Non-native accents in Dutch word stress realisation*

Lilie M. Roosman

Abstract

This study investigates whether speakers of a stress language can realize the stress of another stress language more faithfully than speakers of a non-stress language.

Three perception experiments were run to investigate how well native speakers of two related Indonesian languages, i.e. Toba Batak (a stress language) and Jakarta Malay (a non-stress language), realise Dutch word stress.

The results indicate that the stress realisations of Jakarta Malay speakers are less 'Dutch' to Dutch listeners than those of Toba Batak speakers. The Toba Batak speakers, as native speakers of a stress language, are at an advantage over Jakarta Malay speakers, who do not have word stress in their language.

1. Introduction

In languages with word stress, one syllable is perceived as stronger than the other syllables in the word. On the higher linguistic levels, in phrases or sentences, accent is used to make particular words more prominent than other words. In these languages, phrasal accents 'dock' at the sites that are stressed at the word level (Bolinger 1958). Languages without word stress may also use accent to highlight words in sentences. However, in these languages the position of the accent is not aligned with the one stressed syllable in the word.

This study focuses on the realisation of word prosody of Toba Batak and Jakarta Malay compared to the realisation of Dutch word prosody; in particular the effect of stress/accent on the temporal and melodic structure. Toba Batak is, like Dutch, a stress language (Van der Tuuk 1971, Nababan 1981). Nababan (1981) provides minimal pairs, for example *tíbo* $_{[N]}$ 'height', but *tibó* $_{[ADJ]}$ 'high'. Jakarta Malay is a language that does not have word stress (Muhadjir 1977),

* In Boban Arsenijevic, Noureddine Elouazizi, Martin Salzmann & Mark de Vos (eds.), *Leiden Papers in Linguistics* 1.1 (2004), 63-81. http://www.ulcl.leidenuniv.nl
ISSN 1574-4728

but does have phrasal accent (Wallace 1976). Therefore, it is 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 Toba Batak and Jakarta Malay are influenced prosodically by their native language when they are speaking Dutch, and whether they are sensitive to the prosodic differences in Dutch

1. The first experiment involves native-Dutch listeners evaluating the realisation of Dutch word stress spoken by Toba Batak, Jakarta Malay, and by Dutch speakers.
2. 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.
3. The last experiment investigates the extent to which Toba Batak and Jakarta Malay listeners are able to recognise Dutch-speaking Indonesians, even on the basis of deviant stress realisation only.

2. Background

Prosody is the set of properties in speech that cannot be derived from the segmental sequence of phonemes underlying the utterance (Nooteboom 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 (Nooteboom 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 & 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 & 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 speech than native listeners do (Van Wijngaarden 2001:103).

3. Method

In these 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 Toba Batak (TB) and Jakarta Malay (JM) speakers are compared with data of Dutch speakers.

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.

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 1.

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

(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.'

Table 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			*baan*[1] 'job'
2 syllables		*bami* 'Chinese noodle'	*banaan* 'banana'
3 syllables	*bamibal* 'noodle ball'	*Lambada* 'a dance'	*Balinees* 'Balinese'

3.2. Speakers

Two JM speakers (one male, one female) were involved in this experiment. The JM 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 JM speaker was 26 years old. She studied Dutch for five years in her home country and had been studying Dutch in the Netherlands for six 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 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 (see section 2).

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

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 the stress in the correct position. To help them doing this task, I used the question-answer method as explained in section 3.1. The interviewer read the questions

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

(*Wat zei je?* 'What did you say?') from the sheets. The JM 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 was wrong, the interviewer told the subject where the stress was supposed to be. In addition, 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 had appeared that the Dutch speakers spoke in a rather assertive way, whereas 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 experiment. 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 test were correctly accented and stressed as judged by both experts. The total number of utterances to be manipulated was 108 (6 words × 6 speakers × 3 repetitions).

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 & Weenink 1996).

For the first version of the stimuli, no further manipulation took place; the original sounds of the target words ('origin', Fig. 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’). 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 1b).

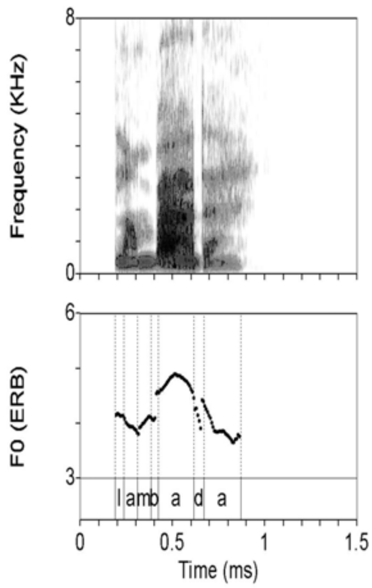

Figure 1a. Spectrogram and original fundamental frequency contour (ERB) of stimulus item *lambada*. Segmentation and phoneme labels are indicated.

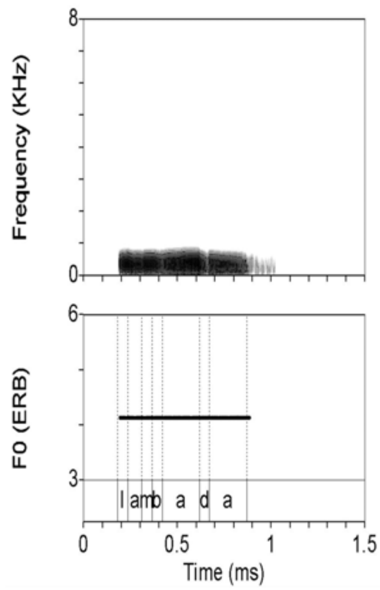

Figure 1b. As figure 1a but for low-pass filtered version (cut-off frequency at 0.35 KHz) and with pitch contour monotonized at 4.2 ERB (150 Hz).

3.5. General procedure

The 18 tokens for every word (6 speakers × 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, lambada, balinees*), and the last word was the monosyllabic word (*baan*). A silent response time of 2.5-seconds separated the words from each other.

The ‘monot’ version was presented first to the listeners, 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 × 6 words × 18 tokens × 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 along with the different sets of stimuli they heard. 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 could get used to the manipulated stimuli before the real experiment started. The whole experiment took about 40 minutes.

4. Experiment 1: Evaluation by Dutch listeners

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 and a column for the marked values. 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 word they heard was spoken by a native or 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, or to leave an item blank. They were not paid for their services.

4.2. Results and discussion

Figure 2 summarizes the mean evaluation scores for the different stimulus versions broken down by speaker group. The figure shows clearly that the Dutch speakers are the most qualified group in this experiment. The JM 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 JM speakers get the lowest scores (3.67).

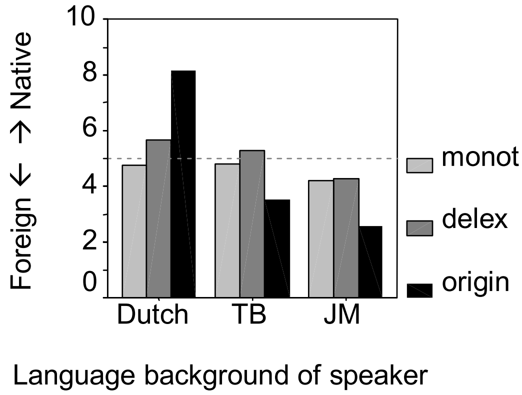

Figure 2. Mean evaluation score (0…10) by 30 Dutch listeners broken down by speaker group and by stimulus version.

An analysis of variance was run on the data with language background of speakers (Dutch, TB, JM) 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 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 drops the scores for the TB and JM speakers. Similarly, the Dutch score in the 'delex' version levels out the JM 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 also 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 JM 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 JM speakers in the 'monot' version and the 'delex' version do not differ from each other ($p = .754$). It seems that the JM speakers did not produce an adequate speech melody so that the Dutch listeners gave the 'delex' stimuli the same scores as the JM 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 result shows 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 JM speakers are significantly poorer than those of the other groups.

Figure 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 Jakarta Malay 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 JM 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 above. In figure 2 we can see that the lowest mean score (2.6) was observed for the JM 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 JM 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.

5. Experiment 2: Identification by Dutch listeners

In the following experiment, I investigate 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 have now to take a categorical yes/no decision on the 'Dutchness' of the speaker. By presenting the subjects with a binary choice I hope to get sharper results than in the previous experiment.

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.

5.2. Results and discussion

Figure 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].

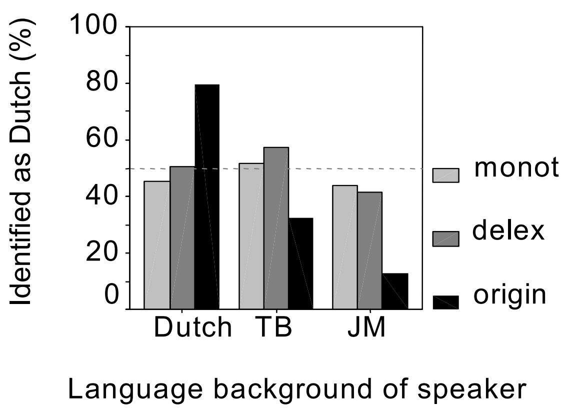

Figure 3. Percentage of the responses by ten Dutch listeners (correctly or incorrectly) identifying the speaker as Dutch, broken down by language background of the speakers and by stimulus version.

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 JM 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 3 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 JM 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 JM speakers the percentage decreases. The 'origin' version is significantly different for all speaker groups. The percentage identification as Dutch for JM 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.

6. Experiment 3: Identification by non-Dutch listeners

This experiment involved only JM 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.

6.1. Subjects and procedure

A group of ten TB listeners and a group of ten JM listeners, who know Dutch, took part in this experiment. All listeners were students or former students of the Dutch Department at the Universitas Indonesia. They listened to the same stimuli in the same order as the Dutch listeners in the previous experiment. It was explained to the listeners that they were going to listen to Dutch words, spoken by native Dutch speakers, TB speakers and JM speakers. They had to identify the mother tongue of the speakers they heard on the tape as either Dutch, Jakarta 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 experiments. All experiments were done individually in Jakarta, Indonesia.

6.2. Results and discussion

Figure 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).

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.

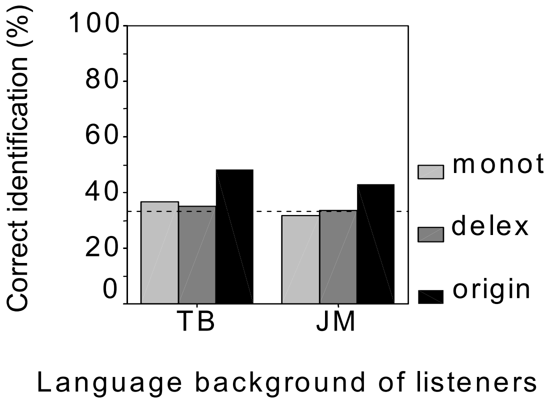

Figure 4. Percentage correctly identified language background of speaker broken down by stimulus version and by language of the listener (Indonesian nationals only).

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.)

Figure 4 shows that, overall, the TB listeners could identify the language of the speakers in all versions somewhat better than the JM 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. The 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 stress system 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, our 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 JM 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 words as a random factor.

6.3. Toba Batak listeners

It was expected that TB listeners, as native speakers of a stress language, could identify the subjects better than JM listeners. Overall, this was indeed the case as is shown in Figure 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 5 summarises the percentage correctly-identified language background of the speakers as identified by TB listeners broken down by stimulus type.

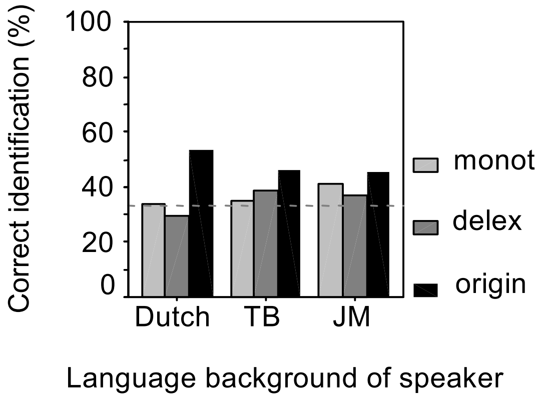

Figure 5. Percent correctly-identified language background for three speaker groups by TB listeners and broken down by stimulus version.

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]. The 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 speaker 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 JM speakers are better identified than the other speakers. This indicates that the listeners are rather familiar with the monotonous speech of JM speakers. 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 JM speakers, which are similar to each other. The melody of the Dutch speakers was often mistaken for the TB melody (see Appendix). 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 JM 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.

6.4. Jakarta Malay listeners

Figure 6 summarizes the results of the correct identification by Jakarta Malay listeners, broken down by the group of speakers and stimulus version.

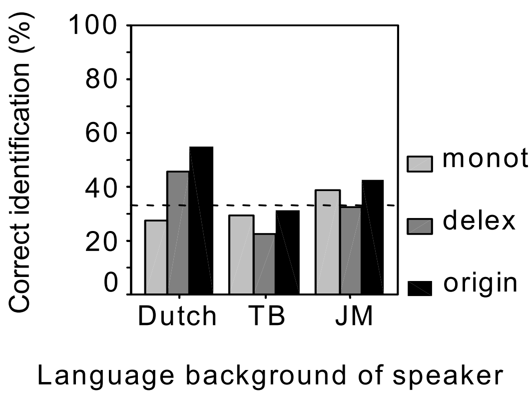

Figure 6. Percent correctly-identified language background for three speaker groups by JM listeners and broken down further by stimulus version.

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 JM speakers second, and the TB speakers worst. Probably, the livelier intonation/accentuation of the Dutch speakers provides a good cue for the identification. The flat prosody of JM speech helps the JM 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 JM identifications, assuming that flat prosody is the hallmark of JM. So we predict that the correct identification of JM 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 JM speakers due to the absence of pitch obtrusions. This prediction is also borne out by the results (for the complete confusion matrix see appendix).

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 JM as opposed to Dutch and TB.

7. Conclusion

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

First, we aimed to evaluate the way in which native speakers of Toba Batak and Jakarta Malay realize Dutch stress by using Dutch listeners. The results of experiment 1 show that the Jakarta Malay realisation of Dutch is not only less Dutch to Dutch listeners than native Dutch, but it also sounds somewhat less Dutch than Toba Batak 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 Toba Batak Dutch is less foreign due to the fact that Toba Batak has more-clearly-marked stress realisations, which is advantageous when Toba Batak 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 (Toba Batak and Jakarta Malay) 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 Toba Batak speakers are more often identified as Dutch speakers than Jakarta Malay 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 Jakarta Malay speakers.

The third and final question asked how well Toba Batak listeners and Jakarta Malay 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 our 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 Jakarta Malay 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, we would argue for a differential curriculum depending on the specific language background of the Indonesian learner of Dutch. If the learner hails from a Toba Batak 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).

Acknowledgments

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). The author gratefully acknowledges Ellen van Zanten and Vincent van Heuven for supervision and comments on earlier versions of this paper.

Lilie M. Roosman
Phonetics Laboratory, Universiteit Leiden Centre for Linguistics,
The Netherlands
e-mail: l.m.roosman@let.leidenuniv.nl
Program Studi Belanda, Fakultas Ilmu Pengetahuan Budaya,
Universitas Indonesia, Depok 16424
Indonesia.
e-mail: lroosman@makara.cso.ui.ac.id

References

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, pp. 109-149.

Bond, Z.S., V. Stockmal & D. Muljani (1998). Learning to identify a foreign language. *Language Sciences* 20, pp. 353-367.

Flege, J.E., O.S. Bohn & S. Jang (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25, pp. 437-370.

Ladefoged, P. (1982). *A Course in Phonetics*. Harcourt, Brace & Jovanovich, New York.

Mayo, L.H., M. Florentine & S. Buus (1997). Age of second-language acquisition and perception of speech in noise. *Journal Speech, Language and Hearing Research* 40, pp. 686-693.

Muhadjir (1977). *Morfologi Dialek Jakarta: Afiksasi dan Reduplikasi*, ILDEP series. Djambatan, Jakarta.

Nababan, P.W.J. (1981). *A grammar of Toba Batak*. Pacific Linguistic Series D–No.37.

Nooteboom, S. (1997). The Prosody of Speech: Melody and Rhythm. Hardcastle, W.J. & J. Lavers (eds.). *The Handbook of Phonetic sciences*. Blackwell, Oxford, pp. 640-673.

Tuuk, H.N. van der (1971). *A grammar of Toba Batak*. Martinus Nijhoff, Den Haag.

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. Keller E. (ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges.* John Wiley, Chichester, pp. 23-40.

Wijngaarden, S.J. van (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication* 35, pp. 103-113.

Appendix

Percentage of identification of the language background of the speakers in three stimulus versions broken down by the speaker group. The bold numbers present the percentages of correct identification.

Toba Batak listeners

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

Jakarta Malay listeners

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

Title:	Non-native accents in Dutch word stress realization
Author:	Lilie M. Roosman Phonetics Laboratory, Universiteit Leiden Centre for Linguistics, The Netherlands, and Program Studi Belanda, Universitas Indonesia, Indonesia.
Postal Address:	Phonetics Laboratory, Universiteit Leiden Centre for Linguistics, Cleveringaplaats 1 PO Box 9515 2300 RA Leiden The Netherlands
E-mail:	l.m.roosman@let.leidenuniv.nl