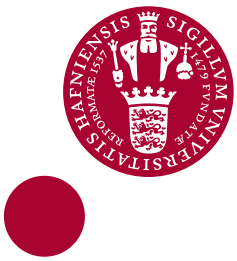


PREPLANNING OF INTONATION IN SPONTANEOUS VERSUS READ ALOUD SPEECH: EVIDENCE FROM DANISH

Tøndering, John

Published in:
Proceedings of the XVIIth International Congress of Phonetic Sciences

Publication date:
2011

Document version
Publisher's PDF, also known as Version of record

PREPLANNING OF INTONATION IN SPONTANEOUS VERSUS READ ALOUD SPEECH: EVIDENCE FROM DANISH

John Tøndering

Department of Scandinavian Studies and Linguistics, University of Copenhagen, Denmark
johnt@hum.ku.dk

ABSTRACT

Preplanning or *look ahead* is one of the basic principles of the current model of Danish intonation. The frequency range spanned by a prosodic phrase is constant regardless of phrase length, ceteris paribus, and the frequency step size between one stressed syllable and the next varies inversely with the number of stressed syllables in the phrase. This empirically well-founded fact is based on read aloud speech. This paper reports on an investigation of whether preplanning is also used in spontaneous speech. Analyses of map task monologues revealed that prosodic phrase contours in spontaneous spoken Danish are not preplanned.

Keywords: preplanning, intonation, Danish

1. INTRODUCTION

The current model of intonation in Danish, which is based on read aloud speech, is a layered, superpositional model where intonation components of smaller temporal scope are superposed on components of larger temporal scope [2, 3]. It follows from the model's hierarchical structure that components of larger temporal scope carry and scale the components of smaller temporal scope. Thus, the stress group pattern — i.e. the fundamental frequency pattern associated with a stressed and all following unstressed syllables — is superposed on the prosodic phrase contour. Utterances with more than 4-5 stressed syllables are divided into a number of prosodic phrases. The prosodic phrase contours ride upon and are constrained by the more global utterance contour.

A stressed syllable always initiates the same F_0 pattern which in standard Copenhagen Danish is described by a brief fall succeeded by a rise to the first unstressed syllable followed by a fall through any succeeding unstressed syllables. The local minimum at the beginning of this stress group pattern coincides with the stressed vowel. The phrase contour is defined by the local minima (cf. Figure 1). The stress group pattern is subject to truncation. A prosodic stress group consisting of only a stressed syllable (and no unstressed syllables) will result in a brief fall and a short rise, and the following local pitch movement (rise-fall) will be absent. This absence of a local high turning point in the F_0 pattern has no influence on the global trend.

The utterance function is signaled globally by the slope of the phrase contour. Different slopes are referred to as intonation types. Terminal declarative sentences are typically accompanied by the most steeply falling contour, and echo-questions are typically accompanied by a horizontal contour. Other questions and non-final clauses will have a slope somewhere between these two extremes.

For a given intonation type the pitch range spanned by the contour is constant across variation in phrase length: In utterances of up to 5 stress groups (above which an utterance is typically decomposed into a succession of prosodic phrases), the stressed syllables are equidistantly spaced in frequency. The frequency intervals are inversely proportional to the number of stressed syllables in the phrase. This fact is illustrated in Figure 1.

Figure 1: The slope of the prosodic phrase contours depends on the number of stressed syllables in the utterance, ceteris paribus. See text.

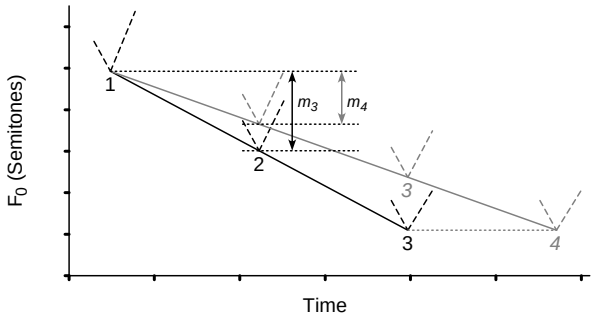


Figure 1 shows the prosodic phrase contours of two utterances with 3 and 4 stressed syllables respectively. The onset and offset of each prosodic phrase is defined by the first and last stressed syllable. The dashed lines in the figure depict simplified F_0 patterns associated with mono-

syllabic stress groups. As can be seen from the figure, the frequency ranges of the two utterances are identical, so that the slope of the prosodic phrase contour associated with the utterance with 4 stress groups (m_4) is shallower than the slope of the contour of the utterance consisting of only 3 stress groups (m_3). That is, longer utterances show smaller frequency intervals between each stressed syllable and accordingly less steep gradients than shorter ones. According to [3], this feature is the result of *preplanning* or *look-ahead* – the speaker simply anticipates the length of the utterance (where length is defined as the number of stressed syllables). The empirical evidence for the model is substantial [7, 9].

Another type of evidence for preplanning which has been reported for other languages is a higher pitch level at the onset of longer utterances [8].

Based on the evidence of preplanning from Danish and other languages, this investigation tests the following three hypotheses regarding preplanning of intonation in spontaneous spoken Danish: 1) Prosodic phrases will have a constant F_0 range across phrase length, 2) the frequency steps between the stressed syllables correlates inversely with the number of stressed syllables in the phrase, and 3) longer utterances will have a higher F_0 starting point.

2. MATERIAL

2.1. Material and speakers

The material was chosen from the publicly available *Danish Phonetically Annotated Spontaneous Speech* corpus (DanPASS) [4, 5]. 45 recordings of monologues spoken by 15 different speakers of standard Copenhagen Danish were selected for the analyses. The recordings of monologues in DanPASS were elicited using three different tasks: 1) describing a geometric network, 2) guiding an unseen and silent listener through four different routes in a virtual city map, and 3) explaining to an unseen and silent listener how to assemble a house from its individual pieces. Since the recordings are a form of one-way communication, questions do not occur in the material. 12 of the speakers are men, and 3 are women. At the time of the recording 3 speakers were older than 40 years (all men), and 12 were younger.

DanPASS is segmented into prosodic phrases, words and syllables, and the annotation includes a phonetic transcription with information about stress. The analysed 45 recordings, lasting a total of 2 hours and 18 minutes, contain 17.135 running words, divided into 24.489 syllables of which 8.590 are stressed. The words are grouped into 2.604 prosodic phrases (the distribution of phrase lengths is shown in Table 1). Since this investigation concerns slope (in the sense explained in previous section), only phrases containing at least 2 stressed syllables are included, and phrases with more than 7 stressed syllables were excluded because of the low number of incidences. That leaves 2.124 prosodic phrases. The segmentation into prosodic phrases was carried out by two project assistants, and was later proof-read by the project manager [4].

Table 1: The distribution of prosodic phrases on phrase length.

Stressed syllables in the phrase	No.
0	90
1	307
2	597
3	559
4	469
5	270
6	157
7	72
8	43
9	19
9+	21
Total	2.604

2.2. Data extraction

The calculation of F_0 and the extraction of data were done using Praat [1]. All the pitch objects created by Praat were hand corrected, and extreme F_0 values, due mainly to aspiration and the Danish stød (a kind of creaky voice), were removed. The mean F_0 value of each stressed syllable was extracted in Hz along with the ordinal number of the syllable in the prosodic phrase.

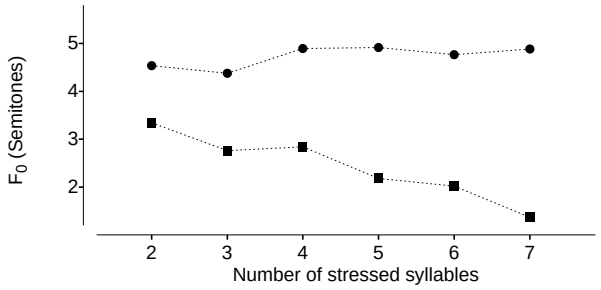
The use of F_0 mean values is not fully consistent with the model's principles. According to the model, the prosodic phrase contour is drawn by a line through all the low turning points in each stress group pattern in the phrase (see Figure 1). But visual inspection revealed that not all stress groups have a fall-rise pattern. The movement can, among other patterns, be purely rising or falling, or rising-falling, i.e. without a low turning point. Therefore, the mean value was chosen.

The measured fundamental frequencies were converted to semitones relative to the tenth percentile of the distribution of F_0 values in the recording in question.

3. RESULTS

If prosodic phrases in spontaneous Danish are preplanned the way it is observed in read aloud Danish, we would expect that the frequency range spanned by prosodic phrases with different length is constant (hypothesis 1). The mean values of phrase onsets and phrase offsets are plotted in Figure 2, and the corresponding values are displayed in Table 2. It is clear that the frequency ranges spanned by phrases with different lengths, are *not* constant. The frequency range spanned by prosodic phrases with 7 stressed syllables is almost three times as high (3.5 semitones) as the range spanned by phrases with 2 stressed syllables (1.2 semitones). In other words, the frequency range increases with increasing phrase length. Thus, there is no evidence for preplanning (of frequency range).

Figure 2: Mean values of onsets (bullets) and offsets (squares) in prosodic phrases. Frequency range increases with increasing phrase length.



The correlation between phrase offset and phrase length has been tested using Spearman's rank order correlation. Spearman's ρ is -0.1421 ($p < 0.0001$, n = 2.124). This may be somewhat surprising, because the mean values displayed in Table 2 and the visual impression given from Figure 2 suggests an even stronger correlation. But this is probably due to the very considerable variation observed in the data, and it indicates that we are dealing with trends, rather than mandatory principles.

Table 2: Mean values of onsets and offsets, and frequency range. d_1 is the mean step size from first to second stressed syllable, and m_2 indicates the average slope of the remaining syllable steps in the phrases.

Phrase length	Onset (ST)	Offset (ST)	Range (ST)	d_1	m_2
2	4.5	3.3	1.2	1.2	-
3	4.4	2.8	1.6	0.9	0.8
4	4.9	2.8	2.0	1.1	0.5
5	4.9	2.2	2.7	0.9	0.6
6	4.8	2.0	2.7	0.9	0.5
7	4.9	1.4	3.5	0.7	0.6

Hypothesis 2 states that the size of the frequency steps between the stressed syllables should correlate inversely with the number of stressed syllables in the phrase. However, the average contours reveal no clear indication of this. The frequency steps for phrases of varying lengths can be seen from Figure 3 and Figure **4**.

In Figure 3a, the average intonation contours are aligned relative to phrase offset. The contours are almost rectilinear, perhaps with a steeper gradient from the first to the second stressed syllable. This impression is confirmed by Table 2. The mean of the first step in the phrases (d_1) is always larger than the mean of the remaining steps in the phrases (m_2). From Figure 3b, in which the perspective is changed as the phrase contours are aligned relative to phrase onset, it can be seen that the phrase contours are almost parallel, and no tendency for a correlation between phrase length and step size can be observed.

Figure 3: Mean intonation contours. In a), the contours are aligned relative to the offset of the phrases. In b), the curves are aligned relative to phrase onset, and each phrase onset is set arbitrarily to 5 semitones.

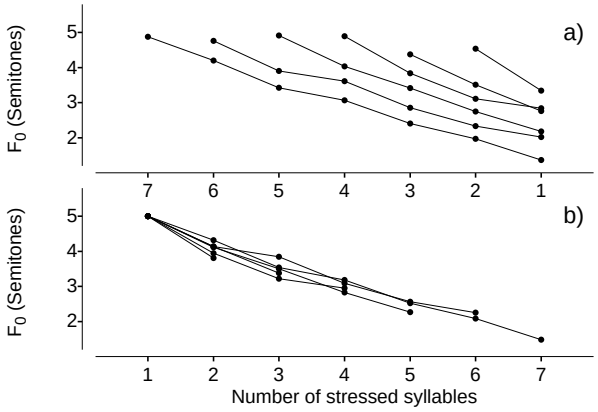


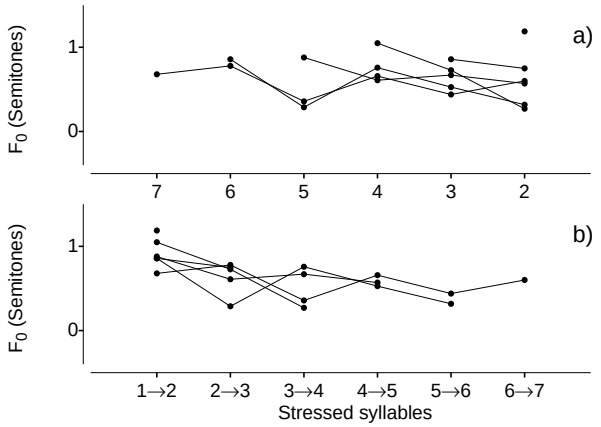
Figure 4 shows the mean frequency steps from one stressed syllable to the next. Again, the same picture is observed: No overall indication of a correlation between phrase length and step size, but a tendency towards larger step size from the first to the second stressed syllable in the phrase.

The correlation between step size and phrase length has further been tested using Spearman's rank order correlation. Spearman's ρ is -0.0637 ($p < 0.0001$, $n = 5.309$). Although the correlation is significant, the correlation coefficient is almost 0, demonstrating that a relationship between phrase length and step size is not present.

In addition, it should be mentioned that the almost rectilinear phrase contours in Figure 3 are

depicted on a logarithmic semitone scale. If the calculations instead were done in Hz and depicted on a linear Hz scale, then the contours would show an asymptotic pattern with decreasing step size from the start to the end of the prosodic phrases. This behaviour is perfectly compatible with a plan-as-you-go model, where each step size represents a fraction of the previous step size. That is, preplanning is not required [6].

Figure 4: Frequency steps from stressed to stressed syllable. In a) the steps are aligned relative to the phrase offset, and in b) the steps are aligned relative to phrase onset.



The last type of evidence for preplanning to be examined is whether longer utterances begin at a higher pitch level (hypothesis 3). It appears from Figure 2 that the mean onsets are quite constant, but a small increase can be observed from phrases with 2-3 stressed syllables to phrases with more than 3 syllables (approx. 0.5 semitones). Although the correlation is significant (p = 0.0026, n = 2.124), the correlation coefficient is very low (Spearman's ρ = 0.0654), which leaves no ground for claiming that speakers look ahead and anticipate phrase length.

4. DISCUSSION

No evidence for preplanning of intonation in Danish spontaneous speech was found. How can we account for this difference between read and spontaneous speech? The explanation seems obvious: When speakers *read aloud* a given sentence, they have a visual representation of an already constructed sentence, and they may on this basis have estimated the number of stressed syllables in the utterance as they start reading aloud. Preplanning could therefore be seen as a consequence of the type of data collected.

5. CONCLUSION

Evidence of preplanning of intonation has been found for read Danish. The purpose of the present investigation was to explore whether evidence for preplanning could also be found in spontaneous Danish. The results showed that the frequency range spanned by a prosodic phrase increases with increasing phrase length. Phrases with different lengths set off from approximately the same height, but they tend to end lower the longer they are. Furthermore, no correlation between phrase length and frequency steps could be found. These observations point to distinct differences between read aloud and spontaneous Danish. In other words, the analyses revealed no evidence of preplanning of intonation in spontaneous Danish.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2011. Praat: doing phonetics by computer. Retrieved from *http://www.praat.org*.

[2] Grønnum, N. 1992. *The Groundworks of Danish Intonation. An Introduction*. Copenhagen: Museum Tusculanum Press.

[3] Grønnum, N. 1995. Superposition and subordination in intonation - a non-linear approach. *Proc. 13th ICPhS* Stockholm, 2, 124-131.

[4] Grønnum, N. 2009. A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication* 51(7), 594-603.

[5] Grønnum, N. 2011. DanPASS – Danish Phonetically Annotated Spontaneous Speech. *http://www.danpass.dk*

[6] Liberman, M., Pierrehumbert J. 1984. Intonational invariance under changes in pitch range and length. In Aronoff, M., Oehrle, R.T. (eds.), *Language Sound Structure*. Cambridge: MIT Press, 157-233.

[7] Petersen, N.R. 1999. Modelling Danish sentence and phrase intonation. *Proc. 14th ICPhS* San Francisco, 925-928.

[8] Rialland, A. 2001. Anticipatory raising in downstep realization: Evidence for preplanning in tone production. *Proc. of the Symp. Cross-Linguistic Studies of Tonal Phenomena* Tokyo, 301-321.

[9] Thorsen, N. 1983. Standard Danish sentence intonation – phonetic data and their representation. *Folia Linguistica, Acta Societatis Linguisticae Europaeae* 17(1-2), 187-220.