

**Phonetic Experiments on
the Word and Sentence Prosody of
Betawi Malay and Toba Batak**

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6000
e-mail: lot@let.uu.nl
<http://www.lot.let.uu.nl/>

Cover illustration: Fragment of a Yogyakarta batik, collection of the KITLV (37B-98), Leiden, The Netherlands.

ISBN-10: 90-76864-98-5
ISBN-13: 978-90-76864-98-3

NUR 632

Copyright © 2006: Lilie M. Roosman. All rights reserved.

**Phonetic Experiments on
the Word and Sentence Prosody of
Betawi Malay and Toba Batak**

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 26 april 2006
klokke 16.15 uur

door

LILIE MUNDALIFAH ROOSMAN

geboren te Jakarta, Indonesië
in 1964

Promotiecommissie

promotor: prof. dr. V.J.J.P. van Heuven
co-promotor: dr. E.A. van Zanten
referent: prof. dr. H. Steinhauer
overige leden: dr. J. Caspers
prof. em. dr. A.M. Moeliono, Universitas Indonesia,
Depok
prof. dr. W.A.L. Stokhof

This research was financially supported (12-months stay in the ULCL phonetics laboratory plus stipend at Universitas Indonesia) by a grant from the Royal Netherlands Academy of Arts and Sciences (KNAW) under program number 95-CS-05 (principal investigators W.A.L. Stokhof and V.J. van Heuven), by the Nederlandse Taalunie (ten-months stay at the ULCL phonetics laboratory), and by the International Institute of Asian Studies (IIAS, travel grant to attend the ISMIL-7 conference).

Contents

Acknowledgements	xi
Chapter I: General Introduction	
1.1 Prosody	1
1.2 Object languages	3
1.3 Prosody and foreign accent	3
1.4 Strategy	5
1.5 Outline of this thesis	6
Chapter II: Background	
2.1 Prosody	9
2.1.1 Definition of prosody	9
2.1.2 Functions of prosody	12
2.1.3 Prosodic domains	15
2.1.4 Focus	16
2.1.5 Intonation	18
2.1.6 Phonetic correlates of stress and accent	22
2.2 Production and perception of L2-word prosody	25
2.3 Language background	
2.3.1 Betawi Malay	26
2.3.2 Toba Batak	31
Chapter III: Temporal and melodic structures in Toba Batak and Betawi Malay word prosody	
3.1 Introduction	35
3.2 Background	36
3.3 Methods	37

3.3.1	Selection of speech materials	37
3.3.2	Speakers and recording procedure	38
3.4	Duration	40
3.4.1	Toba Batak	40
3.4.1.1	Word duration	40
3.4.1.2	Syllable duration	41
3.4.1.3	Segment duration	43
3.4.2	Betawi Malay	46
3.4.2.1	Word duration	46
3.4.2.2	Syllable duration	47
3.4.2.3	Segment duration	49
3.4.3	Conclusion	51
3.5	Pitch analyses	53
3.5.1	Toba Batak	53
3.5.1.1	Stylization	53
3.5.1.2	Results	56
3.5.2	Betawi Malay	59
3.5.2.1	Stylization	60
3.5.2.2	Auditory inspection	61
3.5.2.3	Token frequencies of BM accent-lending pitch movement types	66
3.5.2.4	Acoustical properties of BM accent-lending pitch configurations	69
3.5.2.5	Pitch in [-focus] BM targets	76
3.5.2.6	Pitch accent in Betawi Malay	78
3.5.3	Melodic structures of Toba Batak and Betawi Malay	79
3.6	Conclusion	80
Chapter IV: Non-native accents in Dutch word-stress realization		
4.1	Introduction	85
4.2	Background	86

4.3	Method	87
4.3.1	Preparation of stimulus materials	88
4.3.2	Speakers	89
4.3.3	Recordings	90
4.3.4	Manipulations	91
4.3.5	Procedure	92
4.4	Experiment 1: Evaluation by Dutch listeners	93
4.4.1	Subjects and procedures	93
4.4.2	Results and discussion	94
4.5	Experiment 2: Identification by Dutch listeners	96
4.5.1	Subjects and procedure	97
4.5.2	Results and discussion	97
4.6	Experiment 3: Identification by non-Dutch listeners	99
4.6.1	Subjects and procedure	99
4.6.2	Results and discussion	100
4.6.3	Toba Batak listeners	101
4.6.4	Betawi Malay listeners	103
4.7	Conclusion	104

Chapter V: Acoustical analysis of Dutch word stress as spoken by Dutch, Toba-Batak and Betawi-Malay speakers

5.1	Introduction	107
5.2	Acoustical analysis	108
5.2.1	Temporal structures of non-native Dutch speech	109
5.2.2	Melodic structures of non-native Dutch speech	115
5.3	Correlation between perception and production	123
5.3.1	Temporal perception vs. temporal parameters	123
5.3.2	Melodic perception vs. melodic parameters	125
5.3.3	The contribution of prosodic information in the perception of originals	128

Chapter VI: General Discussion	
6.1 Summary and discussion	135
6.2 Suggestions for further research	140
References	143
Appendices	151
Summary in English	161
Ringkasan (Summary in Indonesian)	167
Curriculum vitae	171

Acknowledgements

Alhamdulillah, this dissertation could not have been finished without the help of many people in the Netherlands and in Indonesia.

Much as I would have liked to express my sincere appreciation to the Leiden University Centre for Linguistics (LUCL, formerly ULCL) staff members who helped me in so many ways – unfortunately Leiden *adat* prohibits me from doing so.

I thank Dr Myrna Laksman, who introduced me to phonetics and taught me much about prosody.

I wish to express my gratitude to Professor Anton Moeliono for his valuable comments and suggestions that made me more confident as *Anak Betawi*.

I am also grateful to Dr Kees Groeneboer, Nederlandse Taalunie advisor at the Dutch Department, Universitas Indonesia, for his continuous support and his efforts to secure scholarships for me.

Thanks are due to the whole ‘family’ of the Phonetics Laboratory of Leiden University, who provided me with such a warm and pleasant environment to work in: Hongyan (my best room mate), Maarten, Rob, Ellen, Jos, Johanneke, Vincent, Jie, Gijs, Josée, you truly made me feel at home in the ‘Lab’.

I also thank the Dutch Department, Fakultas Ilmu Pengetahuan Budaya, Universitas Indonesia, for granting me leave of absence many times to pursue my research activities abroad, and for their encouragement. I am most grateful to Lilie Suratminto, SS MA, Yati Suhardi, SS and Drs Eliza Gustinelly. Thanks also to the Erasmus Taalcentrum in Jakarta for their help with travel documents.

Thanks to all speakers and listeners who participated in the experiments. My Dutch speakers were also most helpful in their other capacities.

Many thanks to Drs Nurhayu Santoso for translating the Summary into Indonesian. I also thank the other PhD researchers in this project, Ruben, Rahyono, Sugiyono and Bert, for their help and friendship.

I am especially grateful to *Mbak* Susi and Marrik Bellen. I was always welcome to stay at their home for many months and over many years, and they made my stays in Leiden/Leiderdorp into a very pleasant experience. They always stimulated me with their warmth and care in all my efforts.

Finally, I thank my family, Mama, Yanto & Selmi, and Imam, for their faithful love, prayers and support.

Jakarta, March 2006

Chapter I

General Introduction

1.1 Prosody

Phonetics has often been called the science of speech sounds. For a long time this definition has been taken literally to indicate that human speech should be seen as a string of sounds each of which can, and should, be described in great detail. Speech, or spoken language in general, is thus seen as an analogy to an alphabetically written text, where syllables, words, phrases and sentences can ultimately be reduced to a sequence of letters. On second thoughts, however, it will be obvious that there is much more to spoken language than just a string of letter-like sounds. Prosody is the general term covering all the formal characteristics of spoken language that cannot be traced back to the simple sequence of sounds (for a more detailed definition see chapter II). Practically, prosody is everything that has to do with the melody and rhythm of spoken language. Melody and rhythm are no properties of individual sounds; only larger linguistic units such as words, phrases, sentences and even paragraphs may have characteristic melodies and rhythmic structures.

Typically, alphabetic writing systems faithfully reflect the string of basic sounds that make up the words in a sentence. Every character or fixed combination of successive letters corresponds to a speech sound, i.e. a vowel or a consonant. Writing systems only very crudely specify the melodic and rhythmic properties of words and phrases. Stressed syllables and accented words are not identified; the melody of speech is indicated at best by symbols such as period, exclamation mark and question mark. Normally, however, there are many different melodies a speaker may choose from in order to mark a sentence as a question, an exclamation or a statement. Each of these melodies adds special meanings to the sequence of words, and yet this melodic information is not expressed in the writing system. It would

seem, therefore, that writing systems are driven by one simple goal, which is to allow the reader to recognize the words on the page. Implicit in this choice is that, normally, identifying the sequence of letters (sounds) is all the reader needs to recognize the word, and recognizing the sequence of words provides sufficient information to understand the sentence. There is a good deal of truth in this view, which is, in fact, corroborated by existing practice in speech-technology products. Commercially available software for automatic speech recognition and speech understanding (such as found in dictation machines and spoken-dialog systems) only use segmental information; melody and rhythm are ignored as these add little or nothing to the performance of the machine.

In the last two to three decades it has become increasingly clear, however, that prosody provides important information to the listener. It helps the listener break up the continuous stream of sounds into smaller chunks that can be readily processed; it identifies the important syllables and words within the chunks, and often provides subtle information on the speaker's intentions, attitudes (towards the verbal contents of the sentence and/or towards the listener) and emotions at the time of speaking. In chapter II I will review these functions of prosody in greater detail.

It will be immediately obvious that languages differ enormously in their inventory of sounds. Some languages have many different sounds; others have only a small set of sounds. Germanic languages, for instance, have some 15 to 25 different vowel sounds, whereas Spanish, Greek, and Indonesian have between five and ten different vowel sounds. And even if two languages have the same number of sounds, it will never be the case that the counterpart sounds in the two languages are exactly the same. The present thesis investigates the extent to which two languages may differ not so much in their segmental structure (inventory of vowels and consonants and their combinatory possibilities) but in terms of the melodic and rhythmic properties. Specifically, I will study two related languages spoken in the Indonesian archipelago and try to establish characteristic differences between these two languages in the way words in a sentence are presented as important to the listener through melodic and rhythmic means.

1.2 Object languages

The subject of this study is word prosody of two regional languages of Indonesia, viz. Toba Batak and Betawi Malay. These two languages differ crucially in that Toba Batak has word stress (van der Tuuk, 1971 [1864]; Nababan, 1981), and Betawi Malay does not (Muhadjir, 1977). I will test the hypothesis that, when speaking a stress language, word stress will be marked more clearly by the speaker, i.e. by larger differences between stressed and unstressed syllables, if stress in the speaker's native language may be used to differentiate between words than in a language which does not have stress. I predict, then, that a Toba-Batak speaker will mark the difference between stressed and unstressed syllables in Dutch more clearly than a Betawi-Malay speaker. The differences should be apparent in the melody on the word in its sentence, and/or in the way the speaker speeds up unimportant (unstressed) syllables and stretches important (stressed) ones.

A more detailed literature survey of the two target languages – with emphasis on their melodic and temporal structure – will be given in chapter II.

1.3 Prosody and foreign accent

Any normal child will learn to speak the language of its caregivers within roughly the first four years of its life. Native-language acquisition during childhood proceeds with apparent ease. It requires no explicit instruction, and even though the input to the child is often incorrect, the result of the acquisition process is perfect. Children will learn to speak their language with a perfect pronunciation and perfect command of grammatical rules.

For reasons that are largely still unknown, this ability to learn a language perfectly diminishes with age. Adults who have learnt to speak a second (or third, fourth, etc., also abbreviated as L2, L3, L4, etc.) language after the age of 20, can nearly always be recognized by native listeners of the target language as non-natives, as foreigners. Their speech has audible properties that deviate from the

implicit norms the target-language community has for the pronunciation of its vowels and consonants. Moreover, the deviations from the native norm are not random but are inspired by the source language (the L1) of the learner. Typically, the learner uses the sounds of his mother tongue, L1, as substitutes for the sounds of the target language, L2. With training, proper instruction and feedback the foreign-language learner may ‘unlearn’ the pronunciation habits of his mother tongue, and acquire the norms for the sounds in the new language. This learning process is often incomplete and the ultimate level attained by the learners may differ widely.

Similarly, when someone has learnt to speak a foreign language after the age of puberty, his (or her) spoken language will have the prosodic properties, i.e. the melodies and rhythmical patters, of the speaker’s mother tongue. Moreover, there is a persistent claim in the literature that learning the prosody of a new language, especially its intonation (speech melody), is even more difficult than learning the correct pronunciation of the vowels and consonants. This is clearly shown in the following quotation from a recent textbook on the learning and teaching of foreign languages:

Intonation [...] is an important aspect of language that seems to be easily, if not automatically, acquired by children in both L1 and L2. Moreover, as observation and experience amply demonstrate, it is easy for adults to maintain and retain in the L1, yet difficult, if not impossible, for adults to learn in an L2. (Chun, 2002:xiii)

I predict, accordingly, that native speakers of Toba Batak and Betawi Malay will have great problems when having to learn the prosody of Dutch as a foreign language. Melodic and temporal structure will still be characteristic of their L1. Dutch is, like Toba Batak, a language that uses stress contrastively and for which the prediction would be that the difference between stressed and unstressed syllables is relatively large. Given the typological similarity between Dutch and Toba Batak, which is lacking in the comparison between Dutch and Betawi Malay, I expect

Toba-Batak learners to have an edge when learning Dutch prosody over learners with a Betawi-Malay background (or Standard Indonesian, for that matter).

1.4 Strategy

This study focuses on the realization of word prosody of Toba Batak and Betawi Malay, in particular the effects of prominence and pre-boundary position of a target word on its temporal and melodic structure. Durations and pitch configurations will be investigated in four types of carrier sentence, in order to create four prominence and boundary conditions such that the same word (string of vowels and consonants) is either presented by the speaker as important ('in focus') or not important ('out of focus') in the discourse, and either occurs in the middle of a phrase or at the end of it, in all four logically possible combinations of focus and boundary position.

Comparing the phonetic correlates of word prosody in a stress language with those of a non-stress language will be more sensible when the evaluation is based on identical segmental structures. In this study, native speakers of Toba Batak and Betawi Malay will therefore not only produce speech in their own language (with different numbers of vowels, and different pronunciation norms) but also in Dutch, which, of course, is a foreign language for both groups of speakers. The results will be evaluated in perception experiments not only by native listeners of Toba Batak and Betawi Malay but also by native listeners of Dutch. In these tests the materials will be presented to listeners in different conditions such that segmental and/or tonal information is eliminated from the stimuli. This elimination technique will allow us to determine the relative contribution of each source of information (segmental pronunciation, speech melody, temporal structure) to the quality of the stress pattern, and to the identification of the speaker's native-language background.

It is scientifically interesting and useful for teaching purposes, to investigate whether speakers of a stress language realise word stress in another stress language (in this case Dutch) more faithfully than speakers of a non-stress language. Three perception experiments were run to investigate how well native speakers of Toba

Batak (stress language) and Betawi Malay (non-stress language) realise Dutch stress. Acoustical analyses of the stimuli that were used in the perception experiments will complete this study. This study will measure the acoustical parameters of Betawi-Malay Dutch and Toba-Batak Dutch, compared to the parameters of native Dutch. Through acoustical measurements I expect to find out in what acoustical aspects the stress/accents realisations are different from each other.

1.5 Outline of this thesis

The general question, then, of this thesis is how speakers of a non-stress language differ from speakers of a stress language in their realisation of stress and/or accent.

Following the present brief introductory chapter, chapter II will give a literature survey of current thinking on prosody at the word and sentence level. After that a description of Betawi Malay and Toba Batak will be given, with special emphasis on previous studies on the prosody of these two languages.

Chapter III describes two production experiments on Betawi Malay and Toba Batak set up to investigate the effects of boundary and prominence on two sets of prosodic parameters, viz. the duration and the fundamental frequency, in both languages. The research aims to give answers to the following questions:

1. What are the effects of sentence boundary (sentence-final versus non-final) and prominence (focus versus non-focus) on the word duration in both languages?
2. What are the effects of sentence boundary and prominence on the duration of the segments and how is the lengthening distributed over the domain (syllables, words)?
3. What are the effects of sentence boundary and prominence on the pitch contours in both languages?

I expect to find similar effects of boundary marking on the duration in both languages. As regards prominence, however, I expect stronger effects in the stress language Toba Batak (especially in the stressed syllable) than in non-stress Betawi Malay. These effects should be apparent when studying the acoustic realisation of the targets in the respective languages, Toba Batak and Betawi Malay. The effects should also be found, and possibly even more clearly, when both groups of speakers produce identical segment strings in a foreign language, viz. Dutch.

Rather than measuring acoustical correlates of (stressed) syllables (at the word level) and/or accented words (at the sentence level), Chapter IV will determine the extent to which the native-language (L1) background (Betawi Malay or Toba Batak) is audible in the production of L2 Dutch. Speakers of Betawi Malay and Toba Batak produced Dutch utterances. These utterances were used as stimuli in three perception experiments which were run to investigate how strongly native speakers of Toba Batak and Betawi Malay are influenced by the prosody of their native language when they speak Dutch, and whether they are sensitive to the prosodic differences in Dutch.

- (i) The first perception experiment involves Dutch native listeners evaluating the realization of Dutch word stress spoken by Toba Batak and Betawi Malay speakers, as well as by Dutch speakers.
- (ii) The second experiment aims to find out whether, and with what (prosodic) cues, Dutch listeners are able to differentiate non-Dutch speakers from Dutch speakers.
- (iii) The last experiment involves Toba Batak and Betawi Malay listeners in an attempt to find out to what extent they are able to recognise Dutch-speaking Indonesians, on the basis of (deviant) stress realisation only.

I expect that native listeners of Dutch will rate the prosodic quality of the Toba-Batak speakers more favourably than that of Betawi-Malay speakers, at least in so far as the realisation of word stress is concerned.

The stimuli involved in the perception experiments will be acoustically analysed in chapter V. Duration and pitch will be measured. Based on these measurements comparisons between native and non-native speech, and between stress and non-stress language, will be made. It is expected that Betawi-Malay speakers deviate more from native Dutch stress realisation than Toba-Batak speakers.

Finally, a general discussion and some suggestions for further research will be presented in Chapter VI.

Chapter II

Background

2.1 Prosody

2.1.1 Definition of prosody

All human languages are characterised by a hierarchical structure such that smaller units are combined into larger units, which in turn constitute the building blocks from which yet larger units are composed. The smallest unit in spoken language is the segment or phoneme, i.e. a single vowel or a single consonant. The segment can be seen as equivalent to the molecule in physics. Although theoretical linguists have proposed even smaller units below the level of the phoneme, similar to the way molecules can be decomposed into atoms in physics, I will not take this step, nor do I have to within the scope of the present thesis.

The segments in a language have to be distinct from each other. Generally, languages have an inventory of some 15 to 75 basically different sounds or 'phonemes'. Each segment is characterized by a set of inherent properties. For instance, some segments are produced with vibrating vocal cords, others are not. Due to different places in the vocal tract where the outgoing flow of air is impeded through a narrowing of the air passage, sounds assume different acoustical properties or resonances, which lead to the perception of distinct phonetic qualities or 'timbre'. Some sounds – such as vowels – have a lot of carrying power, i.e. physical intensity; others do not (consonants). The articulatory and acoustical properties that define a particular segment are called its intrinsic properties. The properties of any speech sound can be decomposed into four subtypes, viz. its length, its loudness, its timbre and its pitch. Roughly speaking a segment's length (or 'quantity') is determined by the physical duration (measured in milliseconds, ms) of the articulatory movements. Loudness primarily corresponds with physical

intensity (conveniently expressed in decibels, dB) which in turn is caused by the force with which air is expelled through the vocal organs. Pitch corresponds to the repetition rate (in hertz, Hz, or cycles per second) of the vocal cords imparting periodicity to the speech sounds. Timbre (also called quality), finally, is brought about by shaping the spectrum of the sound through different resonances (amplification or attenuation of specific frequencies) which are caused by the speaker varying the shapes and sizes of the throat, mouth, opening or closing the air passage through the nose, etcetera.

It is generally taken for granted that the differences in timbre (supralaryngeal filter) and the specific combination of presence/absence of periodicity and noisiness (excitation signal) define the basic properties of a segment. The two sets of properties are represented in a one-to-one fashion in the classical view on the acoustics of speech production which has become known as the source-filter theory (Fant, 1961; Stevens, 1998). In this theory the excitation signal represents the source and the supralaryngeal configuration (shape of throat, mouth, lips, nasal cavity) makes up the filter. The remaining properties, i.e., pitch, intensity and duration, are typically presented in the literature as secondary features of segments. These secondary properties fall out as by-products of the primary features, and only in extreme laboratory conditions can manipulations of secondary features swing the perception of a sound's identity from one category to another. It is well known, for instance, that the degree of openness of a vowel (which affects the resonance of the throat cavity) also affects the vowel's duration, pitch and intensity. The more open the vowel, the longer it takes, *ceteris paribus*, to complete the articulatory gesture. Also, more open vowels have more intensity as the sound is radiated more efficiently from the lips when the mouth is shaped like a bullhorn (as for [a]) than when the sound is directed into a funnel (as for /i, u/). Finally, when the (vowel) sound is produced with a raised tongue posture the vocal cords are involuntarily stretched and tautened so that they vibrate more quickly, yielding higher pitch (see Rietveld and van Heuven, 2001, and references therein).

Not only do sounds have their own defining (primary and secondary) inherent properties, also is it the case that, in connected speech, neighbouring sounds influence each other in predictable ways. For instance, when the vowel has to be

produced with rounded (protruded) lips – such as [y, u] – then preceding and following consonants are likely to be produced with protruded lips as well. Normally, the shape of the lips is immaterial to the identity of consonants so that the speaker is free to initiate the rounding gesture required for the vowel well in advance, and to maintain the lip rounding for some time after the vowel. There is an enormous literature on the mutual influence of neighbouring sounds (for a summary see e.g. Farnetani, 1997). Properties of a sound segment which are predictable from properties of adjacent sounds are called co-intrinsic properties.

Although a large portion of the characteristics of speech can be adequately predicted from the intrinsic and co-intrinsic properties of the string of segments that make up an utterance, there is also a set of characteristics that cannot be derived from the underlying sequence of segments in a straightforward fashion. This ensemble of properties is called prosody.¹ Examples of such properties are the controlled modulation of the voice's pitch, the stretching and shrinking of segment and syllable durations, and the intentional fluctuations of overall loudness (Nooiteboom, 1997:640). Note that these correspond precisely to the secondary segmental features referred to above, viz., pitch, length and loudness. On the surface, then, it appears that there is a neat division of work such that source signal and timbre primarily make up the intrinsic and co-intrinsic properties of speech, while pitch, length and loudness primarily define prosody. Yet, it should be pointed out that such a strict division is unrealistic. In fact, each of the five properties mentioned may function at the level of the segments as well as prosodically. Even the most typical of all inherent segmental properties, the sound's phonetic quality, varies to some extent – in Russian, English and Dutch more than, for instance, in Spanish and Greek – under the influence of stress. Vowels in stressed syllables have a more extreme (or 'peripheral') quality, whereas their unstressed counterparts are centralised (spectrally reduced); they are articulated more closely to the neutral vowel schwa.

¹ The word prosody comes from ancient Greek, where it was used for a 'song sung with instrumental music'. In later times the word was used for the 'science of versification' and the 'laws of metre', governing the modulation of the human voice in reading poetry aloud (Nooiteboom 1997:640).

2.1.2 Functions of prosody

The segments that make up the inventory of basic building blocks in a language, are used to differentiate the words in the lexicon. Given that some languages have considerably smaller segment inventories than others (see above), it follows that – all else being equal – the former type of language tends to build long words whereas the latter type has relatively short words. Polynesian and Austronesian languages generally have small vowel inventories (three to seven distinct vowels) and a limited set of consonants; in so far as these languages do not employ tone (see below) as an extra means to contrast between lexical items, they have to create long words in order to come up with some 50,000 uniquely different segment strings to cover the lexicon. Conversely, languages with large segment inventories, such as the Germanic languages, meet their lexical needs with a huge array of monosyllabic word forms.

Of the speech parameters that primarily serve prosody, duration and pitch (but not loudness), may be used across languages to mark lexical contrasts, i.e. serve to differentiate words in the lexicon. Here duration is almost invariably a property of a single vowel or consonant, i.e. a segmental rather than prosodic phenomenon. Although the majority of the world's languages do not employ length contrasts (Ladefoged and Maddieson, 1996), quite a few differentiate between short and long vowels, or even short ~ long ~ superlong (Estonian, cf. Lehiste and Fox, 1992). Quantity oppositions involving consonants are rare and binary at best; the contrast is a matter of single versus geminate (double) consonants.²

Pitch is used in a lexically contrastive way in so-called tone languages. Typically, the domain of the lexical use of pitch is longer than a single vowel or consonant, and subtends the entire syllable or the voiced/sonorant part of it. This use of pitch is therefore truly prosodic. Mandarin (Chinese) is a good example of such a tone language. In principle, any syllable in Mandarin can be pronounced with four different word melodies (Yip, 2002), viz. high level (H), low rising (LH), low (L) and falling (HL), so that the basic inventory of seven vowels is effectively expanded

² In Scandinavian languages quantity may function at the co-intrinsic level. Long vowels are followed by a single consonant coda whereas short vowels can only be followed by geminate consonants (cf. van Leyden, 2004, and references therein).

to $4 \times 7 = 28$. Since my research will not target any tone languages, I will not go into the matter of lexically contrastive pitch any further.

Whereas the segments of the language are used to differentiate the words in the lexicon, it would seem that the primary function of prosody is another one. It is convenient to distinguish between prosodic functions at the word level and those at the level of the phrase and beyond. At the word level, prosody seems to be geared towards facilitating word recognition. The initial and final segments of words tend to be realised in a way that is different than word-medial segments. Segmental enhancement at the word edges is unpredictable from the mere sequence of sounds that make up the utterance. One has to know that a word boundary intervenes between two successive segments in order to be able to predict the segmental enhancement at the word edge (see Keating, 1994 for details). Also, the temporal organisation of the segments within a word can be seen as an overall characteristic of the larger unit. Generally, the last vowel-plus-coda of a word is lengthened (word-final lengthening), and individual segments are spoken faster as the word is longer (i.e. contains more segments, cf. Nootboom, 1997:656-658). Languages can be subdivided into three word-prosodic categories, viz. tone languages and stress languages, and languages which have neither tone nor stress. In an (idealized) simple tone language (see above) with just two word-tone levels H (high) and L (low), every syllable in the word can be pronounced with H and L, yielding four disyllabic tone words: HH, HL, LH and LL. It is not the case that the syllable bearing the H tone is in any way stronger or more basic to the identity of the word than a syllable carrying an L tone. This type of word-prosodic system is fundamentally different from a stress system. In a language with word stress, one syllable within a (polysyllabic) word is felt to be stronger, more basic to the word's identity, than the other syllables. The dominant syllable is called the 'prosodic head' at the word level, or simply 'the stress'. Stress is a culminative property (Trubetskoy, 1969[1939]; Garde, 1968), i.e., only one syllable can be the strongest in the word, and – more generally – for any prosodic domain, there can be only one prosodic head. Although stress may sometimes be used to mark lexical contrasts, like in so-called minimal stress pairs as English *forebear* 'ancestor' ~ *forbear* 'endure' (stressed syllables underlined) or in *import* (noun) ~ *import* (verb), such minimal

pairs are comparatively rare in English, and in fact, stress is not used systematically to mark lexical contrasts in any language. Rather it seems that stress serves as an aid to the listener who has to break up connected speech into a sequence of individual words. If all the words in the language have the stress in the same position (e.g. all Hungarian words have stress on the first syllable, virtually all Polish words have stress on the penultimate), the stress signals to the listener that a new word has just begun (Hungarian), or will begin after the next syllable (Polish). In less predictable systems, stress at least serves as a word counter. Every time a stress is heard, the listener will know that another word has gone by even though the exact location of the boundary between the successive words yet remains to be determined. For a recent survey of the possible roles of stress for the process of word recognition see Cutler and van Donselaar (2001).

At the higher levels of the prosodic hierarchy, such as the phrase, sentence, and even paragraph levels, prosody functions as a guide to the parsing of continuous speech into chunks of information that can readily be processed (boundary marking), marking the clause type of the chunk (statement, question, command, exclamation, non-final part of a larger array of chunks), the highlighting of important information within the chunks (attentional marking, Gussenhoven, 1984; Hirschberg and Pierrehumbert, 1986), and the expression of the speaker's intentions and status of referents in the discourse (intensional marking, Grosz and Sidner, 1986, 1998). Prosody may also contribute to the expression of paralinguistic information such as the attitude of the speaker towards the hearer or the verbal contents of the message (e.g. sincerity, irony, sarcasm) and emotion (e.g. fear, happiness, sadness, joy, cf. van Bezooijen, 1984; Mozziconacci, 1998, and references therein).

In the present thesis I will not be concerned with the latter three functions of sentence prosody, i.e. the marking of intention, attitude and emotion. Rather I will concentrate on the boundary-marking and attentional function of sentence prosody and its interaction with word-level prosody, specifically with the effects of word stress (or its non-existence, depending on the target language) on the temporal and melodic marking of focus domains and prosodic heads within the domain.

2.1.3 Prosodic domains

It has been widely acknowledged that the hierarchical structure of language extends in two modes, viz. the morpho-syntactic mode as opposed to the phonological mode (e.g. Nespor and Vogel, 1986). The morpho-syntactic structure is concerned with units that carry meaning, i.e. morphemes, and larger structures built upon them. Phonological structure is not based on meaningful units but is defined exclusively on audible aspects of sound structure. Although the two sets of structure are often isomorphic, they are not necessarily so, and in fact, diverge in crucial cases. Morpho-syntactically the compound *blackbird* is composed of two morphemes *black* and *bird* which together make up a new, longer word. Within the compound the morpho-syntactic head is *bird*; this is the unit that determines the part of speech of the compound, viz. a noun, and it also expresses that the compound refers to a particular kind of bird rather than a kind of colour. The element *black* is the dependent; it does not determine the part of speech of the compound, and merely qualifies the meaning of the head: it is a bird which happens to be black. Phonologically the compound is a phonological word (Pw), comprising two smaller units, viz. the (mono-syllabic) ‘feet’ *black* and *bird*. However, the head of the prosodic word is *black* and the dependent is *bird*. In the spoken version of the compound *black* carries the stress, i.e. is pronounced more forcefully, and felt to be stronger by the native English listener, than the second element *bird*. So, even though the division of the compound into smaller units is the same in the morpho-syntactic and phonological hierarchies, the position of the heads and dependents differ crucially. At a higher level of linguistic structure a sentence like *John felt a sharp pain* is analysed into its two basic morpho-syntactic constituents *John* (the NP embodying the subject of the sentence) and *felt a sharp pain*, the VP expressing the predicate. Prosodically, however, the primary cut is between *John felt*, which is not even a proper morpho-syntactic constituent (as *felt* is a necessarily transitive verb) and *a sharp pain*. The chunking of larger utterances into smaller units typically uses prosodically motivated constituent boundaries. In our study I will deal with chunks (or ‘prosodic domains’) at two levels, the Intonational Phrase (I) and the next-higher domain, the Utterance (U). Both domains are bounded by prosodic breaks that are

marked temporally and melodically. The boundaries are optionally signalled by the presence of a pause, a silent interval between 200 ms for an I-boundary and 500 ms for a U-boundary (cf. Klatt, 1985). Whether or not the boundary is signalled by a pause, the segments immediately preceding the boundary are stretched by up to 50%. This temporal expansion of segments is greater as the segment is closer to the boundary. Also, the stretching is more pronounced before a U-boundary than before an I-boundary (see Cambier-Langeveld, 2000 for Dutch). Finally, prosodic boundaries are often signalled by boundary tones, such as the presence of an H% target associated with an utterance-medial I-boundary and an L% target at the end of a U-domain (for an explanation of the H% and L% symbols see section 2.1.5). The H% target corresponds with a rise in pitch before the I-boundary followed by a lower pitch after the break; the L% target is the lowest pitch in the utterance, and is followed by a higher pitch at the onset of the next utterance.³

2.1.4 Focus

Focus is a semantic notion which refers to the relative status of constituents in a spoken sentence. Certain words (or larger or smaller morpho-syntactic units) are said to be ‘in focus’ or [+F] if the speaker wants to instruct the hearer to consider these units as communicatively important: the speaker wishes to focus the hearer’s attention on these units. Any materials that are not presented in focus are called ‘out of focus’ or [-F]. The reasons for a speaker to focus a constituent are manifold. Very often is it because the constituent introduces a new referent into the discourse (‘new’ information). Alternatively, a constituent is worthy of focus because the speaker chooses between two (or more) known but contrasted referents, as in:

Q. Would you like coffee or tea?

A. I prefer [tea]_{+F}

³ When the utterance is a question, the U-boundary is often an H% in languages such as English and Dutch. High pitch at the end of a domain is thought to signal ‘appeal’ by the speaker to the listener, viz. either a request for continued attention (‘I have not finished yet, please hear me out’) or to provide an answer or some non-verbal compliance to a request (Caspers, 1998; van Heuven and Kirsner, 2004).

Typically, [-F] materials involve those parts of the sentence that contain referents and concepts that were introduced into the discourse in the preceding context ('old' information).

In the present study I will manipulate the focal status of constituents such that the same segmental materials (words) will be spoken once in focus and a second time out of focus. The reason for this manipulation is that focus is likely to be marked through prosodic means; this is what was referred to earlier as the attention-marking function of prosody. In most researched languages focus is signalled both by temporal and by melodic means. In West-Germanic languages (English, German, Dutch) the speaker produces a perceptually prominent change in pitch on the prosodic head of the [+F] constituent, that is, on the stressed syllable of the most important word in the constituent. When the word is not in focus, such a pitch change is absent. The prominence-lending pitch configuration is called an 'accent', more specifically a 'pitch accent' or 'focal accent' (cf. Bolinger, 1958). Temporally, a focussed constituent is marked by lengthening. It is rather unclear at this time what the domain of focal lengthening is. Evidence for Dutch indicates that the entire word that carries the focal accent is stretched by some 10 percent; in this accentual lengthening all the segments are stretched by the same percentage – it is not the case that segments in the stressed syllable are treated differently than those in unstressed syllables (Eefting, 1991; Eefting and Nootboom, 1991). Moreover, if the [+F] domain is longer than just the accented word, only the latter is stretched; the duration of non-accented words in a [+F] domain is not affected (Eefting and Nootboom, 1991; van Heuven, 1998). Research on English shows that the domain of accentual lengthening in that language is not the entire word carrying the focal accent but only the segments contained by and following the stressed syllable, excluding segments in syllables that precede the stress (Turk and Sawush, 1997). Since the target words in the Dutch study were invariably stressed on the first syllable, the Dutch and English results are not necessarily in conflict.

Within the context of the present research it is important to discuss one issue which inevitably comes up when Indonesian languages are involved. There is ample evidence that the focal and boundary-marking functions of intonation (see above) are not clearly separated in many Indonesian languages. In these languages only the

last word within an I-domain is accented, and whenever a word is accented it is obligatorily followed by a boundary. As a consequence the focus-marking and boundary-marking function of the pitch movements cannot be separated. There is no word-based stress in Indonesian. In Indonesian the accent tends to be on the pre-final syllable (unless this syllable contains schwa). Due to the complication that only domain-final words can be accented through melodic means, I claim that in systems such as Indonesian, accent and boundary marking coincide.

2.1.5 Intonation

Intonation or speech melody is the pattern of rises and falls of pitch over the course of a spoken sentence. Unlike lexical tone, which is a word-level phenomenon, intonation belongs to the realm of sentence-level prosody. Intonation is a universal phenomenon: not a single human language is known that does not have sentence melody. Moreover, languages differ substantially in their repertoires of melodies. It seems safe to say that no two languages have the same melodic system, and even dialects belonging to the same language may differ markedly in their choice of melodies (cf. work on English dialects by Grabe, Post, Nolan and Farrar, 2000; van Leyden, 2004 and on Dutch (as well as English) dialects by Gooskens, 1999). Also, there is a growing body of results showing that the melodic differences between languages and language varieties are audible, and allow native listeners to reliably differentiate between foreign and native accents. The aim of the present dissertation is to study these and related phenomena for two regional languages spoken in the Indonesian area, to wit Toba Batak and Betawi Malay.

The melody of speech is determined by the repetition rate of the vocal cord vibration. The faster the vocal cords open and shut again, the higher the pitch of the voice. For a typical male speaker the repetition rate is between 70 and 200 Hz, for female speakers the rate of vocal cord vibration is roughly twice that of the males. The sex-related difference is largely caused by anatomical and physiological differences; during puberty the male vocal cords grow longer, heavier and thicker so that they vibrate more slowly than those of female speakers.

Clearly, speech is not produced on a monotone. In the large majority of spoken sentences pitch tends to be rather high at the beginning, but gradually drops down to a lower frequency as the utterance develops in time. This ‘downtrend’ in pitch is probably language universal, and is caused by the gradual reduction of subglottal air pressure over the course of an utterance due to the fact that air trapped inside the lungs is used up during speech (see also chapter III). However, the speaker may deviate locally from this overall ‘global’ trend by executing rises and falls in pitch by tightening and relaxing various muscle structures in and around the larynx, i.e. the cartilaginous structure that encloses the vocal cords (for detailed information on the anatomy and physiology of vocal cord vibration see, for instance, ’t Hart, Collier and Cohen, 1990; Hirose, 1997; Lieberman and Blumstein, 1988 and references therein). It appears that languages differ melodically not so much in global downtrend as in the shapes and sequencing of the local rises and falls.

Several models have been proposed to account for the melodic structure and differences between such structures across languages. Within the scientific community of the Netherlands two approaches are prominent, (i) the approach taken at the Institute for Perception Research (IPO) at Eindhoven (’t Hart et al., 1990) and (ii) the more recent autosegmental approach (e.g. Ladd, 1996).

The IPO approach models a sentence melody as a sequence of rises and falls within a set of two or three reference lines. The reference lines represent the bottom, (mid) and highest pitches between which the rises and falls may extend. The reference lines do not run horizontally but decline at a rate of – roughly – 1.5 semitones per second.⁴ Local movements may differ parametrically in their direction (rise, fall), size (full size, half size, quarter size), steepness (abrupt change, gradual change) and alignment (early, middle, late relative to vowel onset or to end of voicing). Functionally, some movements lend prominence to a particular syllable (accent), others mark a prosodic boundary, or simply connect the end of one movement to the beginning of an other. The IPO approach embraces a so-called superposition model, that is to say that the local rises and falls are superposed onto,

⁴ The actual declination rate, however, is variable and depends on the length of the utterance, such that longer utterances start at a higher pitch and decline to the terminal value at a slower rate. For details see ’t Hart et al. (1990), Rietveld and van Heuven (2001).

i.e. added to, the baseline which is provided by the global declining reference lines. Also, the IPO model is hierarchical in the sense that it decomposes local movements into a small number of primitives or distinctive features (see above), and combines individual rises and falls into a larger set of configurations (frequently recurring fixed combinations of simple movements), which in turn are combined into more complex melodies.

The IPO approach was originally developed to cope with the melody of Dutch sentences. In more recent years the same methodology was applied at IPO towards a description of English (Willems, 1982; de Pijper, 1983; Willems, Collier and 't Hart, 1988; Sanders, 1996), German (Adriaens, 1992) and Russian (Odé, 1989) intonation. A description of a non-Western language within the IPO tradition was made by Ebing (1997) for Indonesian. Outside of the Netherlands, the IPO methodology has been applied to the description of American English intonation (Maeda, 1976) and of French (Beaugendre, 1996).

The autosegmental approach is considerably more abstract. The primitives (or smallest units) are a set of just two tone targets, high (H) and low (L). The targets may be (but do not have to be) associated with boundaries (symbolised as %) either at the beginning or at the end of prosodic domains (%T and T%, respectively, where T stands for either H or L) or with focal accents, in which case the tone letter representing the target carries the diacritic '*'. A following H* accent within an utterance usually has a lower pitch value than the H* preceding it. This universal characteristic is modelled as downstep; a downstepped high target is preceded by the diacritic '!'. Formal operations can be carried out on the abstract tonal targets or on sequences of such targets, in much the same way as is done in other parts of the grammar. Targets can spread, be deleted and copied as in segmental phonology. Clearly, the autosegmental model is well integrated into the mainstream (generative) phonology in current linguistic theory.

At the phonetic (observable) level the autosegmental model assumes as a default that targets are connected by smooth interpolation, i.e. are connected by straight lines. Rises and falls (i.e. the basic descriptive units in the IPO model), are seen as phonetic implementations of a sequence of targets, viz. LH and HL,

respectively.⁵ Whenever an H target should not be connected smoothly with the following L target, the diacritic ‘+’ is added to it, which instructs the phonetic implementation to execute a steep rather than a gradual fall. In earlier versions of the autosegmental model the existence of declination was explicitly denied. Downtrend was held to apply to the high targets only, and could adequately be accounted for by the mechanism of downstep. However, more recent developments acknowledge that not only the high but also the low targets show a tendency to assume lower pitch values as they occur later in the utterance; for this reason declination has been added to the model. For a recent and fairly comprehensive survey of current views on autosegmental intonology I refer to Gussenhoven (2004).

Since I will be concerned mainly with the more fine-grained detail of phonetic implementation of the melodies of two Indonesian language varieties, I will not exclusively adopt one specific theory. Rather I will describe the melodies of the utterances in my target languages in terms of movements, i.e. rises and falls implemented as straight-line interpolations between H and L targets or ‘pivot points’. In doing so, I follow the example set by Stoel (2005) for Manado Malay.

2.1.6 Phonetic correlates of stress and accent

Taking a cue from Lindblom’s (1990) Hyper & Hypo (H&H) theory of speech interaction, I predict that the speaker will spend more effort on the production of linguistic materials which are more essential for the listener in order to reconstruct the speaker’s message. From this view I predict, for example, at the sentence level that materials that are presented in focus will be pronounced in hyper-mode. Hyper-speech (also called ‘clear’ speech) is spoken more deliberately, more slowly, more clearly articulated and with greater loudness, than materials that are out of focus, which are then articulated in hypo-mode. And, indeed, the literature provides experimental data bearing out these predictions (see van Heuven, 1998 and references therein). At the lower level of the word a similar line of argumentation can be followed. Given that the stressed syllable contributes more to the identity of a

⁵ In this respect, again, the autosegmental model treats segmental and prosodic phenomena in a similar fashion: in segmental phonology diphthongs are analysed as sequences of a short vowel and a glide.

word than the unstressed syllables, the H&H theory predicts that the speaker will realise the stressed syllable in hyper-mode and the unstressed syllables in hypo-mode.

There is massive experimental support for this view. Across languages the acoustical correlates that tend to be associated with stressed syllables have greater intensity (in decibels), greater loudness (i.e. intensity weighed by different sensitivities of the hearing system to different frequencies, in Sones), longer duration, and more extreme phonetic quality of the segments (see section 2.1.1 above). When a single syllable is produced in hyper-mode and is surrounded by unstressed syllables pronounced in hypo-mode, it makes sense that the articulatory gestures belonging to the hyper-mode largely overlap with the abutting gestures of the unstressed syllables. As a result of this, it is predicted that the effects of coarticulation from the stressed syllable onto the adjacent syllables are stronger than the other way around, leading to what has come to be called the stressed syllable's resistance to coarticulation (Dogil, 1999; de Jong, Beckman and Edwards 1993). Very often pitch has been mentioned as a further acoustical correlate of stress. The claim is that the stressed syllable typically has higher pitch than its unstressed counterpart. I take the view, however, that the effect of pitch is not directly a correlate of stress per se but is mediated through the sentence-level prosodic phenomenon of accentuation. Only when a word in focus is accented will the speaker realise a prominence-lending pitch movement, which will be executed on or quite near the stressed syllable (the prosodic head) of the accented word. Normally, the accent-lending movement will be a rise in pitch (a movement towards an H* target) which reaches its maximum somewhere in the stressed syllable. As a consequence of this, the average pitch of the stressed syllable will be higher than that of a syllable without an H* target. However, many languages, including English, Dutch and German, allow for the possibility that also L* accents occur. These accents are signalled not by a rise in pitch but by a stretch of low pitch in the stressed syllable. It seems safer, therefore, to list as a correlate of stress not 'high pitch' but 'a change of pitch relative to the pitch of the neighbouring syllables'.

It was realised, ever since the ground-breaking work by Fry (1955, 1958), that some phonetic correlates of stress are stronger than others. Moreover, the relative

strength of the correlate need not be the same in speech production as it is in speech perception. Fry (1955), for instance, showed that both (relative) duration and the difference in peak vowel intensity are acoustical correlates of stress in English minimal pairs of the type *import* ~ *import*. Along each of these two dimensions the two groups of tokens could be separated with near-perfection. For the listener, however, the difference in duration proved to be a much more influential stress cue than the difference in peak intensity. Fry (1958) then varied the shape and size of a pitch movement on either the first or the second syllable of minimal stress pairs, as well as the durations of the two syllables. His results indicated that some manipulations of the pitch were extremely effective stress cues, even stronger than durational differences. In Fry (1965) another pair of potential stress cues were varied, viz. duration and vowel quality (spectral expansion versus reduction). Here vowel quality proved much less effective than duration. Stress correlates have been studied for many other languages, such as Dutch (van Katwijk, 1974; Sluijter, van Heuven and Pacilly, 1997; Sluijter, 1995; Rietveld and Koopmans-van Beinum, 1987), Indonesian (Halim, 1974; Laksman, 1994), Japanese (Beckman, 1986), and ‘exotic’ languages like Samate Ma`ya (a language spoken at the border between the Austronesian and Papuan language area, Remijsen, 2001) and Curaçao Papiamentu (a Creole language of the Dutch West Indies, Remijsen and van Heuven, 2005). There has not been a single study that has attempted to vary all the relevant stress cues for the simple reason that the number of variations in the experimental design is so large that the experiment is unfeasible. Therefore, rather than studying the perceptual effects using stimuli with artificially manipulated stress properties, phoneticians have taken recourse to just studying the strength of acoustical properties of stress as statistical correlates of stress patterns. The results of a large number of studies have revealed that there is not a single, language-universal ranking of stress cues. In one language duration may outrank pitch, in another language the reverse may be the case. Some authors have come up with attempts to predict the relative importance of acoustical cues in the marking of stress from phonological properties of the language. This is a functional approach to the problem, based on the idea that an acoustical property that does work in one part of the phonology of the language cannot be used equally effectively to mark a contrast

elsewhere in the system. For instance, if a language uses duration (at the segmental level) to contrast short versus long vowels, duration will be less effective in the cueing of stress, and will therefore be lower in the hierarchy of stress cues for that language. Although the hypothesis is both attractive and plausible, no convincing experimental data are available to support it (Berinstein, 1979; Potisuk, Gandour and Harper, 1996).

Also, some languages would appear to mark the contrast between stressed and unstressed syllables more forcefully than other languages. The claim has been made, for instance, that the difference between stressed and unstressed syllables is very small in Javanese (Ras, 1985) but is much more noticeable in languages such as Dutch or English. A fairly recent claim is that in languages in which stress is used contrastively (even though the primary function of stress is not to signal contrasts at the word level, see above) stress is marked more clearly than in languages in which stress cannot be used contrastively (Dogil, 1999; van Heuven, 2002). My own work presented in the present dissertation directly speaks to this issue. I have studied the acoustical (and perceptual) correlates of stress as marked in a foreign language (Dutch) when spoken by learners with either a Jakarta-Malay or a Toba-Batak L1 background. These two languages belong to the Austronesian family and are closely related (see below) and yet they have radically different stress systems. In Betawi Malay stress can never be used contrastively (Muhadjir, 1977), and in fact, it can be argued that stress does not even exist in this language, whilst stress is clearly contrastive in Toba Batak and serves to distinguish many minimal stress pairs both lexically and morphologically (van der Tuuk, 1971 [1864]; Nababan, 1981). From this typological difference I predict that Toba-Batak speakers mark the difference between stressed and unstressed syllables more clearly than speakers of Betawi Malay, not only in their respective native languages, but also when they speak a foreign language.

2.2 Production and perception of L2-word prosody

Second-language speakers may be fluent in a given language, but I usually subjectively find their speech less intelligible than that of native speakers. Non-native listeners have more difficulty understanding speech than native listeners do (van Wijngaarden, 2001:103). Van Wijngaarden pointed out that non-native speakers could often be immediately identified by two factors that may reduce intelligibility: speech sounds are produced in an unusual, unexpected way ('distorted' phoneme inventory), and sentences are intoned in an unusual fashion.

A study on the production of non-native French spoken by Japanese learners indicated that prosody plays an important role in the evaluation of the naturalness by French listeners. The results pointed at the significant effects of duration and F0 on the perception of foreign accent (Kamiyama, 2004).

Listeners' ability to recognise different types of speech is affected to some extent by the sound system their language has. For instance, stress in French does not carry lexical information, while stress in Spanish does. A perception experiment involving listeners from both language groups shows that French listeners have difficulties in discriminating stress contrasts, while Spanish listeners have less or no difficulty (Dupoux, Pallier, Sebastian, and Mehler, 1997). A study considering the role of L1 in the production and perception of L2 is found in McAllister, Flege and Piske (2000). They found that non-native speakers of Swedish, who have the most prominent effect of duration contrast in their native language, were the most successful group in discriminating duration contrast.

Production and perception of foreign speech depend on the experience that subjects have in a foreign language, while also the age of acquisition is of importance, leading to a distinction between early and late bilinguals. Piske, Mackay and Flege (2001) describe factors affecting degree of foreign accent in an L2; they found that age of learning a foreign language is the most important predictor of degree of L2 foreign accent. In the present study all foreign-language speakers are late bilinguals.

2.3 Language background

Two related Indonesian languages are chosen as subjects of this research: Betawi Malay and Toba Batak. The latter is a language that has word stress (van der Tuuk, 1971; Nababan, 1981). The former, Betawi Malay, on the other hand, is a language that, like Indonesian, does not have word stress (Muhadjir, 1977) but it does have phrasal accent (Wallace, 1976).

2.3.1 Betawi Malay

Betawi Malay (BM) belongs to the Malayic subgroup of the Western Malayo-Polynesian branch of the Austronesian language family (Adelaar, 2005). BM is genealogically very close to Standard Indonesian (SI). These language varieties certainly seem to resemble each other prosodically, but very little research has been done on the prosodies of both languages. For both SI and BM there is at least some discussion on whether they have lexical stress or phrasal accent.

The language that is spoken in Jakarta is a dialect of Malay (Ikranagara, 1980:142). It can be distinguished into two dialects, modern Jakarta Malay and traditional Jakarta Malay (Wallace, 1976). Modern Jakarta Malay is spoken by the young generation living in Jakarta. Traditional Jakarta Malay is the first language of the ethnic group *anak Betawi* that nowadays has become a small minority group in the city of Jakarta (Grijns, 1991a, b). It is usually referred to as Betawi Malay by the Betawi themselves. It comprises two dialects, the dialect of the central part ('Dialek Kota' or 'Jakarta Kota') and the dialect of the border region ('Dialek Pinggiran' or 'Jakarta Pinggiran'). Jakarta Pinggiran has undergone many influences from other regional languages, e.g. Sundanese and Javanese, because it is spoken in the outskirts of Jakarta where these other languages are also spoken. For an overview of the four subdialects see also Chaer (1976).

For my production research, I concentrate on Betawi Malay (BM), the dialect of the central part of the city (Dialek Kota) because it is used by a homogeneous ethnic group, the Betawi and it has had comparatively little influence from other languages. However, for my perception experiments I needed subjects who also

knew Dutch. It turned out to be impossible to find sufficient Dutch-speaking native speakers of ‘Dialek Kota’. Therefore I had to use speakers of both traditional BM dialects (Dialek Kota and Dialek Pinggiran). Although the majority of my listeners were young Betawi, whose language might have been influenced by modern Jakarta Malay, I prefer the term Betawi Malay (BM) rather than Jakarta Malay in this dissertation

Betawi Malay (BM) differs from Standard Indonesian (SI) in various aspects. Some BM words, like *enteng* [ɛntɛŋ] ‘light, of little weight’, *nggak* [ŋgɑ̃ʔ] or *kage* [kagɛ] ‘no, not’ do not exist or are totally different in SI. Other examples are BM *kaye* [kayɛ] (SI *seperti*) ‘like’, BM *ame* [amɛ] (SI *dengan, oleh*) ‘with, through’, and the BM personal pronouns *gue* [gʷɛ] or *saye* [sayɛ] (SI *saya*) ‘I’ and BM *lu* [lu] (SI *engkaulkamu*) ‘you’. Finally, the typical BM phatic particles like *koq* [kɔʔ], *dong* [dɔŋ], *si* [si(h)] and *ah* [ah] should be mentioned here.

There are morphological differences in verb forms, for instance in BM the suffix *-in* is used where in SI the suffixes *-kan* or *-i* occur. SI imperative verbal forms such as *tuliskan* ‘write down’, *lupakan* ‘forget it’ are *tulisin, lupain* in BM. Also, the SI active verbal prefix *me(N)-* is *N-* or *ng-* in BM. Thus, SI *menakutkan* ‘frighten’ and *melamar* ‘solicit’ are *nakutin* and *ngelamar* in BM.

Phonological differences occur at the end of words. SI *a* in final syllables which are open or closed with consonant *h* corresponds to *e* [e] in BM, for instance SI *iya* ‘yes’, *dosa* ‘sin’, *Jakarta* ‘Jakarta’, *pisah* ‘separate’, *salah* ‘mistake’, *renyah* ‘crispy’, are *ive, dose, Jakarte, pise, sale, renye*, respectively in BM (Wallace 1976, Muhadjir, 1977). The diphthong *ai* in the last syllable in SI words corresponds to *e* [e] in BM; examples are *sampe* ‘arrive’ (SI *sampai*) and *cere* ‘divorce’ (SI *cerai*). Monophthongization of *ai* is, however, rather widespread and not restricted to BM. The frequent occurrence of the vowel schwa is typical for BM. Whereas SI has *a* in closed final syllables, BM has schwa in many lexemes, for instance SI *cepat*, BM *cepat* [çəpət] ‘quick, fast’; SI *senang* ‘happy’, BM *seneng* [sənəŋ]; SI *dengar* ‘hear’ BM *denger* [dɛŋɚ]. Furthermore, some typical BM words show the use of schwa, such as *demen* [dɛmən] ‘like’, *bareng* [barəŋ] ‘together’, *kelelep* [kələləp] ‘be drowned’. Notwithstanding the differences just mentioned, BM can still be understood by those who are familiar with SI (Ikranagara, 1980).

On the strength of the claim that the prosodic systems of BM and SI are essentially the same I will draw on publications on either language variety for a short overview of stress and accent of both languages.

Gerth van Wijk (1985, first published in 1883) observed that stress in Indonesian is usually very weak. All syllables are pronounced with approximately the same emphasis. Stress generally falls on the pre-final syllable of a root, which might be slightly lengthened. If the pre-final syllable is an open syllable and contains a schwa, the stress falls on the final syllable, unless the onset of the final syllable is *ng* [ŋ] in which case stress falls on the pre-final syllable with schwa. Words with schwa in the pre-final syllable are thus pronounced as follows: *déndam*, *sémpit*; *terús*, *besár*; *déngan*, *béngis* (Gerth van Wijk, 1985:45-46).

Fokker (1895) claimed that – phonologically – there is no word stress in Malay. Phonetically, in two-syllable stems, both syllables have almost the same amount of stress. However, Malay does have accent, which is signalled by duration. Accent is on the penultimate syllable, except if this syllable contains a schwa. Importantly, melodic variations are not analysed by Fokker as a reflection of prominence either at the word or at the sentence level.

Samsuri (1971) did research on the prosody of SI spoken by speakers from different language backgrounds. He also claims that SI has no distinctive stress; whatever the position of the prominent syllable in the word, the meaning of the word is the same. However, he found that the last syllable in a word or phrase is the most prominent one. On the other hand, in two- or three-syllable words without schwa the penultimate syllable is in general higher than the other syllables (i.e. *náma* ‘name’, *méja* [méja] ‘table’, *móbil* ‘car’, *usía* ‘age’, *seléra* [sələra] ‘appetite’).⁶ When the penultimate syllable contains a schwa and the final syllable does not, the last syllable is higher in two-syllable words (*senáng* [sənaŋ] ‘happy’, *jémú* [jəmu] ‘bore’). But in three-syllable words, the first syllable can also be higher. Besides *karená* [karəna] ‘because’, *majemúk* [məjəmu:k] ‘plural’, also *sútera* [sutəra] ‘silk’ and *pútera* [putəra] ‘son’ occur.

⁶ In all examples quoted from Samsuri (1971) the acute accent denotes ‘high pitch’. Most likely, high pitch should also be taken as stressed.

According to Halim (1974:111-113), prominence depends on the position of the word in the sentence: before a sentence-internal boundary the stress falls on the final syllable of the word preceding the boundary, whereas sentence-final stresses fall on the penultimate syllable of the last word of the sentence.

Moeliono and Dardjowidjojo (1988) state there is always one word in an utterance that is accented. That word is then highlighted by loudness, duration and pitch movement. Alieva, Arakin, Oglöblin and Sirk (1991:34) also claim that there is no phonological word stress in SI. However, there are always syllables in sentences that are highlighted or pronounced with higher intensity and thus are louder and clearer than the other syllables in the sentence, or that have a particular melody and a higher pitch, or that are longer. The ways in which those accented syllables are realised depend on the intonation pattern of the sentences. Zubkova, (1971, in Alieva et al., 1991:62) observes the way in which syllables are highlighted in disyllabic words. She concludes that pitch and vowel intensity are not important for word stress. Also, differences in duration between both vowels are small and inconsistent. A production experiment done by Pavlenko (1969, in Alieva et al., 1991:62-63) shows that intensity is not important.

Most authors thus seem to claim that stress in SI is either weak or non-existent. Nevertheless, there is a group of authors who formulated rules for the placement of word stress in (Standard) Indonesian. These rules have, in fact, recently been reiterated by Cohn (1989) and Cohn and McCarthy (1994), working in a metrical framework: stress is on the penultimate syllable, unless this syllable contains a schwa, regardless of the morphological structure of the word. However, experimental work by Laksman (1994) provides evidence that schwa can be stressed. Experiments by van Zanten and van Heuven (1998, in press) found no preferred stress position in SI. Similarly, van Zanten, Goedemans and Pacilly (2003) conclude on the basis of experimental evidence that SI does not have word-based stress, but has phrase-level accent only.

The following description of BM prosody is mainly based on Wallace (1976). Wallace notices that the domain of the accent is the phrase rather than the word. His impression is that there is no word stress in BM. Wallace has the impression that accent in BM is realised with a rising pitch; longer duration and an increased

loudness are secondary cues. According to Wallace (1976:56-59), the accent is usually on the penultimate syllable of the last word in a phrase in BM.

<i>tu buku mére</i> ⁷	<i>buku báru</i>
‘That book is red’	‘new book’

The accent goes to the final position if the penultimate has schwa (a), or if the last word of the phrase is made up of a monosyllabic stem preceded by a prefix (b). A monosyllabic word is always accented (c).

(a) <i>Rumenye gedé</i> [gədəɛ]	(b) <i>ubinnye dipél</i>	(c) <i>masukin di bák</i>
‘The house is big’	‘the floor is mopped up’	‘put into the bin’

The prefix *dí-* (passive voice) apparently does not receive accent in BM, i.e. *dicét* ‘be painted’. The same happens with the prefix *nge-* [ŋə-] (active voice) in *ngepél* ‘mop up’, *ngécét* ‘paint’, but here the reason could be the vowel schwa that does not receive accent. Again, Wallace underlines that schwa is unstressed in the examples *kecepatán* /kəçəpətan/ ‘to be fast’ and *itemín* /itəmín/ ‘to make black’. He mentions that the suffixes *-in* and *-an* can bear accent but gives one exception, viz. *kebakáran* ‘to burn’, in which the accent is on the penultimate syllable.

In one case he finds that schwa can be accented, namely when it precedes the unaccented suffix *nye* [ɲe], such as in *itémnye* [itəmɲe] ‘the black, being black’, *sambélnye* [sambəlɲe] ‘the chilli sauce’. That the accent shifts to the penultimate syllable in these instances (*item* → *itémnye*, and *sámbel* → *sambélnye*) is in line with the general rule that accent is penultimate, but it is at odds with the rule that accent goes to the final position when the penultimate contains a schwa.

Wallace did not consider words with schwa in both penultimate and final syllable, like *dekət* [dəkət] ‘close to’, *senəŋ* [sənəŋ] ‘happy’, *kelelep* [kələləp] ‘be drowned’.

⁷ Wallace’s (1976) example is ‘tu buku mérah’. This must be a mistake. Similarly, in the next example, Wallace has ‘Rumahnje gedé’, instead of the correct BM ‘Rumenye gede’.

Laksman (1994) describes SI as spoken by a Jakartan speaker. She found that schwa can be accented as well as any other vowel. Finally, in van Zanten, Goedemans and Pacilly (2003) Indonesian listeners with a Jakartan background were involved. This study suggests that for Jakartan listeners, prominence may freely occur on any of the last two syllables (and sometimes even the antepenult) of any word. Summarizing, the literature seems to indicate that BM does not have a word-based but rather a phrase-based accent.

2.3.2 Toba Batak

Batak belongs to the West Malayo-Polynesian languages (van der Tuuk, 1971 [1864]). The Batak dialects are divided into the northern dialects (Karo, Dairi), and the southern dialects (Toba, Angkola, Mandailing, Simalungun) (Adelaar, 1981, 2005; Woollams, 2005; Sibeth, 1991). Toba Batak (TB) is the most common spoken dialect among the Batak dialects. It is also used as the means of communication among the (southern) Batak people in the region and elsewhere. If people talk about Batak, they usually refer to TB. TB is spoken by about two million people living on Samosir Island and to the east, south and south west of Toba Lake in North Sumatra. TB speakers also live in the other districts in North Sumatra. According to Pelly (1989), in 1980 the TB speakers living in Medan, the capital of North Sumatra, were the second-largest ethnic group.

In my research I studied native speakers of TB who live in Jakarta, but still have a strong relationship with the TB community in Sumatra.

Contrary to BM, all authors agree that TB is a (distinctive) stress language. The first reference to TB is van der Tuuk (1971 [1864]). He wrote about the sound system, the word formation, and the grammar of TB. According to Van der Tuuk TB has five monophthongal vowels and no central vowels (i.e. no schwa /ə/).

TB has lexical stress (van der Tuuk, 1971 [1864]; Nababan, 1981). The description of the TB stress system below is mainly based on Nababan (1981). Stress in TB is penultimate for nouns and verbs containing two or more syllables and final for predicatively used adjectives (Nababan, 1981; Emmorey, 1984). There is a clear difference, for instance, between N *tíbo* 'height' and A *tíbó* 'high' (stressed

syllable indicated by acute accent mark). However, attributive adjectives, and adjectives following *na* (relative pronoun), do not have final stress; they have penultimate stress, e.g. *na tíbo* ‘which is high’. Stress distinctions based on morphological composition are also found, for example *ítom* ‘your brother/sister’ versus *itóm* ‘black’.

If words with penultimate stress are suffixed, stress shifts to the next syllable in order to keep the stress penultimate.

dálan	N	‘road, way’
mardálan	V intransitive	‘to walk’
daláni	V transitive	‘to walk through’
dalánna	N	‘its/her/his way’

Compound words have one main stress (written here with an acute accent) that falls on the stressed syllable of the final constituent word (the prosodic head of the compound). The stress on the first constituent word weakens and becomes a secondary stress (indicated here with a grave accent on the vowel).

hóda ‘horse’	→	hòda pácu ‘race horse’
mángaj ‘celebrate’	→	màngaj júhut ‘celebrate a wedding’

The same process takes place at the phrase level.

In the comparative grade of adjectives stress is final. Stress is not on the stem but on the comparative suffix, e.g. *úli* ‘beautiful’, but *ulíán* ‘more beautiful’.

Nababan claims that the difference between stressed and unstressed syllables is realised as a difference in loudness. Also, if a word has more than two syllables, one syllable has the strongest stress; this does not mean that the unstressed syllables are all equally weak. In *borumuna* ‘your daughter’ the strongest stress is on *mu*, but *bo* is certainly louder than *ru* or *na*. It seems louder not only because it was produced with a particular power, but also because the vowel was lengthened. The syllable *bo* is thus longer than *ru* and *na*.

Consonant length is distinctive in TB (Nababan, 1981; Cohn, Ham and Podesva, 1999). The word *pitu* [pitu] means 'seven', but *pittu* [pit:u] means 'door'. Nababan adds that words with intervocalic geminate consonants such as *pittu* have to be segmented as /VC + CV/. Acoustically it will be difficult to separate this consonant into two syllables. Cohn et al. (1999) found that vowels preceding geminate consonants are shorter than vowels preceding singleton consonants: the *i* in *pittu* is shorter than that in *pitu*.

Emmorey (1984) investigated the TB intonation system. Her research is limited to basic sentence types and a few constructions from one native speaker. Sentences were presented in isolation and in context. She found that in declarative sentences, the nuclear pitch accent is aligned with the stressed syllable of the last word of the phrase. Emphatic stress has a higher nuclear pitch accent than non-emphatic stress.

An experimental study was done by Chen (1984), who also claimed that TB is a stress language. The article discusses the relation between stress and its acoustic parameters based on acoustic measurements. Chen (1984) shows that in TB stress is realized by a rising fundamental frequency. The difference in fundamental frequency between stressed and unstressed syllables is less obvious in connected speech than in isolated words, while the difference in duration between stressed and unstressed syllables is more obvious in connected speech than in isolated words. If target words are not at the intonation peaks (i.e. out of focus and therefore not accented), stress is signalled by longer duration. In contrast to the above, Podesva and Adisasmito-Smith (1999) found no duration-stress relationship for TB vowels. They did, however, find a relation between pitch (but not intensity) and stress.

Van Zanten and van Heuven (1997) used TB as well as Sundanese and Javanese speakers to investigate the influence of substrate language on the temporal organisation of Indonesian words. They found that stress influenced the duration of syllables more strongly for TB Indonesians than for other Indonesian (viz. Sundanese and Javanese) speakers. TB speakers realised the stressed syllable in Indonesian words about 40% longer than the unstressed syllables; for Sundanese and Javanese speakers the lengthening effect was 25%. Van Zanten, Goedemans and Pacilly (2003) found similar lengthening effects for TB speaking Indonesians. Duration thus seems to be a possible cue for stress in TB. The present research

investigates whether duration indeed signals stress in TB, and also which other stress cues may be used.

Chapter III

Temporal and melodic structures in Toba Batak and Betawi Malay word prosody

3.1 Introduction

In languages with word stress, one syllable is perceived as stronger than the other syllables in the same word. On the higher levels, in phrases or sentences, accent is used to make particular words more prominent than other words. In stress languages, the sentence accent typically coincides with the word stress. Languages without word stress may also use accent to highlight words in sentences but then the sentence accent is not restricted to a particular syllable in the word.¹

This study focuses on the production of word prosody in Toba Batak (TB) and Betawi Malay (BM). As indicated in the previous chapter, TB is a language that has word stress (van der Tuuk, 1971; Nababan, 1981). BM, on the other hand, is a language that does not have word stress (Muhadjir, 1977) but it does have phrasal accent (Wallace, 1976). Therefore, it is interesting to compare these two different languages in their realization of accent.

Accent is primarily marked by two prosodic features, i.e., duration and pitch. When a word is prominent (in focus), it will have a perceptually prominent change in pitch and a longer duration than when it is non-prominent (out of focus)

¹ Indonesian has been claimed to be one such language (Van Zanten and van Heuven, 1997; Van Zanten, Goedemans and Pacilly, 2003).

(Edwards, Beckman and Fletcher, 1991). In many stress languages, the stressed syllable is affected most by lengthening.

Next to accent, I will investigate boundary marking in TB and BM. The marking of linguistic boundaries also affects word duration and pitch. Boundaries may be signalled by boundary tones and words in sentence-final position are usually longer than words in other positions (van Heuven, 1994 and references given there). Generally, the effects of pre-boundary lengthening are stronger on segments as these appear closer to the end of the word (Cambier-Langeveld, 2000)².

In this research I expect to find similar effects of boundary marking for both languages. As regards focus, however, I expect stronger effects in the stress language TB (especially in the stressed syllable) than in non-stress BM.

3.2 Background

The languages. The two Indonesian languages used in this research have different prosodic systems, as mentioned in Chapter II. All sources report that TB is a stress language (van der Tuuk 1971 [1864], Nababan 1981, Chen 1984). BM, on the other hand, does not have word stress (Muhadjir 1977), but it does have accent (Wallace 1976; see also Chapter II).

Theoretical background. Speakers use specific variations of duration, pitch, loudness, as well as vowel quality to highlight certain constituents within linguistic domains (Fry, 1958), but usage of these prosodic properties varies across languages and speakers. This highlighting can be used to phonetically realize the abstract phonological property called stress. In stress languages pitch, duration, loudness and vowel quality are the important phonetic correlates of stress (Lehiste, 1970). Also, higher pitch, higher intensity and longer duration tend to characterize prominence in a sentence (Vaissière, 1983). Stress languages like Dutch and English use accent to

² A more complete overview was presented in Chapter II.

mark prominence (Van Heuven, 1994). A prominent word, typically a word in focus, i.e., a word the speaker highlights as more important than other words in a sentence, bears a sentence accent; its stressed syllable carries correlates of both stress and accent. Consequently, I expect focus to have more influence in stress languages, in particular the stressed syllables, than in non-stress languages.

Pre-boundary lengthening affects duration as well. Speech slows down towards the end of a prosodic domain. Words in sentence-final position are thus longer than other words (*ceteris paribus*) in earlier positions in the sentence (Cruttenden 1997:33). In Dutch, boundary lengthening was found to be strongest in the final syllable, and within the final syllable in the final segment (cf. Cambier-Langeveld, 2000). More generally, the effects of pre-boundary lengthening are stronger as the segment concerned is closer to the boundary.

3.3 Methods

3.3.1 Selection of speech materials

For TB I selected eight words with penultimate stress for this study. I suspected that the consonant length distinction might interact with the durational effects of focus and boundary. For this reason I chose four words with intervocalic geminate consonants, *dakka* ['dak:a] 'branch', *pittu* ['pit:u] 'door', *jabukku* [ja'buk:u] 'my house' and *pagatti* [pa'gat:i] 'be exchanged by mistake'. Further, I selected four words with intervocalic singleton consonants, *pitu* ['pitu] 'seven', *suga* ['suga] 'thorn', *jabu* ['jabu] 'house', and *kareta* [ka'reta] 'carriage'.

For BM I also selected eight words consisting of two or three syllables. Three words containing a schwa vowel in the pre-final syllable *pete* [pəte] 'stinking bean', *dekət* [dəkət] 'nearby', *rejeki* [rəjəki] 'fortune', were chosen to investigate whether the schwa behaves differently than the full vowels under the influence of focus and boundary marking. A further five BM words containing full vowels in the last two syllables were used: *kaga* [kaga] 'no, not', *kutu* [kutu] 'louse', *belaga* [bəlaɡa]

'pretend', *pipi* [pipi] 'cheek' and *pepet* [pɛpɛt] 'overtake rashly'.

The target words were embedded in fixed carrier sentences, in order to create four focus and boundary conditions. Four question sentences were devised to elicit these four sentence types. Table 3.1 lists examples of TB and BM materials.

3.3.2 Speakers and recording procedure

Four native TB speakers (two male, two female) and four native BM speakers (two male, two female) took part in the experiments. At the time of recording the four TB speakers (aged between 30 and 50 years) were living in Jakarta. They had come to Jakarta from North Tapanuli (a TB region) after the age of puberty and lived among the TB community in Jakarta, so that they still used TB in their daily life. The BM speakers (30 – 55 years of age) were living in Sawah Besar, Central Jakarta. They used Betawi Malay in their daily life.

All questions and answer sentences were presented to the speakers in a fixed order. Another speaker of the same language read out the question sentences, and the subject then responded by reading the answers. The recordings were made in a quiet room onto a Sony TC-D5 PRO II tape recorder through head-worn Shure SM-10A microphones. Every speaker spoke all the materials three times. The total number of utterances was 384 per language. All speech materials were then digitized (16 kHz sampling frequency, 16 bits amplitude resolution).

Table 3.1. Examples of Toba-Batak and Betawi-Malay speech material in four sentence conditions (question sentences in parentheses).

		Boundary		
Prominence	+focus	+final	TB	(Aha didokkon ibana?) 'What did he say?' Didokkon ibana [dakka] 'He said [dakka]'
			BM	(Die bilang ape?) 'What did he say?' Die bilang [kutu] 'He said [kutu]'
		-final	TB	(Aha didokkon ibana nattoari?) 'What did he say yesterday?' Didokkon ibana [dakka] nattoari 'He said [dakka] yesterday'
			BM	(Die bilang ape tadi?) 'What did he say just now?' Die bilang [kutu] tadi 'He said [kutu] just now'
	-focus	+final	TB	(Nandigan didokkon ibana [dakka]?) 'When did he say [dakka]?' Nattoari didokkon ibana [dakka] 'Yesterday he said [dakka]'
			BM	(Kapan die bilang [kutu]?) 'When did he say [kutu]?' Tadi pagi die bilang [kutu] 'This morning he said [kutu]'
		-final	TB	(Aha [dakka] didokkon ibana?) 'What [dakka] did he say?' Didokkon ibana [dakka] na togu. 'He said [dakka] which is straight'
			BM	(Die bilang [kutu] ape?) 'What [kutu] did he say?' Die bilang [kutu] buku. 'He said [kutu] of books'

3.4 Duration

All target words were segmented and measured with PRAAT speech processing software (Boersma and Weenink, 1996). The section dealing with word duration presents the results of the eight target words of TB and BM. The target words contain two or three syllables. In the analysis of syllable and segment durations, I will only concentrate on the last two syllables of the target words. In the next section I will first deal with the TB speech materials. The analysis of the BM speech materials follows in § 3.4.2.

3.4.1 Toba Batak

3.4.1.1 Word duration

Figure 3.1 presents the duration of the TB target words broken down by prominence (+ vs. – focus) and broken down further by boundary (+ vs. – final position). The figure shows clearly that both focus and boundary affect TB word duration. Prominent words are longer than non-prominent words, (10% longer in sentence-final position and 14% in sentence-medial position). Further, sentence-final words are longer than sentence-medial words (13% longer for prominent words, 17% for non-prominent words).

An analysis of variance was run with focus and boundary as fixed factors and word duration as the dependent variable. The ANOVA indicates that the effects of focus and of boundary on TB word duration are both significant. A subsequent oneway ANOVA treating the four focus-by-boundary conditions as a single factor shows that the mean durations of the target words differ significantly from each other [$F(3,329) = 31.39, p < .001$]. Boundary effects are stronger than focus effects, to the extent that sentence-final non-prominent words are longer than prominent words in sentence-medial position. A post-hoc test (Scheffé procedure) indicates that the differences between prominent ([+focus]) and non-prominent ([-focus]) words are significant ($p = .003$ for sentence-final words, and $p < .001$ for sentence-medial

words). Also, words in sentence-final position are significantly longer than words in sentence-medial position in the same focus condition ($p < .001$).

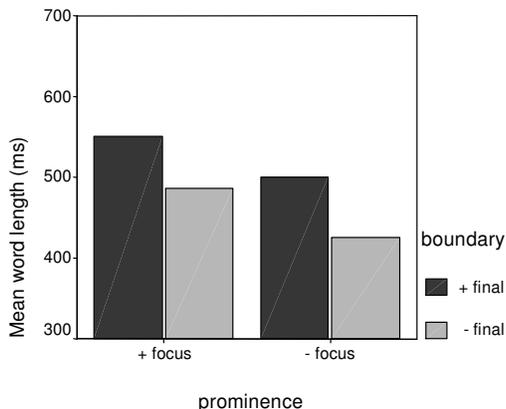


Figure 3.1. Mean durations of eight TB words, broken down by boundary and prominence conditions.

3.4.1.2 Syllable duration

According to Nababan (1981) words with intervocalic geminate consonants have to be segmented as /VC + CV/. But, it is acoustically difficult to separate this consonant into two syllables. Therefore, in this section I only present the group of words containing intervocalic singleton consonants.

Figure 3.2 shows the effects of focus and boundary on the mean durations of pre-final and final syllables in the CVCV words. A [+focus] pre-final syllable is about 15% longer than its [-focus] counterpart (n.s.), both sentence medially and sentence finally. The effect of focus on final syllables is strong for sentence-medial words (31%, significant at $p < .001$), but weak in sentence-final words (n.s.).

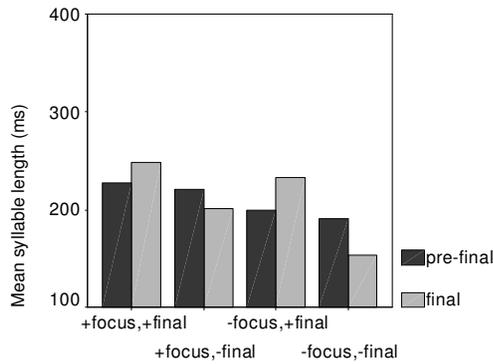


Figure 3.2. Mean durations of pre-final and final syllables of TB words with singleton C, broken down by sentence types.

Differences due to boundary condition are small and insignificant in penultimate syllables; they appear larger in ultimate syllables (24% pre-boundary lengthening of final syllables in prominent words, 51% in non-prominent words). These differences are highly significant in the same focus condition ($p < .001$). This is in line with my expectation: final syllables are strongly influenced by pre-boundary position. The focus effect on final syllables is only significant sentence medially.

It seems that the duration of the pre-final syllable is hardly influenced by focus or by boundary. The one-way ANOVA shows that only the combined effects of focus and boundary conditions on the duration of pre-final syllables are significant [$F(3,166) = 5.13, p = .002$], and highly significant on the duration of the final syllable [$F(3,166) = 32.27, p < .001$].

My finding that the (stressed) pre-final syllable is only marginally affected by the variation of focus is at odds with the assumption that the stressed syllable should be particularly affected by prominence (cf. Cambier-Langeveld, 2000; Vaissière 1983). In the next section I will try to find out whether this unexpected finding has anything to do with the segmental structure of TB.

3.4.1.3 Segment duration

In this section I examine only the last four segments of all eight target-words. Geminate consonants were segmented as one consonant. In this way we can compare the segmental duration of words with geminate consonant (CVC:V) with that of words with singleton consonants (CVCV). Figures 3.3 and 3.4 illustrate the mean segment durations of words with singleton C and words with geminate C.

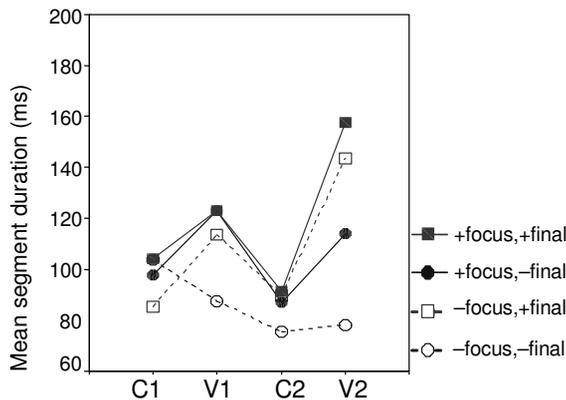


Figure 3.3. Mean durations of the last four segments of words with singleton C, broken down by sentence type.

CVCV-words. Lengthening effects of focus and boundary appear in all segments, although they fail to reach significance in C1. A one-way ANOVA showed that there are significant effects of focus and boundary on the last three segments (V1, C2, V2). Both effects occur especially on both vowels [$F(3,166) = 21.23$ for V1; $F(3,166) = 25.36$ for V2, $p < .001$ for both]. V1 is longer in prominent ([+ focus]) words than in non-prominent ([−focus]) words, but the lengthening is significant only in sentence-medial words (about 40%, Scheffé with $p < .001$). The lengthening effect of focus on C2 is small. There is only one significant difference ($p = .017$) between the two most extreme conditions, i.e. [+focus, +final] vs. [−focus, −final]. V2 in prominent words is longer than in non-prominent words. As was the case with

V1, the difference is large in sentence-medial words (46%) but very small in sentence-final words (n.s.). Lengthening effects attributable to boundary are strong in V2s (38% lengthening for prominent words and 83% for non-prominent words; both significant at $p < .001$). Figure 3.3 shows also some boundary effects on C1, V1, and C2, but here the effects are smaller. Sentence-final words have longer V1s than sentence-medial words, but the differences are significant in non-focussed words only ($p < .001$).

Since the effects of focus on the duration of the (stressed) pre-final syllable are small or absent at the segmental level, there is also no effect, of course, on syllable duration. The strong effect of pre-boundary lengthening on the final syllable is largely due to longer vowel duration. This fits in well with similar lengthening effects in stress languages like Dutch (Cambier-Langeveld, 2000).

CVC:V-words. As is shown by Figure 3.4, lengthening effects are weak in V1, but become stronger in C2 and V2. A one-way ANOVA reveals significant effects of sentence type on V1 [$F(3,159) = 4.86, p < .005$], C2 [$F(3,159) = 13.47, p < .001$], and V2 [$F(3,159) = 28.44, p < .001$]. A post-hoc test (Scheffé) confirmed that some lengthening effects appear in V1 due to focus (sentence-medial only, $p = .008$). Further, no significant differences occur between V1 in sentence-final words and V1 in sentence-medial words. C2 differences due to focus are small and insignificant. V2 in prominent words is generally longer than that in non-prominent words. The difference is significant in sentence-medial words ($p = .021$), but not in sentence-final words.

There are no significant boundary effects for C1 and V1. The C2s are significantly longer in sentence-final words than in sentence-medial words ($p = .001$ in prominent words, $p = .011$ in non-prominent words). Large differences due to boundary are also seen in V2 (44% lengthening for prominent words, 80% for non-prominent words), with a significance level at $p < .001$. These lengthening effects are similar to the effects found in the final vowel of CVCV-words. The geminate C dominates the temporal organization of the word, and seems to block the spreading of the final lengthening to the preceding vowel. Similarly, Cohn, Ham, and Podesva

(1999) found that vowels preceding geminates were substantially shorter than vowels preceding singletons. The geminate C itself is lengthened by boundary. On a more abstract linguistic level it seems that the effects of final lengthening are restricted to a window of three skeletal positions. When the target contains singleton C only, the domain of final lengthening includes the final V1C2V2 sequence. When the word contains a geminate, then the lengthening is restricted to just the C2:V2 sequence.

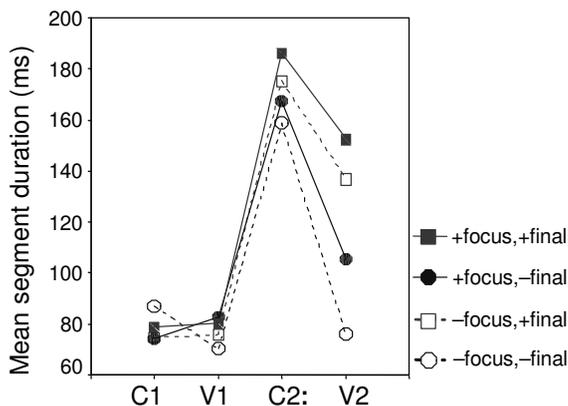


Figure 3.4. Mean durations of the last four segments of words with geminate C, broken down by sentence type.

Accentual lengthening in the stressed pre-final syllable is, unexpectedly, weak in TB CVCV as well as in CVC:V words. Podesva and Adisasmith-Smith (1999) also found no stress-duration correlation in TB vowels, without offering an explanation. A possible explanation may be the fact that TB has phonemic consonant length that makes the duration parameter not available as a stress parameter (cf. Berinstein, 1979).

3.4.2 Betawi Malay

3.4.2.1 Word duration

BM does not have stress. The accented element does not bear the phonetic correlates of stress. Therefore I expected word duration to be less affected by focus in BM than in TB.

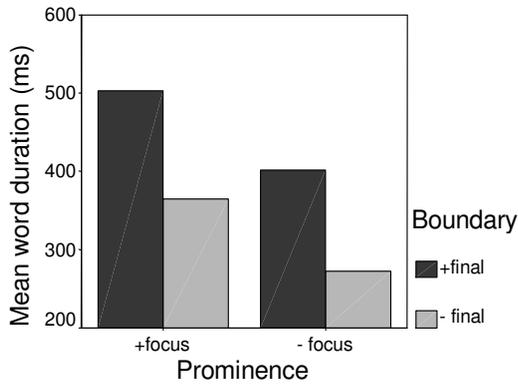


Figure 3.5. Mean duration of eight BM words spoken by four BM speakers, broken down by prominence and boundary conditions.

Figure 3.5 shows that in the four conditions word durations differ strongly from each other. A one-way ANOVA shows significance for both effects [$F(3,369) = 74.42$, $p < .001$]. Prominent words have a longer duration than their non-prominent counterparts. In sentence-final position the difference is 25%, and in sentence-medial position ca. 33%. The differences are highly significant within the same boundary condition ($p < .001$). In addition, sentence-final words are longer than sentence-medial words. The difference is 37% for prominent and 46% for non-prominent words. Significant differences occur in the same focus condition ($p < .001$).

Focus and boundary effects on word duration are stronger in BM than in TB. This disagrees with my expectations that word duration in BM would be less affected by focus than in TB, and that the effects of boundary marking would be similar in both languages.

3.4.2.2 Syllable duration

In this section, I split the target words into two groups, i.e. one with full vowels only (CVCV-words) and another with schwa in the pre-final syllable (CəCV-words).³ The group of words with schwa is analysed separately to find out whether the words with schwa are affected differently by focus and boundary than the words with full vowels only. Figures 3.6 and 3.7 show the mean durations of final and pre-final syllables of both groups.

CVCV-words. The large effects that were found on the word level are also visible in Figure 3.6. The ANOVA notes significant effects of focus and boundary on the length of both syllables [$F(3,229) = 39.38$, for pre-final syllable; $F(3,229) = 60.05$ for final syllable, both $p < .001$]. The focus effects are stronger for pre-final syllables than for final syllables. For pre-final syllables the difference is 29% in sentence-final words and 57% in sentence-medial words. For final syllables the difference is 24% in sentence-final words, and 14% in sentence-medial words. A post-hoc test shows significant differences between pre-final syllables for both positions ($p < .001$), and between final syllables only in sentence-final words ($p < .001$).

On the other hand, the boundary effects are stronger for final syllables than for pre-final syllables. Sentence-final words have longer final syllables than sentence-medial words (66% for prominent words, 52% for non-prominent words). Both

³ This group contained one word with schwa in the final syllable as well. This word behaved in a similar way to the other words in the same group. Also, both groups contained a word with a closed final syllable. All else being equal, vowels are shorter in closed syllables than in open syllables. However, the lengthening effects discussed here are of the same order for closed and open final syllables.

differences are highly significant ($p < .001$). For pre-final syllables the difference is 14% for prominent words ($p = .034$) and for non-prominent words 39% ($p < .001$).

The result indicates that focus affects the duration of the pre-final syllable more than boundary, and boundary affects the final syllable more than focus.

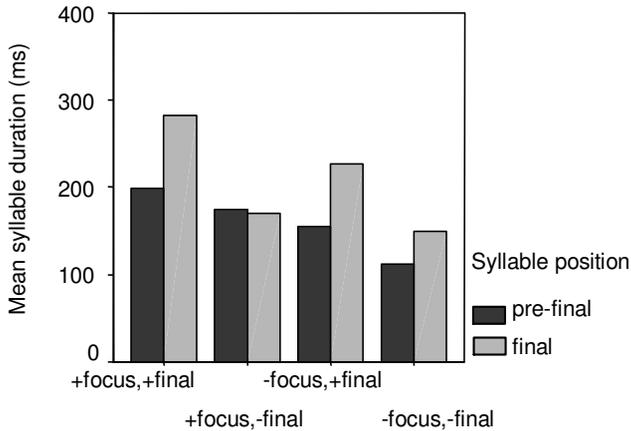


Figure 3.6. Mean durations of pre-final and final syllables of target words with full vowels, broken down by sentence type.

CəCV-words. Pre-final syllables (i.e. the syllables containing schwa) seem to be affected neither by focus nor by boundary. Final syllables, however, show strong effects. A one-way ANOVA notices a significant effect of sentence type on the final syllable [$F(3,136) = 46.75, p < .001$], but not on the pre-final syllable [$F(3,136) = 2.50, p = .062$]. A post-hoc test confirms that the pre-final syllables are similar to each other. Lengthening effects on the final syllables due to focus are 36% in sentence-final words, and 60% in sentence-medial words ($p < .001$ for both conditions). The effect of boundary on final syllables is even stronger (60% lengthening in prominent words, 88% in non-prominent words).

As shown in Figure 3.7, neither lengthening effects occur on the pre-final syllables with schwa. They occur, however, strongly on the final syllable. Both effects are stronger in the final syllable of CəCV-words than in the final syllable of CVCV-words. It would seem true, then, that BM words with pre-final syllables

containing schwa have indeed the accent on the final syllable. As a result the final syllable is simultaneously lengthened by the effects of focus and of pre-boundary position.

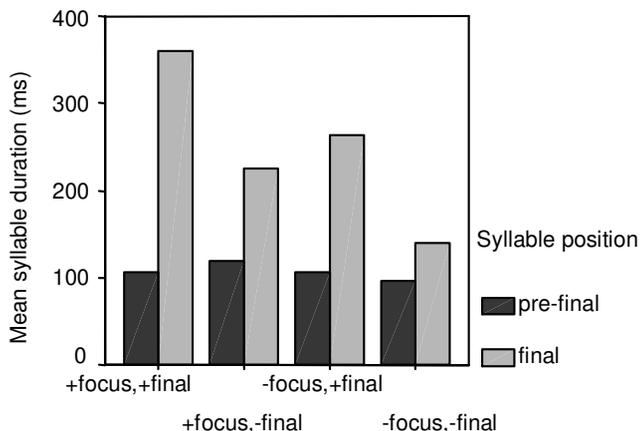


Figure 3.7. Mean durations of pre-final and final syllables of target words containing schwa, broken down by sentence type.

3.4.2.3 Segment duration.

In the following section I investigate the segment durations in the last two syllables of the target words. Two words had a final consonant. The final consonant was not included in the measurements. Figures 3.8 and 3.9 illustrate the findings.

CVCV-words. Figure 3.8 shows that segment durations are increasingly different from each other towards the end of the word. A one-way ANOVA with segment durations as dependent variables, and sentence type as a fixed factor shows significant effects of focus and boundary on all segments [$F(3,229) = 12.59$ for C1, $F(3,229) = 27.68$ for V1, $F(3,229) = 45.14$ for C2, and $F(3,229) = 30.25$ for V2; with $p < .001$ for all segments]. Focus effects show up in both segments of the pre-final syllable and decrease on the last two segments. The differences due to focus in C1s are about 40% in both boundary positions ($p < .005$). In V1 the differences are

24% in sentence-final words ($p = .007$), and 67% in sentence-medial words ($p < .001$). The lengthening for C2s is 22% in sentence-final words ($p = .017$), and 27% in sentence-medial words ($p < .001$). Prominent words have also longer V2s than non-prominent words, but only in sentence-final words (26%; $p = .011$). In sentence-medial words the lengths of V2s are similar to each other, and the difference is not significant.

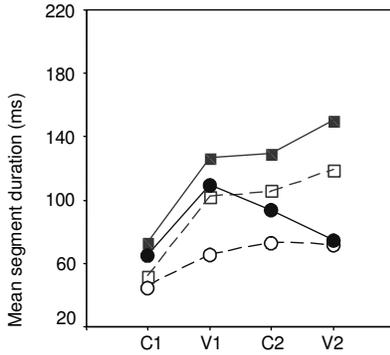


Figure 3.8. Mean segment durations of BM CVCV-words, broken down by sentence type.

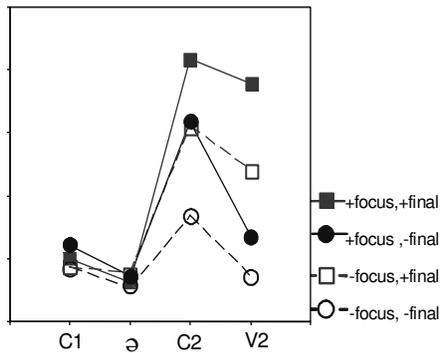


Figure 3.9. Mean segment durations of BM CəCV-words, broken down by sentence type.

The boundary effect is particularly strong in the final syllable. In sentence-final words V2 is longer than in sentence-medial words ($p < .001$). In prominent sentence-final words V2 is twice as long as in prominent sentence-medial words. In non-prominent words the lengthening due to boundary is almost 66%. The pre-boundary lengthening in C2 is about 40% (in both focus conditions $p < .001$). Differences between V1 in sentence-final words and in sentence-medial words are only found for non-prominent words (56%, $p < .001$), but not for prominent words (n.s.). The differences between the C1s are even smaller and insignificant.

Accentual lengthening seems to affect C1 and V1 most strongly, and to decrease near the end of the word. Pre-boundary lengthening starts possibly on V1 and becomes strongest on the last segment.

CəCV-words. There are hardly any differences between the durations of C1, nor between schwas in the four conditions. The ANOVA reveals no significant effects of sentence type on the lengths of C1 and schwa. Significant effects do, however, occur on C2 [$F(3,136) = 57.56, p < .001$] and V2 [$F(3,136) = 21.82, p < .001$].

Prominent words have longer C2 durations than their non-prominent counterparts (29% in sentence-final words, and 68% in sentence-medial words, both $p < .001$). Prominent words also have longer V2s than non-prominent words (47%). This difference is only significant in sentence-final position ($p = .011$), but not in sentence-medial position ($p = .530$).

The C2s of sentence-final words are longer than those of sentence-medial words. This boundary effect is 27% for prominent words, and 68% for non-prominent words ($p < .001$). Sentence-final words also have longer V2s than sentence-medial words in the same focus conditions ($p < .001$). The boundary effect is stronger in V2 than the focus effect.

The duration of the schwa is affected by focus nor by boundary. The schwa apparently blocks lengthening effects on the preceding segment as well. Strong duration effects of focus occur on the segments following schwa. This result is in line with the statement by Wallace (1976) that accent shifts to the final syllable when the pre-final syllable contains a schwa.

3.4.3 Conclusion

I expected that focus effects would be stronger in TB (a stress language) than in BM (a non-stress language), and that boundary effects would be similar in both languages. The results of my research show that effects of focus and boundary are significant at the word level in both languages. But, in contrast to my expectations, both effects are stronger in BM than in TB. Secondly, in both languages boundary effects are stronger than focus effects.

The focus effects in TB are significant at the word level. Sentence-finally, the lengthening effect is 10%, which is similar to Dutch (Eefting, 1991; Eefting and Nootboom, 1991). Sentence-medially, the focus-induced lengthening is stronger:

14%. This is due to the lengthening of both vowels in [+focus, -final] position.

At syllable level only final syllables in sentence-medial words are significantly lengthened in the [+focus] condition. The pre-final (stressed) syllable, especially its consonant, is not significantly affected by focus. This is different from Chen (1984), who did find stressed syllable lengthening, especially in out of focus position.

The lengthening of consonants is small or non-existent. This may be explained by the fact that consonant length is phonemic in TB, which may prevent duration from being used as a correlate of stress (cf. Berinstein, 1979).

Podesva and Adisasmito-Smith (1999) found no duration-stress relation for TB vowels. However, both Podesva and Adisasmito-Smith (1999) and Chen (1984) did find a relation between pitch (but not intensity) and stress. In the next section I will investigate these accent/stress correlates in my material. I expect to find accent-leading pitch movements similar to the ones found in (western) stress languages.

Pre-boundary lengthening in TB does occur in the geminate consonant, but not in the singleton consonant. Possibly the duration of the singleton consonant is limited so as not to be confused with the geminate consonant.

In BM focus and boundary effects influence the word duration stronger than in TB. In BM words with full vowels focus affects the pre-final syllable strongly, especially the vowel. This indicates that duration is a prosodic cue for accent in BM (Wallace, 1976). Pre-boundary effects occur primarily in the final syllable, in particular the vowel.

In the schwa words, the pre-final syllables are not influenced by focus, nor by boundary. In contrast, the final syllables of schwa words are strongly affected by both focus and boundary. This indicates that the accent shifts to the final syllable when the pre-final syllable contains a schwa.

For BM, too, I will study the relevant pitch movements. As BM is a non-stress language I expect these to be smaller and more variable in shape and position than the TB accents.

3.5 Pitch analyses

For the melodic analysis each utterance in the database was subjected to a pitch extraction algorithm (autocorrelation method as implemented in the Praat software, Boersma and Weenink, 1996). Upper and lower frequency bounds were set manually for each speaker. Raw pitch curves were visually inspected and corrected by hand when the algorithm had erred.

3.5.1 Toba Batak

3.5.1.1 Stylization

For the analysis of the TB materials four pitch points in each target word were then located by eye, and their time/frequency coordinates were stored in a database for off-line statistical processing. The pitch points were found as the result of a data reduction technique that was developed at the Institute for Perception Research. In this so-called analysis-by-synthesis method ('t Hart, Collier and Cohen, 1990) the researcher replaces the original raw pitch curve of the target utterance by a straight-line stylization (fundamental frequency expressed in semitones or ERB – see below – as a function of linear time) such that perceptual equivalence is obtained between the original and the stylization (see also Nooteboom, 1997) using the smallest possible number of straight-line segments. The comparison between original and stylization is done by virtue of the PSOLA (Pitch Synchronous Overlay and Add, see e.g. Moulines and Verhelst, 1995) signal processing technique that affords the interactive manipulation of the fundamental frequency of an utterance (and even complete replacement or exchange of melodies between utterances) while good to excellent sound quality is maintained in the resynthesis. The result of the stylization is the reduction of the original, capricious pitch curve to a sequence of straight-line rises and falls. The point in the stylization where a rise changes into a fall (or vice versa) is called a pivot point, or just pivot. The stylization procedure is exemplified in figure 3.10 below. It should be noted that the overall trend of the sentence melody

is not level but slopes down gently. This so-called downtrend is indicated in figure 3.10 by a dotted line fitted by hand through the lower pivot points in the stylization (i.e. where a fall ends and a rise begins). Downtrend is a universal characteristic of human speech. It is most likely caused by the gradual reduction of subglottal air pressure over the course of an utterance (e.g. 't Hart et al., 1990) even though the speaker has a choice to reinforce or to counteract the effect through laryngeal muscle activity (e.g. Strik, 1994). The downtrend line as drawn in figure 3.10 acts as a baseline. Note that the sentence-terminal pitch, especially in statements and commands, tends to go below the baseline. As a result of this, sentence-final falls are often larger than earlier falls; if the vocal pitch reaches the baseline before the last syllable, there will be a noticeable drop in pitch during the last syllable. These effects are generally subsumed under the term 'final lowering'.

The pitch points were:

- p1 a low pitch at the beginning of a rise located in the penultimate syllable in the target word (i.e. the first syllable of a disyllabic target) or in the antepenultimate syllable of the target word (i.e. the first syllable in a tri-syllabic item). P1 is defined as the F0 minimum (i.e. lowest F0) from the beginning of the utterance onwards preceding the pitch peak on the penultimate syllable;
- p2 the peak F0 located in the penultimate syllable of the target word;
- p3 a pivot point between p2 and p4 that affords the stylization of pitch fall in terms of two straight-line segments, the first of which drops off at a modest rate whilst the second part embodies a steep fall. In a fair number of cases, and in fact in all non-final targets without focus, such a point could not be found; p3 was then left undefined;
- p4 end of the fall or F0 minimum between p2 and the end of the utterance. When the target was utterance final, p4 is typically the terminal pitch; in non-final targets without focus p4 could easily be located as the pivot point between the fall after p2 and the large rise marking the focused constituent following the non-focused target.

Figure 3.10 gives an example of an original F0 curve (capricious lines) and a close-copy stylization of an utterance in TB. The dotted line represents the baseline (downtrend, see above).

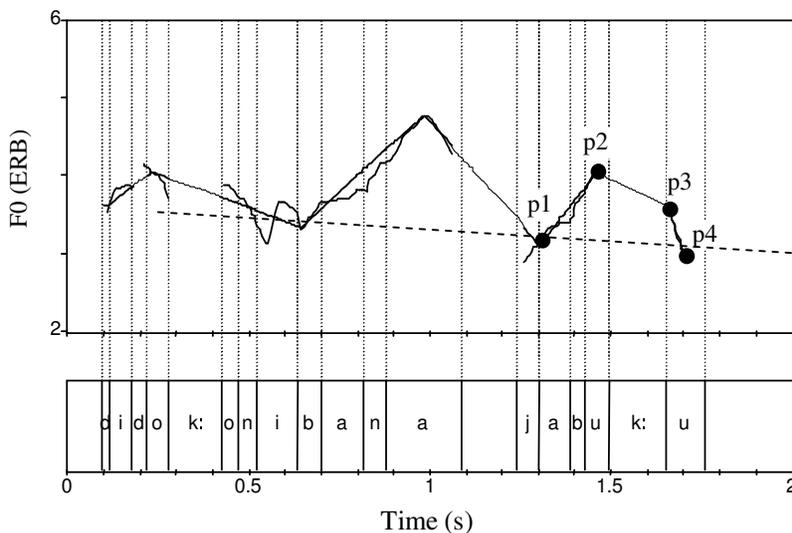


Figure 3.10. Original F0 curve (capricious lines) and close-copy stylization (solid straight lines) of the TB target word *jabukku* ('my house') in [+focus,+final] condition, spoken by a male TB speaker. The dotted line represents the downtrend or declination (see text).

As a first approximation I applied a minimal normalization procedure to the raw pitch data (the four pitch points). Given that two speakers were male and two female, some basic form of speaker normalization was unavoidable. The raw pitch data in hertz were first rescaled to Equivalent Rectangular Bandwidths (ERB units, cf. Hermes and van Gestel, 1991; Nooteboom, 1997; Ladd and Terken, 1995; van Heuven 2004), which is currently held to be the psychophysically most valid scale for comparing vocal pitches in intonation languages across registers. Conceptually, the ERB corresponds to the distance between locations of maximal excitation in groups of hair cells along the basilar membrane. Pitch intervals of equal sizes when

expressed in ERB should be perceptually equivalent regardless of their absolute frequency in hertz. As a rough indication, the typical male vocal pitch range in speech is between 3 and 5 ERB, and that of women between 5 and 7.

Inspection of raw pitch measurements revealed that the lowest recurrent pitch that could be found in the materials, was pitch point #4 (p4) in sentence-final position in [-focus] constituents. This point consistently has the lowest F0 in TB. All pitches were therefore rescaled to ERB and then expressed relative to the minimal reference pitch at p4. This allows straightforward comparison of pitch differences within and between utterances.

3.5.1.2 Results

In the selected TB target words the accent-lending pitch movement occurs on the penultimate syllable, i.e. the stressed syllable. In non-focused words pitch movements occur on the penultimate syllables as well, but there the excursions are rather small. Stress in unaccented words is realised with a smaller pitch obtrusion than in accented words. This agrees with the findings of Chen (1984) and Podesva and Adisasmito-Smith (1999) who both mention a relation between pitch and stress in TB. Pitch movements in the TB targets are generally realised with a rise-fall movement.

Figure 3.11 illustrates the pitch contour of all TB words in normalized F0 (ERB), broken down by sentence type. The x-axis shows the time scale relative to the onset of the penultimate vowel.

One-way ANOVAs with sentence condition as a four-level fixed factor indicated significant effects for several acoustic parameters. For instance, the timing of the peak [$F(3, 163) = 8.7, p < .001$], the height of the peak [$F(3, 163) = 23.5, p < .001$], the size of the rise [$F(3, 163) = 12.2, p < .001$] and of the fall [$F(3, 163) = 62.0, p < .001$] are significant. No significant effect of sentence condition was found for the beginning of the rise [$F(3, 163) = 1.9, p = .136$].

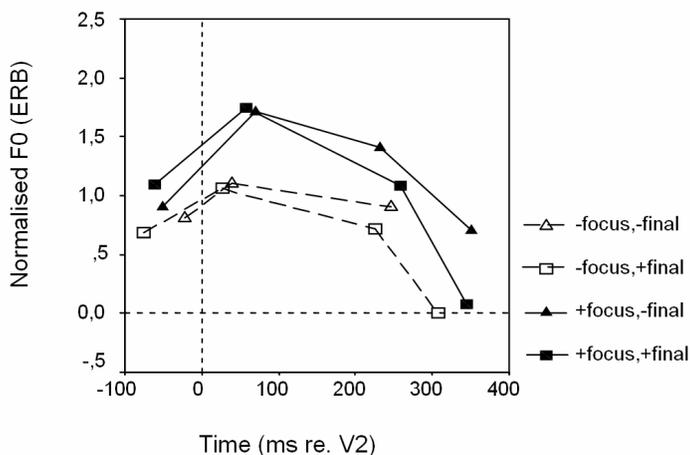


Figure 3.11. Pitch contour of all target words in normalized F0 (ERB), with the time scale relative to the onset of penultimate V, broken down by sentence type.

Two-way analyses with focus and final conditions as fixed two-level factors show that the effects of focus are often significant but that the effects of finality are not. The interactions between focus and finality are hardly significant. Focus condition affects many aspects of the melody, except for the F0 minimum at the end of the target word, which effect is insignificant [$F(1,163) = 3.6, p = .059$], and a small effect on the onset of the rise [$F(1,163) = 4.1, p = .044$]. The effect of focus is highly significant for the timing of the peak [$F(1,163) = 21.3$], the height of the peak [$F(1,163) = 66.0$], the size of the rise [$F(1,163) = 29.0$], the size of the fall [$F(1,163) = 91.8$], and the slope of the fall [$F(1, 163) = 52.9$], all with $p < .001$.

Most pitch movements are not much affected by finality. The effects of final condition on the rising movements are not significant, for the onset of the rise [$F(1,163) < 1$], for the peak timing, [$F(1,163) = 1.7, p = .194$], for the peak height [$F(1,163) < 1$], for the size of the rise [$F(1,163) = 3.6, p = .060$], and for the slope of the rise [$F(1,163) = 3.7, p = .057$]. However, there are highly significant effects on the fall movements for the size of the fall [$F(1,163) = 94.8$], for the slope of the

fall [$F(1,163) = 53.0$] and for the F0 minimum at the end of the target word [$F(1,163) = 156.8$], all with $p < .001$. The interaction between the two effects is significant only for the final F0 with [$F(1,163) = 10.0, p = .002$].

In Table 3.2 the results of the various measurements (means and standard deviations) are given for each of the four sentence conditions.

The results show that focus affects the pitch movements. In sentence-final position the rise starts later in the [+focus] words, and reaches the highest pitch later as well, than in the -focus words. Sentence-medially, the rise starts earlier in focused words, and reaches the highest pitch later than in the [-focus] words. Accented words have significantly higher pitch peaks than unaccented words, the difference amounting to some 0.6 ERB. The excursion sizes of the [+focus] rises are about twice as large (in ERB) as those of their [-focus] counterparts. Also, the slopes of the [+focus] rises are considerably steeper. There is no systematic difference between plus and minus focus falls in terms of final pitch, but the [+focus] falls generally have larger excursions and steeper slopes.

Table 3.2. Mean of pitch accent measurements in eight TB words per sentence condition with standard deviation (in parentheses), and the mean across all conditions.

Measurements	+focus, +final	+focus, -final	-focus, +final	-focus, -final	Mean
Onset timing (ms)*	-61,50 (52,00)	-51,2 (48,00)	-74,70 (35,00)	-21,60 (31,00)	-61,30
Peak timing (ms)*	66,60 (36,00)	83,1 (60,00)	36,90 (29,00)	40,70 (21,00)	64,80
F0 peak (norm. ERB)	1,76 (0,51)	1,73 (0,47)	1,04 (0,28)	1,11 (0,34)	1,57
Rise exc. (ERB)	0,66 (0,36)	0,82 (0,40)	0,36 (0,26)	0,44 (0,18)	0,65
Slope rise (ERB/s)	5,48 (3,10)	7,02 (4,70)	3,59 (2,60)	4,57 (2,80)	5,60
Final F0 (norm. ERB)	0,09 (0,31)	0,70 (0,48)	0,00 (0,17)	1,03 (0,31)	0,37
Fall (ERB)**	1,67 (0,54)	1,02 (0,44)	1,04 (0,34)	0,08 (0,15)	1,20
Slope fall (ERB/s)	-6,52 (2,70)	-4,19 (2,30)	-4,19 (1,90)	-0,42 (0,81)	-4,77

*) Relative to the onset of the penultimate vowel

**) From the peak (p2) to the F0 terminal (p4)

Accent-lending rises start on average 57 ms before the onset of the pre-final vowel, with the slope of around 6 ERB/s. The peak is reached 74 ms after the onset of the vowel in the pre-final syllable, with the F0 maximum at 1.74 ERB. After the peak, pitch goes down gradually to the final-syllable and then drops again to the end of the word.

Boundary marking affects the fall movements. In sentence-final position, the fall excursions are significantly larger than in their [-final] counterparts. The falls in sentence-final words are also steeper than the falls in sentence-medial words. The pitch movements in the penultimate syllables, which are stressed, depend on the focus condition of the words. Irrespective of focus the presence of a final boundary determines the fall movements in the final syllable. At the end of the utterance, the fall is larger and reaches the baseline.

3.5.2 Betawi Malay

Preliminary auditory and visual inspection of the BM materials revealed that pitch movements were in general completely absent when the target word was not in focus. Pitch movements were observed on [+focus] targets but these could occur in either the final or the pre-final syllable in the target word, primarily depending on the target's position in the sentence. When the focused target occurred in sentence-final position, the pitch movement seemed to coincide with the final syllable; accented sentence-medial targets typically carried the pitch movement on the pre-final syllable. The shape of the accent-lending movement could be a rise, a fall or a rise-fall combination. Simple rises always occurred on final syllables, simple falls on pre-final syllables; rise-fall combinations were found in both final and pre-final positions (depending on the position of the target in the sentence). These findings seem to be in line with the view that BM has no word stress but phrasal accent only (Wallace 1974); the distributional details were in fact predicted by Kähler (1966). A detailed presentation of the auditory analysis will be provided below (§ 3.5.2.2).

3.5.2.1 Stylization

In view of the variability in the occurrence and shapes of the pitch movements, some refinements of our stylization point p1 and p2 (as defined above for TB contours) were in order. The definitions for pitch points p3 and p4 were left unchanged.

- p1 As before, this is the low pitch at the beginning of a rise associated with an accented target word. It is defined as the latest F0 minimum (i.e. lowest F0) preceding the pitch peak on the target. However, when no pitch rise could be seen on the target word (as happened in [-focus], i.e. unaccented words) p1 was considered to be absent.
- p2 This point is defined as the F0 maximum (i.e. the highest F0) in the target word. However, in unaccented BM targets (without any pitch rise) p2 was located at the onset of the target.

Figure 3.12 gives an example of a stylization of a BM speech spoken by a BM male speaker in [+focus, +final] condition.

The raw pitch measurements were again normalized (in ERB re. the terminal F0 in [-focus, +final] targets) and aligned relative to the onset of the target in exactly the same way as was done for TB.

However, before we turn to a presentation and discussion of the acoustical analysis we will first present the results of an auditory screening of the BM materials, such that the acoustical analysis can be done separately for the three types of pitch movement (rise, fall, rise-fall) that we found in the BM recordings (see above). The next section describes the procedure and results of the auditory screening by a panel of expert listeners.

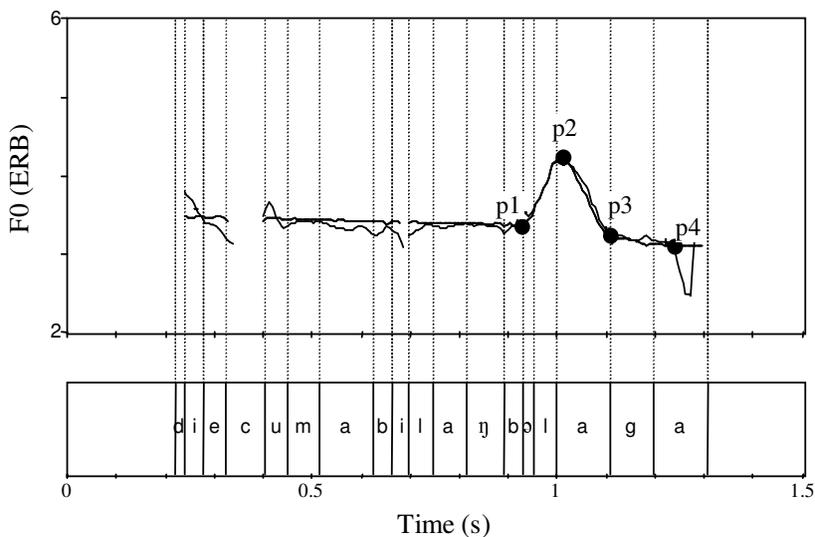


Figure 3.12. Original F0 curve (capricious lines) and close-copy stylization (solid straight lines) of the BM target word *belaga* 'pretend' in [+focus,+ final] condition, spoken by a male speaker.

3.5.2.2 Auditory inspection

Preliminary inspection of the [+focus] utterances revealed a variety of positions of accent-lending pitch movements. These occurred either on the pre-final or on the final syllable but it was not always straightforward which of these two syllables was accented, nor which part of the pitch movement should be seen as accent lending.

In order to be able to make reasoned decisions in this matter, a formal listening test was conducted. In a preliminary screening of the recordings, the present author, who is a native speaker of BM, listened to the materials and decided for each of the 192 utterances (4 speakers \times 2 conditions \times 8 words \times 3 repetitions) whether the speaker had produced an accent on the [+focus] target word. In all, 33 out of the total of 192 utterances contained target words on which a pitch movement was clearly absent. There was no pattern in the distribution of absent pitch movements; these were scattered more or less at random over the various conditions and

repetitions, such that, in fact, every condition in the design was still represented with at least two exemplars by each speaker. Table 3.3 summarises the result of the preliminary screening for the valid cases of the [+focus] words, done by the author. The table specifies the number of [+focus] target words in sentence-final and medial position for each of the four speakers separately, broken down by perceived presence versus absence of a prominence-lending pitch movement.

Table 3.3 shows that the male BM speakers quite consistently realized accent-lending pitch movements on the targets. The female speakers were less consistent: in particular female 2 omitted pitch movements. The BM male speakers were better in doing their task compared to the BM female speakers. Both female speakers dropped the pitch movements in some 20 to 35 percent of sentence medial targets; female 2 even dropped her pitch movements in sentence-final position in more than 50 percent of the cases. I decided to exclude the 33 utterances without audible accent on the [+ focus] target word from further analysis.

Table 3.3. Number of [+focus] Betawi Malay target words realized with and without an audible accent, broken down by position in the sentence (final, medial).

sentence condition	speaker	N perceived as	
		Accented	Non-accented
+focus,+final	male1	22	2
	male2	22	2
	female1	22	2
	female2	11	13
+focus,-final	male1	23	1
	male2	24	0
	female1	16	8
	female2	19	5

As a next stage in the auditory screening, a panel of two Dutch experts on prosody and the present author, listened to the remaining 159 [+focus] utterances and gave their judgments with respect to syllable prominence. They indicated, independently of each other, for each correctly spoken utterance (i.e. with target words bearing an audible accent only) whether they found the pre-final or the final syllable of the target the most prominent or whether they considered both syllables equally prominent. The 159 correctly pronounced utterances (192 – 33 tokens without an audible accent) judged by three listeners yielded 477 valid cases (231 [+final] and 246 [-final] cases).

It occurred to us while performing the auditory analysis that the judgment of prominence was more difficult than the earlier determination of the presence versus absence of a pitch movement on the target. Apparently, the presence of a pitch movement – which was easy to detect as such – did not necessarily contain sufficient cues to enable listeners to accurately pinpoint a syllable that is made prominent by the pitch movement. Therefore I will first briefly analyse the degree of agreement among the judges.

Table 3.4 contains three matrices, one for each pair of judges that can be formed from the group of three. In each matrix the number of responses ‘prominence on pre-final syllable’, ‘prominence on final syllable’ and ‘equal prominence’ is cross-tabulated for the two listeners. The three cells along the main diagonal of the matrix contain the cases where the judges agree; off-diagonal cells contain the number of cases in which the judges disagree. A convenient statistical measure used to express the degree of agreement between two judges is Cohen’s kappa coefficient κ (Cohen, 1960, Streefkerk, 2002). This is a number that ranges between 0 (no agreement at all) and 1 (perfect agreement).

Table 3.4 shows the agreement in prominence assignment between the listeners. Listener 1 and listener 2 are the Dutch experts, and listener 3 is the native BM speaker. Although the kappa coefficients are not very high, they appear to be well within the range that can be expected for this type of judgment. Earlier work by Streefkerk (2002: 28, table 2.8 upper half) on the perception of prominence in connected Dutch speech (in terms of a simple dichotomy between prominent versus

non-prominent) by lay listeners (students at the humanities faculty) revealed kappa values between $\kappa = .21$ and $\kappa = .75$ with a mean of $\kappa = .48$ (standard deviation .14). Moreover, our listeners were required to make a ternary decision rather than a dichotomy, which makes it more difficult to obtain high agreement among the judges.

Table 3.4. Cross-tabulation of judgements 'stress on pre-final syllable', 'stress on final syllable' or 'equal stress' by pairs of expert listeners for 159 Betawi Malay targets bearing an audible pitch movement. Kappa values are indicated.

Listener 1	Listener 2				$\kappa = 0.57$
	pre-final	Final	equal	Total	
Pre-final	66	13	10	89	
final	0	47	5	52	
equal	2	11	5	18	
Total	68	71	20	159	

Listener 1	Listener 3				$\kappa = 0.48$
	pre-final	Final	equal	Total	
Pre-final	71	14	4	89	
final	11	40	1	52	
equal	1	15	2	18	
Total	83	69	7	159	

Listener 2	Listener 3				$\kappa = 0.49$
	pre-final	Final	equal	Total	
Pre-final	60	5	3	68	
Final	15	52	4	71	
equal	8	12	0	20	
Total	83	69	7	159	

In view of the above we argue that all three pairs of judges reacted to the materials with sufficient agreement. At the same time, however, it is clear that one cannot rely on a single judge; therefore the best solution is to use multiple judges and compute some mean or aggregate score. Still, it is clear that the two Dutch judges are in stronger agreement with one another than either of these agrees with the present author (and native listener of BM). The results seem to show that the native listener is more prone to make a decision for perceiving prominence on either of the two final syllables than the non-native experts, who are relatively often undecided in their prominence judgment. Be this as it may, the level of agreement is such that I felt safe to accumulate the judgments over the three experts, as is done in table 3.5. This summarises the results of the prominence test for the [+focus] targets, for full-vowel words and schwa words, broken down by position of target in the sentence.

Table 3.5. Relative frequency (%) of perceived prominence on pre-final versus final syllable of accented targets for BM target words in sentence final and medial position.

Target position	Vowel type	Prominence perception (%)		
		pre-final syll	final syll	equal in both syll
sentence-final	V – V	48	45	7
	ə – V/ə	9	86	5
sentence-medial	V – V	80	11	9
	ə – V/ə	42	41	17

It is obvious from Table 3.5 that in sentence-final position the perception of prominence in words with full vowels is distributed equally over the pre-final and final syllables. Words with schwa in the pre-final syllable, however, are more prone to have accent on the final syllable, regardless of whether this final syllable does or does not contain a schwa. Sentence-medially the pre-final syllable with schwa can also be accented; pre-final syllables with full vowels tend to be accented.

According to Wallace (1970), the accent is usually on the penultimate syllable of the phrase-final word in BM. The results of the auditory perception test can supplement this as follows:

- I. Accent falls on the penultimate syllable (100%).
- II. If the penultimate syllable contains a schwa, accent shifts in 50% to the ultimate syllable (50% penultimate, 50% ultimate).
- III. If the target is in sentence-final position, accent shifts in 50% to the ultimate syllable (50% penultimate, 50% ultimate).

With these rules accent should fall 100% on the ultimate syllable if the target has a schwa in the penultimate syllable and is in sentence-final position.

The rules might seem to reveal a propensity for BM to shift accent position due to the type of the vowel in the syllable. However, accent shifts because of the position of the target in the sentence. This shift of perceived accent shows that BM has phrasal accent only; it has no word stress with which the phrasal accent can be aligned.

The following step in the auditory analysis (in § 3.5.2.3) was to establish the frequency of occurrence of specific types of accent-lending pitch movements on the prominent syllables. We will then be in a position to determine the shapes of the various pitch configurations in acoustical terms. This will be described in § 3.5.2.4.

3.5.2.3 Token frequencies of BM accent-lending pitch movement types

For the next part of the data analysis we will make a distinction between four types of pitch movement on the targets using visual criteria. These are the simple rise and simple fall, and complex rise-fall configurations which were subdivided into early versus late alignment. For early alignment the pitch peak (pitch point p2) should be located within the confines of the penultimate syllable, for late alignment the peak finds itself in the final syllable. The next tables present a cross-tabulation of the four shapes of the pitch contour over the final two syllables of the [+focus] targets

(collapsed over sentence-medial and final positions as well as over all stimulus words and speakers) against the position of the syllable that bears the accent-lending pitch movement (table 3.6a) and against the position of prominent syllable (table 3.6b), as determined by the listening panel.

Table 3.6. Number of accent-lending pitch configurations (a) and perceived prominences (b), heard by two Dutch experts and the present author in final and pre-final syllables in Betawi Malay [+focus] target words broken down by type of movement (F: simple fall, R: simple rise, RF pre-final: rise-fall with peak in pre-final syllable, RF final: rise-fall with peak in final syllable).

a. Perception of pitch-accented syllable.

Shape	Pitch movement heard				Total
	None	pre-final	Final	both/equal	
F	13	131	9	15	168
R	10	17	103	20	150
RF pre-final	2	68	4	4	78
RF final	3	5	64	9	81
Total	28	221	180	48	477

b. Perception of prominent syllable.

Shape	Prominence heard			Total
	pre-final	Final	both/equal	
F	134	17	17	168
R	27	100	23	150
RF pre-final	72	4	2	78
RF final	7	71	3	81
Total	240	192	45	477

The results show, first of all (Table 3.6a), that in 28 cases the panel of listeners could not detect a pitch movement on the target word – even though the author had judged earlier that the target did bear an accent. These 28 cases probably make up a separate category of audible accents that are not marked melodically but, for instance, temporally; these cases will not be analysed as part of the present study. Next, there is a relatively small group of tokens that were judged to bear equal prominence on the final and pre-final syllables (less than 10% of the prominence judgments are in this category); these, too, will not be analysed any further.

For the remaining cases, there is a very strong association between the type of pitch movement and the position of the prominent syllable. If the movement is a simple fall, the prominence is on the pre-final syllable, if it is a simple rise, then the prominence is typically on the final syllable. When the pitch movement is a complex rise-fall configuration, about half of the tokens are perceived with prominence on the pre-final syllable, and the other half with final prominence. From the stylization it was found that the rise-fall could indeed occur in the pre-final syllable (26) and final syllable (27). In sentence-final position there are more rise-fall configurations found in the final syllable (26) than in the pre-final syllable (7). However, in sentence-medial position rise-fall occurs more often in the pre-final syllable (19) than in the final syllable (1).⁴

The position of the prominent syllable depends not only on the position of the target in the sentence, but also on the type of vowel in the target word, and certainly on the shape of the curve. Final prominence, of course, is predicted when the pre-final syllable contains schwa; pre-final prominence is what we typically find when the pre-final syllable contains a full vowel. As a consequence of this, simple falls typically occur on pre-final full vowels, and simple rises are found on final vowels after schwa.

However, a simple fall in a pre-final schwa-word *pete* [pəte] is perceived as having prominence on the final syllable.⁵ Prominence, therefore, tends to be perceived sooner on the syllable that contains a full vowel. From the temporal

⁴ See appendix 1a for the table of shapes.

⁵ See appendix 1b for the complete cross table of shapes against prominence perception.

investigation it was found that the length of the final syllable in words with schwa in the pre-final syllable is strongly affected by both focus and boundary. In such words the final syllable is lengthened strongly when words are accented and in sentence-final position. In the same condition, accent in *pete* shifts thus to the final syllable. On the other hand, simple rises in full-vowel words were not always perceived as having prominence on the final syllable in pre-boundary targets. The results of the temporal investigation indicated that in words with full vowels focus affects the duration of the pre-final syllable strongly. The lengthened pre-final syllables are thus sometimes perceived as more prominent.

In the following section we will present the acoustical analysis of the pitch contours on the BM utterances. The acoustical analysis will first concentrate on the pitch configuration as found on accented targets (as judged by the author) only. In the final subsection we will also consider the (basically flat) pitch pattern that was found on [-focus] targets.

3.5.2.4 Acoustical properties of BM accent-lending pitch configurations

The means comparisons with word position, vowel type and movement shape in the independent list and pitch parameters in the dependent list were run. The ANOVA tables⁶ show that the effects of word position in the sentence are highly significant on the peak timing⁷ [$F(1,157) = 25.9, p < .001$], on the height of the peak [$F(1,157) = 41.6, p < .001$], and on the fall excursion [$F(1,107) = 30.4, p < .001$]. A significant effect of word position is also found on the rise onset [$F(1,101) = 5.2, p = .025$] and on the slope of the rise [$F(1,101) = 4.1, p = .045$]. In sentence-final words the rises start 162 ms before the onset of the vowel; sentence-medially rises start later, about 117 ms before the onset of the vowel. In sentence-final words, the F0 peak is reached 69 ms after the vowel onset and in sentence-medial words earlier, about 33 ms after the vowel onset. However, the peak in sentence-medial words is about .66 ERB steeper than that in sentence-final words. The slope of the rise is thus

⁶ See appendix 2 for the complete table of significance.

⁷ Relative to the onset of the vowel in the accented syllable.

significantly larger in sentence-medial words, with a difference of about 2 ERB/s. Also, the fall excursion in sentence-medial words is larger (.69 ERB) than the fall excursion in sentence-final words.

Effects of vowel type are significant only on the slope of the rise [$F(1,101) = 4.4, p = .039$] and the slope of the fall [$F(1,107) = 4.3, p = .040$]. Words with full vowels have on average steeper rises (2.1 ERB/s) and steeper falls (1.9 ERB/s) than words with schwa.

The effects of movement shape are highly significant on the rise onset [$F(3,99) = 10.2, p < .001$], on the peak timing [$F(3,155) = 18.7, p < .001$], and on the rise excursion [$F(3,99) = 5.6, p = .001$]. Furthermore, significant effects of movement shape are found on the steepness of the rise [$F(3,99) = 2.9, p = .038$], and on the fall excursion [$F(3,105) = 3.0, p = .036$].

Based on these results I will analyse the sentence-final words and the sentence-medial words separately. In every section, the F0 parameters of every shape will be analysed. I will not separate the target words with a schwa vowel from the target words with only full vowels since significance effects of vowel type occur only on two parameters and the significance levels are not high.

I will present first the F0 contours of target words in sentence-final position. Figure 3.13 and figure 3.14 (target words in sentence-medial position) plot the stylised F0 contour in normalised ERB (see § 3.5.2.1) as a function of time, such that the timing of the movements is expressed relative to the onset of the final vowel in the target word. In these figures rises, falls and rise-fall combinations are plotted separately. Rises are always aligned with final syllables, falls with pre-final syllables. Rise-fall combinations were separated into two alignment groups: those that were heard as imparting prominence to pre-final syllables and those that were heard with prominence on the final syllable. As a consequence of this redefinition of movement types, subsequent data analysis, again using ANOVA, will involve a four-level factor for movement type.

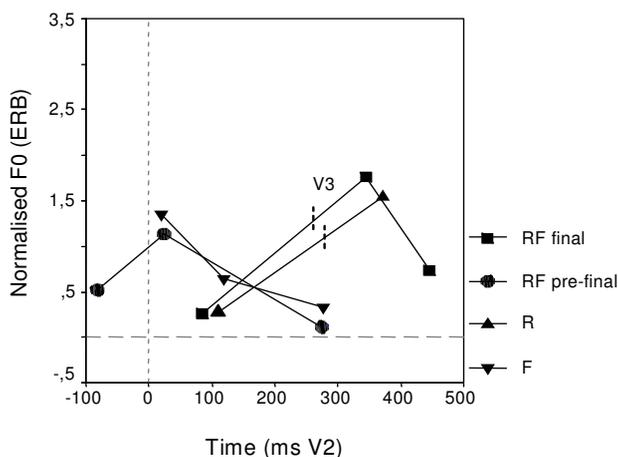


Figure 3.13. F0 contours of BM [+focus] words in sentence-final position in normalised ERB with on the x-axis a time scale relative to the onset of the penultimate vowel. The vertical dashes on the RF-final and the R contours are the onset of the final vowel.

Figure 3.13 shows that sizes of the different shapes are to some extent different from each other. Typically, a movement – whether fall or rise-fall – on the pre-final syllable has a small excursion size and does not reach a high peak frequency. In contrast to this, movements in final syllables, whether rise-fall or just a rise, are characterized by a very large excursion size leading to a high F0 peak. Notice that in the case of the rise, and even in the case of a rise-fall on the final syllable the pitch does not drop down to the baseline but remains 0.7 (F0 final for rise-fall) and 1.5 (F0 peak for rise only) ERB above it. Movements aligned with the pre-final syllable, however, end at baseline level. Analyses of variance for the sentence-final words, with shape of movement as a (four-level) fixed factor showed significant effects on the peak timing [$F(3, 73) = 13.4, p < .001$], the rise excursion [$F(3, 58) = 5.9, p = .001$] and on the slope of the fall [$F(3, 44) = 5.4, p = .003$]. T-tests with two movement shapes, the rise-fall in pre-final and that in final syllable as independent variables, show that there are significant differences between the two movements in

terms of peak height [$t(31) = -2.05$, $p = .049$], peak timing⁸ [$t(31) = -3.47$, $p = .002$], rise excursion [$t(31) = -4.04$, $p < .001$], and slope of the fall [$t(31) = -2.74$, $p = .010$]. In the pre-final syllable the F0 peak is reached 25 ms after the onset of the vowel. In the final syllable the peak is reached later, at about 84 ms after vowel onset. Also, the peak is about .63 ERB higher in the final syllable than when it occurs on the pre-final syllable. The rise excursion of the final-syllable rise-fall is thus larger, by about .88 ERB, than the pre-final syllable rise-fall. However, the fall of the rise-fall movement in the final syllable is 5.7 ERB/s steeper than the fall in the pre-final syllable. In the final syllable the pitch goes down very quickly from the highest point to the next point before the utterance is ended. The complete results of the measurements are summarised in table 3.7.

It seems that the canonical shape of the accent-lending pitch configuration in BM consists in a rise-fall combination, which can occur either on the pre-final or on the final syllable of the [+focus] target word. When the rise-fall is on the pre-final syllable it imparts prominence to that syllable. The rise portion may be absent from the contour (i.e. if the preceding context ends in a high pitch) but the temporal alignment of the fall is not affected by the presence or absence of the rise. Importantly, the fall is always complete and reaches the low declination line around the onset of the final syllable. Due to the severe time constraints on the rise and fall of the configuration on pre-final syllables, the excursion size of the movements is small: the configuration is scaled down. When the rise-fall is executed on the final syllable – which is then perceived as prominent – there seems to be no time constraint. The final syllable, also as a consequence of pre-boundary lengthening (see § 3.2) provides ample space for large movements. Typically the rise portion of the configuration takes up some 200 ms, and during this time interval the pitch rises by roughly a full ERB. The rise is often, but by no means always, followed by a fall, which, however, never reaches the lower declination line (and final lowering seems to be conspicuously absent). Apparently, BM speakers choose to truncate, rather than scale down, rise-fall configurations in final position.

⁸ The peak timing on the final syllable is measured relative to the onset of the vowel in the final syllable.

Table 3.7. Means of pitch accent measurements in BM words per movement shape, in sentence-final position, with standard deviation (in parentheses), and the means across all movements.

Measurements	Fall	Rise	RF pre-final	RF final	Mean
Onset timing (ms)*		-168,10 (115,00)	-81,10 (24,00)	-176,40 (90,00)	-161,70
Peak timing (ms)*	19,10 (12,00)	92,60 (54,00)	24,70 (21,00)	84,10 (43,00)	69,20
F0 peak (norm. ERB)	1,35 (0,50)	1,54 (0,53)	1,13 (0,58)	1,76 (0,75)	1,54
Rise exc. (ERB)		1,26 (0,46)	0,62 (0,27)	1,49 (0,55)	1,29
Slope rise (ERB/s)		5,71 (3,70)	5,83 (2,00)	6,27 (2,60)	6,00
Final F0 (norm. ERB)	0,33 0(,61)		0,11 (0,14)	0,73 (0,97)	0,52
Fall (ERB) **	0,91 0,41)		1,00 (0,59)	1,02 (0,60)	0,98
Slope fall (ERB/s)	5,33 (3,70)		4,77 (2,80)	10,54 (5,30)	8,10

*) Relative to the vowel onset of the accented syllable.

**) From the peak (p2) to the next lower point.

Generalizing further, the small scaled-down rise-fall configurations occur on pre-final syllables that have mostly full vowels. Pre-final syllables with schwa do not carry prominence; prominence is then pushed onto the final syllable. The only exception is *rejeki* [rəjəki], where the pre-final syllable could carry the prominence due to a clear rise-fall pitch movement. If the final syllable contains a full vowel, it usually provides ample space for a large rise-fall configuration, which however, is truncated halfway during the fall portion. When both the pre-final and the final syllables contain schwa (in *deket* 'nearby'), there is much less space for the rise-fall configuration; as a result the entire fall portion is dropped.

In sentence-medial position (figure 3.14) there is only one case of a rise-fall configuration in the final syllable (in *rejeki*), which is different from the pre-final rise-fall. Figure 3.14 illustrates the four pitch movements in sentence-medial position.

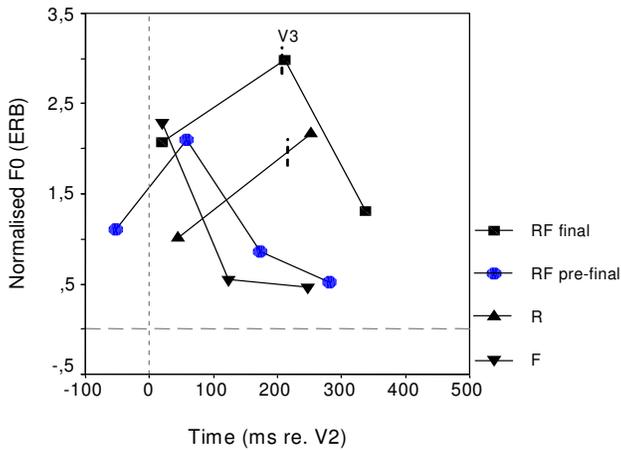


Figure 3.14. F0 contours of BM [+focus] words in sentence-medial position in normalised ERB with on the x-axis a time scale relative to the onset of the penultimate vowel. The vertical dashes on the RF-final and the R contours are the onset of the final vowel.

No statistical comparisons could be made for the rise-fall in the final syllable (one case only). One-way ANOVA's with the shape of the movement (excluding the rise-fall in the final syllable) as a fixed factor showed that effects of shape of movement are significant only for the rise-onset timing [$F(3,37) = 9.0, p < .001$] and for the peak timing [$F(3,78) = 5.9, p = .001$]. However, a post-hoc test (Scheffé, with $p = .05$) indicated that the peak timing is significantly different only between the fall and the rise-fall in the pre-final syllable. Table 3.8 summarises the pitch measurements of the accented words in sentence-medial position.

The pitch movements in sentence-medial words differ to some extent from the pitch movements in sentence-final position. The falls in sentence-medial words start from significantly higher pitches than those in sentence final words [$F(1,54) = 22.1, p < .001$], with a mean difference of ca. .93 ERB. Sentence-medial falls are significantly larger [$F(1,54) = 21.9, p < .001$], around .87 ERB, and 3.9 ERB/s steeper [$F(1,54) = 10.9, p = .002$], than the sentence-final falls.

Rise movements differ only in the peaks: the sentence-final rises have later peaks ($\Delta = 54$ ms) than the sentence-medial rises [$F(1,48) = 15.9, p < .001$]. Also,

the sentence-medial rises have significantly higher peaks ($\Delta = .63$ ERB) than the sentence-final rises [$F(1,48) = 18.0, p < .001$].

Pre-final rise-fall movements are affected by boundary only in terms of peak height and slope of the fall part. Peaks in sentence-final words are significantly lower than in sentence-medial words [$F(1,24) = 8.6, p = .007$], with a difference of roughly a full ERB. The slope of the fall portion of the pre-final rise-fall is 3.7 ERB/s steeper in sentence-medial words than in sentence-final words [$F(1,24) = 4.7, p = .041$].

Table 3.8. Means of pitch accent measurements of BM words per movement shape, in sentence-medial position, with standard deviation (in parentheses), and the means across all movements.

Measurements	Fall	Rise	RF pre-final	RF final	Mean
Onset timing (ms)*		-171,30 (81,00)	-53,20 (60,00)	-187,50 -	-117,00
Peak timing (ms)*	20,60 (14,00)	38,20 (28,00)	57,30 (57,00)	5,40 -	33,00
F0 peak (norm. ERB)	2,28 (0,70)	2,17 (0,58)	2,09 (0,78)	2,98 -	2,22
Rise exc. (ERB)		1,15 (0,49)	0,98 (0,63)	0,91 -	1,06
Slope rise (ERB/s)		5,91 (2,80)	10,46 (9,20)	4,70 -	8,00
Final F0 (norm. ERB)	0,46 (0,64)		0,52 (0,52)	1,31 -	0,50
Fall (ERB)**	1,78 (0,41)		1,47 (0,84)	1,67 -	1,68
Slope fall (ERB/s)	9,25 (4,00)		8,46 (4,10)	13,37 -	9,10

*) Relative to the vowel onset of the accented syllable.

***) From the peak (p2) to the next lower point

Accent-lending pitch movements in sentence-medial words seem larger and steeper than those in sentence-final position. The peaks of the accent-lending pitch movements in sentence-medial words are higher than in sentence-final words. The declination effect explains the lower pitches in pre-boundary position. On the other hand, sentence-final rises need more time to reach the centre of the accented (final) syllable, which are longer than that in sentence-medial words as a consequence of pre-boundary lengthening. Rise movements in sentence-final words are thus less steep than in sentence-medial words.

3.5.2.5 Pitch in [-focus] BM targets

Listening tests on [-focus] targets were not deemed necessary as preliminary auditory and visual inspection indicated that there were no pitch movements in the [-focus] targets. The pitch contours of [-focus] words show only slight falls. Figure 3.15 illustrates the pitch movements in [-focus] targets compared with the [+focus] pitch movements; in the figure the horizontal axis has been time-normalised by plotting equidistant steps between successive pitch pivot points.

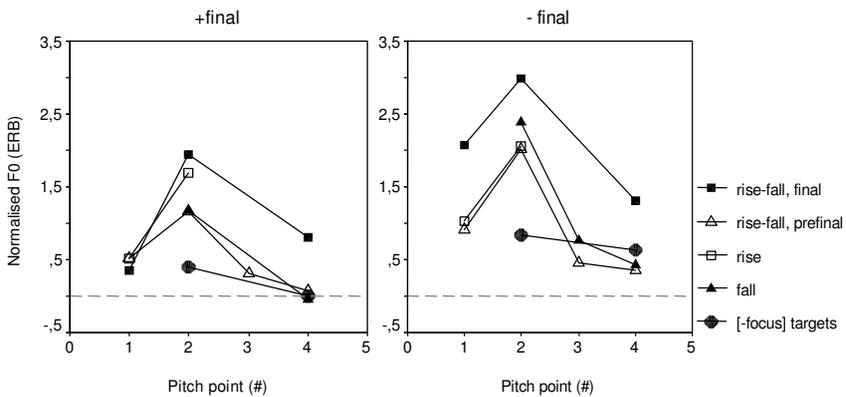


Figure 3.15. Pitch curves of [-focus] (black circles) and [+focus] targets (other marks) in [+final] position (left-hand panel) and in [-final] position (right-hand panel), with pitch points as a time-normalised x -axis.

Figure 3.15 shows clearly that there is no accent-lending pitch movement in [-focus] targets. In [-focus] words pitch point #2, which is the highest F0 in the target word, is lower than in [+focus] targets. To show that, I compare the pitch point #2 and #4 of the non-focus targets with those of the lowest pitch curve of the focus targets, i.e. the pre-final rise-fall. An independent t-test was therefore executed, separated into the finality conditions, with the F0 of point #2, F0 of point #4 and the fall excursion as test variables and the two curves (non-focus fall and pre-final rise-fall in focus targets) as the grouping variables. The results show that the values of the F0 and the fall excursions are significantly different from each other. In sentence-final position

the difference between F0 point #2 in [-focus] and the lowest peak in [+focus] targets, the peak of the pre-final rise-fall, is about .66 ERB [$t(88) = 3.65, p < .001$]. Sentence medially this difference is even larger, 1.25 ERB [$t(104) = 8.55, p < .001$]. The F0 terminal (pitch point #4) in [+focus] words with falling pitch or with a pre-final rise-fall is equal to the F0 terminal of the [-focus] words [$t(88) = .46, p = .64$ in sentence final position; $t(104) = -.84, p = .40$ in sentence medial position]. The fall excursion in the non-focused words is significantly smaller than that in the focused words, with a difference around .55 ERB sentence finally [$t(88) = 2.58, p = .012$], and 1.37 ERB sentence medially [$t(104) = 9.44, p < .001$].

The melodic structure of unaccented words is also affected by boundary marking. Figure 3.16 plots the pitch configuration of the [-focus] words in sentence-final and sentence-medial position. Figure 3.16 shows that unaccented words in sentence-final position are lower by about .5 ERB than in sentence-medial position. The excursion size of the fall is very small and suggests that the fall is entirely due to declination. There is no difference in the steepness of the fall slope between medial and final targets. The larger excursion on final falls is caused by the fact that final falls are longer than medial falls. This state of affairs is born out by a series of one-way ANOVAs which show significant effects of finality on the fall excursion [$F(1,166) = 5.1, p = .025$]. The fall excursion is about .5 ERB for the [+final] words, and .2 ERB for the [-final] words. However, finality does not affect the slope of the fall [$F(1,166) = 1.2, p = .269$].

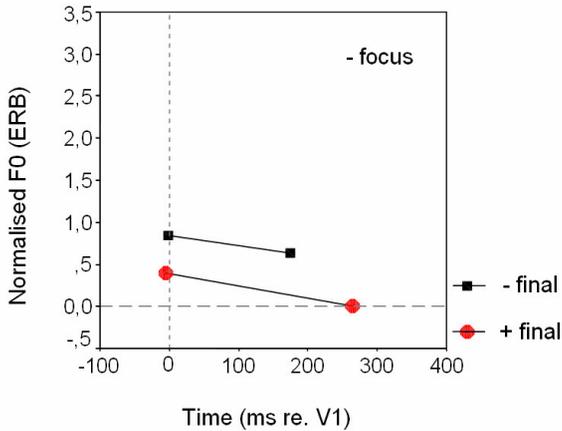


Figure 3.16. Pitch contours of unaccented words in sentence-final and sentence-medial position. The *x*-axis refers to the time (in ms) relative to the onset of penultimate vowel.

3.5.2.6 Pitch accent in Betawi Malay

The results of the pitch analyses of the [+focus] targets show clearly that accents in BM can fall on pre-final or final syllables. Pitch movements are either rises, falls, or rise-fall configurations. These findings indicate that stress position in the word is free in BM – at least within a two-syllable window at the end of the word.

The melodic structure of BM words is influenced by both focus condition and pre-boundary position. Focus yields accent-lending pitch movements; non-focus suppresses pitch movements. The presence of a sentence boundary following the target word determines the position of the accent-lending pitch movements within the target: it attracts the accent to the ultimate position; without a following sentence boundary the accent remains on the penult. Wallace's claim that accent is on the pre-final syllable is in line with my findings for sentence-medial words only, but clearly clashes with my results for sentence-final words. On the other hand, Wallace's claim that words with schwa in the pre-final syllable are generally accented on the final

syllable is found true in sentence-final words only. My results show that accent shifts not only due to the type of the vowel (full versus schwa), but also due to the position of the word in the sentence.

Pre-boundary position affects the pitch of the words significantly. Target words in sentence-medial position are generally realised at higher pitches than target words in pre-boundary position, regardless of the focus condition of the targets. The F0 terminal of target words in pre-boundary position typically reaches the base line.

3.5.3 Melodic structures of Toba Batak and Betawi Malay

As explained in the introduction, focus effects were expected to be stronger in TB than in BM whereas boundary effects should be of roughly equal magnitude in both languages. The results of our analyses show that focus and boundary affect the melodic structures of words in both languages in different ways.

The results of the pitch analyses show primarily that pitch movements occur in TB, even on non-prominent words – in which case the sizes of the pitch movements were smaller but by no means absent. In BM pitch movements do not occur at all in non-prominent words; BM pitch movements occur only on prominent words, and even then their occurrence depends on the position of the target word in the sentence. We interpret the fact that TB, as opposed to BM, has pitch movements on out of focus words as an indication that TB has stress at the word level and BM has not.

The finding by Podesva and Adisasmito-Smith (1999) that there is a relation between pitch and stress in TB, is confirmed in the present research. Prominent words have clearly larger pitch movements than non-prominent words do. Pitch movements occur in TB as rise-fall configurations, with accent-lending rises in the stressed pre-final syllables followed by fall movements towards the final syllables. Boundary marking also affects the melodic structure of TB. Pitch movements are larger in sentence-final words than sentence-medially, regardless of the focus condition of the word.

Focus and boundary effects on the melodic structures of BM are highly significant. Both effects play an important part in defining the position of the accent-lending pitch movements within the words, and in determining the shape of these movements as well. The effect of focus on the melody of the word is, unexpectedly, very strong. Whereas the shape of pitch movement in [-focus] words is a fall throughout, the size of the movements is large in [+focus] words. Fall movements in non-focus words are very small and merely an effect of declination. As in stress-languages, sentence-final boundaries in BM may be marked by pitch lowering. However, in accented words the pitch does not necessarily drop to the base line; it may appear as a rise or a truncated rise-fall combination in the final syllable.

3.6 Conclusion

Production experiments on TB and BM revealed significant and cross-linguistically different effects of focus and boundary marking on the temporal and melodic structures of words. Although the number of speakers was not very large, viz. two males and two females for each language, I feel confident that the effects would be the same when more subjects were involved. The main findings in the comparison of TB and BM are shown in table 3.9 for temporal structure and in table 3.10 for melodic structure.

Temporal structure. In both languages word duration is significantly affected by focus, and even more so by pre-boundary position. Unexpectedly, the effects of focus and pre-boundary position on the temporal structures show up more strongly in BM than in TB. Both effects on word duration are more than twice as strong in BM as in TB.

Unexpectedly, in TB, the duration of the pre-final (stressed) syllable is not affected significantly by focus. In BM, conversely, this effect is rather high in words with full vowels (not schwa) in this syllable. For both languages, boundary effects

are stronger on ultimate syllables than on pre-final syllables. In BM words with penultimate schwa vowels, effects of focus and boundary are accumulated on ultimate syllables.

Table 3.9. Main findings in the comparison of Toba Batak and Betawi Malay in terms of temporal structures at the word, syllable and segment level for all combinations of focus and boundary condition. Effects of focus and of boundary are specified as a duration increment (in percent) when the increase is statistically significant.

		Toba Batak			Betawi Malay			
Word level								
Focus	+final	+10%			+25%			
	-final	+14%			+33%			
Boundary	+focus	+13%			+40%			
	-focus	+17%			+46%			
Syllable level		Words with singleton C only						
		C ₁ V ₁		C ₂ V ₂	C ₁ V ₁		C ₂ V ₂	
Focus	+final	n.s.		n.s.	+29%		+36%	
	-final	n.s.		+31%	+57%		+60%	
Boundary	+focus	n.s.		+24%	+14%		+60%	
	-focus	n.s.		+51%	+39%		+88%	
Segment level		Effects only on last 3 segments			Effects only on last 3 segments			
		V ₁	C ₂	V ₂	C ₁	V ₁	C ₂	V ₂
Focus	+final	n.s.	n.s.	n.s.	+40%	+24%	+22%	+26%
	-final	+25%	n.s.	+46%	+42%	+67%	+27%	n.s.
Boundary	+focus	n.s.	n.s.	+38%	n.s.	n.s.	+40%	+100%
	-focus	30%	n.s.	+83%	n.s.	+56%	+40%	+66%
		V ₁	CC ₂	V ₂	C ₁	ə ₁	C ₂	V ₂
Focus	+final	n.s.	n.s.	n.s.	n.s.	n.s.	+29%	+46%
	-final	+18%	n.s.	+39%	n.s.	n.s.	+68%	n.s.
Boundary	+focus	n.s.	+11%	+44%	n.s.	n.s.	+27%	+130%
	-focus	n.s.	+10%	+88%	n.s.	n.s.	+68%	+135%

At the segmental level, focus effects are strong in BM full-vowel penultimate syllables. In TB, focus condition affects the duration of both vowels equally when words are in non-final position, but not sentence-finally. Boundary effects are for both languages strong on all ultimate-syllable segments except TB singleton consonants; geminate consonants in TB were slightly lengthened in pre-boundary targets. In BM the schwa in the penultimate syllable is not affected by focus nor by boundary. It apparently blocks the lengthening of the preceding consonant as well. Accentual lengthening starts in BM full-vowel words on the penultimate consonant, it becomes larger on the penultimate vowel and then weakens gradually sentence-medially; in pre-boundary position the lengthening spreads out to the end of the word. In TB, accentual lengthening occurs on the penultimate and ultimate vowels, but only sentence-medially.

Pre-boundary lengthening starts in BM on the penultimate (full) vowel and continues until the end of the word, becoming very strongest towards the end of the word. In TB pre-boundary lengthening occurs on the penultimate [focus] vowel (but not if it precedes a geminate consonant). It is largest on the ultimate vowel.

The temporal structure of BM schwa-words may be compared to that of TB words with geminate consonants between the penultimate and ultimate vowels. In both cases the lengthening of pre-final syllables seems to be blocked, and lengthening effects are accumulated on the ultimate syllable.

Melodic structures of words are influenced in both languages by focus and boundary condition. The effects are strongest in BM. In TB, both effects determine the size, but not the position, of the accent-lending pitch movements in the word; the position of the movements is determined by word stress position. All TB words carry an accent-lending pitch movement, but [+focus] words have larger pitch movements than [-focus] words. This is not the case in BM, where focus provides target words with accent-lending pitch movements, but pitch movements are suppressed in non-focussed words.

Boundary marking is in both languages realised by final lowering. In TB, it is also signalled by larger and steeper fall. Conversely, boundary in BM delays the

pitch movement, and makes it less steep. Furthermore, in BM boundary marking primarily influences the position of the accent-lending pitch movement in the word.

Table 3.10. Main findings in the comparison of Toba Batak and Betawi Malay in terms of melodic structures.

	Toba Batak	Betawi Malay
Basic melodic Structures:	One rise-fall (RF) melody (also on [-focus] targets). RF configuration spreads over two syllables: <ul style="list-style-type: none"> ● rise on penult, ● fall on ultimate ● peak always on penult 	Free stress position on focus targets, either penultimate or ultimate. 3 configurations: <ul style="list-style-type: none"> ● fall when accent on penult ● rise when accent on ultimate ● rise-fall on penult/ultimate, viz. <ul style="list-style-type: none"> - (R)F compact, low on ultimate - R(F) broad, F stays rather high on ultimate
focus effect:	<ul style="list-style-type: none"> ● larger rise ($\pm 100\%$) ● later & higher peak ● larger & steeper fall 	<ul style="list-style-type: none"> ● [+focus] accepts pitch movement ● [-focus] rejects pitch movement, declination only
boundary effect:	<ul style="list-style-type: none"> ● larger & steeper fall ● final lowering 	<ul style="list-style-type: none"> ● stretches pitch to lower end position ● longer, smaller & more gradual rise

The melody on the target words in the TB materials is always a rise-fall contour, with the rise beginning at the onset of the penultimate syllable, and the F₀ peak rather early in the penultimate vowel. However, focus and boundary affect the pitch contour on the target word in distinctive ways: prominent words have considerably larger pitch movements than non-prominent words, and sentence-final words have larger and steeper falls than sentence-medial words.

On the surface my results reveal the existence of four distinct pitch configurations in BM. These are:

- (i) a pitch fall on the penultimate syllable of the target word (after a preceding context that ends on a high pitch);
- (ii) a compact rise-fall contour on the penultimate syllable;
- (iii) a pitch rise on the ultimate syllable;

- (iv) a rise-fall contour on the ultimate syllable; when used domain-finally, the fall part is broken off before the baseline is reached; when used domain-medially the fall goes down to the baseline (see figures 3.13-14).

Duration and pitch movements are apparently prosodic cues for accent in BM. Both accent and boundary yield significant lengthening effects. Boundary determines the position of the accent-lending pitch movements: sentence-medially, the accent is on the pre-final syllable, but accent shifts to the ultimate syllable if the word is in sentence-final position. Contrary to expectation, the lengthening effects of focus on (stressed) penultimate segments in TB are small. On the other hand, when we compare the accent shapes which resemble each other most in both languages, i.e. pre-final rise-fall in [+focus, +final] condition, we find that pitch accents tend to be larger and steeper in TB than in BM: rises are larger in TB (0.66 ERB) than in BM (0.62 ERB), and falls are nearly 2 ERB/s steeper in TB than in BM. This research shows that stress in TB is realised not so much by lengthening, but by other factors such as pitch movements.

Chapter IV

Non-native accents in Dutch word-stress realisation¹

4.1 Introduction

This study focuses on the realisation of word prosody of Toba Batak (TB) and Betawi Malay (BM) compared to the realisation of Dutch word prosody; in particular the effect of stress/accent on the temporal and melodic structure. TB is, like Dutch, a stress language (Van der Tuuk 1971; Nababan, 1981). BM is a language that does not have word stress (Muhadjir, 1977), but does have phrasal accent (Wallace, 1976). Chapter III provided evidence for both word-based stress in TB and phrasal accent in BM. It will be scientifically interesting and useful for teaching purposes, to investigate whether speakers of a stress language can realise the stress of another stress language (in this case Dutch) more faithfully than speakers of a non-stress language.

Rather than measuring acoustic correlates of stress and/or accent, in the present paper, we will first determine the audible consequences of the Indonesian L1 background for the production of L2 Dutch. Three perception experiments were run to investigate to what extent native speakers of TB and BM are influenced

¹ This chapter was published as Roosman (2004). Part of it was presented at ISMIL-7, held in Nijmegen, The Netherlands, 27-29 June, 2003.

prosodically by their native language when they are speaking Dutch, and whether they are sensitive to the prosodic differences in Dutch.

The first experiment involves native Dutch listeners evaluating the realisation of Dutch word stress spoken by TB, BM, and Dutch speakers. The second experiment aims to find out whether, and by means of what (prosodic) cues, Dutch listeners are able to differentiate non-Dutch speakers from Dutch speakers. The last experiment investigates the extent to which TB and BM listeners are able to recognise Dutch-speaking Indonesians, even on the basis of deviant stress realisation only.

4.2 Background

Prosody is the set of properties in speech that cannot be derived from the segmental sequence of phonemes underlying the utterance (Nooiteboom, 1997:640-641). Three important parameters in prosody are pitch, duration and intensity. These properties are often called suprasegmental properties of speech. Ladefoged (1982:14) described vowels and consonants as the segments of which speech is composed. Suprasegmental features are aspects of speech that involve more than single consonants or vowels.

On the perceptual level, prosodic properties of speech lead amongst other things, to perceived patterns of relative syllable prominences coded in perceived melodic and rhythmic aspects of speech (Nooiteboom, 1997:640). We can hear the high and low pitch of utterances, the lengthened and shortened syllables. Typically, words with high pitch and lengthened syllables are presented by the speaker as more important than other words in a sentence.

Werner and Keller (1994:26) noted that, at the level of perception, it is common to classify prosodic phenomena in terms of the hearer's subjective experience, such as pauses, length, pitch/melody and loudness. Ladefoged (1982:104) mentioned also that from the listener's point of view a stressed syllable is often louder than an unstressed syllable. Stress is perceived usually, but not always, as a higher pitch. A

stressed syllable frequently has a longer vowel. However, listeners probably perceive the stress that other people are making by integrating all the cues available in a particular utterance, in order to deduce the motor activity we would use to produce the same stress.

Listeners use multiple sources of information – words, segment inventory, rhythm, pitch pattern and perhaps others – in identifying or discriminating between native and foreign languages. Apparently, these characteristics are also important to listeners when they attempt to refine their ability to identify foreign languages (Bond, Stockmal and Muljani, 1998:355).

Second-language speakers may be fluent in a target language. Nevertheless, we usually find their speech less intelligible than that of native speakers. Van Wijngaarden pointed out that non-native speakers could often be immediately identified by two factors that may reduce intelligibility: speech sounds are produced in an unusual, unexpected way ('distorted' phoneme inventory) and sentences are intoned in an unusual fashion. The influence of non-nativeness is then expected at both segmental and supra-segmental levels (Van Wijngaarden, 2001:104). Also, non-native listeners have more difficulty understanding L2 speech than native listeners of that language do (Van Wijngaarden, 2001:103).

4.3 Method

In the following experiments, I want to find out to what extent prosodic information (in particular pitch and duration) is used in language-background identification and acceptability judgement. The production data of TB and BM speakers are compared with data of Dutch speakers in perception experiments.

I start this section with the general aspects of the experiments, i.e. the composition of the stimuli; thereafter I will describe the details of the individual experiments and the results in separate sections.

4.3.1 Preparation of stimulus materials

In order to obtain a sufficiently varied range of stress patterns, target words with one, two and three syllables were selected with stress varying over all possible positions. To make sure that there would be no gaps in the pitch contour of the target words, voiceless consonants were avoided. Six target words were selected as exemplified in Table 4.1.

Table 4.1. Dutch stimulus words used in the experiments, broken down by word length and stress position.

Word length	Stress position		
	Antepenultimate	Penultimate	Final
1 syllable			<i>baan</i> ² 'job'
2 syllables		<i>bami</i> 'Chinese noodles'	<i>banaan</i> 'banana'
3 syllables	<i>bamibal</i> 'noodle ball'	<i>Lambada</i> 'a dance'	<i>Balinees</i> 'Balinese'

All target words were embedded in a carrier sentence, such that they were in final position and focused (accented). Accent was forced onto the target words by manipulating the focus distribution of the sentences through a precursor question asking for the target word, as in the following examples:

² It is actually not correct to say that a one-syllable word has final stress because there is no other syllable in the word. For practical reasons, I put this word in the final stress column.

(1) Wat zei je? Ik zei baan.
 What say-PAST you? I say-PAST baan.
 'What did you say? I said baan.'

(2) Wat zei je? Ik zei lambada.
 What say-PAST you? I say-PAST lambada
 'What did you say? I said lambada.'

4.3.2 Speakers

For my perception experiments I needed subjects who knew Dutch as well as BM/TB. It turned out to be impossible to find sufficient Dutch-speaking native speakers BM of the Dialek Kota. Therefore I used speakers and listeners of both Betawi Malay dialects (Dialek Kota and Dialek Pinggiran). Two BM speakers (one male, one female) were involved in the experiments. Both speakers were born and raised in Jakarta in Betawi families. The BM male speaker was 35 years old at the time of recording, and had been living in the Netherlands for more than ten years. He never took Dutch classes, but he understands and speaks a little Dutch. The female BM speaker was 26 years old. She had studied Dutch for five years in her home country. At the time of recording she had been studying Dutch in the Netherlands for four months.

This experiment also involved two speakers of TB. The TB male speaker was 37 years old and had been living in the Netherlands for eleven years. He speaks Dutch with his Dutch wife and with his children. The female TB speaker was a 25-year old nurse, who had been living in the Netherlands for one year. She had followed a Dutch intensive course for three months in her home country before she came to the Netherlands. She had to speak Dutch at work.

All foreign-language speakers were late bilinguals, i.e. individuals who acquired their Dutch after puberty. Late bilinguals are expected to be more affected by foreign accent in an L2 than early bilinguals.

Finally, one male and one female native speaker of Dutch served as a control group.

4.3.3 Recordings

All sentences were first presented to the speakers on sheets of paper without any indication of the placement of the stress on the target words. The non-Dutch speakers were asked to read the sentences while taking care to put accent and stress in the correct positions. To help them doing this task, I used the question-answer method as explained in section 4.3.1. The interviewer read the questions (*Wat zei je?* ‘What did you say?’) from the sheets. The BM or TB speaker then gave the answer (*Ik zei baan.* ‘I said *baan*’). They were asked to answer the question clearly. If the speaker did not read the answer in the correct way, the interviewer read the question again. If the position of the stress in the second try was wrong, the interviewer told the subject where the stress was supposed to be, and the speaker read the answer again. Finally, to make sure that sufficient correct utterances would be available for the perception tests, the subject was presented with a list of sentences in which the stress position was indicated by a stress mark (accent aigu, as in *bámi*) above the stressed syllable. Every subject had to read every sentence three times correctly.

All non-native speakers had difficulties with their reading task. The three-syllable words with their various stress patterns proved especially difficult. The most difficult words were *lambada*, which has penultimate stress, and *balinees* with final stress.

The native speakers of Dutch performed the same task. However, in an informal listening test, it appeared that the Dutch speakers spoke in a rather assertive way, whereas the Indonesian speakers had pronounced the speech materials more hesitantly and quietly. Therefore, the Dutch native speakers were asked to speak less assertively and rather slowly, so that in the following perception experiments the listeners would not be able to distinguish the groups of speakers on account of their degree of assertiveness or speaking rate.

All recordings were made in the Netherlands in a quiet room on a Sony TC-D5 PRO II tape recorder through head-worn microphones (Shure SM-10A).

All speech materials were then digitised (16 kHz sampling frequency, 16 bits amplitude resolution) and stored on computer disk to be manipulated for the perception experiments. For each speaker, the best three utterances for each target word were selected, on the criteria that the target words had to be accented, and the position of the stress in the target words had to be correct. The author and a phonetically trained native listener of Dutch indicated – independently of each other – for each recorded token whether the word was correctly accented in the sentence and whether the stress was in the appropriate position. All tokens selected for the listening tests were correctly accented and stressed as judged by both experts. The total number of utterances to be manipulated was 108 (6 words \times 6 speakers \times 3 repetitions).

4.3.4 Manipulations

First, all carrier sentences of the selected speech material were removed so that only the target words would be audible to the listeners. The target words were then manipulated to get three versions of the stimuli. All stimuli were generated by computer with the PRAAT speech processing software (Boersma and Weenink, 1996).

For the first version of the stimuli, no further manipulation took place; the original sounds of the target words ('origin', Fig. 4.1a) were presented to the listeners. In the second version, the target words were delexicalised by low-pass filtering at 350 Hz, with a smoothing of 100 Hz; this version will be called 'delex' (see the spectrogram of Fig. 1b). The filtering procedure was carried out to make all verbal information inaudible. This was done in order to assess whether listeners are able to differentiate between the speech of native and non-native speakers of Dutch without lexical segmental information, but with (virtually) all prosodic information intact. In the third version, the filtered stimuli of the second version were made monotonous ('monot'; Figure 4.1b). This was done to assess whether listeners are

able to identify the speaker's language background without lexical and melodic information, but with durational and loudness information only..

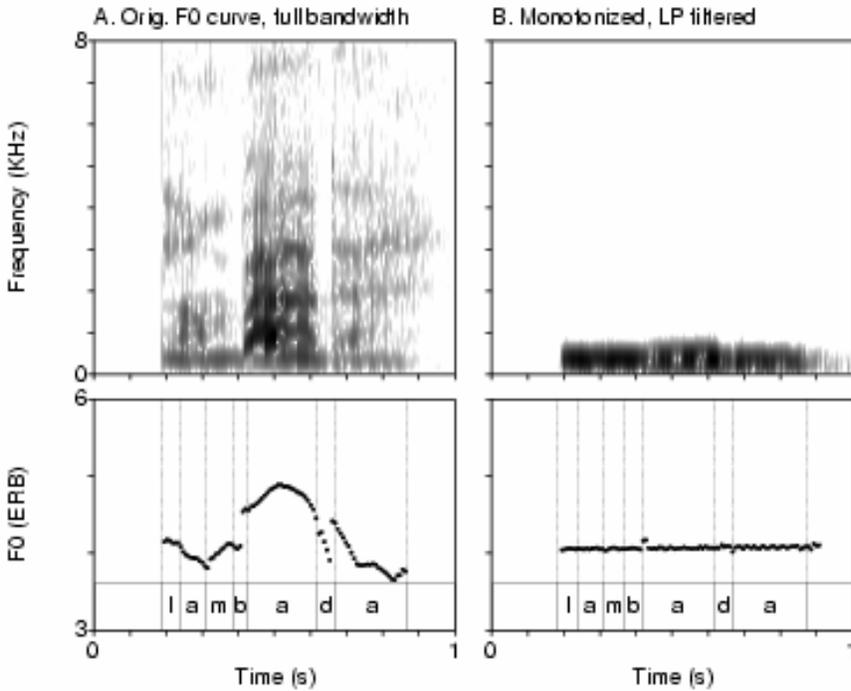


Figure 4.1a. Spectrogram (top) and original fundamental frequency contour (ERB; bottom) of stimulus item *lambda*. Segmentation and phoneme labels are indicated.

Figure 4.1b. As figure 4.1a but low-pass filtered (cut-off frequency at 0.35 KHz; top) and with pitch contour monotonized at 4.2 ERB (150 Hz; bottom).

4.3.5 Procedure

The 18 tokens for every word (6 speakers \times 3 repetitions) were randomly ordered, and each token was made audible twice. The two-syllable words were presented first (*bami*, *banaan*), followed by the three-syllable words (*bamibal*, *lambda*, *balinees*), and the last word was the monosyllabic word (*baan*). A silent response time of 2.5 seconds separated the words from each other.

In order to not give away the identity of the target words, the ‘monot’ version was presented first to the listeners. It was followed by the ‘delex’ version and finally the listeners listened to the ‘origin’ version. 5-second intervals separated the versions. The total number of stimuli was 648 (3 versions \times 6 words \times 18 tokens \times 2 repetitions). Every version was preceded by some examples of stimuli (ten examples before the ‘monot’ version, and five examples before the ‘delex’ and the ‘origin’ versions).

It was briefly explained to the listeners that they had to pay attention to different kinds of prosodic information such as rhythm or melody while listening to the different sets of stimuli. For the first set, the ‘monot’ version, they had to pay attention to the rhythm of the sounds. Second, for the ‘delex’ version, they had to pay attention to the rhythm and the melody of the sounds. For the last version they were able to use all information available in the sounds. They were instructed to listen carefully to the example stimuli so that they would get used to the manipulated stimuli before the real experiment started. The whole experiment took about 40 minutes.

4.4 Experiment 1: Evaluation by Dutch listeners

4.4.1 Subjects and procedures

Thirty Dutch native listeners took part in the experiment. They listened to the stimuli on tape through headphones. They were given a list of target words with a column to indicate their scores (cf. Appendix 2). Subjects were given standardised, written instructions to mark the target words they heard along an 11-point judgment scale ranging from 0 (‘undoubtedly foreign speaker of Dutch’) to 10 (‘undoubtedly native Dutch speaker’).

The subjects were told that score 5 (i.e. the midpoint of the scale) was the boundary between ‘native’ and ‘non-native’. It indicated that the subject was unable to decide whether the stimulus was spoken by a native or a non-native speaker of

Dutch. Subjects had to indicate a choice in all cases (forced choice). They were not allowed to mark two positions on the scale, nor to leave an item blank. They were not paid for their services.

4.4.2 Results and discussion

Figure 4.2 summarizes the mean evaluation scores for the different stimulus versions broken down by speaker group. The information on which the listeners could base their identification increases from left to right along the horizontal axis. In the ‘monot’ version there is only temporal information. In the ‘delex’ version melodic information is added, and in the ‘origin’ version the listener has access to the full signal, including the segmental quality.

The figure shows clearly that the Dutch speakers are the most qualified group in this experiment. The BM speakers are the least qualified group. The TB group assumes an intermediate position. The Dutch speakers get the highest mean scores (6.20), followed by the TB speakers (4.53). The BM speakers get the lowest scores (3.67).

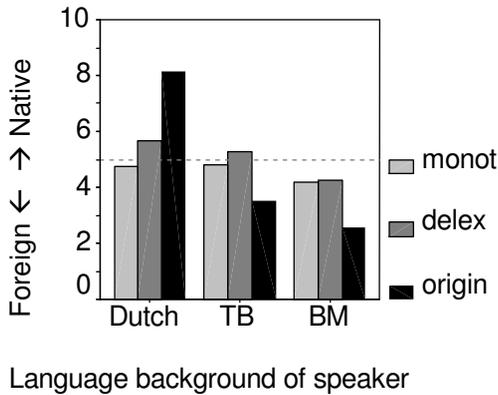


Figure 4.2. Mean evaluation score (0..10) by 30 Dutch listeners broken down by speaker group and by stimulus version.

An analysis of variance was run on the data with language background of speakers (Dutch, TB, BM) and stimulus version ('monot', 'delex', 'origin') as fixed factors with the evaluation scores as dependent variables and target words as a random factor. The results show that a highly significant effect occurs for the language background of the speakers [$F(2,10) = 34.26, p < .001$]. There is no significant effect of the stimulus version [$F(2,10) = 2.43, p = .138$]. The average scores for all stimulus versions are rather similar, as the high score for Dutch speakers in the 'origin' version levels out the low scores for the TB and BM speakers. Similarly, the Dutch score in the 'delex' version levels out the BM score for that version. However, the interaction between language of speakers and stimulus version reaches significance [$F(4,20) = 34.19, p < .001$].

Based on these findings, a series of two-way ANOVA's for each language group was run with stimulus version as a fixed factor, evaluation score as dependent variable, and target word as a random factor, to investigate whether there are significant effects of stimulus version within each group of speakers. The effect of stimulus version on the evaluation score is highly significant for the Dutch speakers [$F(2, 10) = 54.0, p < .001$] and for the TB speakers [$F(2,10) = 20.36, p < .001$]. This effect is also significant for the BM speakers [$F(2,10) = 7.77, p = .009$]. Post-hoc tests of homogeneity (Scheffé procedure with $\alpha = .05$) show, however, that the mean scores of the BM speakers in the 'monot' version and the 'delex' version do not differ from each other ($p = .754$). It seems that the BM speakers did not produce an adequate speech melody so that the Dutch listeners gave the 'delex' stimuli the same scores as the BM artificially monotonized speech.

A series of two-way ANOVAs was also run on each stimulus version with the mean evaluation scores as the dependent variable, speaker groups as fixed factors and target word as a random factor. This was done to find out whether, and for which stimulus version, the differences between the three speaker groups are significant. The results show that there are significant effects of speaker group in all stimulus versions [$F(2,10) = 6.73, p = .014$ for the 'monot' version; $F(2,10) = 7.70, p = .009$ for the 'delex' version; and $F(2,10) = 46.14, p < .001$ for the 'origin' version]. Post-hoc tests were then run on the three speaker groups within each

version. In all versions the mean scores of the BM speakers are significantly poorer than those of the other groups.

Figure 4.2 shows that without verbal and melodic information ('monot' version) all groups of speakers had rather low mean scores: below 5, i.e. below the boundary between 'native' and 'non-native'. Native listeners of Dutch perceive monotonized speech as foreign. However, they heard that BM speakers realized the stress in a significantly different manner from the other two speaker groups [$p < .001$], which were similar to each other. This implies that BM speakers have a different temporal structure than the other groups and that this deviant temporal structure sounds more foreign to the Dutch listeners.

The addition of the melodic information to the duration structure in the 'delex' version does improve to some extent the listeners' ability to differentiate the groups from each other. In the 'delex' condition the three groups of speakers were perceived as significantly different from each other. Finally, the listeners were more outspoken when they were listening to the 'origin' version. This is shown by the highly significant effect mentioned above. In figure 4.2 we see that the lowest mean score (2.6) was observed for the BM speakers in the 'origin' version. This indicates that the pronunciation of the vowels and consonants of these speakers is clearly foreign to Dutch listeners. Compared to the results of TB Dutch, it seems that, in general, the BM speakers sound more foreign to Dutch listeners. This is in line with the assumption that speakers of a non-stress language realise stress more poorly than speakers of stress languages do.

4.5 Experiment 2: Identification by Dutch listeners

In the following experiment, I investigated to what extent Dutch listeners are able to identify non-native and native listeners using the available information. Rather than responding along a scale from foreign to native, listeners now had to take a categorical yes/no decision on the 'Dutchness' of the speaker. By presenting the

subjects with a binary choice I hoped to get sharper results than in the previous experiment.

4.5.1 Subjects and procedure

In this experiment, ten Dutch native listeners were involved. They listened to the same stimuli as in the first perception experiment, played to them through a tape recorder over headphones. The listeners had to determine whether the speaker was a native Dutch speaker or not. Listeners were presented with a list of target words. For each word on the tape the listeners indicated, by ticking one of two response boxes provided on their answer sheets, whether they thought the speaker was Dutch (Ja 'yes') or not (Nee 'no'). A short explanation of the different versions the listeners were going to listen to was given, as in Experiment 1, before the experiment started; cf. Appendix 3.

4.5.2 Results and discussion

Figure 4.3 presents the identification of language background (native vs. non-native) expressed in percent identification as 'Dutch' broken down by speaker's language and by version of stimuli.

An ANOVA was run on the percentage of identification as the dependent variable, with speaker group and stimulus version as fixed factors, and target word as a random factor. The results indicate that, as in experiment 1, there is no significant, overall effect of stimulus version [$F(2,10) = 2.63, p = .121$]. However, the interaction between the stimulus version and the language background of the speakers is highly significant [$F(4,20) = 25.33, p < .001$]. The effect of the language background is also significant [$F(2,10) = 9.65, p = .005$].

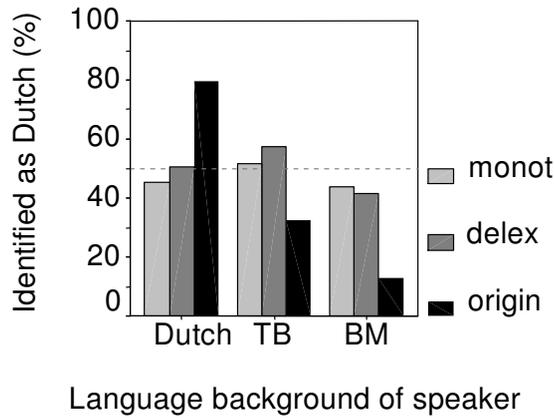


Figure 4.3. Percentage of the responses by ten Dutch listeners (correctly or incorrectly) identifying the speaker as Dutch, broken down by language background of the speakers and by stimulus version.

One-way ANOVAs were run on the percentage of the identification for each group of speakers separately, with stimulus version as a fixed factor. There are significant effects of stimulus version [$F(2,10) = 25.02$, $p < .001$ for the Dutch speakers; $F(2,10) = 6.41$, $p < .016$ for the TB speakers; and $F(2,10) = 32.52$, $p < .001$ for the BM speakers]. Post-hoc tests show that there are no significant differences between ‘monot’ and ‘delex’ versions for all groups of speakers but that the ‘origin’ version always differs significantly from the other two.

Figure 4.3 strongly resembles figure 4.2. It shows that in the ‘monot’ version, the identification as Dutch is roughly the same for all speaker groups. In fact, Dutch speakers were as unconvincing as BM speakers were. In the ‘delex’ version the percentage of Dutch speakers identified as Dutch improves, as does the percentage for the TB speakers, but for the BM speakers the percentage decreases. The ‘origin’ version is significantly different for all speaker groups. The percentage identification as Dutch for BM speakers is much lower than for the other two speaker groups. In this version TB speakers are also less frequently identified as Dutch than in the other versions.

These results indicate that Dutch listeners cannot identify the language background of the speakers (as native or non-native) unless segmental (verbal) information is provided in addition to prosodic information. Prosody by itself does not provide enough information to distinguish the non-native from the native speakers. This result differs from the result of the first experiment in which prosodic cues do, to some extent, distinguish the different groups of speakers. Contrary to expectations, the scaling task was more sensitive than the binary identification task.

4.6 Experiment 3: Identification by non-Dutch listeners

This experiment involved only BM and TB listeners in order to find out whether they could identify the language background of the speakers. It was expected that these listeners could not only distinguish the Dutch speakers from the non-Dutch (Indonesian) speakers, but that they could also distinguish the Indonesian speakers from each other.

4.6.1 Subjects and procedure

A group of ten TB listeners and a group of ten BM listeners who know Dutch, took part in this experiment. All listeners were students or former students of the Dutch Department at the Universitas Indonesia. Although I only selected BM listeners whose parents were Betawi, their language might have been influenced by Modern Jakarta Malay, the language of the younger generation living in Jakarta. The TB and BM listeners listened to the same stimuli in the same order as the Dutch listeners in the previous experiment. It was explained to them that they were going to listen to Dutch words, spoken by native Dutch speakers, TB speakers and BM speakers. They had to identify the mother tongue of the speakers they heard on the tape as either Dutch, Betawi Malay, or Toba Batak, with forced choice. They were presented with a printed list of target words on which they had to tick the language background of the speakers. They got the same explanation about the stimuli as the

Dutch listeners in the previous experiment. All experiments were done individually in Jakarta, Indonesia.

4.6.2 Results and discussion

Figure 4.4 presents the percentages of correct identification broken down by listener group and stimulus type (across all three speaker groups and across all word types). As in figures 4.2 and 4.3, the information on which the listeners could base their identification increases from left to right along the horizontal axis. In the ‘monot’ version there is only temporal information. In the ‘delex’ version melodic information is added, and in the ‘origin’ version the listener has the full signal at his or her disposal, including the segmental quality.

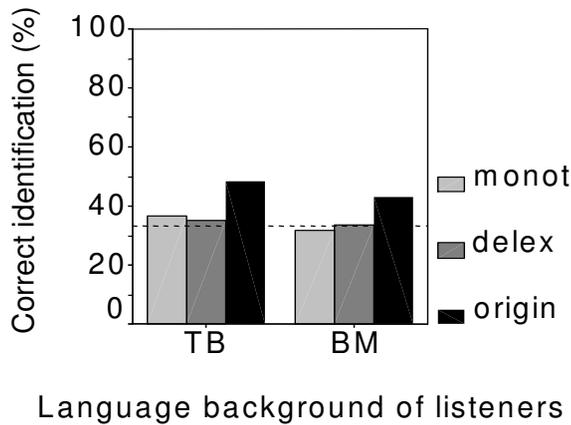


Figure 4.4. Percentage correctly identified language background of speaker broken down by stimulus version and by language of the listener (Indonesian nationals only).

An ANOVA with stimulus version, listener group, and speaker group as fixed factors, and target word as a random factor, was run with percentage correct identification as the dependent variable. The result shows a highly significant effect of stimulus type [$F(2,10) = 19.08, p < .001$]. The effect of the language background of the listeners is also significant [$F(1,5) = 10.0, p = .025$], but there are no

significant effects of speaker group [$F(2,10) = 3.64, p = .065$] (not shown in figure 4.4.)

Figure 4.4 shows that, overall, the TB listeners could identify the language of the speakers in all versions somewhat better than the BM listeners could. The test revealed no interaction between the listener group and the version of stimulus [$F(2,10) = 1.02, p = .396$]. For both groups of listeners, there are no significant differences between the ‘monot’ and the ‘delex’ versions in terms of percentage correct identification. A post-hoc test shows that these two versions form one group. Language background is reported correctly significantly more often in the ‘origin’ version ($p < .001$).

The task of identifying the language background of the speakers from all versions of stimuli was difficult to both groups of listeners. Apparently, neither group of listeners knew the word prosody of Dutch well, nor did they know the proper segmental pronunciation of the Dutch words. Nevertheless, they could identify the speakers better from the ‘origin’ version. In fact, the Indonesian listeners could only identify the language background of the speakers better than chance if the vowels and consonants were recognisable.

There is, however, a significant interaction between the groups of listeners and the groups of speakers [$F(2,10) = 14.47, p = .001$]. The interaction between stimulus version and speaker group also reaches significance [$F(4,20) = 4.70, p = .008$]. In the following section, I will therefore examine the results of the TB listeners and the BM listeners separately. Separate ANOVAs were run on the percentages of correct identification per listener group, with speaker group and stimulus version as fixed factors and target word as a random factor.

4.6.3 Toba Batak listeners

It was expected that TB listeners, as native speakers of a stress language, could identify the subjects better than BM listeners. Overall, this was indeed the case as is shown in Figure 4.4. But which speakers are better recognised depends on the stimulus version: the interaction between listeners and speaker group is significant,

and also the interaction between speaker group and stimulus version. Figure 4.5 summarises the percentage correctly identified language background of the speakers as identified by TB listeners broken down by stimulus type. For a complete overview of the results see Appendix 4a.

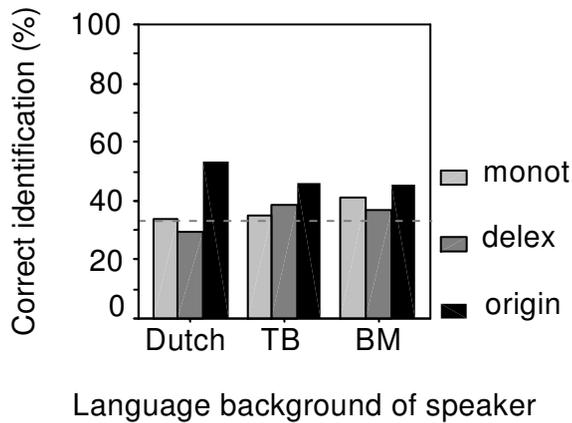


Figure 4.5. Percent correctly identified language background for three speaker groups by TB listeners and broken down by stimulus version.

The analysis of variance for TB listeners indicates that there is a highly significant effect of stimulus version [$F(2,10) = 18.05$, $p < .001$]. A post-hoc test shows that the percentages of correct identification of the 'monot' and 'delex' versions are similar to each other ($p = .751$). TB listeners could identify the speakers' background in the 'origin' version rather better than in the other versions ($p < .001$). There is no significant effect of the language background of the speakers [$F(2,10) = .22$, $p = .807$]. The correct identification percentages across all stimulus versions are similar for all groups of speakers. However, the interaction between stimulus version and language background of the speakers just reaches significance [$F(4,20) = 3.64$, $p = .022$].

The ability of the TB listeners to identify the language background of the Dutch and the TB speakers is the same in the monotonous version. The BM speakers are better identified than the other speakers. This seems to indicate that the TB listeners

consider speech of BM speakers to be rather monotonous. When the melodic information is added the listener's ability to identify the language background of the Dutch speakers is poorer than the ability to identify the TB and the BM speakers, which are similar to each other. The melody of the Dutch speakers was often mistaken for the TB melody (see Appendix 4a). Listeners' ability to identify the language background in the 'origin' version, in which the full information is available, is higher for the Dutch speakers than for the TB and the BM speakers, which have the same scores. This result indicates that the pronunciation of the vowels and consonants by the Dutch native speakers is very noticeable to TB listeners.

4.6.4 Betawi Malay listeners

Figure 4.6 summarizes the results of the correct identification by BM listeners, broken down by the group of speakers and stimulus version. For a complete overview of the results see Appendix 4b.

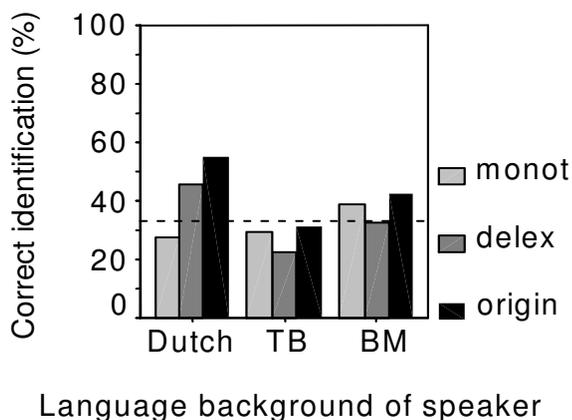


Figure 4.6. Percent correctly identified language background for three speaker groups by BM listeners and broken down further by stimulus version.

If we ignore the ‘monot’ condition for the moment, the results are straightforward. The identification of the speakers’ language background is somewhat better if the phonetic quality of the segments is audible, but the increment is rather small. A much larger contribution is made by the melody: the Dutch speakers are identified best, the BM speakers second, and the TB speakers worst. Probably, the livelier intonation/accenuation of the Dutch speakers provides a good cue for the identification. The flat prosody of BM speech helps the BM listeners pick out their own kind with some measure of success, but the slightly more lively prosody of the TB speakers is often mistaken for Dutch, hence the poorest identification of the TB speakers.

Now, what happens when we replace the natural pitch patterns by a monotone? There will be a bias towards BM identifications, assuming that flat prosody is the hallmark of BM. So we predict that the correct identification of BM speakers will rise in the ‘monot’ version, which is indeed what is found in the results. At the same time we predict that Dutch, and to a smaller extent TB speakers, will be mistaken for BM speakers due to the absence of pitch obtrusions. This prediction is also borne out by the results (for the complete confusion matrix see appendix 4b).

The results were submitted to an ANOVA with stimulus version and language background of the speaker as fixed, and with lexical word as a random factor. There are significant main effects of stimulus version [$F(2,10) = 9.6, p = .005$] and of language background of the speaker [$F(2,10) = 11.1, p = .003$]. The highly significant interaction between the two factors [$F(4,20) = 6.4, p = .002$] is the result of the differential effect of the removal of pitch for the identification of BM as opposed to Dutch and TB.

4.7 Conclusion

In the introduction I posed three questions, which I will now try to answer on the basis of the experimental results.

First, I aimed to evaluate the way in which native speakers of TB and BM realize Dutch stress by using Dutch listeners. The results of experiment 1 show that the BM realisation of Dutch is not only less Dutch to Dutch listeners than native Dutch, but it also sounds somewhat less Dutch than TB Dutch. Although the largest source of 'Dutchness' is in the phonetic quality of the segments (pronunciation of vowels and consonants), there is a clear indication that TB Dutch is less foreign due to the fact that TB has more clearly-marked stress realisations, which is advantageous when TB speakers produce Dutch words.

The second question was if, and on the basis of what (prosodic) cues, Dutch listeners are able to differentiate non-native (TB and BM) speakers from Dutch speakers. The results of experiment 2 show that Dutch listeners very clearly differentiate Indonesian speakers of Dutch from native Dutch speakers, especially when the fully specified original speech signal is available. However, when the phonetic quality is obliterated through filtering, the difference between the Dutch and Indonesian speakers is negligible. Overall, the TB speakers are more often identified as Dutch speakers than BM speakers, indicating that their prosody, especially their temporal organisation and to a lesser extent their pitch pattern, approximates the Dutch norms better than those of the BM speakers.

The third and final question asked how well TB listeners and BM listeners can pick out their own language background (amidst other non-native speakers of Dutch and native Dutch speakers) when speaking Dutch as a foreign language. Experiment 3 shows that the non-native listeners were rather poor in picking out their own group of speakers, even when they were provided with the complete, original speech samples. Nevertheless, the results provide indications that a flat, monotonous prosody is seen as a characteristic of the BM speaking style, which is transferred to Jakartan-accented Dutch.

Although these are just preliminary results of a relatively small-scale study, the results may have implications for the teaching of Dutch as a foreign language to Indonesian nationals. Specifically, I would argue for a differential curriculum depending on the specific language background of the Indonesian learner of Dutch. If the learner hails from a TB background (or from any other linguistic community

speaking a stress language), less specific attention is needed in the area of prosody than for Jakartan learners of Dutch (or any other Indonesian language group who speak an essentially stress-less language).

Chapter V

Acoustical analysis of Dutch word stress as spoken by Dutch, Toba-Batak and Betawi-Malay speakers

5.1 Introduction

In the preceding chapter I have found that Dutch and Toba-Batak (TB) listeners proved able to identify – better than could reasonably be expected on the basis of pure chance – the native-language background of speakers of Dutch stimuli produced by speakers of the same three languages. Betawi-Malay (BM) listeners, while correctly identifying their own speech and Dutch above chance, often misjudged the TB speakers as Dutch. I am not so much interested, in the present research, in the contribution that is made to successful identification of a speaker's native language by the segments (vowels and consonants). The perception experiments in chapter IV showed quite clearly that prosody makes an independent contribution to the perception of foreign accent: if the segmental information was obliterated through low-pass filtering, enough information remains available in the melody and temporal structure of the words to permit above-chance identification of the speaker's native language background, or at least of the stress versus non-stress character of the native language. Even after removing both the melodic and the segmental information, such that only some temporal structure remains audible, did the listeners differentiate between the various speaker groups.

In the present chapter I aim to determine what phonetic aspects of the melodic and temporal structure information might account for the perceptual identification of the speaker's native-language background. To this effect, I carried out a so-called

stimulus analysis of the materials used in the perceptual identification tests described in chapter IV. Given that I am only interested in the language-background information contained in the prosody, the acoustic-phonetic characteristics measured were limited to temporal and melodic parameters.

Acoustical analyses were carried out for the six target words that were used in Chapter IV, i.e. *baan*, *bami*, *banaan*, *bamibal*, *lambada*, *balinees*. In § 5.2, I will outline and motivate the temporal and melodic parameters that I have selected as potential correlates of native-language background of the speakers, and then present a detailed survey of the measurements. In § 5.3, I will attempt to correlate the acoustic parameters with the perceptual behaviour of the listeners. To this effect, I will use multiple regressions, trying to predict the perceptual identification of native-language background from selected acoustic-phonetic parameters. Such a correlational approach to the problem uncovers potential perceptually relevant measures, but provides no conclusive answer to the question what information was actually used by the listeners in performing their task. Perceptual relevance can only be shown through manipulation studies; these, however, fell beyond the scope of the present dissertation.

5.2 Acoustical analysis

The measurements of the prosodic parameters of the target words were done on the stimuli used in the perception experiments. Artificially monotonized ('monot') stimuli were used to investigate the perception of the temporal aspect of the non-native speech in word-stress realization. Therefore, duration measurements were done. A low-pass filtered ('delex') version of the stimuli was used to investigate the melodic aspects of native and non-native speech. Therefore, F_0 parameters were also analysed. Duration and F_0 measurements of the target words will be compared with the results of the perception experiments. I expect that the results of the comparisons will provide some explanations for the findings of Chapter IV.

For the temporal analyses, the target words were first segmented by hand. The segment boundaries were stored in so-called Praat TextGrids. Durations of the time intervals between successive segment boundaries were then automatically determined by the Praat speech analysis software. The duration (in ms) of each vowel and consonant in each stimulus word (6 speakers \times 6 words \times 3 tokens = 108) was stored in a database for off-line statistical analysis. Syllable duration and word duration could be computed from the raw segment durations, as were a number of relational temporal parameters (e.g. the ratio of the duration of the stressed syllable to the mean duration of the unstressed syllables in the same word).

For the melodic analyses, I used as my raw measurements the F_0 values (frequency-time coordinates) of the pivot points P1 to P4 in the target words. From these raw measurements, pitch intervals (in ERB) and time intervals of rises and falls could be computed, as well as the steepness of the rises and falls (in ERB/s). Since timing information was available of both the onsets and offsets of segments as well as of the rises and falls, alignment measures could be defined which might reveal systematic differences between the three groups of speakers in the temporal location of the rises and/or falls relative to some segmental landmark (e.g. onset of stressed vowel).

5.2.1 Temporal structure of the targets

In chapter IV, I found that monotonous speech tends to sound foreign to the Dutch listeners. However, in spite of this Dutch listeners could hear that BM speakers realised the words with a significantly different temporal structure than the other two speaker groups, which were similar to each other. I expect, therefore, that the duration analysis will reveal temporal structures for the BM realisations that are different from the realisations of the other two groups. Besides, I will also try to investigate whether there are significant differences per individual speaker.

First, I will focus my investigation on the overall duration of the target words. Then, I will go down to the smaller units that can bear accents, i.e. the stressed syllable and the vowel within the stressed syllable.

Separate analyses of variance were run with overall duration of target word, duration of the stressed syllable, and duration of the stressed vowel as dependent variables, and with language background of the speaker as a fixed factor. The result shows that there is no effect of language background of the speakers on the duration of the target words [$F(2, 105) < 1$], nor on the duration of the stressed syllable [$F(2, 105) < 1$]. However, there is a significant effect of language background on the duration of the vowel in the stressed syllable [$F(2, 105) = 4.34, p = .015$].

Figure 5.1 illustrates the mean durations of the target words, the stressed syllables, and the vowels in the stressed syllables as spoken by the three groups of speakers.

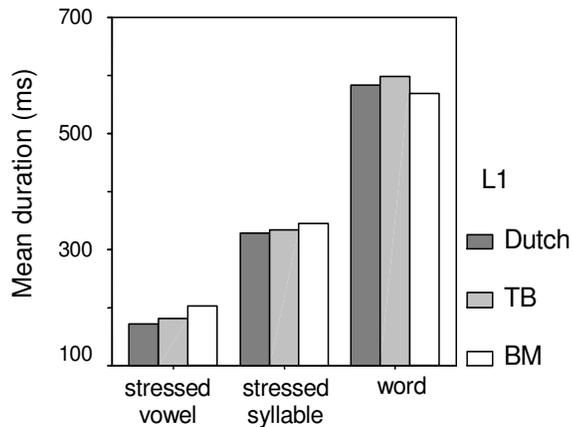


Figure 5.1. Mean duration (ms) of stressed vowel, stressed syllable and target words broken down by native language background of the speaker.

On average, the words spoken by the two Betawi-Malay speakers are shorter than the words spoken by the Toba-Batak and the Dutch speakers, but the differences are not significant. The mean duration of all words spoken by BM speakers is the shortest (569 ms), and that of the TB speakers the longest (598 ms), which are closely approximated by the Dutch speakers (583 ms). The duration of the stressed syllables spoken by BM speakers (344 ms) is in general longer than the stressed syllables spoken by TB speakers (334 ms), which again are longer than those of the Dutch speakers (328 ms). As shown by the ANOVA, the differences between the

mean stressed syllable durations spoken by the different language groups are not significant. The mean duration of the stressed vowel of the native Dutch speakers (171 ms) is shorter than that of both groups of Indonesian speakers of Dutch. The TB speakers (181 ms) come close to the native Dutch stressed vowel duration; the difference is insignificant (Scheffé post-hoc test, $p = .147$). Only the BM speakers realise significantly longer stressed vowels (203 ms) than the Dutch speakers do ($p = .019$).

To investigate whether the temporal organisation of words differ per speaker I ran an ANOVA with the same dependent variables as before with the individual speakers as a fixed factor. The result shows also that there is no effect of the different speakers on the duration of the target words [$F(5,102) = 1.46$, $p = .211$], nor on the duration of the stressed syllable [$F(5,102) < 1$]. However, the effect of speaker is significant for the duration of the stressed vowel [$F(5,102) = 3.75$, $p = .004$].

Tables 5.1a-b are adapted from the post-hoc tests (Scheffé procedure) and list homogenous subsets of the vowel duration per speaker group and per individual speaker. Speaker differences in the duration of the stressed vowel are clearer if we classify the speakers by native language rather than by individual.

This result is rather unexpected, since the vowels in the stressed syllables in the six stimulus words are always phonologically long vowels (either /a:/ or /e:/). Given that the interfering Indonesian languages have no vowel length contrast, one would expect the Indonesian approximations to the Dutch long vowel duration to be shorter (rather than longer) than the native Dutch norm. Be this as it may, the duration of the stressed vowel does provide a statistical correlate for the perceptual determination of the speaker's language background: the shorter the stressed vowel, the more likely the speaker will be Dutch.

Table 5.1. Homogeneous subsets (Scheffé) of stressed vowel duration (ms) per speaker group (a) and per individual speaker (b).

a.

L1	N	Subset for $\alpha = .05$	
		1	2
Dutch	36	171.42	
TB	36	181.20	181.20
BM	36		202.58
Sig.		0.666	0.147

b.

Speaker	N	Subset for $\alpha = .05$	
		1	2
TB male	18	161.29	
Du female	18	169.22	169.22
Du male	18	173.61	173.61
BM male	18	191.44	191.44
TB emale	18	201.11	201.11
BM female	18		213.72
Sig.		0.217	0.120

The longer BM stressed vowel duration and the shorter BM word duration are in contradiction. I suspected that the duration distribution within the word is different for the different speaker groups. In addition, syllable duration depends on its position within the word. Therefore, I calculated the ratio of the stressed-to-unstressed syllable duration and the ratio of the stressed-to-unstressed vowel duration. Figure 5.2 illustrates the stressed-to-unstressed ratio for syllable and vowel duration, broken down by speaker group.

Figure 5.2 shows the differences in ratios between the Dutch speakers and the Indonesian speakers. The figure indicates that the ratios of the Dutch speakers are the smallest (1.86:1 for syllable duration and 1.87:1 for vowel duration), and the ratios of the BM speakers are the largest (2.31:1 for syllable duration, 2.88:1 for vowel duration). The syllable ratio of the TB speakers resembles that of the Dutch speakers (1.90:1), and the vowel ratio of TB speakers (2.07:1) is also closer to the Dutch than to the BM group.

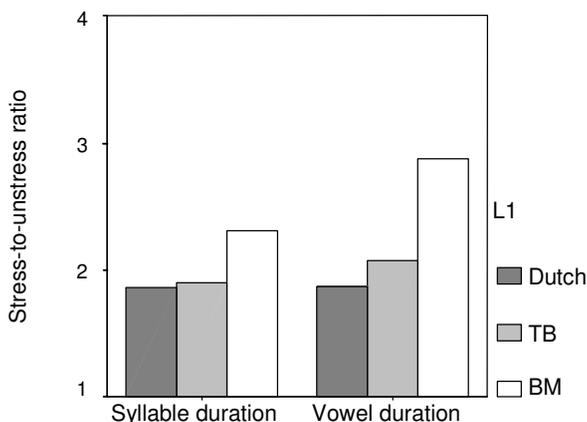


Figure 5.2. The stressed-to-unstressed ratio of syllable and vowel duration per speaker group.

Separate one-way ANOVAs with syllable duration ratio and vowel duration ratio as the dependents were run with the L1 of the speaker group as a fixed factor. The results show that there is a significant effect of language background on the vowel duration ratio [$F(2,87) = 4.79$, $p = .011$], but not on the syllable duration ratio [$F(2,87) = 1.12$, $p = .332$]. Similar results were found when we ran the ANOVA with the six individual speakers as a fixed factor. There is no significant effect of speaker on the syllable duration ratio [$F(5,84) < 1$], but there is a significant effect on the vowel duration ratio [$F(5,84) = 2.77$, $p = .023$]. However, the post-hoc test (Scheffé procedure) indicates that the differences between the individual speakers are not significant, not even in the case of the stressed-to-unstressed vowel duration ratio. The differences are significant, however, if we collapse the male and the female speakers with the same native-language background. Tables 5.2a-b are adapted from the post hoc tests (Scheffé procedure) and show homogenous subsets of vowel duration ratio per speaker group and per individual speaker.

A post-hoc test for the groups of speakers indicated that there is a significant difference between the Dutch speakers and BM speakers ($p = .017$). There are no

significant differences between the Dutch speakers and the TB speakers, nor between TB and BM speakers (see Table 5.2a.). The post-hoc test for the individual speakers yields no significant differences among the speakers. The only thing that we can observe is that male speakers have a larger stressed-to-unstressed vowel ratio than their female counterparts, and that the BM-male speaker has the largest ratio.

Table 5.2. Homogeneous subsets (Scheffé) of the stressed-to-unstressed vowel duration ratio per L1 speaker group (a) and per individual speaker (b).

a.

L1	N	Subset for $\alpha = .05$	
		1	2
Dutch	30	1.868	
TB	30	2.073	2.073
BM	30		2.875
Sig.		0.838	0.071

b.

Speaker	N	Subset for $\alpha = .05$
		1
Du female	15	1.705
TB female	15	1.903
Du male	15	2.030
TB male	15	2.243
BM female	15	2.442
BM male	15	3.308
Sig.		0.062

The comparisons of the stressed-to-unstressed vowel duration ratio confirm that the shorter the stressed vowel, the more likely the speaker will be Dutch. Dutch speakers have the smallest ratio due to the short stressed vowel. On the other hand, BM speakers have the largest ratio (longest stressed vowels) and they might be evaluated as less Dutch. Apparently, BM speakers realised the stressed vowels with an extreme lengthening so that their temporal structure did not sound properly Dutch. This result explains why BM speakers scored significantly lower to Dutch listeners than other speakers did in the monotonous version (Chapter IV), while TB speakers scored similar to Dutch speakers for all three listener groups.

BM realisations of vowel duration differ significantly from the Dutch realisations. This suggests that the temporal structure of BM differs significantly from the Dutch temporal structure as far as differences in vowel duration are concerned. On the other hand, TB temporal structure does not differ much from Dutch. In many cases, TB realisations resemble Dutch realisations. This finding explains why in Chapter IV, when the verbal and melodic information were deleted and only temporal realisation was presented, Dutch-native listeners could not differentiate well between all three groups. Monotonous speech samples produced by TB speakers were evaluated as similar to those of Dutch speakers, while BM speakers were perceived as more 'foreign' to Dutch native listeners than TB speakers were.

5.2.2 Melodic structures of the targets

In Chapter IV, I found that when only the verbal information was deleted, so that not only temporal but also melodic information was available, Dutch listeners could differentiate the speaker groups rather better. In this section, I will analyse how different the melodic structures of the three groups of speakers are. The pitch data were processed in a similar way as the pitch data from the production experiment (Chapter III). Preliminary inspection showed that all accents were realised by a rise-fall pitch movement. Up to five pitch pivot points p1 to p5 were located in the target word and aligned relative to the onset of the stressed syllable. The pivot points were:

- p1 a low pitch before the onset of the rise located in the syllable before the stressed syllable or in the onset of the stressed syllable. This point is optional, and may therefore be absent;
- p2 a low pitch at the beginning of the stressed syllable. This point is defined as the onset of the rise;
- p3 the peak F0 located in the stressed syllable of the target word;
- p4 a lower pivot point following the peak. This point is usually the end of the fall;
- p5 the terminal pitch of the utterance.

female speakers, realised the peaks after the stressed syllable. The male speakers have also higher peaks than their female counterparts do.

In the following paragraphs, the analyses will be done over all the target words. The speakers will first be classified according to their language background because in the previous section (duration analyses) language background yielded significant differences and gender did not.

Figure 5.4 illustrates the pitch configuration over all the target words, broken down by the language background of the speakers. To normalise over speakers, the pivot points (in ERB) are vertically aligned relative to the starting point of the rise (p2). Timing of the pivot points is expressed relative to the onset of the stressed syllable in the target word.

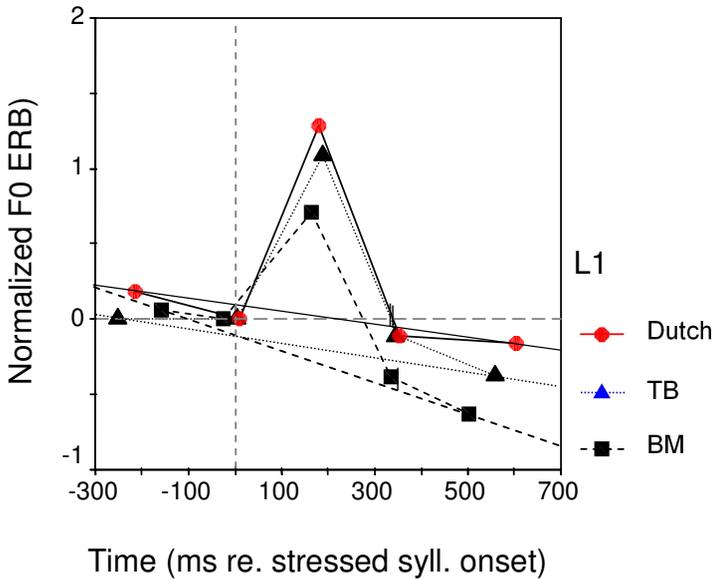


Figure 5.4. Pitch configurations of 8 Dutch words spoken by Dutch, TB and BM speakers. Time scale relative to the onset of the stressed syllable, pitches rescaled relative to p2. The ']' symbols indicate the end of the stressed syllable. The dotted lines are the declination lines of each group.

Figure 5.4 shows that the pitch movements are different for each group. Dutch accent-lending pitch movements are steeper, and span a larger interval, than the other speakers' movements. Pitch movements spoken by speakers with a BM language background are less steep than those of the two other speaker groups. The rises of the Dutch and TB speakers start at the same time, while BM speakers start the rise earlier, i.e. before the onset of the stressed syllable. The peak of the Dutch speakers is the highest, followed by the TB speakers. BM speakers have the lowest peak. The most striking difference, of course, which sets the BM speakers apart from the TB and Dutch speakers, is the much steeper declination function.

The excursion size of the accent-lending rise-fall configuration is considerably smaller for BM than for the other two groups. The TB and Dutch patterns are roughly the same, with one systematic difference, however. Although the onset and terminal pitches are the same for the two groups, the peak is about 0.2 ERB higher for the Dutch speakers than for the TB speakers.

Oneway ANOVAs with the language background of the speakers as a fixed factor indicate that there are significant effects of language background on the parameters of the pitch movements. The parameters per speaker group and their significance levels are listed in table 5.3.

From figure 5.4 and table 5.3 we see that the BM speakers start the rise significantly earlier than the other speakers, who are similar to each other. However, all speakers reach the peak (p3) at the same time. In figure 5.4, we see that the rises reach the peak on different heights. It is clear from table 5.3 that BM speakers have significantly smaller rise excursions than the Dutch and TB speakers, who are similar to each other. BM rises are thus significantly less steep than Dutch and TB rises. In addition, fall movements are also different from each other.

Nevertheless, differences in melodic realisation could also appear for each individual speaker. A one-way ANOVA for the pitch parameters with individual speaker variable as a fixed factor shows that effects of speaker occur highly significant on all parameters, except for the onset time of the rise. Table 5.4 illustrates the significant levels of speaker effects on all parameters.

Table 5.3. Acoustical measurements of the melodic parameters and the levels of significance across speaker group. Overall means with the standard deviations are also indicated.

	Dutch	TB	BM	Mean	s.d.	F(2,105)	Sign.
Rise onset (ms)*	9	5	-26	-4	62	3.74	0.027
Peak alignment (ms)*	179	189	163	177	57	2.00	0.140
F0 peak norm. (ERB)	1.28	1.09	0.71	1.03	0.44	21.36	<0.001
Rise excursion (ERB)	1.28	1.09	0.71	1.03	0.43	23.68	<0.001
Slope rise (ERB/s)	8.00	6.70	4.00	6.20	3.00	22.59	<0.001
F0 end fall norm (ERB)	-0.11	-0.11	-0.38	-0.20	0.42	5.68	0.005
Fall excursion (ERB)	1.39	1.20	1.09	1.23	0.42	4.99	0.008
Slope fall (ERB/s)	-8.90	-7.80	-6.80	-7.80	3.60	3.08	0.050
Declination (ERB/s)	-0.03	-0.04	-0.07	-0.05	0.05	5.53	0.005

* Relative to the onset of the stressed syllable.

Table 5.4. Speaker effects on the melodic parameters and levels of significance.

	F (5,102)	Significance
Rise onset (ms)*	1.68	0.145
Peak alignment (ms)*	5.06	< 0.001
F0 peak (normalized, ERB)	13.48	< 0.001
Rise excursion (ERB)	15.26	< 0.001
Slope rise (ERB/s)	9.34	< 0.001
F0 end fall norm (ERB)	4.84	0.001
Fall excursion (ERB)	3.55	0.005
Slope fall (ERB/s)	5.91	< 0.001
Declination (ERB/s)	7.83	< 0.001

* Relative to the onset of the stressed syllable.

To investigate to what extent effects of language background differ per speaker gender, I separated the pitch configurations into two groups based on gender (male and female), and then compared them according to their language background. Figures 5.5a-b plot the pitch configurations of Dutch words as spoken by speakers with Dutch, TB, and BM language backgrounds for male (left-hand panel a) and female (right-hand panel b) speakers separately. In Figures 5.5 time and frequency have been normalized per speaker in the same way as above.

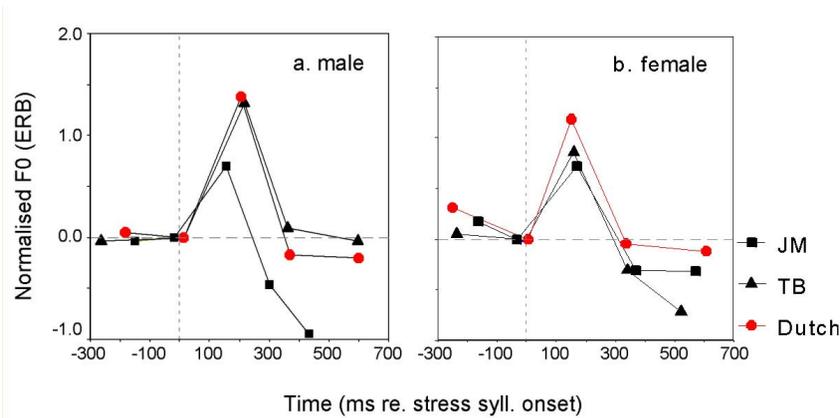


Figure 5.5. Melodic structures of eight Dutch words as spoken by Dutch, TB and BM speakers split into a. male and b. female speakers.

Figures 5.5a-b show that melodic variation differs per gender group. Figure 5.5a shows that the BM male differs clearly from the other (TB and Dutch) male speakers, who are very similar to each other. Figure 5.5b shows that the Dutch female differs from the other (Indonesian) females, who are similar to each other.

An ANOVA indicates that effects of language background on pitch parameters differ per gender group. However, in both gender groups effects of language background are highly significant on the height of the peaks, the rise excursion and the slope of the rises. Table 5.5 lists the significances of language background on pitch parameters in gender groups separately. A post-hoc test indicates that the BM-male pitch curve has a significantly lower peak and a smaller rise excursion than the

pitch curves of the Dutch male and the TB male (both, $p < .001$), which are similar to each other.

The slope of the rise of the BM male speaker is significantly less steep than that of the Dutch ($p = .001$) and the TB male ($p = .007$), who are similar to each other. No significant differences are indicated between the three speakers on the fall parameters, but the declination of the BM male is significantly steeper than that of the Dutch male ($p = .034$) and the TB male ($p < .001$), who are similar to each other ($p = .175$).

As shown in figure 5.5b, the pitch heights of the female speakers are also different from each other. A post-hoc test indicates that the Dutch female has a significantly higher peak than the TB ($p = .004$) and the BM females ($p < .001$), who are similar to each other ($p = .309$). As a result, the Dutch female has a larger rise excursion than the TB female ($p = .002$) and the BM female ($p < .001$), who are, again, similar to each other. As for the male speakers, no significant differences are indicated between the female speakers on the fall parameters. However, significant differences on declinations occur between the Dutch and the TB females ($p = .002$); no significant differences are indicated between the other female speakers.

Once more, this analysis demonstrates that in general, the BM speakers, whether male or female, differ from the Dutch speakers, and to some extent from the TB speakers. The BM speakers realised the melody with a significantly smaller rise movement than the Dutch and the TB. The TB male speaker realised the stress with a melody rather similar to that of the Dutch male speaker, but the TB female did it with a significantly smaller rise excursion than the Dutch female.¹

The results of the melodic analysis explain why the BM realisation of Dutch stress was evaluated as rather poor by the Dutch listeners when only the prosody was audible (i.e. the 'delex' version in Chapter IV). The melodic structures of BM differ significantly from the Dutch melodic structures, and clearly more so than the TB structures. The TB speakers received higher scores for nativeness than the BM speakers. In the yes-no perception experiment the Dutch listeners could not very

¹ In two instances (*baan* and *bamibal*, each one time), the TB female speaker started the rise at a comparatively high pitch. The rise itself was then fairly small.

well detect the non-nativeness in the TB melody (i.e. the ‘delex’ version in Chapter IV): they identified TB speeches as Dutch more often than Dutch speeches.

Table 5.5. Acoustical measurements of the melodic parameters and the levels of significance across speaker group. Means and standard deviations are indicated.

Male	Dutch	TB	BM	Mean	s.d.	F (2,51)	Sign.
Rise onset (ms)*	12	12	-19	1	62	1.53	0.227
Peak alignment (ms)*	205	217	156	193	63	5.53	0.007
F0 peak norm. (ERB)	1.38	1.32	.70	1.14	0.51	14.57	< 0.001
Rise excursion (ERB)	1.38	1.32	.70	1.14	0.49	16.63	< 0.001
Slope rise (ERB/s)	7.50	6.70	4.30	6.30	2.80	9.14	< 0.001
F0 end fall norm (ERB)	-0.17	-0.09	-0.46	-0.18	0.53	5.81	0.005
Fall excursion (ERB)	1.55	1.23	1.16	1.32	0.51	3.19	0.049
Slope fall (ERB/s)	-10.90	-08.50	-8.20	-9.20	4.20	2.37	0.104
Declination (ERB/s)	-0.05	-0.02	-0.09	-.05	0.05	10.63	< 0.001
Female							
Rise onset (ms)*	7	-2	-33	-9	62	2.23	0.118
Peak alignment (ms)*	152	161	169	161	44	>1	0.523
F0 peak norm. (ERB)	1.18	0.86	0.72	0.92	0.33	13.63	< 0.001
Rise excursion (ERB)	1.18	0.86	0.72	0.92	0.32	15.08	< 0.001
Slope rise (ERB/s)	8.40	6.40	3.70	6.20	3.30	13.68	< 0.001
F0 end fall norm (ERB)	-0.05	-0.30	-0.31	-0.22	0.27	6.67	0.003
Fall excursion (ERB)	1.23	1.16	1.02	1.14	0.28	2.76	0.073
Slope fall (ERB/s)	-06.90	-7.00	-5.50	-6.40	2.10	2.96	0.061
Declination (ERB/s)	-0.02	-0.06	-0.04	-0.04	0.04	7.20	0.002

* Relative to the onset of the stressed syllable.

5.3 Correlation between perception and production

The production analyses in §§ 5.1 and 5.2 seem to have yielded some explanations for the results of the perception experiments in Chapter IV. Therefore, I will now

examine the correlation between the perceptual scores and the production parameters. Bivariate correlation analysis (Pearson's correlation coefficient) of the perceptual scores in Chapter IV and the present acoustical parameters was done to investigate whether Dutch listeners used a particular parameter, or group of parameters, when evaluating the speech samples that were produced by Dutch, TB, and BM speakers. The perceptual scores to be correlated with the array of prosodic (both temporal and melodic) acoustic parameters, are those that were collected in the first experiment of Chapter IV, where Dutch listeners were asked to rate monotonized, delexicalized and original versions of words along a foreignness ~ nativeness scale.

I will start this section with the scores of the monotonized stimuli that were presented to determine the perceptual effect of temporal structure only. Next I will examine the correlations for the delexicalized stimuli that were designed to test the perceptual effects of melodic structure (additional to purely temporal structure). Finally, I will look at the correlation between the scores of the original stimuli and all acoustic parameters that have been measured.

5.3.1 Temporal perception vs. temporal parameters

Bivariate correlation analysis between the perception scores obtained for the monotonized stimuli and the temporal parameters shows that the perception scores correlate significantly only with the duration of the stressed syllable and of the stressed vowel for the total group of speakers. However, analysing data per speaker reveals that the correlations are not always significant. In fact, the two parameters are negatively (but insignificantly) correlated for some speakers. Table 5.6 gives the complete list of correlations between the perceptual scores and the temporal parameters, i.e. stressed syllable duration, stressed vowel duration, stressed-to-unstressed syllable duration ratio, and stressed-to-unstressed vowel duration ratio.

The perception score obtained for the monotonized stimuli is significantly correlated with the duration of the stressed syllable when the six speakers are collapsed, but the correlation is negative (but insignificant) for the TB female speaker. The same holds for the relationship between perceptual scores and the

duration of the stressed vowel; here negative (but insignificant) correlations for the Dutch female and the TB female contradict the significant positive correlation for the total speaker group.

Table 5.6. Correlations between perception scores obtained for monotonized stimuli and the temporal parameters per speaker and for all six speakers. The correlation (r), and the number of cases (N) are indicated. Bold numbers indicate a significant correlation.

Speaker	Stressed syllable		Stressed vowel		stressed:unstressed syllable ratio		stressed:unstressed vowel ratio	
	r	N	r	N	r	N	r	N
Dutch male	.082	18	.060	18	-.097	15	-.050	15
Dutch female	.245	18	-.015	18	.084	15	.071	15
TB male	.370	18	.372	18	.148	15	.110	15
TB female	-.067	18	-.045	18	-.392	15	-.214	15
BM male	.708	18	.539	18	.597	15	.609	15
BM female	.434	18	.381	18	.184	15	.123	15
All speakers	.311	108	.259	108	.023	90	-.074	90

The scatterplot in figure 5.6 illustrates the correlation between the monotonous scores and the mean duration of the stressed syllable.

Figure 5.6 shows that there is a linear relationship between the perception scores for the monotonized stimuli and the stressed syllable duration for the six speakers together, but, as mentioned in table 5.6, the correlation between these two parameters is not evenly distributed across all speakers.

Apparently Dutch listeners evaluated the monotonized stimuli by paying attention to the duration of the stressed syllable and/or the stressed vowel. A multiple regression of these two temporal parameters onto the perceptual scores reveals an $R = .313$. This means that the stressed syllable duration determines the multiple correlation; there is no further improvement of the correlation after the

inclusion of other temporal parameters. Overall, the correlation analysis indicates that perception scores obtained for the monotonized versions do not correlate strongly with any temporal parameters. This demonstrates that Dutch native listeners hardly used the temporal parameters when evaluating the monotonized utterances that were spoken by native and non-native speakers of Dutch.

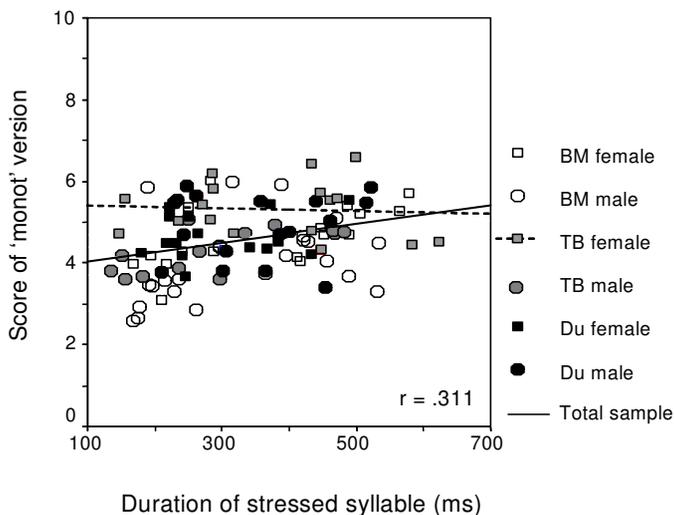


Figure 5.6. Relation between perception scores for monotonized stimuli and stressed syllable duration, broken down by individual speakers. The black dashed line refers to the correlation of the total sample; the solid line refers to that of the TB-female.

5.3.2 Melodic perception vs. melodic parameters

In Chapter IV it was shown that our listeners' ability to evaluate the speech samples improved when melodic information was added to the monotonized delexicalized utterances. Dutch listeners were able to evaluate the Dutch speakers as more native than the non-Dutch speakers when they had access to the melody of the speech. They could also hear that certain non-native speakers were better than others, depending on their language background. The scores of the 'delex' version differ per language group. In addition, in § 5.2.2 significant effects of language background

were found on the melodic parameters of the speech, for instance, the excursion size of the rise and the fall of the accent-lending pitch configuration.

Therefore, in this subsection I will try to determine whether the scores of the ‘delex’ versions correlate with acoustic measurements of the melodic parameters. Table 5.7 lists the correlations between the evaluation scores for the ‘delex’ stimuli and the melodic parameters.

Table 5.7. Correlations between perception scores for delexicalized stimuli and the melodic parameters, per speaker and for all speakers. The correlation (r), and the number of cases (N) are indicated. Bold numbers indicate a significant correlation.

Speaker	Rise onset	Peak align.	F0 peak	Rise exc.	Slope rise	F0 end fall	Fall exc.	Slope fall	Decl.	N
Dutch male	-.283	-.131	-.355	-.242	-.380	-.301	.016	-.121	.006	18
Dutch female	.122	.316	.510	.104	-.088	.024	.334	-.188	-.456	18
TB male	.028	-.162	-.158	-.111	.074	-.099	-.018	.021	-.029	18
TB female	-.076	.232	-.424	.213	-.069	-.590	.095	.299	.166	18
BM male	.457	.436	.182	.422	.333	-.084	.216	-.051	-.216	18
JM female	.084	.149	.086	.278	.219	.054	.043	.131	.463	18
All speakers	.085	.012	.036	.123	.178	-.105	.142	-.018	.030	108

Across all six speakers, the ‘delex’ evaluation scores correlate significantly with the slope of the rise. However, again, the correlations between these two variables are not distributed uniformly over the speakers. Analysing data separately for individual speakers yields different outcomes. No significant correlations are found between criterion and predictor variables for any speaker. The significant correlation between these variables in the collapsed condition does not reflect individual speakers. The ‘delex’ scores and the rise slope are negatively correlated for the Dutch male, the Dutch female, and the TB female (although insignificantly so). Figure 5.7 plots the correlation between the ‘delex’ scores and the rise slope.

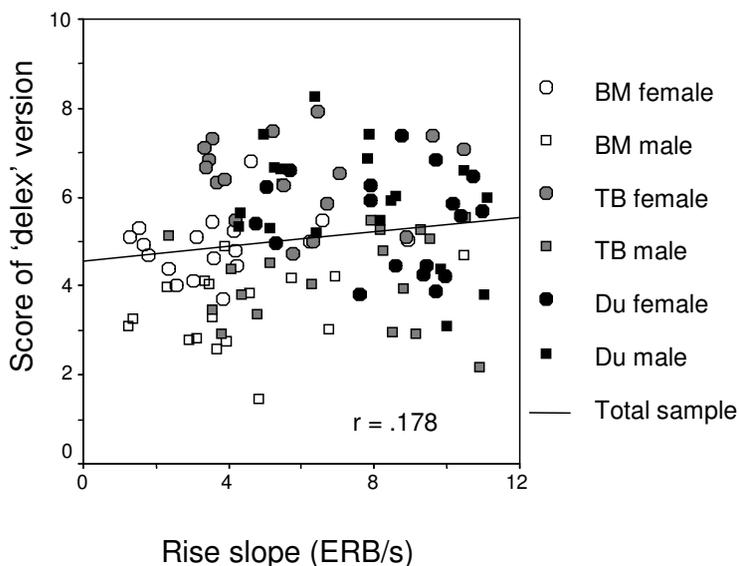


Figure 5.7. Relationship between perception scores of 'delex' stimuli and slope of the rise, per individual speaker and across all speakers. The dashed line refers to the correlation in the full set of six speakers.

The results of the correlation analysis do not clarify how the listeners evaluated the speech on the basis of melodic information. The slope of the rise is the melodic parameter that contributes most to the melodic perception, as is also indicated in § 5.2.2. Individual differences in melodic structures are also found mostly in this parameter.

A correlation analysis between the 'delex' scores and the temporal parameters is also possible because the listeners also heard the temporal information when they were listening to the delexicalized versions. This correlation analysis reveals a significant relationship between the 'delex' scores and the stressed-to-unstressed vowel-duration ratio ($r = -.211$, $p = .023$). However, this correlation is not distributed uniformly across individual speakers. Nevertheless, a multiple regression that combines the melodic and temporal parameter as predictors yields a better total correlation ($R = .270$). This result would seem to indicate that the listeners

considered both melodic and temporal information when asked to evaluate the delexicalized versions of the stimulus utterances.

5.3.3 Contribution of prosodic information in the perception of the originals

Listeners' ability to evaluate speakers was better when they listened to the original stimuli. Dutch listeners could differentiate the non-native speakers from the Dutch native speakers when all information available in the utterances was presented.

In the previous section, it was found that listeners used temporal and melodic information when they were evaluating the nativeness of the delexicalized version. In this section, I will also investigate which acoustical parameters from the temporal and melodic domains might have been used in the evaluation of the original versions. Table 5.8 lists the results of the correlation analysis between the 'original' scores and the temporal parameters.

Table 5.8. Correlations between perception scores for original stimuli and the temporal parameters, per speaker and for the all speakers. The correlation (r), and the number of cases (N) are indicated. Bold numbers indicate a significant correlation.

Speaker	Stressed vowel		Stressed syllable		stressed:unstressed syllable ratio		stressed:unstressed vowel ratio	
	r	N	r	N	r	N	r	N
Dutch male	-.196	18	.612	18	.568	15	.298	15
Dutch female	.101	18	-.162	18	-.343	15	-.490	15
TB male	.301	18	.315	18	.373	15	.689	15
TB female	.117	18	.174	18	.041	15	-.053	15
BM male	.066	18	-.059	18	.077	15	.252	15
BM female	-.522	18	-.476	18	-.458	15	-.509	15
All speakers	-.204	108	-.037	108	-.138	90	-.254	90

Correlation analysis between the temporal parameters and the original scores across all speakers demonstrates a weak but significant relationship between the 'original'

scores and the stressed vowel duration ($r: -.204, p = .017$), and the stressed-to-unstressed vowel-duration ratio ($r: -.254, p = .008$). However, the correlation is not distributed uniformly across the individual speakers. The negative relationships in the collapsed data are in contrast to the positive relationships for some individual speakers.

In general, Dutch listeners seem to use vowel duration as their primary source of information when evaluating the nativeness of original speech samples spoken by native Dutch, and TB and BM speakers. Differences in temporal structure are also found mostly in stressed-vowel duration and stressed-to-unstressed vowel-duration ratio (§ 5.2.1.). Figure 5.8 shows the correlation between the scores of the original stimuli in the perception experiment and the stressed-to-unstressed vowel-duration ratio with a regression line across all speakers.

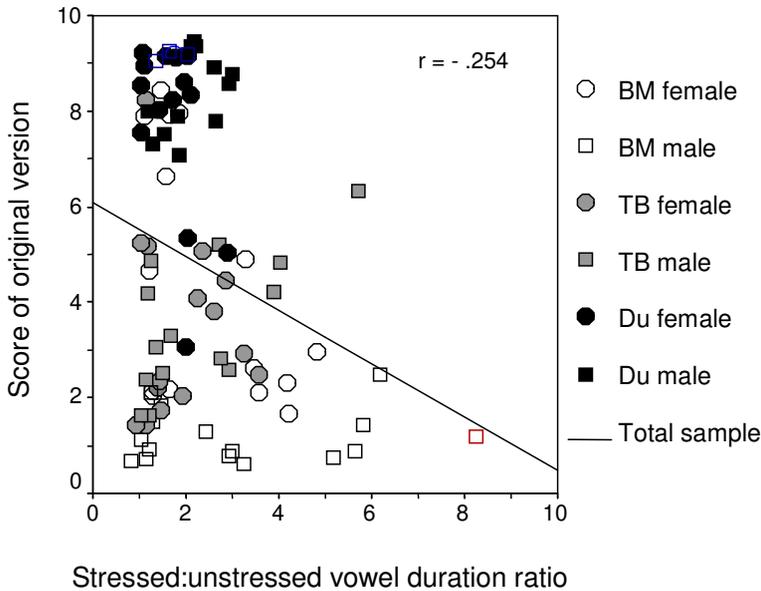


Figure 5.8. Relationship between perception scores of original stimuli and stressed:unstressed vowel duration ratio, per individual speaker and across all speakers. The regression line has been drawn through the collective data of all six speakers.

The negative correlation for the stressed-to-unstressed vowel-duration ratio is in line with our earlier prediction based on the acoustical measurements that the smaller the ratio the more the speaker would be judged as Dutch. The same relationship is found for the stressed vowel duration; the shorter the stressed vowel the more the speaker was judged as Dutch.

Nevertheless, the listeners primarily used the melody to evaluate the complete speech samples. Correlation analysis between melodic parameters and scores for the original stimulus versions demonstrates that all melodic parameters correlate significantly with the 'original' scores, especially when the data are pooled over all speakers. Table 5.9 lists the correlations between the melodic parameters and the scores for the original stimulus versions.

Table 5.9. Correlations between perception scores for original stimuli and the melodic parameters, per speaker and for all speakers. The correlation (r), and the number of cases (N) are indicated. Bold numbers indicate a significant correlation.

	Rise onset	Peak align.	F0 peak	Rise exc.	Rise slope	F0 end fall	Fall exc.	Fall slope	Decl.	N
Dutch male	.186	-.018	.077	.076	.158	.476	-.530	-.286	.446	18
Dutch female	.359	.484	.625	.379	.136	.121	.338	-.503	-.216	18
TB male	.123	-.077	-.166	-.251	.027	-.100	-.022	.029	-.314	18
TB female	.155	.226	.033	.379	.127	-.316	.359	-.063	.444	18
BM male	-.093	.137	-.369	-.243	-.168	-.077	-.316	.110	.111	18
BM female	.588	.426	.412	.288	.389	-.051	.457	-.313	-.240	18
All speakers	.267	.175	.426	.446	.411	.184	.268	-.204	.254	108

Table 5.9 shows that in the collapsed data the scores of the original versions correlate significantly with all melodic parameters, especially the height of the peak, the rise excursion and the slope of the rise. The levels of significance reflect the results of the acoustical analysis of the melodic parameters in § 5.2.2. The height of the peak, the rise excursion and the slope of the rise are the three most significant parameters to differentiate the language background of the speakers. Apparently, listeners use melodic information mostly from the original speech samples to differentiate non-native speakers from native speakers. However, analysis of individual speakers reveals different outcomes. The relationships across all speakers are in contrast to the relationship of some individual speakers. Figure 5.9 plots the correlation between the scores of the original version and the rise excursion.

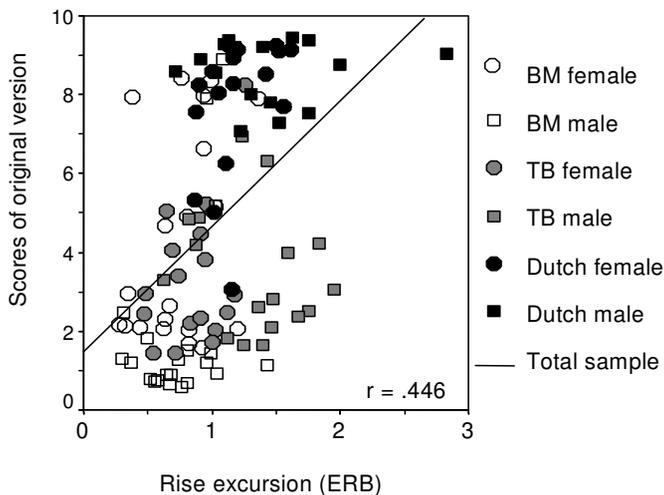


Figure 5.9. Relationship between perception scores of original stimuli and rise excursion, per individual speaker and across all speakers. The line refers to the correlation in the full set of six speakers.

Figure 5.9 shows clearly that the larger the excursion size of the rise, the higher the nativeness score a speaker can get for his/her original version. Dutch listeners gave high scores for original stimuli that were produced by Dutch speakers because the

Dutch speakers realised the stressed syllable with a larger rise excursion than other (non-Dutch) speakers did.

Figure 5.10 illustrates the correlation between the nativeness scores of the original versions and the slope of the fall across all speakers.

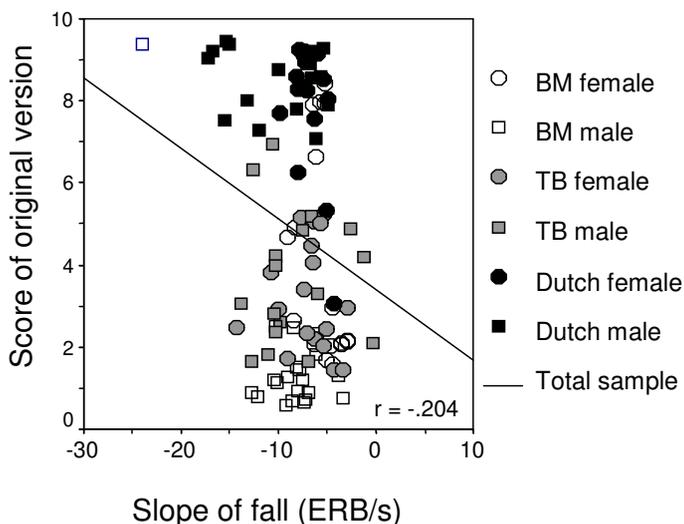


Figure 5.10. Relationship between perception scores of original stimuli and the slope of the fall, per individual speaker and across all speakers. The solid line refers to the correlation in the full set of six speakers.

Figure 5.10 indicates that the steeper the slopes of the fall are, the higher the nativeness scores of the ‘original’ stimuli. The falls realised by Dutch speakers are in general steeper than those of TB and JM speakers. As a result, the original utterances spoken by Dutch speakers scored higher than the original utterances of non-Dutch speakers.

A multiple regression between temporal and melodic parameters as predictors and ‘original’ scores as the criterion yields a high multiple correlation ($R = .525$). Compared to the regression analyses of the monotonized and delexicalized stimulus versions, in the previous sections, the magnitude of the present correlation

coefficient is substantially greater. Apparently, prosodic information is used effectively only if the prosodic cues are presented in the company of the full segmental information.

5.4 Summary

The acoustical analyses of temporal parameters of utterances spoken by Dutch, TB and BM speakers demonstrate that stressed vowel duration and stressed-to-unstressed vowel-duration ratio contribute most to differentiate the temporal structure of a particular group of speakers. BM speakers realise the stressed vowel significantly longer than the Dutch and the TB speakers do, who are close to each other. BM speakers realised stress with an extreme lengthening of the vowel in the stressed syllable. Furthermore, the stressed-to-unstressed vowel-duration ratio of the BM speakers is the largest followed by the TB speakers, and that of Dutch speakers is the smallest.

The results of the temporal analyses indicate that the shorter the stressed vowel, the more likely the speaker will be Dutch. Unfortunately, this finding was not confirmed by the correlation analysis that yields a positive correlation between the perception scores of the monotonized stimuli and the stressed vowel duration. Moreover, also the multiple correlation between the monotonized scores and the ensemble of temporal parameters is low. Therefore, it is clear that listeners could hardly differentiate the language background of the speakers just by listening to the monotonized stimuli. The results of the identification experiment (Chapter IV) revealed that in the 'monot' version, the identification as Dutch is roughly the same for all speaker groups.

On the other hand, temporal analysis of the stressed-to-unstressed vowel-duration ratio per speaker group shows that Dutch speakers have the smallest ratio followed by the TB speakers, and the BM speakers have the largest ratio. This seems to constitute a perceptual cue given the significant (inverse) correlation between the 'delex' scores and the stressed-to-unstressed vowel-duration ratio. The smaller the stressed-to-unstressed vowel-duration ratio is, the higher the score of the 'delex' version the speaker gets. It seems that listeners could differentiate the temporal

structure of the speakers better when the melodic information was added to the monotonous speech samples.

The analysis of the melodic parameters shows that the melodic structures of the BM speakers differ in general from those of the Dutch speakers and to some extent from those of the TB speakers. The differences are clearly observed in the rise movements. This confirms the low nativeness scores for the BM speakers when melodic as well as temporal information was presented ('delex' version). The melodic structures of the TB speakers resemble to some extent those of the Dutch speakers. This explains why the TB speakers received higher nativeness scores than the BM speakers in the delexicalized version.

The correlation analysis of the melodic parameters showed that only the slope of the rise correlates significantly with the delexicalized scores. According to the acoustical analyses, the slope of the rise is the melodic parameter that contributes most to the differentiation of language background of the speakers (§ 5.2.2). However, the low significance of this relationship clarifies why listeners could hardly differentiate the non-native from the native speakers when listening to the delexicalized versions. As a consequence of this, in the second experiment of chapter IV, where listeners were instructed to make a binary native/foreign decision, the 'delex' versions of the TB speakers were identified as Dutch more readily than the native Dutch speakers were themselves.

Higher significant correlations were found between the temporal and melodic parameters as predictors and the nativeness scores obtained for the original stimuli. These findings correspond to the results of the perception experiment in Chapter IV showing that the listeners could evaluate the speakers better when all information was accessible. The temporal and melodic realisations of the non-native speakers of Dutch sounded clearly foreign to the Dutch listeners when they heard the pronunciation of the vowels and consonants as well. This in itself should not come as a surprise, as foreign accent is generally heard most clearly in the realisation of the segments. The interesting effect is that the strength of the correlations between nativeness judgments and prosodic parameters increases as the segmental quality is better.

Chapter VI

General Discussion

6.1 Summary and discussion

The general question of this thesis was how speakers of a non-stress language and those of a stress language differ in their realisation of stress and/or accent. Toba Batak (TB) and Betawi Malay (BM) differ crucially in that TB has word stress and BM has not. After a brief discussion of prosody and its functions in **Chapter II**, I focused on the production and perception of word prosody of BM and TB, in particular on the effects of focus and pre-boundary position on temporal and melodic structure. I expected TB speakers to mark the differences between stressed and unstressed syllables in focussed words more clearly than BM speakers. As regards boundary marking I expected to find similar effects for both languages.

My findings in the comparisons of focus and pre-boundary effects on the temporal and melodic structure of words did not always reflect what I expected. The production experiments in **Chapter III** proved that boundary effects were indeed almost similar for both languages; these effects were, as expected, strongest on the ultimate syllables. However, my hypothesis that focus effects would be stronger in a stress language than in a non-stress language, was not confirmed. As regards duration, focus effects were small or even insignificant in TB, and occurred on vowels only, if at all; possibly the duration of singleton consonants is limited in this language, so as not to be confused with the geminate consonant (cf. Cohn et al., 1999 and Berinstein, 1979). In BM, on the other hand, duration effects of focus were highly significant.

Focus effects on the pitch contours were highly significant in both languages, but stronger in the non-stress language, i.e. BM. It is, however, important to notice that in this language, boundary affected both the shape and the position of pitch

movements in the words considerably. In TB only the size of the movement, but not the shape of the accent-lending pitch configuration changed as a function of focus, such that the larger movements occurred on focussed words; in contrast to BM, the pitch movements in non-focussed words were smaller but never deleted in TB.

On aggregate, it would appear that my hypothesis that a language with contrastive word stress (such as TB) marks its stresses/accents more clearly than a language without word stress (such as BM) is not supported by my results. In fact, in terms of temporal structure the results would rather support the opposite, as the effects of presence versus absence of focus (signalled by accentuation) on vowel duration were larger and more systematic in BM than in TB. Also in the melodic domain did I find large effects of focus in BM, larger at least at first sight than in TB. There was, however, an important difference in the effects of focus on melodic structure in BM as opposed to TB. In BM pitch movements occurred only on prominent words, and their shapes depended on the position of the target word in the sentence. In TB, the shape and the position of the pitch configuration (always a rise-fall) was constant, i.e. unaffected by focus, but only the size of the rise-fall movement varied as a function of focus: the movements were larger in focussed than in non-focussed words.

It is not clear, then, how these results should be interpreted. I would argue, first of all, that the existence of (contrastive) word stress in TB corresponds to the finding that the position of a prominence-lending rise-fall configuration is always in the syllable that carries the word stress. In BM the position of the pitch configuration varies over the syllables in the word, and interacts with the presence of an utterance boundary following the target word. The variability in shape and position of the pitch configuration seems to indicate that there is no need to tie the prominence marking to one specific syllable, which is fully compatible with a system that has no word stress.

That TB speakers marked word stress more clearly was also indicated by the perception experiments in **Chapter IV**. Prosodically, TB non-native speakers of Dutch were significantly better accepted by Dutch listeners than BM speakers of Dutch as far as the realisation of word stress was concerned. TB speakers were

evaluated as being more like Dutch speakers, while BM speakers were always less acceptable. This indicates that the non-stress language (BM) is prosodically different from both stress languages (TB, Dutch). However, the non-nativeness of the Indonesian pronunciations of Dutch was better detected by Dutch listeners when all information, in particular segmental, was available. This indicates that there is some, rather weak, information on non-nativeness in the prosody of the Indonesian learners of Dutch, but that there are clearer cues in the segmental structure, i.e. the pronunciation of the vowels and the consonants. Interestingly, when the two groups of Indonesians were exposed to the same stimuli, they could not differentiate the language background of the speakers.

In **Chapter V**, duration and pitch of the stimuli used in Chapter IV, i.e. Dutch words spoken by BM, TB and Dutch speakers, were acoustically measured and analysed. The results indicated that BM speakers produced the stressed syllable with a significantly longer vowel than the other speakers did. They also had the largest stressed-to-unstressed vowel-duration ratios, which differed significantly from those of the Dutch speakers, which were the smallest, while the TB speakers' ratios were in between. This comparison indicated that the shorter the stressed vowel and the more even the stressed-to-unstressed vowel-duration ratio, the more likely the speaker will be Dutch. In a study on Indonesian word stress realized by TB, Sundanese and Javanese speakers, van Zanten and van Heuven (1997) found that duration was a possible cue for stress in TB. The duration effects that I found in TB were smaller than expected. More surprisingly, they were stronger in the non-stress language BM.

As regards melodic parameters, the BM speakers had significantly earlier rises, and smaller and less steep pitch movements, than Dutch and TB speakers. Native Dutch accent-lending pitch movements were larger and steeper than those of the BM and TB speakers. The TB speakers' rises started in the same position as the Dutch rises and led to roughly similar melodic configurations, but their pitch peaks were rather smaller.

Finally, correlation analyses of the perceptual nativeness scores (Chapter IV) and acoustical prosodic parameters for the three groups of speakers (**Chapter V**)

showed low overall correlations between perception and acoustics when monotonized versions were presented. This reflects a poor ability of listeners to differentiate non-native from native speakers when segmental and melodic information was removed from the stimulus. On the other hand, a significant (negative) correlation was found between the stress-to-unstressed vowel-duration ratio and the nativeness scores for delexicalized stimuli, which contained melodic information in addition to purely temporal information. This indicates that the (Dutch) listeners differentiated the three language groups better when the melody was added to the monotonous speech. Importantly, the correlation between perceived nativeness and temporal structure was better when melodic information was added (in the ‘delex’ condition) than when temporal structure was the only cue made available to the listeners (as in the ‘monot’ condition).

The correlation coefficient of the ‘delex’ scores and the melodic parameters revealed one significant relationship, i.e. the slope of the rise, which was also the acoustic parameter that most clearly differentiated between the three language groups. However, the low correlation coefficient of the melodic parameters and the ‘delex’ scores indicated that the listeners could not clearly differentiate between the three language backgrounds when listening to the delexicalised versions. This explains why listeners confused TB and Dutch speakers in the binary choice experiment (native ~ foreign) when presented with delexicalized versions.

Better correlations were found between the prosodic parameters and the nativeness scores for the original stimuli, which confirmed that the listeners recognised the speakers the best when they also heard the pronunciation of the vowels and consonants. Remarkably, here again the correlations between temporal and melodic parameters and perceived nativeness increased substantially relative to the stimulus versions with only melodic and/or temporal information.

Generalizing from the above summary of results, it seems as if prosodic, i.e. temporal and melodic, acoustical properties contribute more to perceived nativeness as the stimulus quality is better, i.e. approaches human speech more closely. One possible interpretation of this result would be that listeners are more proficient in detecting traces of foreignness as the stimulus is more natural, i.e. closer to

everyday-life experience. A possible alternative explanation of the effect would lie in statistical artefact. As can be seen in figure 4.2 the spread of the nativeness scores is very small and tightly clustered around the midpoint of the scale for the monotonized versions. Perceived nativeness is more differentiated for the delexicalized condition and is clearly largest for the original stimuli. It is a general characteristic of bivariate distributions that correlation coefficients are reduced when one of the variables has little variance.

The tendency of the BM speakers to lengthen the stressed vowel in Dutch words (Chapter V) may reflect the strong lengthening effect on (full) vowels that are realised in accented words in BM (Chapter III). On the other hand, the smaller stressed-unstressed vowel-duration ratio in TB Dutch could be seen as a reflection of the absence of accentual lengthening in pre-boundary position in TB.

More generally, the findings of the melodic analyses of Dutch words spoken by the two BM and the two TB speakers (as found in the stimulus analysis in Chapter V) basically reveal the same temporal and melodic structures of words as reported in Chapter III for four speakers of each of these languages.

When we compare the penultimate rise-fall contours in focused pre-boundary words from both Indonesian speaker groups, we find that in both experiments BM speakers start the rise about 20 ms earlier than TB speakers. The pitch peak in BM is about 0.6 ERB lower than its counterpart in TB, so that the rise-fall pitch movements in BM are generally smaller than pitch movements in TB. It is thus reasonable that BM speakers realised the melody of the Dutch word stress differently than the TB speakers. The convergent results of the acoustical analyses of Chapter III and those of Chapter V indicate that the native language clearly affects the realisation of structures in the second language.

To sum up, the non-stress BM speakers systematically used duration and pitch in order to realise accent and boundary marking; the stress language TB speakers used duration almost exclusively for boundary marking, but used pitch non-uniquely to mark accent and stress as well as boundaries. Finally, the stress-language listeners were rather better at identifying the speakers' language background than the non-stress language listeners.

6.2 Suggestions for further research

In the description of the prosodic systems of BM and TB, in Chapter III, I analysed durational and pitch effects of stress/accent. Due to lack of time, I restricted the analysis to the most important prosodic features, i.e. pitch and duration. I did not measure other factors such as vowel quality, intensity and loudness. It is known from the literature that unstressed syllables not only tend to be shorter than stressed syllables, but also that vowels in unstressed syllables are often reduced in quality, intensity and loudness. In the future, I intend to research these correlates of perceived prominence, to see whether a stress language like TB differs from a non-stress language like BM also in these respects.

In the study of BM, more attention should be paid to words with the central vowel schwa. My findings so far show that a penultimate syllable containing schwa is perceived as accented if it occurs sentence-medially. However, this vowel and its syllable were never lengthened. In the pitch analysis, I did not pay attention to words with only schwas. A separate study seems called for to analyse the prosodic parameters of words with only schwas, like *seneng* [sənəŋ] ‘happy’, *denger* [dəŋəŋ] ‘hear’, *kelelep* [kələləp] ‘be drowned’. In such research, we may investigate the prosodic characteristics of words with only schwas, and compare these to words with only full vowels or both full vowels and schwas.

Further, I did not differentiate the BM dialects from each other. These dialects are spoken in two different regions; the dialect of the border region is more influenced by other regional languages, while the inner-city dialect is less influenced. Comparative studies on these two dialects would be a valuable contribution to the description of prosody of Indonesian languages. This would also present a welcome opportunity to describe an endangered language; indigenous users of traditional BM dialects have been a shrinking minority for several decades (Grijns 1991a, b).

As regards TB, I used word pairs with distinctive consonant length like *pitu* – *pittu*, but I did not pay special attention to the effect of gemination. It would be interesting to compare these two word types to know how long a consonant must be in order to be perceived as a geminate, and whether their melodic structures are different. Further, experimental research on word pairs with contrastive stress, such in *tibo* [N] ‘height’ and *tibo* [A] ‘high’, would be a good contribution to the description of TB prosody.

The contribution of the present study to the description of prosody in Indonesian languages is limited to word-level prosody only. Further studies are also needed to describe the prosody of Indonesian vernaculars completely, i.e. on levels higher than word, such as the phrase and the utterance. As regards TB, research beyond the word level would be all the more interesting, as TB has several levels of stress in compound words and in phrases as well (Nababan, 1981).

The results of the acoustical analyses of the effects of focus and boundary on the temporal and melodic structure of utterances should be complemented by a series of perception experiments. The purpose of such follow-up studies would be to ascertain whether the acoustic differences observed for the variations of focus and boundary in these two Indonesian languages are perceptually relevant. For instance, I found that the shapes as well as the location within the target word of the pitch configurations differed systematically depending on the focus and boundary conditions of the targets. By systematically exchanging the shapes and positions of the pitch configurations in the original stimuli (presented in their full context so that the speaker’s intended focus distribution would be transparent) we could determine whether certain shapes and positions are obligatory for a specific combination of focus and boundary, or that any shape and position would be acceptable. This would then allow us to develop a clearer view of the constraints that apply to the prosodic structures and their communicative functions in the languages of Indonesia.

In Chapter IV, I researched to what extent stress realizations by non-native speakers sounded ‘Dutch’ to native speakers of Dutch. I did not look at the contribution of segmental information on its own, i.e. speech from which the prosodic information has been removed. Thus, I have not been able to compare the

relative contributions of prosody and segments to foreign accent. Varying the full set of cues, temporal, melodic and segmental, in order to determine their relative contribution to the perception of non-nativeness for Dutch listeners would constitute a desirable research agenda by itself.

During the recording sessions in which the basic materials for the listening tests were collected, it appeared that the Dutch speakers spoke in a rather assertive way, whereas the Indonesian speakers pronounced their materials rather more hesitantly and quietly. This extra-linguistic information was so obvious, that I suspected that because of this information alone, listeners would be able to detect the language background of the speakers, even when only the durational structure was audible. Therefore, the Dutch native speakers were requested to speak less assertively and more slowly than they would habitually do. In spite of this precaution, we cannot exclude the possibility that factors like speaking rate and assertiveness may have contributed to successful identification of language background, especially when speakers from very different cultural backgrounds, like Indonesia and the Netherlands, are involved. If this is indeed the case, teaching such non-linguistic behaviour would deserve a place in second-language teaching.

The results of the perception experiments have implications for the teaching of Dutch as a foreign language to Indonesian nationals. For teaching purposes, more attention is needed in the area of prosody, especially to the learners hailing from non-stress vernaculars (such as BM), who clearly constitute the majority in the Indonesian archipelago. Furthermore, a longitudinal study on Dutch as an L2, with the students of the Dutch departments in Indonesian universities as subjects could reveal useful information on the problems of the Dutch speaking skills, especially Dutch prosody, among the Indonesian students. With this information, a better teaching method could be devised in order to improve the speaking skills of the students.

References

- Adelaar, A. (1981) Reconstruction of Proto-Batak Phonology. *NUSA* 10, 1-20.
- Adelaar, A. (2005) The Austronesian languages of Asia and Madagascar: a historical perspective. In A. Adelaar & N.P. Himmelmann (eds) *The Austronesian languages of Asia and Madagascar*, 1-42. London/New York: Routledge.
- Adriaens, G. (1992) From COGRAM to ALCOGRAM: Toward a Controlled English Grammar Checker. In Ch. Boitet (ed.) *Proceedings of the 15th International Conference on Computational Linguistics*, 595-601. Nantes: COLING.
- Alieva N.F., V.D. Arakin, A.K. Ogloblin & J.X. Sirk (1991) *Bahasa Indonesia: deskripsi dan teori [Indonesian: description and theory]*. Seri ILDEP, Yogyakarta: Kanisius.
- Beaugendre, F. (1996) Modèles de l'intonation pour la synthèse. In H. Meloni (ed.) *Fondements et Perspectives en Traitement Automatique de la Parole*. AUPELFUREF, 97-107.
- Beckman, M. E. (1986) *Stress and non-stress accent*. Dordrecht/Cinnaminson: Foris.
- Berinstein, A. E. (1979) *A Cross-Linguistic Study on the Perception and Production of Stress*. UCLA Working Papers in Phonetics 47. University of California, Los Angeles.
- Bezooijen, R. van (1984) *Characteristics and recognizability of vocal expressions of emotion*. Dordrecht/Cinnaminson: Foris.
- Boersma, P. & D. Weenink (1996) Praat: A system for doing phonetics by computer. *Report of the Institute of Phonetic Sciences, University of Amsterdam* 132.
- Bolinger, D.L. (1958) A theory of pitch accent in English. *Word* 14, 109-149.
- Bond, Z.S., V. Stockmal & D. Muljani (1998) Learning to identify a foreign language. *Language Sciences* 20, 353-367.
- Cambier-Langeveld T. (2000) *Temporal Marking of Accents and Boundaries*. Utrecht: LOT (LOT dissertation series 32).
- Caspers, J. (1998) Who's next? The melodic marking of question versus continuation in Dutch. *Language and Speech* 41, 375-398.
- Chaer, A. (1976) *Kamus Dialek Jakarta [Dictionary of the Jakartan Dialect]*. Ende, Flores, Indonesia: Nusa Indah.
- Chen, P. (1984) Stress in Toba Batak. In P. Schachter (ed.) *Studies in the Structures of Toba Batak*. UCLA Occasional Papers in Linguistics 5, 1-8.
- Chun, D.M. (2002) *Discourse Intonation in L2: From Theory and Research to Practice*. Amsterdam: John Benjamins.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37-46.
- Cohn, A. (1989) Stress in Indonesian and Bracketing Paradoxes. *Natural language & linguistic theory* 7, 167-216.

- Cohn, A., W. Ham & R.J. Podesva (1999) The phonetic realization of singleton-geminate contrasts in three languages of Indonesia. *Proceedings of the 14th International Congress of Phonetic Sciences*, 587-590.
- Cohn, A. & J. McCarthy (1994) Alignment and Parallelism in Indonesian. *Working papers of the Cornell phonetics laboratory* 12, 53-137.
- Cruttenden, A. (1997) *Intonation*. 2nd edition. Cambridge: Cambridge University Press.
- Cutler, A. & W. van Donselaar (2001) Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44, 171-195.
- Dogil, G. (1999) The phonetic manifestation of word stress. In Harry van der Hulst (ed.) *Word Prosodic Systems in the Languages of Europe*. Berlin: de Gruyter, 273-334.
- Dupoux, E., C. Pallier, N. Sebastian & J. Mehler (1997) A Destressing “Deafness” in French? *Journal of Memory and Language* 36, 406-421.
- Ebing, E. (1997) *Form and function of pitch movements in Indonesian*. Leiden: Research School CNWS.
- Edwards J., M.E. Beckman & J. Fletcher (1991) The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89, 369-382.
- Eefting, W. (1991) The effect of ‘information value’ and ‘accentuation’ on the duration of Dutch words, syllables, and segments. *Journal of the Acoustical Society of America* 89, 412-424.
- Eefting, W. & S.G. Nootboom (1991) The effect of accentedness and information value on word durations: a production and a perception study. *Proceedings of the 22nd International Congress of Phonetic Sciences*, Aix-en-Provence, 302-305.
- Emmorey, K. (1984) The intonation system of Toba Batak. In P. Schachter (ed.) *Studies in the structures of Toba Batak. UCLA Occasional Papers in Linguistics*, 5, 37-58.
- Fant, G. (1961) The acoustics of speech. *Proceedings of the 3rd International Congress on Acoustics*, 188-201.
- Farnetani, E. (1997) Coarticulation and connected speech processes. In W.J. Hardcastle & J. Laver (eds) *The Handbook of Phonetic Sciences*, 371-404. Oxford: Blackwell.
- Fokker, A.A. (1895) *Malay Phonetics*. Leiden: Brill.
- Fry, D.B. (1955) Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustic Society of America* 27, 765-768.
- Fry, D.B. (1958) Experiments in the Perception of Stress. *Language and Speech* 1, 126-152.
- Fry, D.B. (1965) The Dependence of Stress Judgements on Vowel Formant Structure. *Proceedings of the 6th International Congress of Phonetic Sciences*, 306-311.
- Garde, P. (1968) *L’accent*. Paris: Presses Universitaires de France.
- Gerth van Wijk, J. (1985, first published in 1883) *Spraakleer der Maleische Taal [Malaysian grammar]*. Batavia: G. Kolff & Co.

- Gooskens, C. (1999) *On the role of prosodic and verbal information in the perception of Dutch and English language varieties*. PhD dissertation, Catholic University Nijmegen, The Netherlands.
- Grabe, E., B. Post, F. Nolan & K. Farrar (2000) Pitch accent realisation in four varieties of British English. *Journal of Phonetics* 28, 161-185.
- Grijns, C. D. (1991a) *Jakarta Malay: A multidimensional approach to spatial variation*. 2 vols. VKI 149. Leiden: KITLV.
- Grijns, C. D. (1991b) *Kajian bahasa Melayu-Jakarta*. (Seri ILDEP) Jakarta: Grafiti. [Indonesian translations by Rahayu Hidayat et al. of the 1976, 1979, and 1981 articles by Grijns.]
- Grosz, B. & C. Sidner (1986) Attention, Intentions, and the Structure of Discourse. *Computational Linguistics* 12, 175-204.
- Grosz, B. & C. Sidner (1998) Lost Intuitions and Forgotten Intentions. In M. Walker, A. Joshi & E. Prince (eds) *Centering in Discourse*. Oxford: Oxford University Press, 39-51.
- Gussenhoven, C. (1984) *On the grammar and semantics of sentence accents*. Dordrecht/Cinnaminson: Foris.
- Gussenhoven, C. (2004) *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Halim, A. (1974) *Intonation in relation to syntax in Bahasa Indonesia*. Jakarta: Djambatan.
- Hart, J. ʔ, R. Collier & A. Cohen (1990) *A perceptual study of intonation*. Cambridge: Cambridge University Press.
- Hermes, D.J. & 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 (1994) Introducing prosodic phonetics. In C. Odé & V.J. van Heuven (eds) *Experimental studies of Indonesian prosody*, 1-26. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University. (Semaian 9).
- Heuven, V.J. van (1998) Effects of focus distribution and accentuation on the temporal and melodic organisation of word groups in Dutch. In S. Barbiers, J. Rooryck & J. van de Weijer (eds) *Small words in the big picture*, 37-42. Leiden: HIL (HIL Occasional Papers 2).
- Heuven, V.J. van (2002) *Boven de klanken [Beyond the Segments]*. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen.
- Heuven, V.J. van & R.S. Kirsner (2004) Phonetic or phonological contrasts in Dutch boundary tones? In L. Cornips & J. Doetjes (eds) *Linguistics in the Netherlands*, 102-113. Amsterdam/Philadelphia: John Benjamins (AVT Publications 21).
- Hirose, K. (1997) Disambiguating recognition results by prosodic features. In Y. Sagisaka, N. Campbell & N. Higuchi (eds) *Computing prosody – Computational Models for Processing Spontaneous Speech*, 327-342. New York: Springer.
- Hirschberg, J. & J. Pierrehumbert (1986) Intonational Structuring of Discourse. *Proceedings of the 24th Meeting of the Association for Computational Linguistics*, 136-144.
- Ikranagara, K. (1980) *Melayu Betawi grammar*. Jakarta: Universitas Atma Jaya.

- Jong, K. de, M.E. Beckman & J. Edwards (1993) The interplay between prosodic structure and coarticulation. *Language and Speech* 36, 197-212.
- Kähler, Hans (1966) *Wörterverzeichnis des Omong Djakarta*. Berlin: Dietrich Reimer.
- Kamiyama, T. (2004) Perception of foreign accentedness in L2 prosody and segments: L1 Japanese speakers learning L2 French. Conference paper presented at Speech Prosody 2004, the International Conference for ISCA; Nara, Japan.
- Katwijk, A. van (1974) *Accentuation in Dutch, An experimental linguistics study*. Assen/Amsterdam: Van Gorcum.
- Keating, P.A. (1994) *Phonological structure and phonetic form*. Cambridge: Cambridge University Press.
- Klatt, D.H. (1985) *From text to speech*. Cambridge, MA: MIT Press.
- Ladd, D. R. (1996) *Intonational phonology*. Cambridge: Cambridge University Press (Cambridge Studies in Linguistics 79).
- Ladd, D.R. & J. Terken (1995) Modelling inter- and intra-speaker pitch range variation. In K. Elenius & P. Branderud (eds) *Proceedings of the 13th International Congress of Phonetic Sciences*, 386-389.
- Ladefoged, P. (1982) *A Course in Phonetics*. New York: Harcourt, Brace & Jovanovich.
- Ladefoged, P. & I. Maddieson (1996) *The Sound of the world's languages*. Cambridge, MA: Blackwell.
- Laksman, M. (1994). Location of stress in Indonesian words and sentences. In C. Odé & V.J. van Heuven (eds) *Experimental studies of Indonesian prosody*, 108-139. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University (Semaian 9).
- Lehiste, I. (1970) *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lehiste, I. & R.A. Fox (1992) Perception of prominence by Estonian and English listeners. *Language and Speech* 35, 419-434.
- Leyden, K. van (2004) *Prosodic characteristics of Orkney and Shetland dialects. An experimental approach*. Utrecht: LOT (LOT dissertation series 92).
- Lieberman, P. & S. E. Blumstein (1988) Speech physiology, speech perception, and acoustic phonetics. *Cambridge Studies in Speech Science and Communication*. Cambridge: Cambridge University Press.
- Lindblom, B.E.F. (1990) Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (eds) *Speech Production and Speech Modelling*. Dordrecht: Kluwer.
- Maeda, S. (1976) A characterization of American English intonation. Ph.D. thesis. Cambridge: MIT.
- McAllister, R., J. Flege & T. Piske (2000) Aspects of the Acquisition of Swedish Quantity by Native Speakers of English, Spanish and Estonian. *Proceedings of the 13th Swedish Phonetics Conference (FONETIK 2000)*, Skövde, Sweden, May 24-26, 2000.

- Moeliono, A.M. & S. Dardjowidjojo (1988) *Tata Bahasa Baku Bahasa Indonesia [A comprehensive grammar of Indonesian]*. Jakarta: Balai Pustaka.
- Moulines, E. & E. Verhelst (1995) Time domain and Frequency-domain techniques for prosodic modification of speech. In W.B. Kleijn & K.K. Paliwal (eds) *Speech coding and synthesis*, 519-555. Amsterdam: Elsevier Sciences.
- Mozziconacci, S. (1998) *Speech variability and emotion: production and perception*. PhD dissertation, Technical University Eindhoven, The Netherlands.
- Muhadjir (1977) *Morfologi Dialek Jakarta: Afiksasi dan Reduplikasi [Jakarta dialect morphology: Affixation and reduplication]*. Jakarta: Djambatan. (ILDEP series).
- Nababan, P.W.J. (1981) *A grammar of Toba Batak*. Pacific Linguistic Series D-37.
- Nespor, M. & I. Vogel (1986) *Prosodic Phonology*. Dordrecht/Cinnaminson: Foris.
- Nooteboom, S. (1997) The Prosody of Speech: Melody and Rhythm. In W.J. Hardcastle & J. Lavers (eds) *The Handbook of Phonetic sciences*, 640-673. Oxford: Blackwell.
- Odé, C. (1989) *Russian Intonation: A Perceptual Description*. Amsterdam/Atlanta: Rodopi.
- Pelly, U. (1989) Hubungan antar Kelompok Etnis; Beberapa Kerangka Teoritis dalam Kasus Kota Medan [Relationships among ethnic groups; theoretical approaches in the case of Medan]. In: *Interaksi Antarsuku Bangsa dalam Masyarakat Majemuk*. Jakarta: Departemen Pendidikan dan Kebudayaan Direktorat Sejarah dan Nilai Tradisional (Proyek Inventarisasi dan Dokumentasi Sejarah Nasional), 1-16.
- Pijper, J.R. de (1983) *Modelling British English intonation*. Dordrecht/Cinnaminson: Foris.
- Piske, T., I. Mackay & J. Flege (2001) Factors affecting degree of foreign accent in L2: a review. *Journal of Phonetics* 29, 191-215.
- Podesva, R. J. & N. Adisasmito-Smith (1999) Acoustic investigation of the vowel systems of Buginese and Toba Batak. *Proceedings of the 14th International Congress of Phonetic Sciences*, 535-538.
- Potisuk, S., J. Gandour & M.P. Harper (1996) Effects of Stress on Vowel Length in Thai. *Proceedings of the 4th International Symposium on Language and Linguistics: Pan-Asiatic Linguistics, Vol. 1: Language Description*, 95-103.
- Ras, J.J. (1985) *Inleiding tot het modern Javaans [Introduction to modern Javanese]*. Den Haag: M. Nijhoff.
- Remijsen, B. (2001) *Word-prosodic systems of Raja Ampat languages*. Utrecht: LOT (LOT dissertation series 49).
- Remijsen, B. & V.J van Heuven (2005) Stress, tone, and discourse prominence in the Curaçao dialect of Papiamentu. *Phonology* 22, 205-235.
- Rietveld, A.C.M. & V.J. van Heuven (2001) *Algemene fonetiek [General phonetics]*. Bussum: Coutinho.
- Rietveld, A.C.M. & F.J. Koopmans-van Beinum (1987) Vowel reduction and stress. *Speech Communication* 6, 217-229.
- Roosman, L.M. (2004) Non-native accents in Dutch word-stress realisation. *Leiden Papers in Linguistics* 1, 63-81.

- Samsuri (1971) *Tjiri-tjiri prosodi kalimat bahasa Indonesia [Prosodic features of Indonesian sentences]*. Malang: Tim Publikasi Ilmiah, Fakultas Keguruan Sastra dan Seni, IKIP Malang.
- Sanders, M. J. (1996) *Intonation contour choice in English*. Utrecht: Onderzoeksinstituut voor Taal en Spraak (OTS dissertation series).
- Sibeth, A. (1991) *The Batak, Peoples of the Island of Sumatra*. London: Thames and Hudson.
- Sluijter, A.M.C. (1995) *Phonetics correlates of stress and accent*. The Hague: Holland Academic Graphics (HIL dissertation series 15).
- Sluijter, A.M.C., V.J. van Heuven & J.J.A. Pacilly (1997) Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America* 101, 312-322.
- Stevens, K.N. (1998) *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Stoel, R.B. (2005) *Focus in Manado Malay: Grammar, particles, and intonation*. Leiden: CNWS Publications.
- Streefkerk, B. (2002) *Prominence, Acoustic and Lexical/Syntactic correlates*. Utrecht: LOT (LOT dissertation series 58).
- Strik, H. (1994) *Physiological control and behaviour of the voice source in the production of prosody*. PhD dissertation, Catholic University Nijmegen, The Netherlands.
- Trubetsky, N.S. (1969) *Principles of Phonology*. Berkeley: University of California Press. [first published in 1939 as *Grundzüge der Phonologie*. Göttingen: Vondenhoeck and Ruprecht; translated by Christiane Baltaxe].
- Turk, A.E. & J.S. Sawush (1997) The domain of accentual lengthening in American English. *Journal of Phonetics* 25, 25-41.
- Tuuk, H. N. van der (1971) *A grammar of Toba Batak*. Den Haag: Martinus Nijhoff. [first published in 1864].
- Vaissière, J. (1983) Language-Independent Prosodic Features. In A. Cutler and D.R. Ladd (eds) *Prosody: Models and Measurements*, 53-66. Berlin/New York: Springer.
- Wallace, S. (1976) *Linguistic and social dimensions of phonological variation in Jakarta Malay*. PhD Dissertation, Cornell University.
- Werner, S. & E. Keller (1994) Prosodic aspects of speech. In E. Keller (ed.) *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges*, 23-40. Chichester: John Wiley.
- Wijngaarden, S.J. van (2001) Intelligibility of native and non-native Dutch speech. *Speech Communication* 35, 103-113.
- Willems, N. (1982) *English Intonation from a Dutch Point of View*. Dordrecht/Cinnaminson: Foris.
- Willems, N., R. Collier & J. 't Hart (1988) A synthesis scheme for British English intonation. *Journal of the Acoustic Society of America*, 84, 1250-1261.
- Woollams, G. (2005) Karo Batak. In A. Adelaar & N.P. Himmelmann (eds) *The Austronesian languages of Asia and Madagascar*, 534-561. London/New York: Routledge.
- Yip, M. (2002) *Tone*. Cambridge: Cambridge University Press.

- Zanten, E.A. van, R.W.N. Goedemans & J.J.A. Pacilly (2003) The status of word stress in Indonesian. In J.M. van de Weijer, V.J. van Heuven & H.G. van der Hulst (eds) *The Phonological Spectrum. Volume II: Suprasegmental structure*, 151-175. Amsterdam/Philadelphia: John Benjamins.
- Zanten, E.A. van & V.J. van Heuven (1997) Effects of Word Length and Substrate Language on the Temporal Organization of Words in Indonesian. In C. Odé & W.A.L. Stokhof (eds) *Proceedings of the Seventh International Conference on Austronesian Linguistics*, 63-80. Amsterdam/Atlanta GA: Rodopi.
- Zanten, E.A. van & V.J. van Heuven (1998) Word stress in Indonesian; its communicative relevance. *Journal of the Humanities and Social Sciences of Southeast Asia and Oceania*, 154, 129-149.
- Zanten, E.A. van & V.J. van Heuven (in press) Word stress in Indonesian: Fixed or free? *NUSA* 48.

Appendices

Appendix 1a. Aggregated table of pitch-movement shapes in Betawi Malay.

Target position in sentence	Vowel type	Target	Fall	Rise	Rise-Fall pre-final	Rise-Fall final	Total
+final	ə - V/ə	deket		5		4	9
		pete	2	5		2	9
		rejeki		2	3	4	9
	V - V	kaga	3	4		3	10
		kutu	1	5		5	11
		pipi	5	1		4	10
		pepet	4	2	2	1	9
		belaga		5	2	3	10
	Total			15	29	7	26
- final	ə - V/ə	deket	6	5			11
		pete	6	2	1		9
		rejeki	2	3	5	1	11
	V - V	kaga	4	3	5		12
		kutu	8	1	1		10
		pipi	7	2			9
		pepet	6	3	1		10
		belaga	2	2	6		10
	Total			41	21	19	1
Total			56	50	26	27	159

Appendix 1b. Absolute and relative frequencies of perceived prominence of pitch-movement shapes in Betawi Malay, broken down by word type and word position.

Word position	Vowel	Melodic shape	Prominences			Total
			pre-final	final	equal	
+final	V - V	fall	32 82%	3 8%	4 10%	39 100%
		rise	21 41%	24 47%	6 12%	51 100%
		RF pre-final	12 100%			12 100%
		RF final	7 15%	40 83%	1 2%	48 100%
		Total	72 48%	67 45%	11 7%	150 100%
	ə - V/ə	fall		4 67%	2 33%	6 100%
		rise		36 100%		36 100%
		RF pre-final	7 78%	1 11%	1 11%	9 100%
		RF final		29 97%	1 3%	30 100%
		Total	7 9%	70 86%	4 5%	81 100%
-final	V - V	fall	77 95%	2 2%	2 2%	81 100%
		rise	6 18%	15 45%	12 36%	33 100%
		RF pre-final	39 100%			39 100%
		Total	122 80%	17 11%	14 9%	153 100%
	ə - V/ə	fall	25 60%	8 19%	9 21%	42 100%
		rise		25 83%	5 17%	30 100%
		RF pre-final	14 78%	3 17%	1 6%	18 100%
		RF final		2 67%	1 33%	3 100%
		Total	39 42%	38 41%	16 17%	93 100%

Appendix 2. Instructions and sample response sheet for subjects of evaluation experiment (Chapter IV, Exp. 1)

Beste luisteraar,

U gaat meewerken aan een onderzoek naar verstaanbaarheid van het Nederlands gesproken door Nederlanders en buitenlanders. Het is uw taak om te beoordelen hoe goed hun uitspraak van het Nederlands is.

De schaal van de beoordeling loopt van 0 tot 10. Zet een kruisje onder het waarderingscijfer dat u aan iedere uiting geeft. Geeft u de score **lager dan 5** als u denkt dat de spreker **buitenlander** is, **hoger dan 5** als u denkt dat hij/zij **Nederlander** is. Kruist u onder **0** als het **onmiskenbaar buitenlands** klinkt, onder **10** als het **herkenbaar Nederlands** is.

Het onderzoek bestaat uit drie onderdelen. Het eerste deel is het moeilijkste. De spraak is onverstaanbaar en monotoon gemaakt. Hier moet u niet op de uitspraak van de klinkers en medeklinkers letten. Let alleen goed op het ritme van de spraak. In het tweede deel is de spraak onverstaanbaar maar niet monotoon gemaakt. Hier moet u goed op de melodie letten. En in het laatste deel hoort u de originele spraak.

U krijgt iedere uiting twee keer te horen. U krijgt telkens het laatste woord van een zin te horen. Het begin van de zin is onhoorbaar gemaakt. Bij voorbeeld: “[*Wat zei je? Ik zei*] *Madonna*”; “[*Wat zei je? Ik zei*] *baan*”. Alleen de laatste woorden (*Madonna* en *baan*) zijn dus hoorbaar. De gedeelten tussen de rechte haken zijn onhoorbaar.

Voor ieder onderdeel laten wij u eerst oefenzinnen horen.

Ik dank u alvast voor uw medewerking.
Lilie Roosman.

Wilt u uw naam hieronder invullen en de vraag beantwoorden?

Naam: _____

Hoe goed spreekt u Indonesisch? Goed / Een beetje / Helemaal niet

2. In het volgende proefje hoort u de onverstaanbare, niet-monotone spraak. Let niet op de uitspraak van de klinkers en medeklinkers, maar let op de melodie van de spraak.

(Examples and list of target words as in 1)

3. In het volgende proefje hoort u de originele spraak van het Nederlands.

(Examples and list of target words as in 1)

Appendix 3. Instructions and sample response sheet for subjects of identification experiment (Chapter IV, Exp. 2)

Beste luisteraar,

U gaat meewerken aan een onderzoek naar Nederlandse spraak gesproken door Nederlanders en buitenlanders. Het is uw taak om de taalachtergrond van de spreker te identificeren.

Zet een kruisje onder **Ja** als u denkt dat de spraak door een **Nederlander** gesproken is, en onder **Nee** als u denkt dat de spreker **geen Nederlander** is.

U krijgt het laatste woord van een zin te horen. Het begin van de zin is onhoorbaar gemaakt. Bij voorbeeld: “[*Wat zei je? Ik zei*] *Madonna*”; “[*Wat zei je? Ik zei*] *baan*”. Alleen de laatste woorden (*Madonna* en *baan*) zijn dus hoorbaar. De gedeelten tussen de haken zijn onhoorbaar. U krijgt de woorden telkens twee keer te horen.

Het onderzoek bestaat uit drie onderdelen. Het eerste deel is het moeilijkste. De spraak is onverstaanbaar en monotoon gemaakt. Hier moet u niet op de uitspraak van de klinkers en medeklinkers letten. Let alleen goed op het ritme van de spraak. In het tweede deel is de spraak onverstaanbaar, maar niet monotoon gemaakt. Hier moet u goed op de melodie letten. En in het laatste deel hoort u de originele spraak.

Voor ieder onderdeel laten wij u eerst oefenzinnen horen.

Ik dank u alvast voor uw medewerking.
Lilie Roosman.

Wilt u uw naam hieronder invullen en de vraag beantwoorden?

Naam: _____

Spreekt u Indonesisch? Ja, goed / Ja, een beetje / Nee, helemaal niet.

1. Monotoon en onverstaanbaar Nederlands. Let niet op de uitspraak van de klinkers en medeklinkers, maar let alleen op het ritme van de spraak.

Oefening: 

	Stimulus	Is de spreker Nederlander?		Stimulus	Is de spreker Nederlander?	
		Ja	Nee		Ja	Nee
(Ik zei) Madonna	1.			6.		
	2.			7.		
	3.			8.		
	4.			9.		
	5.			10.		

De echte proef:

	Stimulus	Is de spreker Nederlander?		Stimulus	Is de spreker Nederlander?	
		Ja	Nee		Ja	Nee
(Ik zei) bami.	1.			10.		
	2.			11.		
	3.			12.		
	4.			13.		
	5.			14.		
	6.			15.		
	7.			16.		
	8.			17.		
	9.			18.		

	Stimulus	Is de spreker Nederlander?		Stimulus	Is de spreker Nederlander?	
		Ja	Nee		Ja	Nee
(Ik zei) banaan.	1.			10.		
	2.			11.		
	3.			12.		
	4.			13.		
	5.			14.		
	6.			15.		
	7.			16.		
	8.			17.		
	9.			18.		

Etc.

2. In het volgende proefje hoort u de onverstaanbare, niet-monotone spraak. Let niet op de uitspraak van de klinkers en medeklinkers, maar let op de melodie van de spraak.

(Examples and list of target words as in 1.)

3. In het volgende proefje hoort u de originele spraak van het Nederlands.

(Examples and list of target words as in 1.)

Appendix 4. Percentage of identification of language background in three stimulus versions broken down by the speaker group. Bold numbers represent correctly identified stimuli.

a. Toba Batak listeners

Stimulus version	Language of speaker	Language background identified as		
		Dutch	TB	BM
Monotonized	Dutch	34	30	36
	Toba Batak	33	35	32
	Betawi Malay	29	30	41
Delexicalized	Dutch	30	36	34
	Toba Batak	30	39	31
	Betawi Malay	35	28	37
Original	Dutch	53	31	16
	Toba Batak	21	46	33
	Betawi Malay	26	28	46

b. Betawi Malay listeners

Stimulus version	Language of speaker	Language background identified as		
		Dutch	TB	BM
Monotonized	Dutch	28	28	44
	Toba Batak	37	29	34
	Betawi Malay	32	29	39
Delexicalized	Dutch	46	21	34
	Toba Batak	38	22	39
	Betawi Malay	36	31	33
Original	Dutch	55	29	16
	Toba Batak	26	31	43
	Betawi Malay	31	26	43

Summary

The subject of this study is the word prosody of two regional languages of Indonesia, viz. Toba Batak (TB) and Betawi Malay (BM). These two languages differ crucially in that TB has word stress and BM has not. After a brief discussion of prosody and its functions in Chapter II, I focus on the production and perception of prominence (words in [+focus] position versus words in [-focus] position) and boundary marking (sentence-medial versus sentence-final words) on the temporal and melodic structures of TB and BM. As regards prominence, I expected stronger effects in the stress language TB (especially in the stressed syllable) than in non-stress BM. As regards boundary marking, I expected to find similar effects for both languages.

My production studies (Chapter III) show that the durational effects of prominence and boundary are significant at the word level in both languages. Unexpectedly, both effects were more than twice as strong in BM as in TB. Thus, my hypothesis that durational effects of prominence would be stronger in the stress language TB than in non-stress BM was not confirmed, nor was my expectation that boundary marking would have similar effects in both languages. In both languages, focus affected lengthening less than boundary. However, like in Western languages, boundary effects were strongest on the ultimate syllables.

At the word level the lengthening effect of [+focus] in TB are comparable to those in Dutch: sentence-finally 10%; sentence-medially: 14%. However, contrary to expectation, the pre-final (stressed) syllable, especially its consonant, was not significantly affected by focus in TB. Overall, the lengthening of TB consonants was small or non-existent. This may be explained by the fact that consonant length is phonemic in TB, which may prevent consonant duration from being used as a correlate of stress. Some pre-boundary lengthening did occur in phonemically long ('geminate') consonants, but not in short ('singleton') consonants. Possibly the

duration of the singleton is limited so as not to be confused with the geminate consonant.

Both focus and boundary lengthening effects were much stronger in BM than in TB words. In BM words with full vowels focus affected the pre-final syllable strongly, especially the vowel. This indicates that duration is a prosodic cue for accent in BM. Pre-final syllables containing schwa were not lengthened by focus; instead, the final syllables were strongly affected. As expected, pre-boundary lengthening effects occurred primarily in the final syllable, in particular the vowel.

The pitch contours in both languages were strongly affected by focus, but again more strongly so in the non-stress language BM. However, in this language both the shape and the position of pitch movements on prominent words were considerably affected by the position of these words in the sentence. The variability in shape and position of the pitch configuration indicates that the prominence marking is not tied to one specific syllable, which is fully compatible with a system that does not have word stress, such as BM.

In TB, not the shape (always a rise-fall) but only the size of the accent-lending pitch configuration changed as a function of focus; on [+focus] words large movements occurred, whereas in [-focus] words the pitch movements were smaller, but never deleted. Importantly, the pitch movement was always on the stressed syllable. This fits in well with the existence of (contrastive) word stress in TB.

It would appear, then, that my hypothesis that a language with contrastive word stress (TB) marks its stresses/accents more clearly than a language without word stress (BM) was only partly supported by the results. In terms of temporal structure the results rather support the opposite, as the effect of focus on duration was larger and more systematic in BM than in TB. The rather weak lengthening in TB may, however, at least partly be explained by the phonemic length contrast in the consonants of this language, which seems to demote duration as a stress cue. In view of this, it is remarkable that the prominence-related pitch movements of TB, which are tightly connected to the stressed syllable, occur in [+focus] as well as [-focus] words. This is unlike Western stress languages, where such movements are used

exclusively in [+focus] words. In TB pitch apparently does not only mark focus but also stress position, thereby compensating for the weak stress cue of duration.

In similar vein, boundary, which is only weakly marked by lengthening in TB, is also marked by the size and the shape of the fall: in sentence-final position falls are larger and steeper than in sentence-medial position. The larger but variable pitch movements in BM, on the other hand, serve to simultaneously cue accents and boundaries, but not stress position.

In Chapter IV I investigated to what extent native speakers of TB and BM are influenced by the prosody of their native language when they speak Dutch, and whether they are sensitive to the prosodic differences in Dutch. Three perception experiments were run to determine the audible consequences of native language interference in the production of Dutch stress/accent. As stimulus material Dutch words spoken by BM, TB and Dutch speakers were used. The (digitized) stimuli were presented to the listeners in three different versions: (i) as originally spoken, (ii) delexicalized (no information on consonants and vowels, but all prosodic information intact), and (iii) delexicalized as well as monotonized (only durational and loudness information available to the listeners).

The results indicated that the realizations of the BM speakers were less 'Dutch' to Dutch listeners than those of the TB speakers. Prosodically, TB speakers of Dutch were significantly more acceptable to Dutch listeners than BM speakers of Dutch. This indicates that the non-stress language (BM) is prosodically different from both stress languages (TB and Dutch).

However, the non-nativeness of the Indonesian pronunciations of Dutch was better detected by Dutch listeners when not only prosodic information was audible, but also the pronunciation of the vowels and the consonants. Apparently, there is some, rather weak, information on non-nativeness in the prosody of the Indonesian speakers of Dutch, but there are clearer cues in the segmental structure. When the two groups of Indonesians listened to the same stimuli, they could not differentiate the language background of the speakers, even when the full segmental information was made audible.

In Chapter V duration and pitch of the stimuli used in Chapter IV, i.e. the Dutch words spoken by BM, TB and Dutch speakers, were acoustically analysed. I found that stressed vowel duration and stressed-to-unstressed vowel-duration ratio contributed most to differentiate the temporal structures of the groups of speakers. The speakers of non-stress BM realised the stressed vowel significantly longer than the speakers of the stress languages Dutch and TB did, and the stressed-to-unstressed vowel-duration ratio of the BM speakers was the largest, followed by that of the TB speakers, whilst that of Dutch speakers was the smallest. In other words, the shorter the stressed vowel and the more even the stressed-to-unstressed vowel-duration ratio, the more likely the speaker would be Dutch.

Interestingly, the correlation analysis yielded a positive correlation between the perception scores of the monotonized stimuli and stressed vowel duration. However, the multiple correlation between the monotonized scores and the ensemble of temporal parameters was low; listeners had great difficulties differentiating the language backgrounds of the speakers by listening to the monotonized stimuli.

In the delexicalized version, the larger stressed-to-unstressed vowel-duration ratios resulted in higher nativeness scores. The correlation between perceived nativeness and temporal structure was better when melodic information was included in the stimuli than when temporal structure was the only cue made available to the listeners; apparently, listeners could differentiate the temporal structure of the speakers better when the melodic information was also available in the speech samples.

The melodic structures of the BM speakers differed in general from those of the Dutch speakers and to some extent from those of the TB speakers, especially as regards the pitch rises. The melodies of the TB speakers were judged to be better approximations of the Dutch patterns. This fits in well with the low nativeness scores for the BM speakers and the higher nativeness scores for the TB speakers when melodic as well as temporal information was presented.

The nativeness scores for the delexicalised stimuli were significantly correlated with one melodic parameter only, i.e. the slope of the rise, which was also the acoustic parameter that most clearly differentiated the three language groups.

However, this correlation was rather low. This explains why the listeners had difficulties in differentiating the non-native from the native speakers when listening to the delexicalised versions.

For the original stimuli, higher (significant) correlations were found between the temporal and melodic parameters and the nativeness scores. Temporally and melodically the non-native speakers sounded clearly foreign to the Dutch listeners when they heard the pronunciation of the vowels and consonants as well. Interestingly, the strength of the correlations between nativeness judgments and prosodic parameters increased as the segmental quality was better. It seems as if prosodic properties contributed more to perceived nativeness as the stimulus quality was better. One possible interpretation of this result would be that listeners are more proficient in detecting traces of foreignness as the stimulus is more natural, i.e. closer to everyday-life experience.

The findings of the melodic analyses of Dutch words spoken by BM and TB speakers as found in the stimulus analysis in Chapter V basically reveal the same temporal and melodic structures as the ones reported in Chapter III. These convergent results indicate that the prosodic structure of the native language clearly affects second-language speech production.

Chapter VI, finally, contains a general discussion and some suggestions for further research. It is emphasized that more attention is needed in the area of prosody, especially for the purpose of teaching a stress language to learners hailing from a non-stress language.

Ringkasan

(summary in Indonesian)

Pokok pembahasan dalam disertasi ini adalah prosodi kata dari dua bahasa daerah di Indonesia, yakni bahasa Batak Toba (BT) dan bahasa Melayu Betawi (MB). Kedua bahasa ini sangat berbeda karena BT mempunyai tekanan kata dan MB tidak. Setelah pembahasan singkat mengenai prosodi dan fungsinya pada Bab II, saya akan memusatkan pembahasan pada produksi dan persepsi penanda prominensia (kata-kata dalam posisi [+focus] lawan kata-kata dalam posisi [-focus]) dan sempadan (kata-kata dalam posisi tengah lawan kata-kata dalam posisi akhir kalimat) pada struktur temporal dan melodi dari BT dan MB. Mengenai penanda prominensia, saya berharap akan adanya efek-efek yang lebih kuat dalam bahasa bertekanan BT (terutama pada suku kata yang bertekanan) daripada dalam bahasa MB yang tidak mempunyai tekanan. Mengenai sempadan, saya berharap akan dapat menemukan efek-efek yang sama dalam kedua bahasa tersebut.

Penelitian saya tentang produksi ujaran (Bab III) memperlihatkan bahwa efek durasi prominensia dan sempadan adalah penting pada tataran kata dalam kedua bahasa tersebut. Tidak seperti yang diharapkan, ternyata kedua efek itu dua kali lebih kuat dalam MB daripada dalam BT. Jadi, hipotesis saya yang mengatakan bahwa efek durasi prominensia akan lebih kuat dalam bahasa bertekanan BT daripada dalam bahasa tak bertekanan MB dengan demikian tidaklah terbukti, begitu juga dugaan saya bahwa sempadan pada kedua bahasa itu akan mempunyai efek yang sama ternyata tidaklah benar. Dalam kedua bahasa itu pemanjangan sebagai efek fokus lebih sedikit daripada efek sempadan. Meskipun begitu, seperti halnya di dalam bahasa-bahasa Barat, efek sempadan paling kuat pada suku kata terakhir.

Pada tataran kata efek pemanjangan [+fokus] dalam BT bisa dibandingkan dengan efek yang sama dalam bahasa Belanda: pada posisi akhir kalimat 10%; pada posisi tengah kalimat: 14%. Meskipun demikian, berlawanan dengan yang diharapkan, suku kata (bertekanan) praakhir, terutama konsonan, tidak terlalu dipengaruhi oleh fokus dalam BT. Secara keseluruhan, pemanjangan konsonan dalam BT sangat sedikit atau tidak ada sama sekali. Hal ini mungkin dapat dijelaskan oleh fakta bahwa panjang konsonan dalam BT adalah fonemis, sehingga mungkin mencegah durasi konsonan digunakan sebagai korelasi tekanan. Beberapa pemanjangan prasempadan muncul dalam konsonan panjang fonemis (*geminate*), tetapi tidak muncul dalam konsonan pendek (*singleton*). Mungkin saja durasi konsonan pendek terbatas agar tidak terkacaukan dengan konsonan panjang fonemis.

Efek fokus dan efek pemanjangan sempadan jauh lebih kuat pada kata-kata dalam MB daripada dalam BT. Pada kata-kata MB yang mempunyai vokal penuh, fokus mempengaruhi suku kata praakhir, terutama vokalnya. Hal ini menunjukkan bahwa durasi adalah isyarat prosodis untuk aksentuasi dalam MB. Suku kata praakhir yang mengandung pepet tidak dipanjangkan oleh fokus; sebagai gantinya, suku kata akhir sangat dipengaruhi fokus. Seperti yang diharapkan efek pemanjangan prasempadan terutama muncul dalam suku kata akhir, khususnya pada vokal.

Kontur tinggi nada (*pitch contour*) dalam kedua bahasa itu sangat dipengaruhi oleh fokus, tetapi sekali lagi dalam bahasa tak bertekanan MB hal ini lebih kuat. Meskipun demikian, dalam bahasa ini baik bentuk maupun posisi gerakan tinggi nada pada kata-kata prominensia sangat dipengaruhi oleh posisi kata-kata itu di dalam kalimat. Keragaman dalam bentuk dan posisi konfigurasi menunjukkan bahwa penanda prominensia tidak terikat pada suatu suku kata tertentu, yang kompatibel sepenuhnya dengan suatu sistem yang tidak mempunyai tekanan kata, seperti dalam bahasa MB.

Dalam bahasa BT, bukan bentuk konfigurasi tinggi nada penanda tekanan (selalu dalam bentuk naik-turun) tetapi hanya ukurannya yang berubah sebagai efek fokus; pada kata-kata [+fokus] terjadi gerakan yang besar, sebaliknya pada kata-kata [-fokus] gerakan tinggi nadanya lebih kecil, tetapi tidak pernah hilang sama sekali. Yang paling penting, gerakan dari tinggi nada selalu terjadi pada suku kata bertekanan. Hal ini sesuai dengan adanya tekanan kata (kontrastif) dalam bahasa BT.

Dengan demikian akan terlihat bahwa hipotesis saya yang mengatakan bahwa bahasa yang bertekanan kata kontrastif (BT) menandai tekanannya/aksennya dengan lebih jelas daripada bahasa tak bertekanan kata (MB), hanya sebagian terdukung oleh hasil penelitian ini. Dalam hal struktur temporal, hasil penelitian ini justru lebih mendukung kebalikannya, karena efek fokus pada durasi dalam bahasa MB ternyata lebih luas dan lebih sistematis daripada dalam bahasa BT. Namun, pemanjangan yang agak lemah dalam BT, paling tidak sebagiannya dapat dijelaskan oleh kontras panjang fonemis pada konsonan bahasa ini, yang kelihatannya mengurangi efek durasi sebagai isyarat tekanan. Mengingat hal ini, adalah luar biasa bahwa gerakan tinggi nada yang berhubungan dengan prominensia dalam bahasa BT – yang erat dikaitkan dengan suku kata bertekanan – muncul baik dalam kata [+fokus] maupun [-fokus]. Hal ini tidak seperti halnya dalam bahasa-bahasa Barat yang bertekanan. Dalam bahasa-bahasa ini gerakan seperti itu hanya dipakai pada kata-kata [+fokus]. Dalam bahasa BT tinggi nada rupanya bukan hanya menandai fokus tetapi juga menandai posisi tekanan, dengan demikian sebagai kompensasi untuk lemahnya isyarat tekanan dari durasi.

Dalam nada yang sama, sempadan, yang hanya ditandai secara lemah oleh pemanjangan dalam BT, juga ditandai oleh ukuran dan bentuk penurunan tinggi nada: dalam posisi akhir kalimat penurunannya lebih besar dan lebih curam daripada dalam posisi tengah kalimat. Sebaliknya, gerakan tinggi nada yang lebih besar tetapi beragam dalam bahasa MB adalah isyarat untuk aksent dan sempadan, tetapi bukan untuk posisi tekanan.

Dalam Bab IV saya meneliti sejauh mana penutur asli BT dan MB dipengaruhi oleh prosodi bahasa ibu mereka pada saat mereka mengucapkan bahasa Belanda, dan apakah mereka peka terhadap perbedaan prosodis dalam bahasa Belanda. Tiga eksperimen persepsi dilakukan untuk menentukan konsekuensi interferensi bahasa ibu yang dapat didengar dalam mengucapkan tekanan/aksent bahasa Belanda. Sebagai stimulus digunakan kata-kata bahasa Belanda yang diucapkan oleh penutur MB, BT, dan bahasa Belanda. Stimuli (secara digital) disajikan kepada para pendengar dalam tiga versi yang berbeda: (i) sebagaimana ucapan yang sebenarnya, (ii) yang dideleksikalkan (*delexicalized*) (tanpa informasi konsonan dan vokal, tetapi semua informasi prosodis dipertahankan) (iii) baik dideleksikalkan maupun

dimonotonkan (hanya informasi durasional dan kelantangan suara yang bisa didengar oleh para pendengar).

Hasil penelitian menunjukkan bahwa bagi para pendengar Belanda realisasi penutur MB kurang terdengar “seperti penutur Belanda” dibandingkan dengan realisasi penutur BT. Secara prosodis, penutur BT yang berbicara bahasa Belanda jauh lebih bisa diterima oleh pendengar Belanda daripada penutur MB berbicara bahasa Belanda. Hal ini menunjukkan bahwa bahasa tak bertekanan (MB) secara prosodis berbeda dari kedua bahasa bertekanan itu (BT dan bahasa Belanda).

Namun demikian, ketidakeaslian (*non-nativeness*) cara pengucapan orang Indonesia dalam berbicara bahasa Belanda lebih bisa dideteksi dengan baik oleh pendengar Belanda apabila bukan hanya informasi prosodisnya yang diperdengarkan, tetapi juga pengucapan vokal dan konsonannya. Rupanya, ada beberapa informasi - yang agak lemah - ketidakeaslian tuturan dalam prosodi penutur Indonesia yang mengucapkan bahasa Belanda, tetapi ada isyarat-isyarat yang lebih jelas dalam struktur segmentalnya. Yang menarik, apabila kedua kelompok orang Indonesia mendengarkan stimuli yang sama, maka mereka tidak dapat membedakan latar belakang penutur, bahkan juga apabila informasi segmental penuh diperdengarkan.

Dalam Bab V durasi dan tinggi nada stimuli yang digunakan dalam Bab IV, yakni kata-kata Belanda yang diucapkan oleh para penutur MB, BT, dan bahasa Belanda dianalisis secara akustis. Saya menemukan bahwa durasi vokal bertekanan dan rasio durasi vokal bertekanan - tak bertekanan yang paling banyak berperan dalam pembedaan struktur temporal kelompok-kelompok penutur. Penutur bahasa tak bertekanan MB mengucapkan vokal bertekanan jauh lebih lama dibandingkan dengan penutur bahasa bertekanan BT dan bahasa Belanda, dan rasio durasi vokal bertekanan - tak bertekanan dari penutur MB adalah yang paling besar, diikuti oleh penutur BT, sementara penutur Belanda adalah yang paling kecil. Dengan kata lain, semakin pendek vokal bertekanan dan semakin seimbang rasio durasi vokal bertekanan - tak bertekanan, maka semakin besarlah kemungkinan bahwa penutur itu adalah penutur Belanda.

Yang menarik, analisis korelasi menghasilkan korelasi positif antara skor persepsi stimuli yang dimonotonkan dan durasi vokal bertekanan. Meskipun begitu, korelasi majemuk antara skor yang dimonotonkan dan ensambel parameter temporal adalah rendah; dengan mendengarkan stimuli yang dimonotonkan, para pendengar mendapat kesulitan besar untuk membedakan latar belakang bahasa dari penutur.

Dalam versi yang dideleksikalkan, makin besar rasio durasi vokal bertekanan - tak bertekanan menghasilkan skor keaslian cara pengucapan penutur bahasa ibu (*nativeness*) yang lebih tinggi. Korelasi antara struktur keaslian cara pengucapan yang diterima dan struktur temporal menjadi lebih baik apabila informasi melodi diikutsertakan di dalam stimuli dibandingkan dengan apabila hanya struktur temporal satu-satunya yang bisa didengarkan oleh pendengar; rupanya, para pendengar dapat membedakan struktur temporal penutur dengan lebih baik, apabila informasi melodinya juga disertakan dalam contoh ujaran.

Struktur melodi penutur MB pada umumnya berbeda dengan para penutur Belanda dan dalam tingkatan tertentu juga dari penutur BT, terutama dalam hal naiknya tinggi nada. Melodi para penutur BT dianggap lebih mendekati pola-pola bahasa Belanda. Hal ini sesuai dengan skor keaslian (cara pengucapan) yang rendah

untuk para penutur MB dan skor keaslian (cara pengucapan) yang lebih tinggi untuk para penutur BT, apabila baik informasi melodi maupun informasi temporalnya diikutsertakan.

Skor keaslian cara pengucapan untuk stimuli yang dideleksikalkan berkorelasi hanya dengan satu parameter melodi, yakni kecuraman peningkatan tinggi nada, yang juga merupakan parameter akustis yang paling terlihat membedakan ketiga kelompok bahasa itu. Akan tetapi, korelasi ini agak rendah. Hal ini menjelaskan mengapa para pendengar mengalami kesulitan dalam membedakan penutur tidak asli dari penutur asli pada saat mereka mendengarkan versi yang dideleksikalkan.

Untuk stimuli aslinya, ditemukan korelasi yang (jelas) lebih tinggi antara parameter temporal dan melodi dengan skor keaslian (cara pengucapan). Dalam hal temporal dan melodi, bukan penutur asli kedengaran asing di telinga para pendengar Belanda apabila mereka mendengar pengucapan vokal dan konsonan. Yang menarik, kekuatan korelasi antara penilaian keaslian (cara pengucapan) dan parameter prosodis meningkat apabila kualitas segmentalnya lebih baik. Tampaknya sifat-sifat prosodi membuat aspek keaslian (cara pengucapan) lebih diterima apabila kualitas stimulusnya lebih baik. Salah satu interpretasi yang mungkin dari penelitian ini adalah bahwa para pendengar lebih mahir dalam mendeteksi ciri-ciri asing (bukan penutur asli) apabila stimulusnya lebih alamiah, yakni lebih mendekati pengalaman kehidupan sehari-hari.

Penemuan analisis melodi kata-kata Belanda yang diucapkan oleh penutur MB dan BT seperti yang diketengahkan dalam analisis stimulus dalam Bab V sebenarnya mengungkapkan struktur temporal dan melodi yang sama seperti yang dilaporkan dalam Bab III. Hasil yang konvergen ini menunjukkan bahwa struktur prosodi bahasa ibu jelas mempengaruhi produksi pengucapan bahasa kedua.

Bab VI, akhirnya, memuat diskusi umum dan beberapa gagasan untuk penelitian di kemudian hari. Saya tekankan bahwa perlu sekali adanya lebih banyak perhatian terhadap bidang prosodi, terutama untuk tujuan pengajaran bahasa bertekanan kepada pembelajar yang berasal dari bahasa tak bertekanan.

Curriculum vitae

Lilie Mundalifah Roosman was born on September 20, 1964, in Jakarta. Her father, Roosman Roekmantoro was born in Jakarta in a Javanese family and grew up in a *kampoeng* in Central Jakarta. Her mother, Wirdah Rasyad was born in Jakarta from native Betawi parents. Lilie Roosman started her Dutch study in 1983. After she graduated in Dutch linguistics from the Dutch Department, University of Indonesia in 1989, she got a scholarship to study at the Department of Dutch Studies at Leiden University, The Netherlands. She obtained her Master's degree in 1992 with a thesis: *Glijklanken in het Nederlands [diphthongs in Dutch]*. From 1992 until 1994 she worked as an assistant lecturer at the Dutch Department, Faculty of Letters, University of Indonesia. In 1995 she was promoted to university lecturer in the same department. From February 1999 until January 2003 she was relieved from (most of) her teaching duties in order to pursue her PhD research in the KNAW project 95-CS-05, entitled 'Phonetics and Phonology of (Word) Prosodic Systems in Indonesian Languages' in a joint research program of Leiden University and the University of Indonesia, with the research topic: Word Prosody in Betawi Malay and Toba Batak.