

STRATEGIES FOR PROSODIC PHRASING IN SWEDISH

Gösta Bruce*, Björn Granström** and David House*⁺

*Department of Linguistics and Phonetics
Lund University, Helgonabacken 12, S-223 62 Lund, Sweden

**Dept of Speech Communication and Music Acoustics
Royal Institute of Technology, Box 70014, S-10044 Stockholm, Sweden

⁺Names in alphabetic order

ABSTRACT

This study focuses on the problem of prosodic phrasing in Swedish. A small database of sentences, potentially ambiguous with respect to phrase boundary location, have been recorded and analysed. Considerable variation in phrase and clause boundary realizations was observed. Strategies including both boundary and coherence signalling have been identified.

1. INTRODUCTION

This contribution represents cooperative work on a model for Standard Swedish prosody in the context of a research project on prosodic phrasing in Swedish. The aim of the project is to investigate the phonetic correlates of phrasing using production data, text-to-speech synthesis and automatic prosodic recognition.

In earlier work [2] we have outlined our joint research work on modelling Swedish prosody in a text-to-speech framework. See also [1] for earlier work aimed at developing a model for Swedish prosody, [3] for work directed towards the development of the prosodic component of a text-to-speech system, and [5] for a description of the prosodic parser.

It is widely recognized that grouping - involving the double aspect of coherence (connective) signalling and boundary (demarcative) signalling - is one of the main functions of prosody. Our focus of interest here is particularly in the division of an utterance into prosodic phrases and clauses.

The acoustic-phonetic signalling of prosodic phrasing is assumed to be complex, involving several parameters such as F0, duration, intensity, and voice quality as well as possible silence

(physical pause). The more precise exploitation of these cues for prosodic phrasing in Swedish is, however, not well understood. The aim of the present paper is to explore different phrasing strategies which make use of some of these cues and their possible combinations.

2. SPEECH MATERIAL

In order to gain more knowledge about prosodic phrasing in Swedish [2], we devised speech material specifically designed for this purpose. As a starting point we chose sentences which, for the most part, were syntactically ambiguous. This was done to give us a preliminary idea about phrasing strategies and to enable us to easily test these strategies in the text-to-speech framework. The speech material consisted of 22 sentences, typically occurring as minimal pairs, where the location of the sentence internal clause boundary was varied. Example sentence pairs are the following:

1a. Skolan börjar med samling i klassen. (School begins with a meeting of the class.)

1b. Skolan börjar, när barnen vågar. (School begins when the children dare.)

2a. När pappa fiskar, stör Piper Putte. (When daddy is fishing, Piper disturbs Putte.)

2b. När pappa fiskar stör, piper Putte. (When daddy is fishing sturgeon, Putte peeps.)

3a. När han överlämnade sej, och bonden hälsade kungen med ett leende, så blev det bara så. (When he surrendered, and the farmer greeted the king with a smile, it just happened that way.)

3b. När han överlämnade sej och bonden, hälsade kungen med ett leende;

så blev det bara så. (When he and the farmer surrendered, the king greeted them with a smile; that's the way it happened.)

A male Stockholm Swedish informant read the speech material three times. He was given explicit instructions not to make any pauses at sentence-internal boundaries.

3. SPEECH ANALYSIS

In the present speech corpus, considerable variation in the acoustic-phonetic signalling of phrasing and phrase boundaries was observed. Here we will not aim at giving an exhaustive description of the production data, but rather point to a few possible strategies in the exploitation of acoustic-phonetic cues for prosodic phrasing that we find especially interesting.

3.1. Boundary by duration only

One possible strategy is to use only duration for clause/phrase boundary signalling. This appears in some of the shorter sentences of our test material where there is no marking of the boundary in terms of F0. In these sentences we find segmental lengthening before the clause boundary. An example of this is given in Figure 1 where the final segments of the word "börjar" are clearly lengthened before the clause boundary (sentence 1b) as contrasted with the same word in the context before the prepositional phrase (sentence 1a).

3.2. Coherence by deaccentuation

Another strategy for prosodic grouping represented in our speech corpus is to use F0 and duration (usually in combination) for the signalling of coherence within a speech unit. Exemplification is given with reference to the ambiguous pair of sentences 2a and 2b (Figure 2). In sentence 2b we observe the backgrounding of "fiskar" involving both flattening of F0 (deaccentuation) and segment shortening. The two words - "fiskar" (verb) and "stör" (object) - are produced as a unit with only "stör" being accented (focal accent). This unit accentuation thus serves as a connective signal and may by itself be sufficient for the disambiguation of sentences 2a and 2b. Usually, however, this coherence signalling is accompanied by explicit boundary signalling. A typical F0 correlate here is the terminal F0-fall to a bot-

tom F0 level on "stör" (Figure 2b), which is also combined with segment lengthening.

3.3. Coherence by hat pattern

For the other member of the pair, test sentence 2a, with the intended internal boundary located between "fiskar" and "stör", we encounter another kind of coherence signalling without the use of deaccentuation. Here the F0 rise on "stör" followed by the F0 fall on "piper" together form a hat pattern [4], which serves as a connective signal. In this sentence we do not observe any obvious F0-boundary cues in connection with "fiskar", i.e. no F0-fall to a bottom level, although there are apparent segment lengthenings.

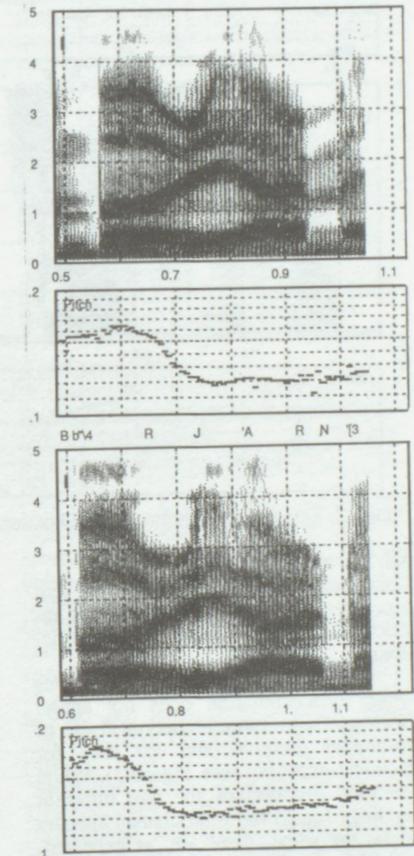


Figure 1. Partial spectrograms and FO of sentences 1a (top) and 1b (bottom)

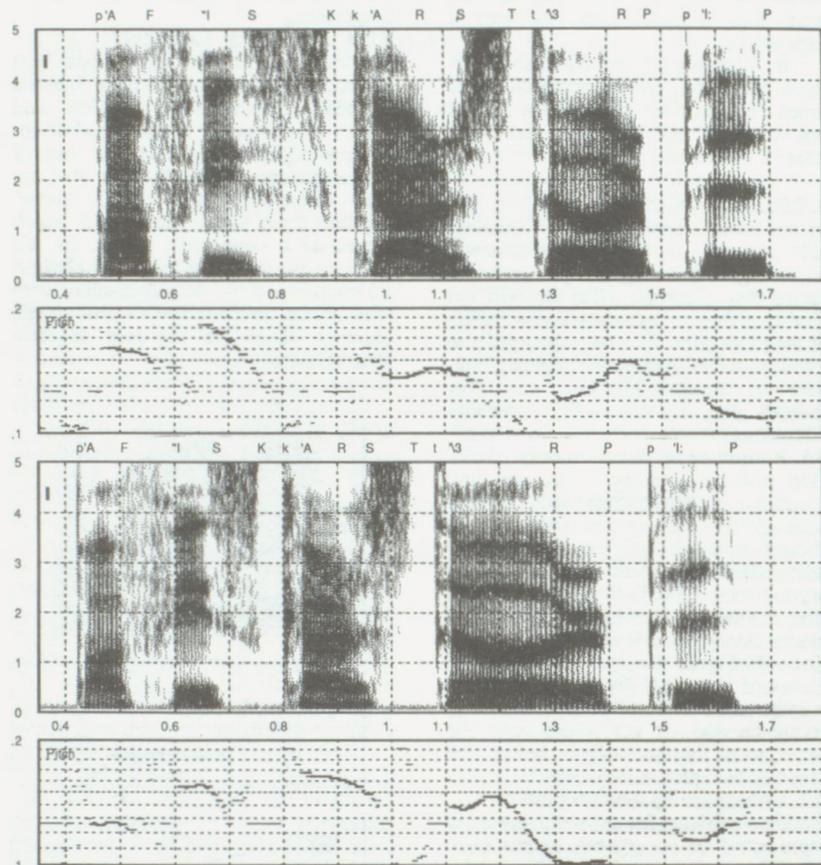


Figure 2. Partial spectrograms and FO of sentences 2a (top) and 2b (bottom)

3.4. Phrasing and syntax

Coherence signalling in the form of unit accentuation as exemplified here is restricted to certain syntactic constructions.

Sentences 3a and 3b are examples where deaccentuation does not apply. Here we find a more archetypical use of combined duration and F0 cues for prosodic grouping (see Figure 3). While the total duration of the two different readings (up to the final clause) appears to be the same, there are, as expected, local lengthenings at different places depending on the location of the internal boundary.

In test sentence 3b, where the boundary occurs after "bonden", the pre-boundary lengthening is combined with a drop in F0 to a bottom level. This F0 drop is also the end of a typical downstepping pattern for the two last accents ("sej" and "bonden") within the first phrase.

In the other member of the sentence pair, 3a, where the boundary is located after "sej", we observe the pre-boundary lengthening as well as a fall to a fairly low F0 level, albeit not a bottom F0 level. When comparing the F0 valleys at "sej" and "bonden" across the boundary,

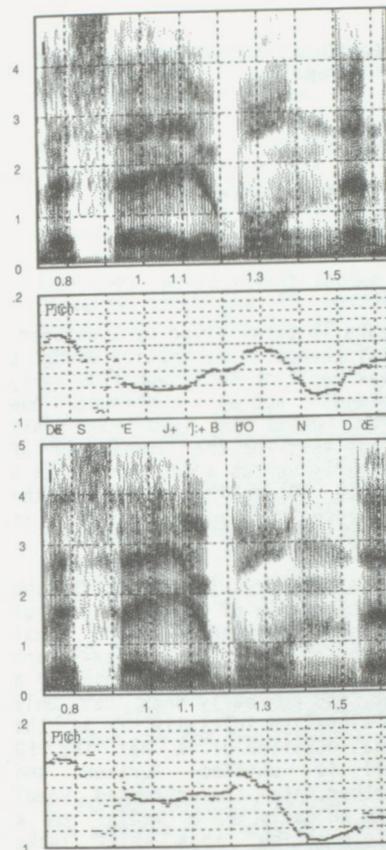


Figure 3. Partial spectrograms and FO of sentences 3a (top) and 3b (bottom)

there is no downstepping (pattern) to be observed.

The moderate F0 drop at "sej" in connection with the boundary in 3a, as compared with the drop to a bottom F0 level at "bonden" (where the boundary is in 3b), invites the following possible account of phrasing strategies. The syntactic structure of the two sentences displays an interesting difference. In 3a we have a coordination of two subordinate clauses before the main clause of the sentence, while in 3b a single subordinate clause precedes the main clause of the sentence, which is then followed by another independent clause.

According to our interpretation the moderate F0 drop at "sej" represents the sign of the continuation of the subordinate clause (in 3a), while the larger F0 drop at "bonden" (in 3b) represents the termination of this syntactic unit (subordinate clause).

4. CONCLUSIONS

We have identified and explored some alternative phrasing strategies in Swedish. Phrase boundaries can be signalled by duration only (pre-boundary lengthening) or by duration in combination with an F0 drop to a low level. Coherence within a unit can be marked by deaccentuation as well as by more complex means involving specific combinations of F0 and duration. Experiments using these strategies in synthetic speech and prosodic recognition will be reported at the congress.

ACKNOWLEDGMENT

This work has been supported in part by grants from the Swedish National Board for Technical Development and the Swedish Telecom. The present work is carried out within the Language Technology Programme under a grant from the Swedish Council for Research in the Humanities and Social Sciences.

REFERENCES

- [1] Bruce, G. (1985), "Structure and functions of prosody", in Guerin & Carré (eds.) Proceedings of the French Swedish Seminar on Speech, 549 - 559, Grenoble.
- [2] Bruce, G. & Granström, B. (1990), "Modelling Swedish prosody in text-to-speech: phrasing", in Wiik & Raimo (eds.), Nordic Prosody V, 26-35, Turku University.
- [3] Carlson, R., Granström, B. & Hunnicutt, S. (1990), "Multilingual text-to-speech development and applications", in Ainsworth (ed.), Advances in speech, hearing and language processing, 269-296, JAI Press, London.
- [4] Hart, J. 't & Cohen, A. (1973), "Intonation by rule: a perceptual quest", J. of Phonetics 1, 309-327.
- [5] House, D. & Bruce, G. (1990), "Word and focal accents in Swedish", in Wiik & Raimo (eds.), Nordic Prosody V, 156-173, Turku University.