FORMAL AND FUNCTIONAL EVALUATION OF A MELODIC MODEL FOR STANDARD INDONESIAN

Ewald F. Ebing, Vincent J. van Heuven[1] & Cecilia Odé

Dept. Languages and Cultures of Southeast Asia and Oceania, Leiden University
[1]Dept. Linguistics/Phonetics Laboratory, Leiden University

ABSTRACT

A model of Indonesian intonation was perceptually evaluated using an improved testing methodology and listener selection. In a second experiment the focus and boundary marking functions of Indonesian intonation were investigated.

1. INTRODUCTION

A model for Standard Indonesian intonation has been developed following an analysis by synthesis methodology [1,2]. Successive versions of the model were perceptually evaluated by having native Indonesian listeners rate melodic versions of utterances (human originals versus model-generated contours, as well as *a priori* less adequate melodies, e.g time-shifted or Dutch contours) along a 10-point scale of formal melodic adequacy [3]. Listeners proved very insensitive to the melodic differences among the versions, so that we decided to re-run the evaluation with (hopefully) improved materials and more carefully selected listeners (section 2). It is difficult in Indonesian to distinguish between the accent-lending and boundary marking function of certain pitch movements [4]. In section 3, therefore, we examine how successfully Indonesian listeners can disambiguate arithmetic expressions with ambiguous focus distribution and internal bracketing.

2. FORMAL EVALUATION

Stimuli were taken from our corpus of quasi-spontaneous monologue by an educated speaker of Indonesian from Riau (East Sumatra) also used in our earlier experiments [2,3]. The stimuli comprised two tokens of the eight perceptually relevant pitch configurations found in our previous experiments. Four melodic versions of each configuration were produced by manipulating F_0 in the resynthesis (for procedural details see [3,6]:

a. *Close-copy* stylizations (COPY) of human originals; these should receive the highest ratings.
b. *Standardized* versions (STAN), i.e. generated according to our model; these should be (almost) as acceptable as COPY.
c. *Dutch*-based versions (DUTCH), generated according to the Dutch intonation grammar [1,3]; these versions should be rated as less acceptable than a or b.
d. *Mirrored* versions (MIRROR). Close-copies were mirrored along the frequency axis: rises became falls and vice versa; these versions should receive low ratings (as c).

The target configurations were now presented in their original contexts (rather than in isolation). To direct the listeners' attention to the relevant pitch configuration, the resynthesized context, but not the target configuration, was voiceless (whispered) throughout. This resulted in 64 stimulus types, each presented twice, yielding 128 judgments per listener.

The experiment was run at Universitas Islam Riau in Pekanbaru with 25 university students. Seventeen spoke Riau Malay as their first language, others had a different mother tongue, e.g. Minangkabau. Listeners rated each utterance along a 10-point scale of melodic adequacy (1: extremely poor; 10: excellent).

The results are summarized in Figure 1. The ordering of the acceptability ratings for the entire group of listeners is as predicted. No difference was found between COPY and STAN, t(783)=.149, ins., nor between STAN and DUTCH, t(777)=1.4, ins. However, the COPY ver-

sions were rated as significantly better than the DUTCH-versions, t(779)=3.12, p<.01. The MIRROR versions were rated as poorer than all other versions. Unexpectedly, STAN and DUTCH versions still do not differ significantly.

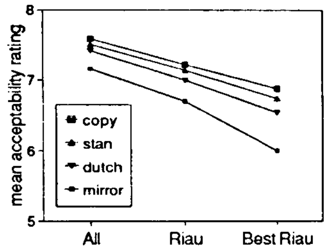

Figure 1. Acceptability of four melodic versions of Indonesian utterances broken down by listener selection.

We decided to enhance the effects by selecting only listeners with (i) the same variety of Indonesian as the speaker of the stimuli, and (ii) who were optimally sensitive to melodic differences.

First, the analysis was repeated for the 17 Riau listeners only. This time COPY and STAN versions do not differ from each other, t(538)=1.0, ins. but STAN and DUTCH do, t(532)=1.7, p<.05 (one-tailed). DUTCH and MIRROR versions differ as before, t(533)= 3.5, p<.01.

As the most sensitive listeners, only those eight Riau listeners were selected who obtained F>1 for the melodic version as a factor in listener-individual ANOVA's. Acceptability ratings are now better differentiated, while retaining the same ordering between conditions. These results show that the standardized pitch movements are perceptually adequate alternatives for close-copy stylizations. Moreover, to the Riau listeners, model-generated contours prove more acceptable than Dutch-based approximations. This confirms our hypothesis that the phonetic properties of the building blocks of Indonesian intonation are indeed language specific. Since the mirrored versions were included as a baseline condition, it is not surprising that they turn out to be the least acceptable. The fact that pitch contours that have been distorted in this manner are still rated in the upper half of the scale, is puzzling. Compared with results of similar experiments on English intonation [7], Indonesian listeners are remarkably tolerant towards deviations.

Finally, the difference between the whole group and the selected listeners suggests that regional and linguistic background does play a role: Riau listeners are more critical and discriminative.

3. ACCENTS AND BOUNDARIES

The aim of our second experiment was to find out to what extent accentuation and boundary marking can be (independently) expressed by means of the pitch movements in our model.

Focus distribution was manipulated by applying metalinguistic contrasts [5,6]. In the same set of test utterances, we also varied the position of a prosodic boundary by forcing the speaker to disambiguate a potentially ambiguous arithmetic expression (cf. e.g. [8]).

A single male native speaker of Indonesian produced eight versions of the same word sequence *dua kali tiga tambah lima*, orthogonally varying the position of the phrase boundary: *2x(3+5)* versus *(2x3)+5*, and focus structure:
(1) narrow focus on the first numeral
(2) narrow focus on the second numeral
(3) narrow focus on the third numeral
Each sentence was prompted by a question sentence to provide a context where one word was placed in focus. By manipulating F_0, model-generated contours were made for each realization.

The 25 subjects mentioned above indicated where they thought the speaker had intended the internal bracket of the expression to be, and - in a second part - which one of the three numerals in each phrase carried the strongest accent.

Table I specifies the percentage of accent responses for each of the three relevant numerals broken down by intended focus condition, and by intended phrase boundary position, first for the model-generated pitch contours (A) and then for the human originals (B).

Table I. Perceived accents (%) for focus on 1st, 2nd and 3rd numeral, broken down by boundary position; (A) human originals, (B) model-generated contours.

A. Human	boundary after num. #1			boundary after num. #2			Δ due to boundary after	
focus on num.	acc perceived on num. #1	#2	#3	#1	#2	#3	#1	#2
#1	82	9	9	55	37	8	27	28
#2	16	80	4	6	93	1	10	13
#3	49	28	23	24	60	17	25	32
Mean	49	39	12	28	63	9	21	24

B. Model	boundary after num. #1			boundary after num. #2			Δ due to boundary after	
focus on num.	acc perceived on num. #1	#2	#3	#1	#2	#3	#1	#2
#1	45	25	35	48	40	12	-3	15
#2	18	54	28	19	62	19	-1	9
#3	25	22	53	16	43	41	9	21
Mean	29	34	37	27	49	24	2	15

In the human originals, accents on the first and second numerals are mostly correctly perceived, although the percentages are lower than we expected, and quite probably lower than what would be obtained with speakers and listeners of English or Dutch. Perception of an accent on the third numeral is strongly disfavoured. Crucially, there is a clear effect of the position of the internal boundary on accent perception: chances of perceiving an accent increase immediately before a phrase boundary. This effect is stronger when focus is on the first syllable than on the second.

For the model-generated contours, the same effects and interactions exist but in a weaker form. When the boundary is after the first numeral, the majority of accents is perceived on the syllables where they were generated, for all three positions: bias disfavouring the third numeral has disappeared. When the boundary is after the second numeral, some bias against perceiving accent on the third numeral remains, but it is clearly weaker than in the human originals. Apparently, our human speaker pronounced very clear accent-lending pitch movements on the first and second, but not on the third numeral. Our model-generated accents were identical for each numeral position, i.e. smaller than the human accents on the first two numerals, but larger than the human accent on the third numeral.

Again, there is an effect of boundary position on accent perception. This time, however, the effect is strongly asymmetrical: a boundary after the second numeral attracts many perceived accents onto the second numeral, but there is virtually no migration of accents to the first numeral when the boundary is after this numeral.

Table II specifies percent boundaries perceived after the first versus second numerals for the human originals (A) and the model-generated contours (B), broken down by intended phrase boundary position and intended focus condition.

Table II: Correctly perceived phrase boundaries (%) broken down by intended boundary position and focus distribution (A) human originals, (B) model-generated contours.

A. Human	boundary correctly perceived after numeral #1	numeral #2	Δ
focus on num.			
#1	47	64	17
#2	27	79	52
#3	41	77	37
Mean	38	73	35

B. Model	boundary correctly perceived after numeral #1	numeral #2	Δ
focus on num.			
#1	49	69	20
#2	34	66	32
#3	41	76	35
Mean	41	70	29

There is a very strong effect, both for human and for model-generated contours, for more (twice as many) boundaries to be perceived after the second numeral than after the first. It is unclear at this time to what extent this is a stimulus effect. A stimulus analysis (not presented) shows clear differences in duration structure as a function of intended boundary position, but the duration effects are in fact stronger for the first numeral than for the second. Therefore, it seems that the effect is due to linguistic expectancy.

There is a smaller effect, both in human and in model contours, to perceive

(10 percent) more boundaries after the first numeral when it is accented, and (10 percent) fewer when the accent is on the second numeral. In human contours there is a complementary effect to perceive fewer boundaries after the second numeral when the accent is on the first, and to perceive more boundaries after a second accented numeral; in the model contours, however, this interaction between accent and boundary position for the second numeral is no longer found.

From the above we conclude that the perception of accentuation and melodic boundary marking are intertwined. Boundaries are more likely to be perceived after accented words, and accents are more likely in pre-boundary position.

Identification of accents and prosodic phrase boundaries is only partly successful, both with human and model-generated pitch contours. However, asymmetries are stronger for the human originals. This may be due to the fact that the pitch movements used by the human speaker show large differences in excursion size as opposed to the standardized movements used in the model.

4. CONCLUSIONS

The formal evaluation of the proposed intonation model has shown that the pitch contours produced by the model are acceptable substitutes for (close-copy stylizations of) the originals. The functional evaluation allows to important conclusions to be drawn:

Firstly, is seems indeed true that the accent and boundary-marking functions are strongly intertwined in Indonesian; nevertheless, listeners were able, much better than at chance level, to distinguish between the functions. It is unclear at this moment whether this degree of interdependence is unusual. We know of no similar experiments, i.e. varying both focus and boundary positions, in other languages, so that we have no basis for comparison. Cross-linguistic experiments are essential for placing the performance of the Indonesian listeners, with both human and model-generated contours, in their proper perspective.

Secondly, formal evaluation of a melodic model (based on quality judgments) in itself is insufficient: it has to be complemented by a functional assessment of melodic adequacy.

ACKNOWLEDGMENT

Research supported by the Netherlands Organisation for Research through the Foundation for Language, Speech & Logic (project # 300-172-018).

5. REFERENCES

[1] Ebing, E.F. (1991). "A preliminary description of pitch accents in Bahasa Indonesia", in: *Proc. 12th Int. Con. Phon. Sc.*, Aix-en-Provence, pp. 258-261.

[2] Ebing, E.F. (1994). "Towards an inventory of perceptually relevant pitch movements for Indonesian", in: C. Odé and V.J. van Heuven (eds.), *Phonetic studies of Indonesian prosody*, Semaian 9, Vakgroep TC Zuidoost-Azië en Oceanië, RU Leiden, pp. 181-210.

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

[4] Ebing, E.F. and Heuven, V.J. van (1994). "Some formal and functional aspects of Indonesian intonation", in: *Proc. 7th Int. Con. Austronesian Ling.*, Leiden (in press).

[5] 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), pp. 76-98.

[6] Heuven, V.J. van, (1994b) "Introducing prosodic phonetics", in: C. Odé, V.J. van Heuven, eds. *Phonetic studies of Indonesian prosody*, Semaian, 9, Leiden (Vakgroep TC Zuidoost-Azië en Oceanië, RU Leiden), pp. 1-26.

[7] Pijper, J.-R. de (1983). *Modelling British English intonation*, Foris, Dordrecht.

[8] Lehiste, I., Olive, J.P. and Streeter, L.A. (1976). "Role of duration in disambiguation syntactically ambiguous sentences", *J. Acoust. Soc. Am.*, 60, pp. 1199-1202.