Modelling Fundamental Frequency in First Post-tonic Syllables in Danish Sentences

Niels Reinholt Petersen
Department of General and Applied Linguistics
University of Copenhagen, Denmark
nrp@cphling.dk

Abstract
The work reported in the paper continues previous research on the description of Danish intonation in mathematical terms using a linear regression model. The present paper focuses on the first post-tonic syllable which constitutes the fundamental frequency peak in the stress group in Standard Danish. The results indicate that—in contradistinction to the stressed syllables—the F0 variation over the sentence in the first post-tonics is explained mainly by their position in the sentence, and to a much lesser degree by their position in the prosodic phrase.

1. Introduction
The work reported in the present paper is part of the development of the intonation rule module for a Danish text-to-speech system. Research on Danish intonation carried out primarily by Nina Grønnum (see e.g. [1] and [2] and references therein) has shown that Danish intonation can be described in terms of a hierarchical model where components of smaller temporal scope are superposed on components of larger temporal scope. So far the following components have been established: 1) a text component, 2) a sentence component, 3) a prosodic phrase component, 4) a stress group component, 5) a 'stød' component, and 6) a microprosodic component.

The work described below focuses attention on the interplay between the sentence and prosodic phrase components and the stress group component.

Sentence function in Standard Danish is signalled globally by the slope of the general fundamental frequency contour: in terminal declarative sentences the contour is steeply falling over the sentence, in continuative sentences and non-echo questions, the fall is less steep, and in echo-questions the contour is horizontal. In terminal declaratives, the slope of the sentence contour has been shown to vary with sentence length, being steeper in short than in longer sentences, i.e. the fundamental frequency range is independent of sentence length. Further, in sentences of more than 4 – 5 stress groups the sentence is divided into a number of prosodic phrases, each having its own declining F0 contour starting with a (partial) resetting and superposed on the sentence contour [3, 4]. The prosodic phrasing of a sentence seems to be the result of a complex—and not very well understood—interaction between a number of factors, such as syntactic structure, semantic content, and a tendency to avoid long (more than 5 stress groups) prosodic phrases.

The Danish stress group consists of a stressed syllable and the following unstressed ones up to the next stressed syllable (or the end of the sentence). In the Standard Danish stress group F0 is relatively low in the stressed syllable followed by a steep rise to a peak in the first post-tonic syllable. After the peak F0 falls gradually over the succeeding unstressed syllables. The steepness and shape of the F0 fall varies with speaker and number of unstressed syllables following the peak, the invariant characteristic of the stress group being the rise from the stressed to the first post-tonic syllable.

Figure 1 summarises the effects of the sentence, prosodic phrase, and stress group components on the fundamental frequency pattern in a Danish sentence consisting of eight stress groups divided into 3 phrases of 3, 3, and 2 stress groups, respectively.

Previous research using multiple regression analysis [6] has shown that—if the prosodic phrasing of a sentence is known—the fundamental frequency of a stressed syllable can be described with a high degree of accuracy as a linear function of two independent variables, viz. the syllable's position in the sentence and its position in the prosodic phrase.

Figure 1: The fundamental frequency pattern in a sentence consisting of eight stress groups divided into 3 phrases of 3, 3, and 2 stress groups, respectively. The stressed syllables are indicated by filled circles and unstressed syllables by empty circles. The heavy lines connect the stressed syllables within a prosodic phrase.
The work reported in the present paper continues the research described in [6]. The aim was to examine whether the regression model used in the description of the stressed syllables can also be applied in the description of the fundamental frequency variation in the first post-tonic syllables. Specifically, the question addressed is whether both independent variables (the stress group’s position in the sentence and in the prosodic phrase) should be included in the regression model, which is what would be expected from the superpositional model of Danish intonation, or whether the stress group’s position in the sentence is sufficient as a predictor of $F_0$ variation over the sentence.

## 2. Method

### 2.1. Material and speakers

The fundamental frequency measurements analysed here were extracted from two previously recorded materials, which were also used in the work on stressed syllables referred to above [6]. One sentence material (henceforth material A) consisted of 4 terminal declarative sentences spoken once each by 8 male speakers of Standard Danish. The sentences were syntactically relatively complex and varied in length between 10 and 12 stress groups (see further [5]). Of the 32 renderings of the material, 9 had to be left out either because of difficulties in determining $F_0$ in the first post-tonic syllable or because of shwa elision in that syllable. All renderings of the sentences contained at least two prosodic phrases. The other material (material B) consisted of two sets of terminal declarative sentences recorded and measured by Nina Grønnum [3, 4], who kindly made the measurements available for the present research. The sentences were syntactically relatively simple and varied in length between 4 and 8 stress groups. Data for two male speakers of Standard Danish (one was the author) was used here. One set of sentences was spoken 6 times by both speakers and the other set was spoken 6 times by one of the speakers. Two renderings of the 4 stress group sentences were left out because they turned out to consist of only one prosodic phrase (cf. section 2.2. below). Altogether 111 sentence items were analysed, viz. 23 in material A and 88 in material B.

### 2.2. Determining prosodic phrasing

The investigation of the interplay between the first post-tonic syllables and the sentence and prosodic phrase intonation contours presupposes that the phrasing of each sentence item is known. This is not the case; the lack of precise knowledge of the complex interaction between several factors makes it highly problematic to predict prosodic phrasing with any certainty from external (higher level) information, nor can the prosodic boundaries be determined unambiguously by visual inspection of fundamental frequency tracings. However, since $F_0$ in the stressed syllables in a sentence can be described as a linear function of their position in the sentence and in the phrase, multiple linear regression analysis can be applied as a tool for determining prosodic phrase boundaries: Each sentence item was analysed using all prosodic phrasings possible for that sentence (including the possibility that the sentence comprises only one phrase), the only constraint being that a phrase must consist of at least two stress groups. The phrasing yielding the highest correlation coefficient was selected as the prosodic phrasing for the sentence item in question, and this was the phrasing employed in the subsequent treatment of the data. In the regression analyses $F_0$ was expressed in semitones relative to the mean $F_0$ of the sentence item analysed; and the sentence and phrase position of a stressed syllable was defined as the stress group’s ordinal number in the sentence and phrase, respectively. The ordinal number was employed because it yields a slightly better (but significantly so) prediction of $F_0$ than does physical time (see [6]).

### 2.3. Regression analyses

Multiple and simple regression analyses were applied to the fundamental frequency values (in semitones) of the first post-tonic syllables of each sentence item. The phrasings used for the multiple regression analyses were those established for the stressed syllables as described above. The independent variables were the ordinal number of the stress group in the sentence and in the phrase for the multiple regression analyses, and the number in the sentence for the simple regression analyses.

In order to compare the behaviour of the first post-tonic syllables to that of the stressed syllables, simple regression analyses were also carried out on the stressed syllables—in addition to the multiple analyses made in connection with the phrase finding procedure described in section 2.2.

One of the statistics produced by the regression analyses—and the one used in the evaluation of the results—is the squared correlation coefficient ($R^2$), which is the ratio of the amount of variance accounted for by the independent variable(s) to the total variance of the dependent variable. On the basis of the squared simple and multiple correlation coefficients it will be possible to estimate the relative explanatory power of the two independent variables in stressed and first post-tonic syllables.

## 3. Results

In Order to illustrate the principles of analysis outlined above the results for one sentence item (of 8 stress groups and 3 prosodic phrases) are given in figure 2 and table 1. Figure 2 displays $F_0$ in semitones in the stressed and first post-tonic syllables as a function of stress group number in a sentence of 8 stress groups and 3 prosodic phrases. The lines are the least squares simple and multiple regression lines.

![Figure 2](image.png)
regression lines. It is seen that $F_0$ in the stressed syllables is more accurately predicted by the multiple than by the simple regression line, whereas in the first post-tonics the two regression lines are almost coinciding, indicating that they predict $F_0$ equally well. Table 1 shows that including the phrase component in the analysis increases the explained variance by more than 5 per cent in the stressed syllables and by less than 1 percent in the first post-tonics.

Table 1: Squared multiple and simple correlation coefficients for stressed and first post-tonic syllables in one sentence (cp. figure 1).

<table>
<thead>
<tr>
<th></th>
<th>stressed syllables</th>
<th>first post-tonic syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>multiple regression</td>
<td>0.975</td>
<td>0.958</td>
</tr>
<tr>
<td>simple regression</td>
<td>0.923</td>
<td>0.951</td>
</tr>
</tbody>
</table>

The results of the regression analyses of all sentence items are summarised in figure 3 and table 2. Figure 3 displays the frequency distributions of the squared simple and multiple correlation coefficients obtained for stressed and first post-tonic syllables in material A and B pooled, and table 2 shows the medians of the squared correlation coefficients and the differences between the medians (the median is used as a measure of central tendency because of the heavy negative skewness frequency distributions). In the case of the stressed syllables there is an appreciable difference between the squared correlation coefficients obtained by simple and multiple regression analysis. The median increase in explained variance when the prosodic phrase component is included in the analysis is more than 13 per cent for material A and B pooled. For the first post-tonic syllables the increase only amounts to 1.4 per cent. It is seen from table 2 that the effect on the amount of explained variance of including the phrase component in the regression analysis differs between the two materials. In material A (the syntactically more complex material) the median increase in explained variance is 23 per cent in the stressed syllables and 8.8 per cent in the first post-tonics. In the less complex material (B) the effect of adding the phrase component is smaller, the increase being 10.8 per cent in the stressed and 0.8 per cent in the first post-tonic syllables. But the relation between stressed syllables and first post-tonics is the same in the two materials: The position of

Table 2: Median simple and multiple squared correlation coefficients in stressed and first post-tonic syllables in material A and B and in the two materials pooled.

<table>
<thead>
<tr>
<th></th>
<th>stressed syllables</th>
<th>first post-tonic syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material A</td>
<td>Simple regr. 0.654 0.718</td>
<td>Multiple regr. 0.884 0.806</td>
</tr>
<tr>
<td>Difference</td>
<td>0.230</td>
<td>0.088</td>
</tr>
<tr>
<td>Material B</td>
<td>Simple regr. 0.826 0.924</td>
<td>Multiple regr. 0.934 0.932</td>
</tr>
<tr>
<td>Difference</td>
<td>0.108</td>
<td>0.008</td>
</tr>
<tr>
<td>Material A+B</td>
<td>Simple regr. 0.790 0.904</td>
<td>Multiple regr. 0.924 0.918</td>
</tr>
<tr>
<td>difference</td>
<td>0.134</td>
<td>0.014</td>
</tr>
</tbody>
</table>

The finding that the differences between squared multiple and simple correlation coefficients are quite small in the first post-tonic syllables could—of course—result from a large random variation of $F_0$ in these syllables or reflect an effect of variables other than those included in the regression analyses. Under such circumstances only slight differences should be expected between multiple and simple correlation coefficients. But the correlation coefficients would also be low, and this is not the case in the present material: in the vast majority of sentence items, the squared correlation coefficient exceeds a value of 0.8. On this basis it seems safe to conclude that it is possible to describe $F_0$ in first post-tonic syllables as a linear function of one independent variable, viz. the stress group’s ordinal position in the sentence.

4. Discussion

The finding that the differences between squared multiple and simple correlation coefficients are quite small in the first post-tonic syllables could—of course—result from a large random variation of $F_0$ in these syllables or reflect an effect of variables other than those included in the regression analyses. Under such circumstances only slight differences should be expected between multiple and simple correlation coefficients. But the correlation coefficients would also be low, and this is not the case in the present material: in the vast majority of sentence items, the squared correlation coefficient exceeds a value of 0.8. On this basis it seems safe to conclude that it is possible to describe $F_0$ in first post-tonic syllables as a linear function of one independent variable, viz. the stress group’s ordinal position in the sentence.

5. References
