SOME FORMAL AND FUNCTIONAL ASPECTS OF INDONESIAN INTONATION[1]

Ewald F. Ebing & Vincent J. van Heuven

1 Analysis-by-resynthesis in descriptive intonation research

1.1 Introduction

Literature on both formal and functional aspects of Indonesian[2] intonation is scarce. Over the past decades, only two publications dedicated specifically to Indonesian intonation have appeared (Samsuri 1971; Halim 1981) prior to the present research, which employs the method of analysis-by-resynthesis.

Older studies on the intonation of specific languages were usually dependent on impressionistic methods of data collection. With the current availability of pitch measuring devices, it has become possible to achieve a much higher degree of accuracy in the acoustic description of pitch phenomena. However, the modern investigator of intonation faces a different problem: how to select the meaningful distinctions from the abundance of data supplied by such devices? Which of the variations obtained by measurements of fundamental frequency (henceforth: F_0) represent intonational 'allophones', 'phonemes' and 'morphemes'? Finding intonational 'minimal' pairs is not as easy as it is in segmental phonemics, because no appeal can be made to lexical distinctions; intonational meaning is notoriously elusive (cf. Ladd 1980). A different strategy is needed to identify the building blocks of an intonation contour.

Stylizing intonation. One strategy which has proved effective is the stylization method ('t Hart, Collier & Cohen 1990). It involves a method of analytic listening, allowing the researcher to quickly compare fragments of speech. The F_0-parameter is measured in the original human utterance, but can be changed by the researcher. The result of this manipulation can be made audible, which makes direct comparison with the original intonation possible. Fluctuations in F_0 which are perceptually irrelevant can thus be eliminated. A stylized pitch contour consisting of the smallest possible number of straight

[2] The term 'Indonesian' as used in this contribution refers to the spoken continuum of linguistic forms reflecting a standardized form of Malay known as Bahasa Indonesia, the official language of the Republic of Indonesia.

lines while still sounding identical to the original, is called a close-copy stylization.

Modelling intonation. Such stylized pitch contours, which are perceptually equivalent to the original F_0-curves, can be characterized by sequences of pitch movements that can be described with a relatively small set of parameters: their direction, excursion size, fundamental frequency of the starting point, and duration. Finally, it must be specified how these movements are synchronized with the segmental structure on which they are superimposed.

These structures can be further analysed by obtaining similarity judgments from native speakers for small chunks of speech, each containing a limited number of pitch movements, preferably only one. This allows us to move up from the intonational 'phonetic' to the 'phonemic' level. The resulting units can then be defined in terms of a set of standardized phonetic characteristics, specifying properties such as excursion size, duration and timing relative to the segmental structure. An intonation model is defined in this context as an algorithm specifying which sequences of intonational units are well-formed. It does not predict which particular contours will occur on a given utterance.

2 Applying the stylization method to Indonesian intonation

2.1 Classification experiment

To date, a preliminary model exists which is based on a classification experiment (Ebing 1991, 1994). Listeners were asked to sort twenty-four short fragments (one to two seconds) of speech on the basis of melodic similarity. Using a hierarchical cluster analysis, the pitch contours were categorized into eight categories of 'configurations', i.e. more or less fixed, recurrent combinations of pitch movements (cf. 't Hart et al. 1990). The resulting categories were broadly characterized in terms of the features shape (rise, rise-fall, etc.), high vs. low register, wide vs. narrow range, and early vs. late timing (Ebing 1994:199).

2.2 Development of a melodic model

The above-mentioned categorization was developed into a preliminary intonation model, consisting of an inventory of perceptually relevant pitch movements which may be combined by rules to form the pitch configurations found in the corpus. Adding more data may lead to future adjustments in the phonetic specifications or the addition of more possible combinations of pitch, especially since most of the research was based on the unprepared speech of a single speaker.

Table 1: Two versions of standardized specifications for Indonesian pitch movements. Durations and synchronization points are expressed in milliseconds (ms), excursion size of pitch movements in semitones (st).

MOVEMENT	FIRST VERSION	(CHANGES IN) SECOND VERSION
Rises		
1. excursion:	6 st	
duration:	100 ms	140 ms
onset:	onset of penultimate syll.	onset of final syll., if preceded by A; 180 ms earlier otherwise
offset:	at least 30 ms before onset of final syll.	middle of final syll., if preceded by A; 40 ms before onset of final syll. otherwise
2. excursion:	3 st	
duration:	100 ms	140 ms
onset:	if followed by B: onset of penultimate; otherwise middle of penultimate	if followed by B: 180 ms before onset final syll.; if followed by C, middle of penult.
offset:	dependent on following movement (B or C)	if followed by B: 40 ms before onset final syll.; otherwise variable
3. excursion:	6 st	
duration:	variable	200 ms
onset:	middle of prefinal syll.	variable
offset:	30 ms after onset of final syll.	40 ms after onset of final syll.
4. excursion:	8 st	6 st
duration:	120-200 ms depending on duration of final syll.	200 ms
onset:	onset of prefinal syll.	variable
offset:	variable	end-of-voicing final syll.
Falls		
A. excursion:	8 st	7 st
duration:	150 ms	if preceded by 1, 180 ms; otherwise variable
onset:	end of preceding rise, if present; otherwise at onset of penultimate syll.	
offset:	variable	
B. excursion:	4 st	
duration:	150 ms	if preceded by A1, variable; otherwise 180 ms
onset:	end of preceding rise	
offset:	variable	
C. excursion:	1 st	
duration:	dependent on duration of final syll.	
onset:	end of preceding rise	
offset:	end-of-voicing final syll.	

Our preliminary inventory of perceptually relevant pitch movements for Indonesian comprises four different rises and three different falls. The parameters involved are outlined below. Onset and offset timing and duration can be restricted by the duration of the syllables. For some movements, timing specifications differ depending on the melodic context in which the pitch movement appears. The values in the first version of the standard were derived from results of our classification experiment (Ebing 1994:200-207), and are presented in table 1.

The perceptual categories established in the classification experiment can be considered configurations of the above-mentioned pitch movements: Category RFL thus consists of 2&B, RFHWRE of 1&A, RFHNRE of 1&B, RFHNRL of 3&B, RFR of 1&A4, R of 4, RHP of 2C, and FRF of A1B.

2.3 Experimental evaluation: two perception experiments

Two experiments were run in order to establish the adequacy of the inventory of standardized pitch movements for Indonesian. In the first experiment we evaluated the first version of the standardization; on the basis of the results we obtained from this experiment, the specification of the standardized movements was adjusted in the way listed in table 1 under 'second version'. In both experiments listeners rated exemplars of Indonesian utterances (resynthesized with various melodic contours) on a 10-point acceptability scale.

2.3.1 Method

Experiment 1: stimuli, listeners, procedure

Eight fragments were selected from the materials previously used in our classification experiment, each fragment representing one of the perceptually relevant pitch configurations. Four versions of the intonation contours were produced by manipulating the F_0-parameter in the resynthesized fragments, namely:

1a. *Close-copy stylizations (COPY)*. These were made using the stylization method mentioned in section 1.1; since these exemplars are by definition prosodically indistinguishable from the original human utterances, they should receive the highest acceptability ratings.

1b. *Standardized versions (STAN)*. Here the close-copy pitch movements were replaced by the standardized pitch movements according to the specifications listed in table 1 under 'first version'. The acceptability of these exemplars should approach that of the COPY-versions; in so far

as they should receive poor ratings, the standardization can be considered a failure.

1c. *Dutch-based versions (DUTCH)*. Here the Indonesian pitch movements were replaced by standardized movements taken from the Dutch intonation grammar, such that the best visual match was obtained between the Indonesian and Dutch movements. This condition was included to examine the extent to which the Indonesian movements are language specific, which would be the case if the Dutch exemplars receive consistently poorer acceptability ratings than the STAN-version.

1d. *Shifted versions (SHIFT)*. In the SHIFT-versions, standardized Indonesian movements were shifted in time to the immediately preceding syllable. This condition served as a baseline: we expected that Indonesian listeners would object to these melodies since they are no longer in accordance with the syntactic/semantic structure of the utterance.

The 32 stimuli were recorded on tape in pseudo-random order, avoiding sequences of two identical fragments or two identical experimental conditions. In the second presentation the order of the stimuli was reversed. The actual test was preceded by five warming-up items.

Twelve native speakers of Indonesian, all students at the Department of Indonesian at the Universitas Islam Riau (UIR), Pekanbaru, Riau, participated as listeners in Experiment 1.[3] They listened to the tape over good quality loudspeakers in a quiet lecture room. The task of the listeners was to rate the melodic acceptability of the intonation contours on a 10-point scale (1 = totally unacceptable, 10 = fully acceptable).[4]

Experiment 2: stimuli

Experiment 2 resembles the previous experiment in most respects; therefore we will only briefly discuss the relevant differences that were introduced. In this experiment, two realizations (rather than one) of each configuration were used, resulting in 64 (2 x 8 x 4) stimulus types. This time, the stimulus words were embedded in a larger portion of their original melodic context than in Experiment 1 as we had reasons to believe that presentation of certain configurations in isolation had resulted in poorer acceptability ratings for other

---

[3] The experiment was run in situ by Drs. Al Azhar of Pusat Pengajian Melayu, UIR Pekanbaru, whose help is gratefully acknowledged here.

[4] In an additional run of the experiment the same listeners were asked to state their preference between (the 48 logically possible) pairs of these stimuli in a forced choice pairwise comparison. The results showed the same tendencies as those for the 10-point scaling test, and will not be reported here.

than prosodic reasons. The following four melodic versions were prepared for each speech fragment:

2a. *Close-copy stylizations (COPY)*. These were the same melodic versions as in version (1a); the added set of eight new speech fragments had been given its close-copy stylizations by the same methods as the existing set.

2b. *Adjusted standard specification (STAN)*. Here the melodies of (2a) were replaced by the standardized movements according to the specifications listed in table 1 under 'second version'. We expected that this version would approach the COPY-versions even more closely than in the first experiment.

2c. *Dutch-based versions (DUTCH)*. See (1c) above.

2d. *Original F_0-curves (ORIG)*. The original F_0-curves measured by the pitch determination programme we used in the resynthesis. This version was included to check whether the close-copy stylizations were indeed perceptually indistinguishable from the original, so as to be able to rule out the possibility that poor ratings in Experiment 1 could have been caused by errors in the stylization process.

In order to highlight the part of each stimulus fragment containing the target pitch configuration, the speech that formed the melodic context surrounding it was attenuated by 3 decibels, and both spectrally and temporally smeared without affecting the prosody, so that it was no longer intelligible. The spectral properties of the target words were left untouched; only their F_0 was changed to produce the four versions.

Thirty listeners took part in Experiment 2; all were native Indonesian linguists attending a workshop on experimental phonetics given by the second author at Universitas Indonesia (UI) in Jakarta. Instructions and procedure were as in Experiment 1.

2.3.2 Results and conclusions

Experiment 1

Figure 1 presents the mean acceptability ratings for the eight different configurations of pitch movements broken down by the four melodic versions (24 responses per data point). The rightmost entry ('mean') in this figure represents the mean acceptability rating accumulated for each melodic version over all eight configurations.

The data in figure 1 were subjected to an analysis of variance with configuration and melodic version as fixed factors and with repeated measures for listener and repetition. Generally, the differences between the four melodic versions are small but they reflect the order of acceptability that was

postulated above. The close-copy versions get the highest mean scores (7.63), followed by the standardized versions (7.61), then the Dutch-based versions (7.39) and finally the time-shifted versions (7.29). Although the differences are small, the effect of melodic version is significant, $F(3,764) = 3.0$, $p < .05$. Moreover, a Student-Newman-Keuls range test ($\alpha = 5$ %) shows that the mean scores for COPY and STAN do not differ from each other but do differ significantly from those of DUTCH and SHIFT, which, again, do not differ from each other.

These results indicate that even the preliminary standardized specifications of our melodic model for Indonesian is reasonably adequate: there is no significant overall difference between the close-copy stylizations and the standard movements. Moreover, the listeners differentiated the acceptability of these two melodic versions from that of the 'strange' versions, i.e. the Dutch approximations and the time-shifted melodies. On the one hand this would mean that the Indonesian movements are indeed language-specific, and that the movements are critically bound to specific syllables on the basis of the syntactic/semantic structure of the fragment. On the other hand, it should be noted that the differences between the favoured and disfavoured melodic versions are very small, and that their ratings are all in the upper half of the rating scale, i.e., are sufficient. When similar 'strange' melodic versions were tested in comparative research on Dutch and British English intonation (de Pijper 1983) they were outright rejected by the listeners. This, then, would imply that the 'Dutch' versions are quite reasonable substitutes for the native Indonesian specifications,[5] and that the exact syllable that bears the pitch movement is not really critical.

Curiously enough, the results in figure 1 show large differences in acceptability between the various configurations. For instance, configurations 1&B and 4 receive clearly poorer ratings than the other configurations. This effect may have been caused by the fact that these configurations cannot occur out of context, which is the reason why we offered the stimuli in the next experiment embedded in a proper melodic context. Moreover, there is quite a bit of interaction between configuration and melodic version. For example, though the time-shifted exemplars are usually the least acceptable versions, they are highly favoured in configuration 2&B and 1A&B. For lack of space we shall not deal with possible explanations of the interactions here.

---

[5] The interchangeability of our Dutch and Indonesian movements may well be deceptive. Ebing (1988) replaced the original speech melodies of longer Indonesian utterances by melodies produced by native speakers of Dutch, and vice versa. This time the differences in acceptability ratings were considerably larger: 8.3 against 7.5 points for Indonesian utterances with native and foreign movements, respectively, and 7.7 against 6.8 points for Dutch utterances with native and foreign pitch movements.
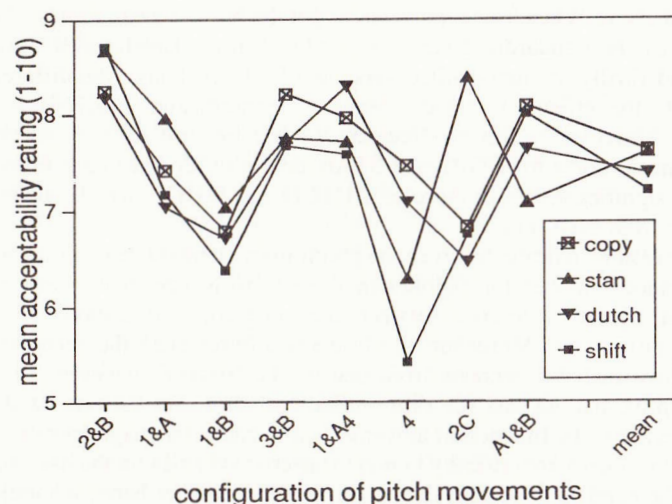
Figure 1: Mean acceptability rating for 8 configurations of pitch movements broken down by melodic version (first model). Rightmost entries represent means across configurations.

Experiment 2

The mean acceptability ratings for the eight different configurations of pitch movements broken down by the four melodic versions (60 responses per data point) are presented in figure 2.

This time there are no differences between the four melodic versions of the pitch movements, $F(3,3768) < 1$. This would indicate that our stylizations were quite adequate, and that the standardizations, whether Indonesian or Dutch, are melodically as adequate as the originals. There is no significant interaction between configuration and melodic version, so that we must conclude that the four versions used are all equally acceptable for all the fragments in the material. Also, there is only one fragment that deviates from the others in terms of general acceptability, viz. 1&A. Apparently, the absence of a proper melodic context to the target configurations in Experiment 1 was the main cause for the rather large discrepancies in acceptability between the various configurations. The addition of melodic context has not led to an improvement for the 1&A configuration: this configuration is still judged at more than a full point less acceptable than average. At this time we have no explanation why this should be so.
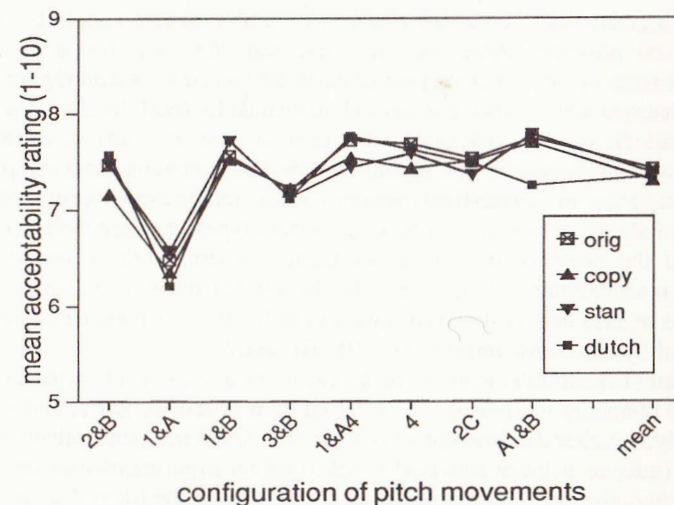

Figure 2: Mean acceptability rating for 8 configurations of pitch movements broken down by melodic version (second, adjusted model). Rightmost entries represent means across configurations.

The results of Experiment 2 do not allow us to conclude that the adjustments to our melodic model for Indonesian are an improvement over the first version, but on the other hand, their acceptability is certainly not less than in Experiment 1.

3 Accent and intonational boundary marking in Indonesian

3.1 Formal distinctions and functional categories: some problems for Indonesian

Two important functions of prosody which are almost universally recognized by researchers of different backgrounds, are (i) accentuation, which is used to highlight certain parts of an utterance in relation to the background, and (ii) segmentation of utterances at various levels, which helps the listener determine the syntactic and/or information structure of an utterance.

In the sentence melody of Indonesian the great majority of melodic phenomena seems to occur at the phrase or sentence boundary, i.e. all pitch movements seem to occur at the end of the phrase or sentence, so that it is extremely difficult to determine whether a particular pitch movement is accent-lending or boundary-marking. In fact, it may well be the case that the distinction is irrelevant to Indonesian. Data reported by Odé (1994:60),

however, indicate that Indonesian listeners were able to disentangle the two functions of prosody: there was no statistical tendency for perceived boundaries to co-occur with perceived prominence (accents). Assuming for the moment that boundary marking is not only signalled by local deceleration but also by melodic means (and Odé's data tend to bear this out), it must be possible to distinguish between accent-lending and boundary-marking pitch movements, even in phrase-final position. Also, our previous explorations seemed to indicate an association between certain types of rising pitch movements and the presence of phrase boundaries, whereas others seemed to represent pitch accents (Ebing 1988:24). In our current nomenclature, the former are represented by 4 and to some extent by 2C, whereas the latter are represented by the configurations 1A, 2B, 1B, and 3B.[6]

The aim of the following experiment was to try to establish to what extent accent and boundary marking can be effected by means of prosody, and more importantly, whether these two functions can be marked independently of each other. In order to achieve this goal we devised an experimental set-up that combines two methodologies. The first is to manipulate the focus distribution in an utterance by having a speaker apply a metalinguistic contrast in order to rectify a potential misunderstanding on the part of his listener. For example, in *(I did not say: 'one plus [NINE]$_{+F}$ equals three'; I said:) 'one plus [TWO]$_{+F}$ equals three'*, narrow focus (indicated by [...]$_{+F}$) is on the second numeral *two*, which will therefore be marked by a pitch accent. However, in *(I did not say: '[NINE]$_{+F}$ plus two equals three'; I said:) '[ONE]$_{+F}$ plus two equals three'* narrow focus is on the first numeral *one*, leaving the second numeral *two* out of focus, which is then de-accented (cf. van Heuven 1994a, b). Second, we manipulated the position of a prosodic phrase boundary by forcing the speaker to disambiguate a potentially ambiguous arithmetic expression, in the way this was done before by Lehiste (1970), Lehiste, Olive & Streeter (1976) and O'Malley, Kloker & Dara-Abrams (1973).

3.2 Method

In this experiment we asked a single speaker to produce the same sentence 14 different ways, orthogonally varying focus structure and position of the phrase

_____

[6] The pitch phenomena associated with boundaries seem to reflect Halim's (1981) description of the pattern /233/, which he claims is linked specifically to the semantic function of topic marking, and /231/ which is the citation form. According to Halim, accent is exclusively associated with the occurrence of the high pitch phoneme /3/. In our material, however, early rise 1 and late rise 4 can occur in adjacent syllables within the same word, and both reach the upper region of the tonal space used by the speaker (RFR or 1&A4). If Halim's analysis applies here, this would mean that both syllables would have an accent, which is problematic, since Halim does not recognize the existence of a secondary accent.

boundary. Indonesian listeners were then asked to indicate for each utterance where they perceived the primary and secondary accents, and where the phrase boundary was located. The responses should reveal to what extent the focus-marking (accent) and boundary-marking functions of prosody can be separated in Indonesian. In a post-hoc stimulus analysis we measured the relevant acoustic properties of our stimulus utterances in order to establish to what degree focus and boundary marking can be automatically recovered from the spoken sentences.

Stimuli. A total of 14 different utterances were recorded on tape in a sound attenuated booth. They comprised 7 versions of each of the arithmetic expressions *2 x (3 + 5)* and *(2 x 3) + 5*, both pronounced *Dua kali tiga tambah lima* 'two times three plus five'.[7] The following focus distributions were imposed on the sentence: (0) broad focus on the entire phrase, (1) narrow focus on the first numeral, (2) narrow focus on the second numeral, (3) narrow focus on the third numeral, (4) double focus on first and second numeral, (5) double focus on first and third numeral, and (6) double focus on second and third numeral. The test sentences were pronounced by a male native speaker of Indonesian, a lecturer of Indonesian who recently joined Leiden University. He was instructed to use as few physical pauses as possible. Each sentence was prompted by a question sentence to provide a context where one or more words were placed in focus. The recorded materials were digitized and stored in a VAX/VMS computer (10 KHz, 12 bits, 4.5 KHz LP). Pitch was measured through the method of subharmonic summation (cf. Hermes 1988) and subsequently stylized using the close-copy method (see above).

Listeners. Six Indonesian subjects, five of whom were graduate students at Leiden University, participated in the perception experiment. The speaker who had read the test sentences for the recording also participated as the sixth Indonesian subject.

Design and task. In the first part of the test, which was concerned with boundary perception, each stimulus token was presented 5 times, making for a total of 5 x 14 = 70 stimuli to be judged by the listeners. The 70 stimuli were presented in random order. The actual test was preceded by 10 warming-up items. The listeners were asked to indicate, with forced choice, for each

_____

[7] A corresponding set of arithmetic expressions *2 x (3 + 5) = 16* and *(2 x 3) + 5 = 11*, pronounced *dua kali tiga tambah lima sama dengan enambelas*, 'two times three plus five equals sixteen' and *dua kali tiga tambah lima sama dengan sebelas*, 'two times three plus five equals eleven', respectively, were also recorded. This procedure yielded 2 x 7 'short' versions and 2 x 7 'long' versions. The long versions were analysed but not used in the perception experiment; our report here is exclusively based on the short versions.

instance which bracketing was appropriate for the sentence they were listening to, by placing a slash at the relevant position in the sentences on their answer forms, as follows:

dua kali / tiga tambah lima representing 2 x (3 + 5), or
dua kali tiga / tambah lima representing (2 x 3) + 5.

In the second part of the test each stimulus was presented twice in succession, so that the test contained 70 pairs or a total of 140 stimuli, preceded by 5 pairs of warming-up items. For each first presentation of a stimulus the listeners ticked the word on their answer sheets which, in their opinion, was most prominent in the utterance; on the second presentation they ticked the word which in their perception was second-most prominent.

3.3 Results

For lack of space, and for the sake of clarity, we will limit our data presentation to only the perception of the phrase boundary and primary accent. Moreover, we shall only consider the results obtained for the three focus conditions that asked for a single contrastive accent. Tables 2 and 3 present the results of this experiment. Table 2 specifies the percentage of primary, secondary and no accent responses for each of the three relevant numerals in each stimulus utterance broken down by intended focus condition, and by intended phrase boundary position. In table 3 the complementary cross-tabulation is given: here the percentage of perceived phrase boundaries after first versus second numeral are broken down by focus, word position and intended focus distribution.

The results indicate, first of all, that our listeners have great difficulty in perceiving a narrow focus accent on the third numeral; they do accurately differentiate between accent on the first and second numeral. Secondly, there is a strong tendency for our listeners to report an accent on a word immediately followed by a phrase boundary. This tendency is stronger when the narrow focus is on the last numeral in the utterance than when it is on the first or second numeral. The intermediate conclusion based on this part of the data would be that our Indonesian listeners heavily confuse the accent-lending and boundary-marking functions of the speaker's pitch movements. We do not know yet, whether this behaviour resides with the listeners themselves or whether it is the speaker who failed to adequately encode the functions in his speech production.

Table 2: Percentage of perceived primary accents on words with narrow focus in first, second and third sentence position, broken down by position of phrase boundary. The increment in number of perceived accents due to the presence of a phrase boundary is given in the rightmost column.

	boundary after numeral #1			boundary after numeral #2			extra accents due to boundary after numeral	
focus on numeral #	[acc] perceived on: #1	#2	#3	[acc] perceived on: #1	#2	#3	#1	#2
1	**97**	3	0	**73**	20	7	24	17
2	23	**73**	3	0	**97**	3	23	24
3	83	7	10	37	63	0	46	56
mean	68	28	4	37	60	3	31	32

Table 3: Percentage of correctly perceived intended phrase boundaries for words with narrow focus in first, second and third position, broken down by position of phrase boundary. The increment in number of perceived boundaries per focus position is given in the rightmost column.

focus on numeral #	phrase boundaries correctly perceived after numeral #1	numeral #2	Difference
1	69	83	14
2	29	74	45
3	31	94	63
mean	43	83	40

Here the results show that phrase boundaries are sooner perceived after the second numeral than after the first. Moreover, there is a sizeable and complex interaction with the position of the focused (accented) numeral and phrase boundary perception: there is only a relatively small effect of focus position on the perception of a boundary after the second numeral; the (confounding) effect of focus is much larger for boundary perception after the first numeral, especially when the focus is on a non-initial constituent. By and large it seems that there is heavier confusion between boundary and accent as either or both accent and boundary occur early in the sentence.

We would like to conclude from these data that the boundary-marking and accentuation functions of prosody are easily confused in the perception of Indonesian listeners. In order to see whether this is due to the speech input or whether the problem resides entirely with the listeners, we have undertaken a stimulus analysis.

3.4 Stimulus analysis

For each of the three numerals occurring in the remaining six (3 single focus conditions x 2 phrase boundary positions) stimuli used in the perception experiment, measurements were made of the following phonetic properties: segmental synchronization, duration and excursion size of the rising and falling pitch movements, the beginnings and endings of the words, and the vowel onsets of the two syllables making up each numeral. Table 4 provides the acoustic measures determined for the 6 sentences (vertically) broken down by the 3 numerals (horizontally). In principle, each numeral contained a rise-fall pitch contour; pitch excursions were measured in semitones (st)[8] with a positive value for rises and a negative value for falls. The segmental synchronization of the movements is expressed in terms of the time interval (in milliseconds, ms) between the onset of the rise and the onset of the first vowel in the numeral, or between the onset of the fall and the onset of the second vowel. A negative synchronization value indicates that the movement starts prior to the relevant vowel onset (early movements).

It appears that the most useful cue to characterize accent and phrase boundary is the excursion size of the pitch movements associated with these functions. In what follows we shall use the mean excursion of the rise-fall configuration as the most reliable estimate of the excursion. The data show clearly that the coincidence of narrow focus and preboundary position prompts the speaker to produce extremely large pitch movements in excess of a full octave. Apparently our listeners took these very large movements as simultaneously cuing accent and preboundary position, since both accent and preboundary status were identified as intended by the speaker.

When the phrase boundary occurs immediately after the first numeral it is not difficult to acoustically distinguish between the focused and the preboundary numeral, and to distinguish each of these from the non-focused phrase-medial numerals. The pitch excursions of the latter type never exceed 3 st; when focus and preboundary position are dissociated, each is characterized by movements of intermediate size, i.e. with a mean excursion between 6 and 10 st. Here it seems that the crucial acoustic factor discriminating between the two functions is the timing of the movements relative to the vowel onset. The difference is most apparent in the timing of the fall: if it precedes the onset of the second vowel by 100 ms or more, it is accent-lending; when its onset occurs 100 ms (or more) after the second vowel onset, it marks a phrase boundary. We would interpret these findings as evidence that the speaker adequately differentiates between boundary-marking and accent-lending pitch movements. The listeners, however, apparently found it difficult to use the available acoustic cues to their fullest advantage. For

reasons unknown to us, our listeners are heavily biassed towards hearing a phrase boundary after the second numeral in spite of the acoustic cues to the contrary.

Table 4: Excursion size (in st), duration (in ms), synchronization of rise and fall (in ms re. vowel onset of first and second syllable, respectively), and duration (in ms) of first and second syllable of numerals in test utterances, broken down by position of focused constituent (+F) and position of the phrase boundary (+B).

	numeral #1 (dua)		numeral #2 (tiga)		numeral #3 (lima)	
	[+F,+B]					
excursion of rise/fall:	13	-12	3	-0	2	-3
duration of rise/fall:	250	190	30	80	70	100
synchronization of rise/fall:	-20	120	110	-20	-120	-200
duration of syllables:	300	130	200	170	160	30
	[+B]		**[+F]**			
excursion of rise/fall:	9	-6	8	-11	2	-4
duration of rise/fall:	160	240	130	200	60	90
synchronization of rise/fall:	-70	100	-120	-140	-50	-100
duration of syllables:	180	100	200	110	170	130
	[+B]				**[+F]**	
excursion of rise/fall:	11	-9	4	-0	7	-12
duration of rise/fall:	230	260	60	60	120	210
synchronization of rise/fall:	-30	-110	90	10	-90	-140
duration of syllables:	210	140	170	70	250	120
	[+F]		**[+B]**			
excursion of rise/fall:	4	-2	3	-5	2	-5
duration of rise/fall:	40	60	50	360	90	240
synchronization of rise/fall:	-10	-100	90	10	-120	-320
duration of syllables:	190	70	160	70	200	110
			[+F,+B]			
excursion of rise/fall:	2	-3	14	-14	4	-3
duration of rise/fall:	130	310	130	210	80	110
synchronization of rise/fall:	-60	0	-270	-120	-80	-150
duration of syllables:	140	90	200	120	220	140
			[+B]		**[+F]**	
excursion of rise/fall:	2	-2	8	-9	2	-4
duration of rise/fall:	170	70	130	220	100	190
synchronization of rise/fall:	-90	10	-120	-130	-90	-130
duration of syllables:	160	100	170	120	200	120

---

[8] A semitone is one-twelfth of an octave, or a 6% increment in frequency.

When the phrase boundary occurs after the second numeral, the speaker's behaviour is far less systematic. With the exception of one case, the accent-lending and boundary-marking functions are no longer characterized by pitch movements in the intermediate 6 to 10 st range: in terms of size they are indiscriminable from those on non-accented phrase-medial numerals. In the absence of clear acoustical cues our listeners' behaviour seems almost completely guided by the kind of biasses we discussed above: they assume a phrase boundary after the second numeral as long as there are no compelling counter-indications. Also, our listeners more or less refuse to hear an accent on the third numeral.

3.5 Discussion and conclusion

The results of the above experiment bear out the intuition voiced in the introduction that there is a problem in Indonesian in separating the accent-lending and boundary-marking functions of intonation. Referring to the stimuli with an intended phrase boundary after the first numeral, we are faced with the unusual situation where a speaker more or less adequately encodes the two distinct functions in his utterances, which his listeners subsequently fail to recover, presumably due to strong perceptual biasses. In the stimuli with the boundary after the second numeral, which seems to be the preferred position for the boundary, the speaker is no longer adequate in his encoding, and the listeners' bias overrides whatever acoustic cues might be left.

It is unclear at this moment whether this unsatisfactory result is due to our infelicitous choice of speaker. We intend to check this matter (i) by recording several more speakers performing the same task as above, and (ii) by synthesizing utterances in which we shall systematically vary the excursion size and timing characteristics of rises and falls and see to what extent these movements are interpreted by native listeners as either accent-lending or boundary-marking.

Finally, more work will be necessary in order to establish the identity of the pitch movements that our speaker used to mark focus and preboundary position. Most of the movements can be related in a straightforward fashion to the standardized pitch movements in our inventory, but some complications have arisen. Most importantly, we shall have to make provisions in our melodic model to allow extra large excursions when accent and boundary marking coincide. An attractive solution would be to consider the extra large movements as compositional: conceivably they can be analysed as the result of stacking a late boundary-marking movement on top of an early accent-lending rise.

REFERENCES

Ebing, E.F.
1988   'Intonatie van het Indonesisch. Naar een analyse door resynthese van perceptief relevante toonhoogtebewegingen in het standaard-Indonesisch'. Leiden University, unpublished M.A. Thesis.
1991   'A preliminary description of pitch accents in Bahasa Indonesia', *Proceedings of the XIIth International Congress of Phonetic Sciences* 3. Aix-en-Provence, 258-261.
1994   'Towards an inventory of perceptually relevant pitch movements for Indonesian', in: C. Odé & V.J. van Heuven (eds.) *Experimental studies of Indonesian prosody*. Semaian 9. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden, 181-210.

Halim, Amran
1981   *Intonation in relation to syntax in Indonesian*. Pacific Linguistics D 36. Materials in Languages of Indonesia 5.

Hart, J. 't, R. Collier & A. Cohen
1990   *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Hermes, D.J.
1988   'Measurement of pitch by subharmonic summation', *Journal of the Acoustical Society of America* 83:257-264.

Heuven, V.J. van
1994a   'What is the smallest prosodic domain?', in: P. Keating (ed.) *Papers in Laboratory Phonology III: phonological structure and phonetic form*. London: Cambridge University Press, 76-98.
1994b   'Introducing prosodic phonetics', in: C. Odé & V.J. van Heuven (eds.) *Experimental studies of Indonesian prosody*. Semaian 9. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden, 1-26.

Ladd, D.R.
1980   *The structure of intonational meaning. Evidence from English*. Bloomington: Indiana University Press.

Lehiste, I.
1970   *Suprasegmentals*. Cambridge, Mass. and London: MIT Press.

Lehiste, I., J.P. Olive & L.A. Streeter
1976   'The role of duration in disambiguating syntactically ambiguous sentences', *Journal of the Acoustical Society of America* 60:1199-1202.

Odé, C.
1989   *Russian intonation: A perceptual description*. A.A. Barentsen et al. (eds.) Studies in Slavic and General Linguistics 13. Amsterdam: Rodopi.

O'Malley, M.H., D.R. Kloker & B. Dara-Abrams
1973   'Recovering parentheses from spoken algebraic expressions', *IEEE Transactions on Audio and Electroacoustics* AU-21:3

Pijper, J.R. de
1983   *Modelling British English intonation*. Dordrecht/Cinnaminson: Foris.

Samsuri
1971   *Tjiri-tjiri prosodi kalimat Bahasa Indonesia*. Malang: FKSS-IKIP Malang.

Proceedings of
the Seventh International Conference
on *Austronesian Linguistics*
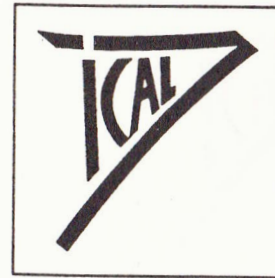
Leiden
22-27 August 1994

Cecilia Odé & Wim Stokhof
Editors