intensity patterns as well. Our results are also similar to those of Laksman (1994), in a study limited to disyllables. Laksman found F0 to be consistently different between the initial and final syllables, as was amplitude for identical vowels. For duration Laksman found an interaction between syllable structure and duration, not observed in the present study, since we restricted syllable type. She also found a difference in duration between tokens recorded with or without a pause, with the duration of syllables in prepausal position being longer. She attributes this to prepausal final lengthening, lending support to the conclusion that the length of the final syllable in the present study is also due to final or prepausal lengthening. Similar patterns to those observed by Laksman were found in our study for the penult vs. final syllables in all our forms, two- five-syllables, showing that these patterns are not restricted to disyllables.

Turning to syllable duration, we can compare the observed durations of penults to those patterns found by van Zanten & van Heuven (1994). (Unfortunately duration by syllable was not presented for unstressed syllables, so a broader comparison is not possible.) Like van Zanten & van Heuven (1994), we found the duration of the penult to across the two- to five-syllable cases to be quite similar for both speakers. We also found duration to be quite constant independent of the number of syllables. This suggests that duration of stressed syllables in Indonesian is quite constant independent of word length and the long duration in monosyllables observed by van Zanten & van Heuven (1994) might be a lengthening effect.

Considering the difference between primary and secondary stress, the duration results for SpeakerW show an interesting contrast with English. The pattern observed here shows the initial syllables patterning together durationally with stressed syllables, while in

English, the secondary stresses have been observed to resemble stressless syllables in terms of duration for some speakers (Nakatani et al. 1981).

Finally, this study highlights the importance of expanding the range of languages investigated to reach a fuller understanding of the phonetic realization of both primary and secondary stress, while providing further evidence of the reliability of the technique of reiterant speech.

6 References

- Berenstein, A. (1979) A cross-linguistic study on the perception and production of stress. *UCLA Working Papers in Phonetics* 47: 1-59.
- Cohn, A. (1989) Stress in Indonesian and bracketing paradoxes. *NLLT* 7: 167-216.
- Cohn, A. and J. McCarthy (1994) Alignment and parallelism in Indonesian phonology. Cornell University and UMass ms. (to appear in *Linguistic Inquiry*).
- Halim, A. (1974) *Intonasi*. Jakarta: Djambatan.
- Halim, A. (1981) *Intonation in Relation to Syntax in Indonesian*. Pacific Linguistics: Series D, No. 36. Canberra.
- Laksman, M. (1994) Location of stress in Indonesian words and sentences. In C. Odé and V. van Heuven (1994).
- Lapoliwa, H. (1981) *A Generative Approach to the Phonology of Bahasa Indonesia*. Pacific Linguistics: Series D, No. 34, Canberra.
- Larkey, L. (1983) Reiterant speech: an acoustic and perceptual validation. *JASA* 73: 1337-1345.
- Liberman, M. and L. Streeter (1978) Use of nonsense-syllable mimicry in the study of prosodic phenomena. *JASA* 63: 231-233.
- MacDonald, R. (1976) *Indonesian Reference Grammar*. Washington, DC: Georgetown University Press.
- Nakatani, L., O'Conner, K. and C. Aston (1981) Prosodic aspects of American English speech rhythm. *Phonetica* 38: 84-106.
- Odé, C. and V. van Heuven (1994) *Experimental Studies of Indonesian Prosody*. SEMAIAN 9. Leiden: Valkgroep Talen en Culturen van Zuidoost-Azië en Oceanië.
- Zanten, E. van and V. van Heuven (1994) Effects of word length and substrate language on the temporal organisation of words in Indonesian. Paper presented at the 7ICAL, Noordwijkerhout, The Netherlands.

Appendix: Word list

A. Experiment 1: normal and reiterant speech 2 & 3 syllable words

B. Experiment 2: reiterant speech 2, 3, 4, & 5 syllable words

2 syllable

cara	'manner'
cari	'search for'
ciri	'feature'
didik	'educate'
gigit	'bite'
katak	'frog'

3 syllable

alamat	'address'
batara	'deity'
bicara	'speak'
nasihat	'advice'
pidana	'criminal'
wilayah	'region'

4 syllable

bijaksana	'wise'
masyarakat	'society'
meditasi	'meditation'
provokasi	'provocation'

5 syllable

antisipasi	'anticipation'
kontinuasi	'continuation'