5

# Prosodic Phrasing in Swedish

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

## Introduction

The present paper is a status report from the initial stage of a new research
project named *Prosodic Phrasing in Swedish* which started in the second
half of 1990. The project represents cooperative work between the
Department of Linguistics and Phonetics at Lund and the Department of
Speech Communication and Music Acoustics, KTH, Stockholm. The starting
point for our research effort within phrasing and prosody has been our
judgement that we possess a fair amount of knowledge about prominence
relations and accentuation but we know relatively little about prosodic
grouping and phrasing (Bruce 1977, 1985; Bruce & Gårding 1978; Carlson
& Granström 1973, 1986; House 1990).

The present project is directed towards basic research, where the
primary goal is to attain new knowledge about phrasing and prosody in
Swedish. The problems to tackle concern both questions of phonology and
phonetics. One of the main phonological issues is to try to understand what
structure could be assumed for prosodic phrasing, particularly what types
of prosodic phrases can be identified as relevant domains between a
'prosodic word' and a 'prosodic utterance' (for a discussion see for example
Selkirk 1984, Nespor & Vogel 1986, Pierrehumbert & Beckman 1988).
The basic phonetic issue is to find out what speech properties (F0, duration,
intensity, phonation type, pausing, etc.) and combinations of them can be
used to signal phrasing. The possibility of a hierarchy among these phonetic
cues will also be explored. Concerning the grouping function of prosody,
our approach is to actively look not only for boundary signals (demarcative
cues) but also for coherence signals (connective cues).

*Dept. of Speech Communication and Music Acoustics, KTH, Stockholm

+Names in alphabetic order

Basically three different methods are being exploited in the project work. The first method is the collection and analysis of speech production data. It involves the construction and recording of specially designed test material, as well as the selection and recording of suitable read text passages (laboratory speech). This method also includes the processing of speech data in the KTH speech data base involving phonetic segmentation and labelling of phonetic segments (Carlson, Granström & Nord 1989). This information can then be used for different kinds of analysis (spectrograms, pitch extraction, statistics) and for searching in the database according to linguistic criteria. This processing also allows for the integration of the speech database and synthesis. The speech production data method will be used primarily to generate tentative hypotheses about important speech properties in phrasing.

The second method is the use of text-to-speech synthesis for the testing of hypotheses about the signalling of prosodic phrasing. In the KTH text-to-speech system there are several ways of interacting with rules and parameters (Carlson, Granström & Hunnicutt 1991). Different types of phrasing can be simulated and selected through the use of special symbols in the phonetic transcription. There are special facilities to interactively change rule variables. Another possibility is to display and modify parameters after rule execution but before parameter interpolation and synthesis. Thus the synthesis method can be used both for the testing of hypothesized generalizations on new, unrecorded speech material and for the design of more specific perception tests.

The third method is prosodic parsing directed towards the recognition of phrasing. The first stage of the prosodic recognition method is the use of a human recognizer, an expert reader of an acoustic record of speech for the identification of potential phrases of an 'unknown' speech signal (cf. House & Bruce 1990). Based on the knowledge used by the expert reader for prosodic parsing, the subsequent procedure will then be to teach the computer to make an automatic analysis of prosodic phrases. Generally, we believe that the prosodic parser is particularly suitable for testing hypotheses about the interaction between different speech variables for the expression of prosodic phrasing.

## Phrasing strategies (analysis)
One part of the project work is to try to identify different phrasing strategies by means of prosody. As an extension of the speech production

data that we have described in earlier papers we have recorded speech material where the test sentences occur as minimal pairs (cf. Bruce, Granström & House 1990, 1991). The typical, distinctive feature of these sentences is the absence vs. presence of an internal clause boundary. Example sentences are:

1.    LÄRARNA BACKAR FÖR POJKARNAS SPARKAR.
           [ "læːɹaɳa 'bàkːaɹ fœɹ 'pɔ̀jkaɳas "spàɹkaɹ ]
           'the teachers back away from the boys' kicks'

2.    LÄRARNA BACKAR, FÖR POJKARNA SPARKAR.
           [ "læːɹaɳa 'bàkːaɹ ‖ fœɹ 'pɔ̀jkaɳa "spàɹkaɹ ]
           'the teachers back away because the boys kick'

Examples will here be taken from six readings by a male, Stockholm Swedish informant, where he was encouraged to make a free and varied interpretation of the test material. A common denominator of the test sentences is that they all contain four accents, the first and the last typically being focal. Our analysis of the speech material demonstrates a fair degree of variation. Test sentence 1 – without an intended internal boundary – displays basically only one prosodic pattern. The whole sentence is made up of one prosodic phrase. A characteristic feature of this pattern is a connective F0 downstepping from the second to the final accent and also the absence of internal pre-boundary lengthening (see Figure 1). Test sentence 2, on the other hand, is represented by at least four different strategies for prosodic phrasing. One of these patterns appears *not* to be distinct from the corresponding pattern of sentence 1 (see Figure 2). This means that sentence 2 can be realized phonetically as one coherent prosodic phrase. This has also been confirmed in an informal listening test, where this particular version of sentence 2 was often confused with sentence 1. The characteristic connective, downstepping F0 course is also apparent in this version of sentence 2.

Three other versions of sentence 2 are clearly distinguishable from sentence 1. All three are divided up into two distinct prosodic phrases. One version of sentence 2 is temporally distinct but not tonally distinct from sentence 1 (see Figure 3). The apparent boundary cue is a marked pre-boundary lengthening preceding a small physical pause, while the F0 downstepping indicates coherence. Auditorily, this is still a relatively strong
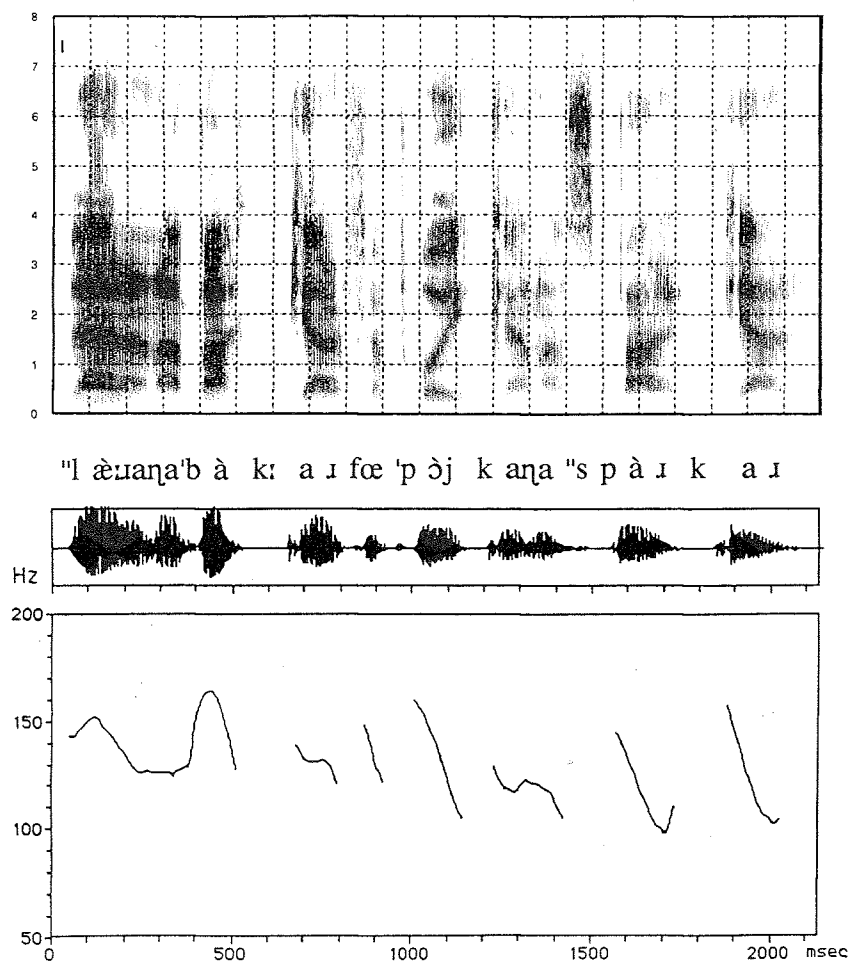
**Figure 1.** Spectrogram, waveform and fundamental frequency contour of test sentence 1 displaying one prosodic phrase with F0 downstepping functioning as a coherence cue.
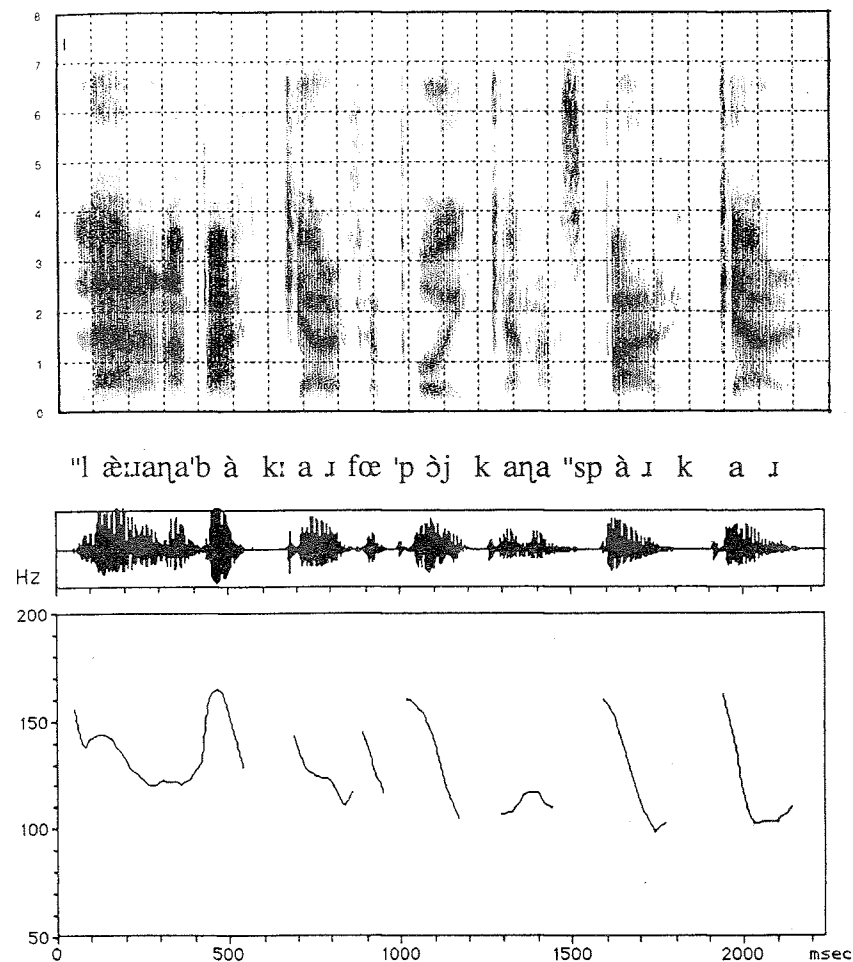


**Figure 2.** Spectrogram, waveform and fundamental frequency contour of test sentence 2 displaying one prosodic phrase with F0 downstepping functioning as a coherence cue.
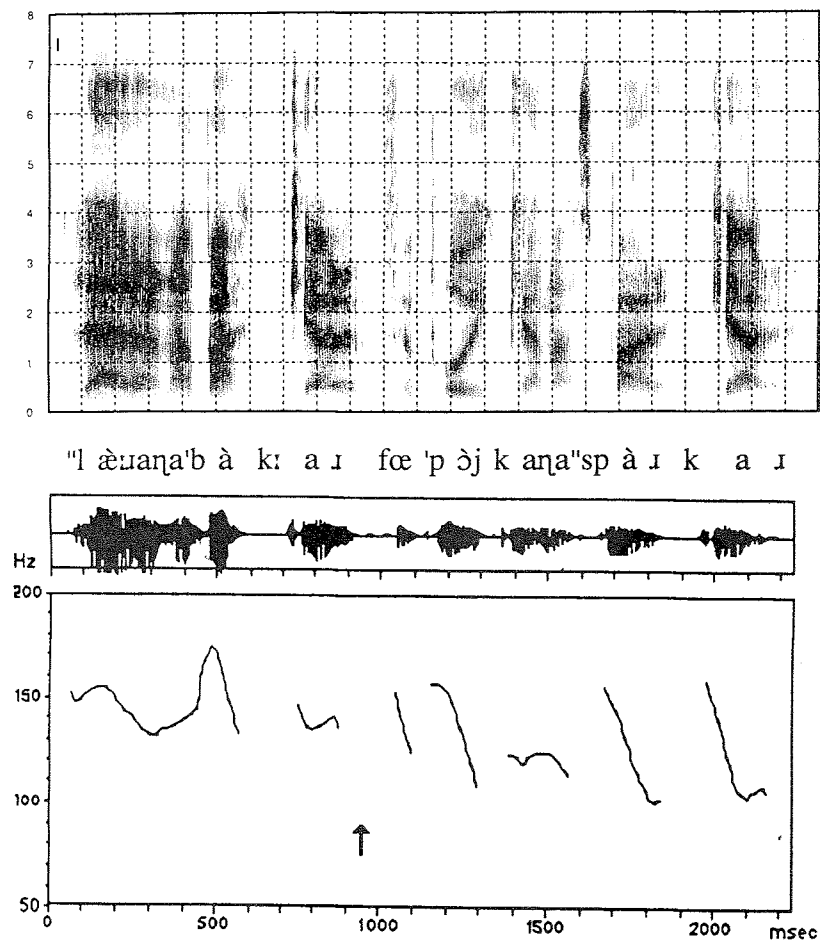
**Figure 3.** Spectrogram, waveform and fundamental frequency contour of test sentence 2 displaying two prosodic phrases with duration and a pause (arrow) functioning as boundary cues.
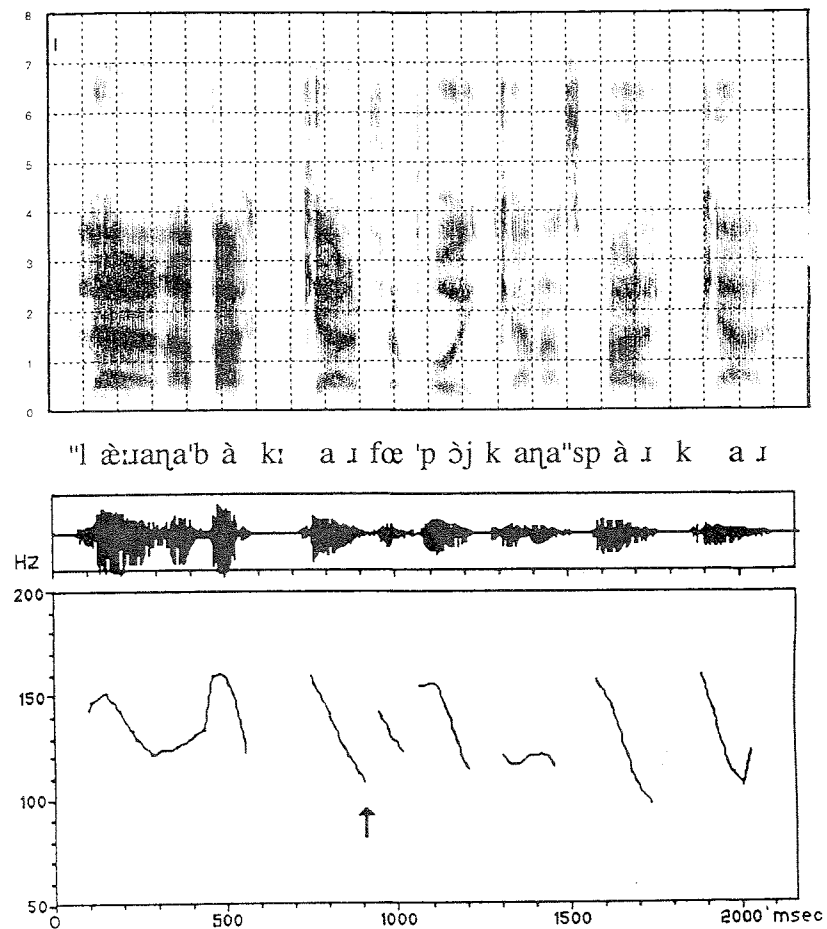


**Figure 4.** Spectrogram, waveform and fundamental frequency contour of test sentence 2 displaying two prosodic phrases. Boundary cues are a focal accent and a terminal fall (arrow).
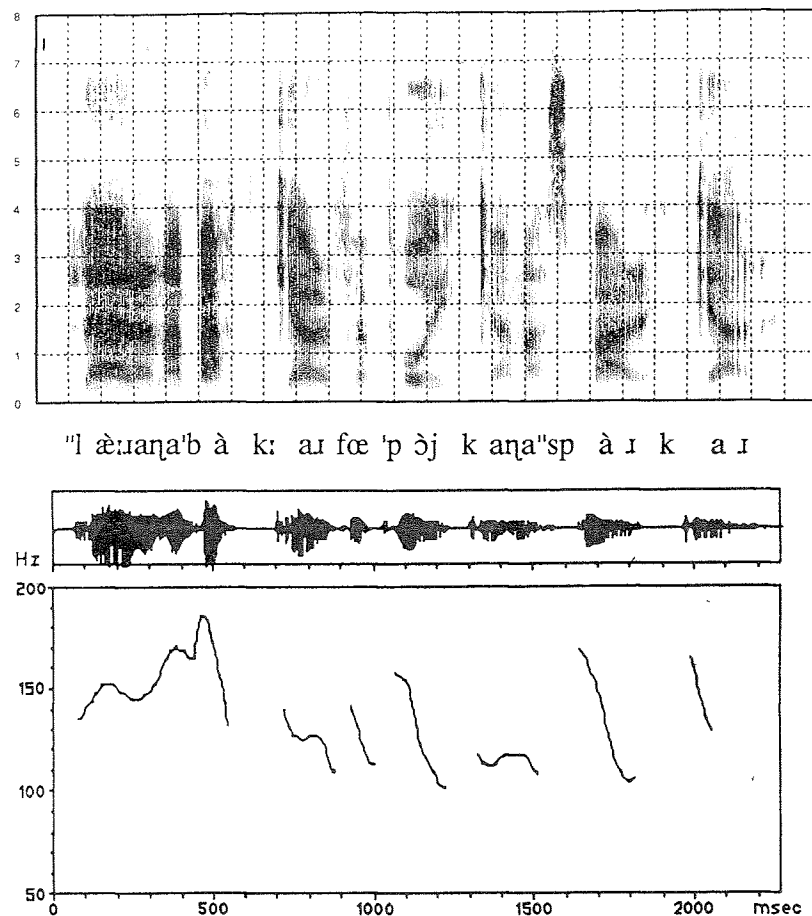
"l æːɹaɳa'b à kː aɹ fœ 'p ɔj k aɳa"sp à ɹ k a ɹ

**Figure 5.** Spectrogram, waveform and fundamental frequency contour of test sentence 2 displaying two prosodic phrases where the phrasing cue is a wide initial F0 range giving emphasis for contrast.

phrasing signal. Another version of sentence 2 – also distinct from sentence 1 – is characterized by the addition of a focal accent to the second accent accompanied by a terminal F0 fall and a moderate pre-boundary lengthening (see Figure 4). This also constitutes a strong boundary signal. Yet another version of sentence 2 is realized as consisting of two prosodic phrases. Characteristic features are an upstepping F0 pattern initially – interpretable as an extra emphasis for contrast on the first accent – and a relatively wide F0 range on the second accent accompanying a moderate pre-boundary lengthening (Figure 5). This also has the effect of favoring the interpretation with the boundary, even if it seems to be a more moderate phrasing signal.

## Intonation phrase (synthesis)

In the synthesis part of the project work we are presently planning to investigate the specific contribution of intonation to phrasing through the use of perceptual testing. For this purpose we are devising test material consisting of a minimal sentence pair similar to test sentences 1 and 2 above. The reason for not choosing these two sentences for perceptual testing is the possibility of sentence 2 being produced as one coherent phrase not distinct from sentence 1 (see above). While in these sentences the distinctive feature is typically the absence vs. presence of an internal boundary, the characteristic difference in the new test material is instead the location of the internal boundary, either a grouping of 2 + 3 accents or 3 + 2 accents:

3.  FAST MAN OFFRADE BONDEN, Å LÖPAREN HÄLSADE KUNGEN.
    [ fast man "ɔfɹadə 'bɷndən ‖ ɔ 'løːpaɹən 'hɛlsadə "kəɳːən ]
    'but he sacrificed the pawn, and the bishop greeted the king'

4.  FAST MAN OFFRADE BONDEN Å LÖPAREN, HÄLSADE KUNGEN.
    [ fast man "ɔfɹadə 'bɷndən ɔ 'løːpaɹən ‖ 'hɛlsadə "kəɳːən ]
    'though he sacrificed the pawn and the bishop, the king greeted him'

As a starting point for the perceptual testing, our male, Stockholm Swedish informant recorded the material (sentences 3 and 4) three times. Figure 6 shows typical examples of sentences 3 and 4 respectively. It is clear that both tonal and temporal cues are combined to signal the difference in phrasing. For the present synthetic testing we are concentrating on the tonal contribution to phrasing. Based on our analysis of speech production
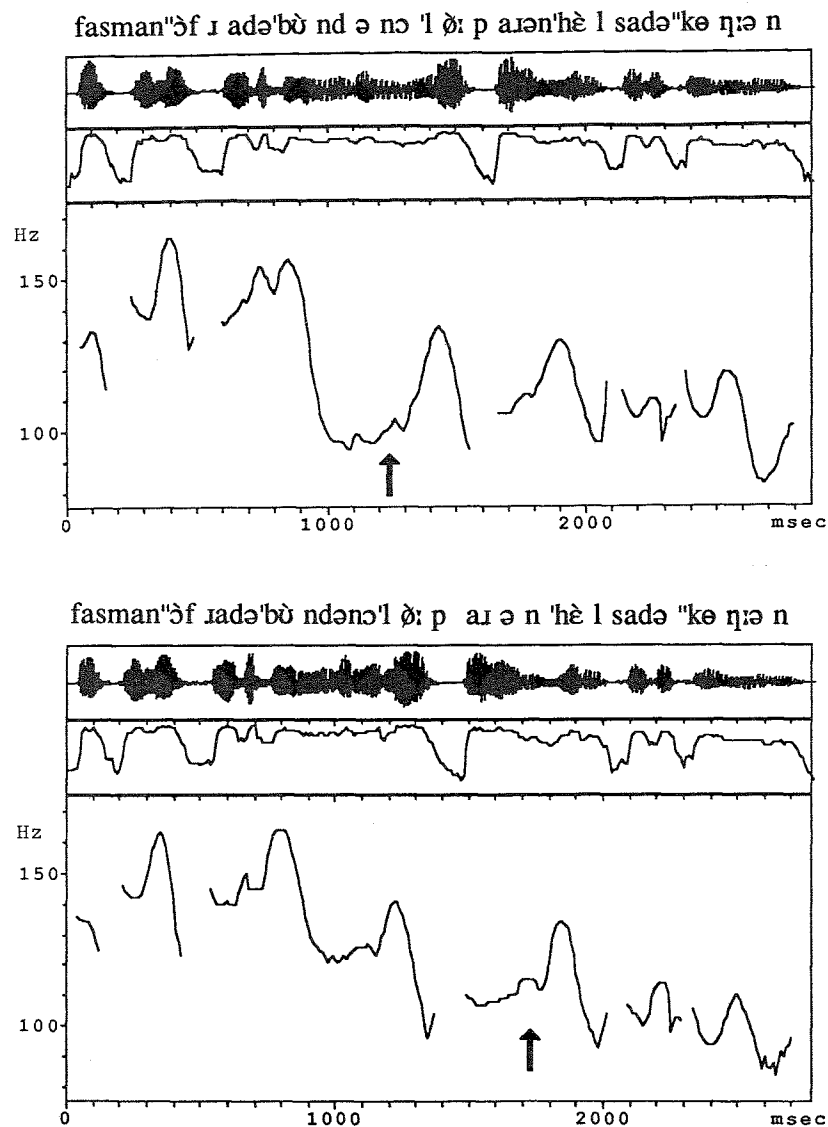
fasman"ɔf ɹ adə'bʊ̀ nd ə nɔ 'l ø̀ː p aɹən'hɛ̀ l sadə"kə ŋɪə n

fasman"ɔf ɹadə'bʊ̀ ndənɔ'l ø̀ː p aɹ ə n 'hɛ̀ l sadə "kə ŋɪə n

**Figure 6.** Waveform, intensity and fundamental frequency contour of a typical example of sentence 3 (upper panel) and of sentence 4 (lower panel). Phrase boundaries are indicated by arrows.
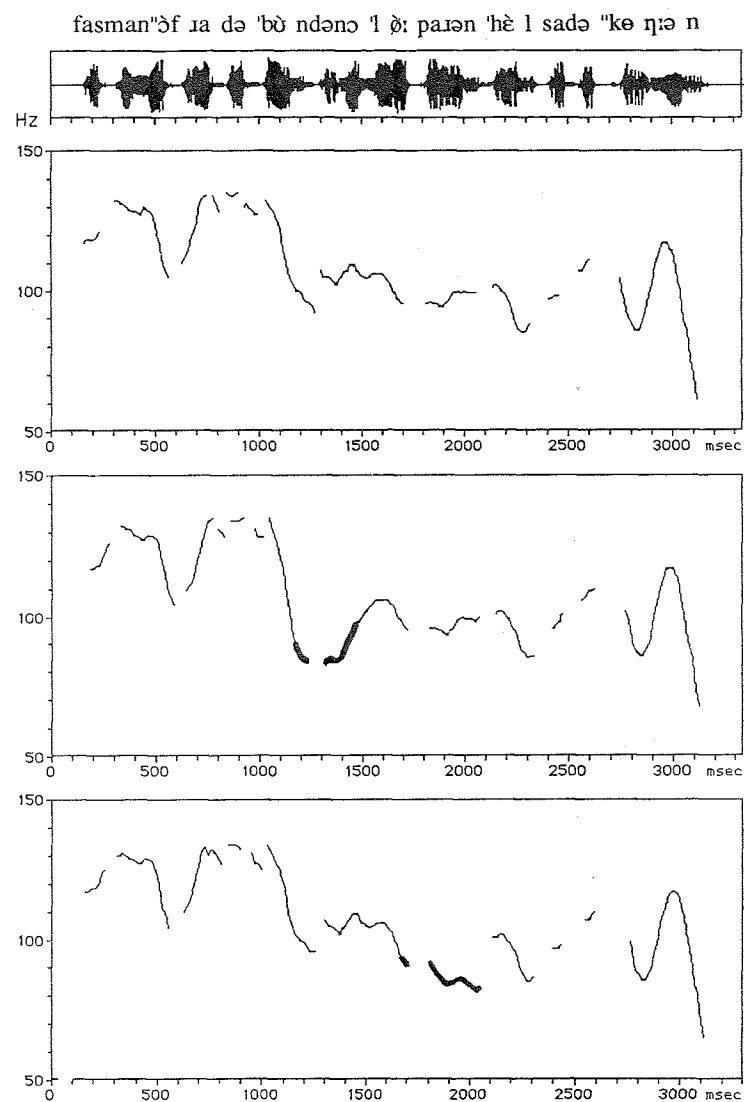
fasman"ɔf ɹa də 'bʊ̀ ndənɔ 'l ø̀ː paɹən 'hɛ̀ l sadə "kə ŋɪə n

**Figure 7.** Three synthetic versions of the sentence *Fast man offrade bonden och löparen hälsade kungen*, showing the default version (upper panel), the version with an intended boundary after *bonden* (middle panel), and the version with an intended boundary after *löparen* (lower panel). See text for details.

data, we will be testing the following hypothesis about tonal relations for phrasing: *F0 downstepping has a connective function signalling coherence within a phrase, while a break in the downstepping trend has a demarcative function signalling a phrase boundary*. As part of our preliminary, informal testing we have set up three distinct versions of the test sentence through the use of RULSYS (the KTH rule synthesis). The three versions, which are temporally identical, are shown in Figure 7. One default version (upper part) appears to be ambiguous between sentence 3 and 4. Another version (middle part) has the intended phrasing of sentence 3; the boundary cue is a lower (by 2 semitones) and slightly longer F0 minimum at *bonden* relative to the default. A third version (lower part) has the intended phrasing of sentence 4; the boundary cue here is a lower (by 2 semitones) and longer F0 minimum at *löparen* relative to the default.

Preliminary, informal perceptual testing suggests that the tonal relations of a downtrend – especially a break in the downtrend – are critical for the impression of phrasing (grouping) but *not* critical for the impression of differences in accentuation (prominence relations). The hypothesis above concerning intonational phrasing will be tested in a formal perception experiment using different synthesized versions of the test sentences (with phrasing as in sentences 3 and 4). Work in progress has also revealed similar strategies of phrasing used by the same and a different speaker in longer text passages. These strategies are also being tested in synthetic versions of longer texts where the strategies can be denoted by different symbols in the transcribed text.

## Acknowledgment

## References

Bruce, G. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Bruce, G. 1985 'Structure and functions of prosody'. In B. Guerin & R. Carré (eds.), *Proceedings of the French Swedish Seminar on Speech*, 549-559. Grenoble.

Bruce, G. & E. Gårding. 1978. 'A prosodic typology for Swedish dialects'. In E. Gårding et al. (eds.), *Nordic Prosody*, 219-228. Department of Linguistics, Lund University.

Bruce, G., B. Granström & D. House. 1990. 'Prosodic phrasing in Swedish speech synthesis'. In *Proceedings of the ESCA workshop on speech synthesis*, 125-128. Autrans (Grenoble), France.

Bruce, G., B. Granström & D. House. 1991. 'Strategies for prosodic phrasing in Swedish'. In *Proceedings of the Twelfth International Congress of Phonetic Sciences*, 4:182-185. Aix-en-Provence, France.

Carlson, R. & B. Granström. 1973. 'Word accent, emphatic stress, and syntax in a synthesis by rule scheme for Swedish'. *Speech Transmission Laboratory - Quarterly Progress and Status Report (STL-QPSR)* 2-3/1973, 31-36. Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm.

Carlson, R. & B. Granström. 1986. 'A search for durational rules in a real-speech data base'. *Phonetica* 43, 140-154.

Carlson, R., B. Granström & S. Hunnicutt. 1991. 'Multilingual text-to-speech development and application'. In W. Ainsworth (ed.), *Advances in speech, hearing and language processing*, 269-296. London: JAI Press.

Carlson, R., B. Granström & L. Nord. 1989. 'The KTH speech data base'. In *Proceedings of the ESCA workshop on speech input/output assessment*, 1.3.1-1.3.4. Phonetics Institute, Amsterdam.

House, D. 1990. *Tonal Perception in Speech*. Lund University Press.

House, D. & G. Bruce. 1990. 'Word and focal accents in Swedish from a recognition perspective'. In K. Wiik & I. Raimo (eds.), *Nordic Prosody V*, 156-173. Turku University.

Nespor, M. & I. Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Pierrehumbert, J. & M. Beckman. 1988. *Japanese Tone Structure*. Cambridge, Mass: The MIT Press.

Selkirk, E. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass: The MIT Press.