Coherence among prosodic phrases

Petra Hansson

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

E-mail: petra.hansson@ling.lu.se

ABSTRACT

In the present paper, we investigate whether decreasing phrase-initial accent peaks over the course of several prosodic phrases is used to signal coherence among phrases in spontaneous Swedish. Since little is known about prosodic phrasing and discourse prosody in spontaneous Swedish, the investigation has an exploratory nature and a number of different methods have been employed in order to answer the research question at hand.

1. INTRODUCTION

In the present paper, we follow up on the finding that speakers do not adjust the F0 starting point in order to accommodate for the upcoming phrase's length in spontaneous Swedish [3]. In other words, speakers do not start longer prosodic phrases higher than short phrases in order to ensure that F0 does not drop below the speaker's F0 floor, or to ensure that some constant or minimum F0 slope can be produced. Nevertheless, measurements of phrase-initial accent peaks reveal a substantial amount of variation. Given that this variation cannot be accounted for by phrase length, i.e. by a feature of the prosodic phrase, it seems reasonable to look for an explanation in the discourse. The present paper is dedicated to investigating tonal coherence among prosodic phrases. 'Phrasal downstep' (downward scaling of a prosodic phrase's pitch range or register relative to the preceding phrase) [5] or 'tonal coupling' [1] provides a possible explanation to some of the variation in our data.

1.1 RESEARCH QUESTION

The research question to be addressed in this study is whether F0 is used in spontaneous speech to signal coherence among prosodic phrases. To answer this question, we need to investigate both acoustic aspects of prosodic phrases in spontaneous speech and the perception of prosodic phrase boundaries between and within the units that demonstrate tonal coherence.

2. METHOD

2.1 SPEECH MATERIAL

The speech material investigated was the speech of ten male and five female subjects who represent the southernmost region of Sweden (*Skåne* 'Scania') in the research project *SweDia 2000*'s public database. The speech sections, approximately one minute long, are typically short anecdotes dealing with the older subjects' youth and descriptions of the young subjects' work.

2.2 PROCEDURE

We attempted to get an insight into the tonal organization of prosodic phrases in discourse by using several methods. Firstly, we recorded a measure that can be expected to vary over the course of the larger speech unit to give the impression of tonal coherence among prosodic phrases: the prosodic phrase-initial accent peak. Each prosodic phrase's position in the audio file was also recorded. Prior to the labeling of the phrase-initial accent peaks, a prosodic segmentation into prosodic phrases was done interactively using the speech analysis program ESPS/Waves+^{TM}. Listening to the speech and observing the F0 tracings, the phrase-initial accent peak in each prosodic phrase's F0 contour were then marked in a label tier. The F0 values of the accent peaks were then extracted automatically using the labels placed in the label tier.

Since little is known about prosodic phrasing and discourse prosody in spontaneous Swedish, the investigation had an exploratory nature and a number of different methods were employed in order to answer the research question at hand. In the first stage of the study, we used the measurements of the phrase-initial accent peaks to group the prosodic phrases in the material into sequences of downstepped /tonally coupled phrases. Secondly, we looked for evidence of perceived tonal coherence within groups of prosodic phrases formed on the basis of the F0 values of the phrase-initial accent peaks. We used the labeling of boundary strength made by two expert transcribers (experienced phoneticians) to determine whether boundaries between prosodic phrases contained in the same sequence of downstepped/tonally coupled phrases in fact are perceived as weak and boundaries between prosodic phrases pertaining to different sequences as strong. Finally, we examined disfluent parts in the material (phrases containing phrase-internal pauses, disfluencies (speech repairs) and signs of syntactic reorganization) in order to get some insight into the amount of preplanning needed to produce tonally coherent sequences of prosodic phrases.

3. RESULTS AND DISCUSSION

First, an analysis of each speech unit separately was made by studying the F0 values of the phrase-initial accent peaks (more details in [3]). Let us examine one speech unit in more detail here (the speech of the young male speaker from Norra Rörum). This speech unit contains 17 prosodic phrases, although, due to vocal fry, we have no reliable records of the seventh and seventeenth phrase's F0 starting points. Traces of some sort of tonal coherence signaling among the phrases can be seen. Three sequences of downstepped phrases can be observed; the six (or seven) first prosodic phrases pertain to one such sequence, the next four (or five) prosodic phrases to another, and the last four to a third sequence of downstepped phrases. It is difficult to say to which sequence of phrases the sixteenth phrase should be grouped since we have no reliable record of the following prosodic phrase's F0 starting point, but given its high peak, it is likely that phrases number sixteen and seventeen constitute a downstepped sequence on their own (or at least the beginning of such a sequence).

Listening to the speech, and comparing our grouping of the prosodic phrases described above (which is based only on the F0 values of the phrase-initial accent peaks) with the boundary strengths perceived by the expert transcribers, we can get an insight into the perceptual relevance of the observed decrease in F0 between successive phrases. It is not difficult to see the similarities between the grouping of phrases made above based on the initial accent peaks' F0 values and the transcribers' transcriptions. The seventh prosodic phrase (which we were not able to group due to the lack of measurements) is perceived to cohere with the first six, and after it, a strong boundary is perceived by both transcribers. As regards the boundaries after the next two stipulated sequences of phrases (4 + 4 phrases), the transcribers show some disagreement; both perceive the boundaries in question but one transcribed them as strong and the other transcribed them as weak.

A comparison between the expert transcribers' labeling of boundary strength indicate that agreeing on boundary strength is more difficult than agreeing on whether or not a given word is followed by a boundary or not. A closer review of the expert transcribers' labeling also reveals that there is no simple relationship between the perceived strength of a boundary and the degree of tonal coherence existing between the prosodic phrases on either side of the boundary (as measured by the extent and direction of the difference in phrase-initial accent height). It is not the case that tonally coherent prosodic phrases are always perceived to be separated by weak (prosodic phrase) boundaries. Nor is it the case that phrases between which the starting point is reset are always perceived as divided by strong (prosodic utterance) boundaries. This is most likely due to the presence or lack of other strong boundary signals that modify or function to complement the degree of perceived boundary strength. Pauses are likely to be one such strong boundary signal. The presence of a pause can be expected to modify and decrease the perceived degree of coherence between two prosodic phrases, whereas the lack of a pause between two phrases probably increases the amount of coherence perceived between the phrases. Note that we are assuming that the degree of perceived boundary strength and the degree of perceived coherence across the boundary are related.

Conflicts between strong boundary signals (such as a long pause) and coherence signaling cues (such as a decrease in the F0 values of initial accent peaks) often arise in spontaneous speech. By 'conflict' we mean that the boundaries in question are not easily classified as to boundary strength within the Swedish base prosody system due to the presence of both coherence and boundary signaling cues. However, the possibility to use conflicting cues causes no problems in the speech situation, but should rather be seen as an asset to the speaker. The speaker can, e.g., make an addition to what (s)he has said although its end has already been marked by a pause. (S)he can do this by making the addition cohere intonationally with the preceding speech, see Figure 1. Here the addition *på vintrarna* 'in the winters' follows a pause and the steep F0 fall to a low end point in the focally accented phrase-final word *säd* 'grain'. It is nevertheless clearly understood as an addition to the preceding utterance rather than as the beginning of a new utterance (which is also possible syntactically). The fact that the F0 decrease of phrase-initial accent peaks between *på vintrarna* and the preceding phrase is similar to the F0 decrease occurring between the parenthetical expression *loftet* 'the loft' and its preceding phrase, as well as between the continuation *där hade vi ju* 'there we had' and the parenthetical *loftet*, indicates to us that intonational coherence signaling can be achieved without lookahead. It seems reasonable to assume that the speaker had a continuation planned when inserting the parenthetical expression. Unlike the addition *på vintrarna*, the planned continuation shows several signs of prosodic coherence with the preceding speech (e.g. a relatively short pause). However, as regards the tonal coherence signaling among the phrases, no evident differences can be found between how the preplanned continuation is made to cohere with the preceding speech and how the postplanned addition is made to cohere.

Analyses of additions like the one above, are interesting from a planning point of view. How are postplanned additions incorporated into the preceding speech? Is the speaker able to make it cohere intonationally in a way similar to prosodic phrases occurring in more fluent parts of speech? If so, then we must assume that the tonal coherence signaling under investigation requires no lookahead.

There are several examples of adjunct elements that are set aside as separate prosodic phrases in the material with pauses preceding them. Like the pauses occurring at sentence boundaries [2], the pauses preceding these additions likely reflect the externalisation of a thought. The strong prosodic marking of the end of the phrase preceding the adjunct element including the presence of a pause indicate to us that these sentences were not entirely planned
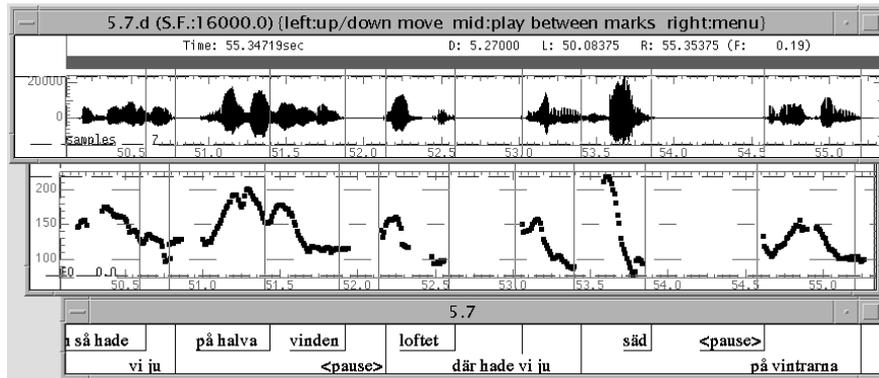
Figure 1: *Speech wave and F0 contour of* och så hade vi ju på halva vinden – loftet – där hade vi ju säd på vintrarna *'and we had in half the attic – the loft – there we had grain in the winters' (elderly female subject from Össjö).*
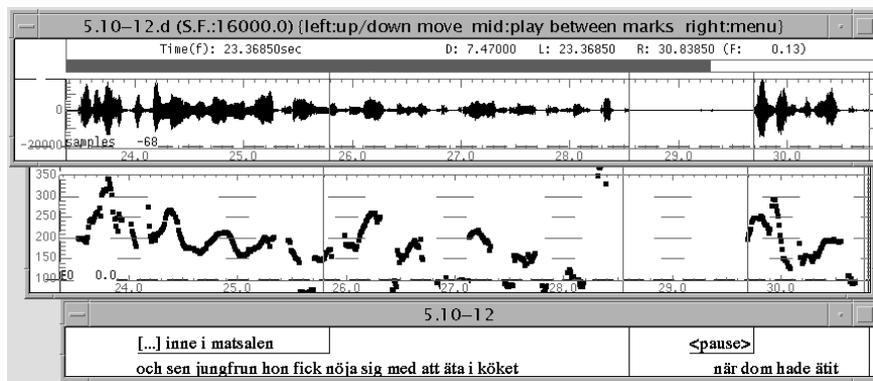
Figure 2: *Wave form and F0 contour of the three prosodic phrases* där åt ju familjen inne i matsalen *'the family ate in the dining-room',* och sen jungfrun hon fick nöja sig med att äta i köket *'and then the maid had to be satisfied with eating in the kitchen' and* när dom hade ätit *'when they had eaten' (elderly female subject from Broby).*

ahead. In many cases, the additions (e.g. prepositional phrases) are syntactically ambiguous in the sense that they could either belong to the preceding or following sentence. In the cases examined so far, prosody clearly helps the listener to disambiguate these structures i.e. to understand to what sentence the addition belongs. However, the plan-as-you-go basis of spontaneous speech and the apparent possibility speakers have to produce tonally coherent sequences of phrases without lookahead, open many possibilities for the speaker. The speaker may e.g. make changes to the original plan as he or she goes along, thereby producing syntactically "impossible" – but communicatively perfectly acceptable – structures. Let us examine an example in detail.

In (1), the speaker talks about the situation of maids in Sweden in her youth: they had to eat in the kitchen while the family got to eat in the dining-room. While the first prosodic phrase is separated from the following by a weak boundary, the end of the second phrase is more strongly marked prosodically. The two prosodic phrases are semantically coherent and syntactically coordinated and the presence of a weak boundary between the clauses and a stronger prosodic marking after them (the sentence boundary) is therefore both semantically and syntactically

motivated. However, the speaker chooses to make an addition (*när dom hade ätit* 'when they had eaten'). She makes clear that the maids had to wait to eat until after the family had eaten. The end of this prosodic phrase is also strongly marked prosodically (the new end of the sentence). The phrase is furthermore tonally coherent with the preceding. The strength of the preceding phrase boundary is thus presumably reevaluated by the listener, and the addition is perceived as a continuation, see Figure 2. The speaker then chooses to clarify the situation further, by adding the prosodic phrase *så fick hon äta resterna* 'then she got to eat the leftovers'. The temporal clause 'when they had eaten' can now be interpreted as either being a part of the preceding or the following sentence. Because the last clause in (1) cannot stand on its own syntactically, only one syntactic analysis is possible: the sentence boundary precedes the temporal clause. Prosodically, however, this analysis is not motivated: each prosodic phrase is a continuation of the preceding, and the temporal clause functions first to modify the first, and then the second sentence (or sentence-like unit).

(1)

$_1$där åt ju familjen inne i matsalen | $_2$och sen jungfrun hon fick nöja sig med att äta i köket | $_3$när dom hade ätit | $_4$så

fick hon äta resterna | *(elderly female subject from Broby)*

'₁the family ate in the dining-room | ₂and then the maid had to be satisfied with eating in the kitchen| ₃when they had eaten | ₄then she got to eat the leftovers |'

We believe that this flexibility of spontaneous speech reflect a plan-as-you-go basis. The tonal coherence that nevertheless can be observed and perceived among the phrases indicates that no lookahead is required to produce downstepped/tonally coupled phrases.

4. SUMMARY

Based on our observations of groups of prosodic phrases with uninterrupted decrease in F0 (observed here in the F0 starting points), we believe that intonation is also used to signal coherence within larger domains than the prosodic phrase in spontaneous Swedish. Whether the downward scaling of phrase-initial accent peaks and register in successive prosodic phrases should be regarded as the result of phrasal downstep, as suggested by van den Berg *et al.* [5], tonal coupling, as suggested by Bruce [1] or some other mechanism, we cannot say. Our main concern in the study reported on here was simply to determine if decreasing F0 over the course of several prosodic phrases is used to signal coherence in spontaneous speech.

By examining a number of groups of downstepped/tonally coupled prosodic phrases in more detail, we concluded that coherence is perceived among the tonally coherent phrases, and consequently that the lowering of phrase-initial accent peaks likely is perceptually relevant. Naturally, there may be other cues that coexist with the decrease in F0 that help, and perhaps are more important in the signaling of coherence among prosodic phrases. The present analysis does not allow us to draw any conclusions as to F0's unique contribution to the perceived degree of coherence. The fact that neither does lowering F0 starting points between prosodic phrases invariably give rise to the perception of coherence across the phrase boundary (i.e. a weak prosodic boundary) nor does resetting invariably give rise to the perception of a strong boundary between the phrases, suggests the presence of other cues that also affect the degree of perceived coherence.

5. IMPLICATIONS

What implications does the finding of tonal coherence signaling among prosodic phrases have on our conception of the amount of preplanning involved in prosodic phrasing in spontaneous speech? Liberman and Pierrehumbert [4] make a distinction between what they term 'hard' and 'soft' preplanning. Hard preplanning refers to such processing that needs to be accomplished prior to the execution of the phrase, while soft preplanning refers to the sort of planning that the speaker may choose to make prior to the execution of the phrase, but can omit.

The adjustment of F0 slope to phrase length as understood and modeled in the original Lund model for intonation requires hard preplanning. An obligatory adjustment of the F0 starting point to phrase length, as modeled in the revised Lund model, would also require some lookahead and hard preplanning. However, no evidence to suggest that our speakers do such hard preplanning could be found in the data. That does not mean that the speakers are prevented from planning ahead. What is important here, however, is the finding that they do not always plan ahead, or even frequently make these kinds of adjustments that require lookahead. When adjustments are made, they must be seen as the result of optional – soft – preplanning.

Thus, although the lack of evidence of lookahead is not evidence against hard preplanning, we nevertheless believe that preplanning in spontaneous speech involves no hard preplanning. Some support against hard preplanning can be found in our analyses of postplanned additions. We believe that they demonstrate that well-formed tonal coherence signaling is produced by speakers even in utterances where the speaker plans his or her speech on a plan-as-you-go basis. That the speaker would have "the choice of continuing a downdrift" in spontaneous speech was hypothesized already in Bruce [1: 285].

REFERENCES

[1] G. Bruce, "Textual Aspects of Prosody in Swedish", *Phonetica* 39, pp. 274-287, 1982.

[2] F. Goldman-Eisler, "Pauses, Clauses, Sentences", *Language and Speech* 15, pp. 103-113, 1972.

[3] P. Hansson, "Prosodic Phrasing in Spontaneous Swedish. Travaux de l'institut de linguistique de Lund 43", 2003.

[4] M. Liberman and J. B. Pierrehumbert, "Intonational Invariance under Changes in Pitch Range and Length", in: M. Aronoff and R. T. Oehrle, eds, *Studies in Phonology Presented to Morris Halle by His Teacher and Students*, pp. 157-233. Cambridge, Massachusetts: MIT Press, 1984.

[5] R. van den Berg, C. Gussenhoven and A. C. M. Rietveld, "Downstep in Dutch: implications for a model", in: G. J. Docherty and D. R. Ladd, eds, *Papers in Laboratory Phonology. Gesture, Segment, Prosody*, pp. 335-359. Cambridge: Cambridge University Press, 1992.