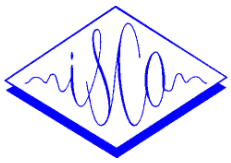

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

# ANALYSIS AND SYNTHESIS OF DIALOGUE PROSODY

*Gösta Bruce and Paul Touati*

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

## ABSTRACT

The object of study is the prosody of spontaneous dialogue. The focus is on the methodology that we are developing, and the exemplification is from Swedish. We have been conducting three types of analysis: analysis of dialogue structure, auditory (prosodic) analysis, and acoustic-phonetic analysis. Dialogue structure analysis concerns textual, interactive and turn taking aspects. The auditory analysis takes the form of a transcription encoding five prosodic features: prominence, phrasing, pitch range, boundary tones and pausing. The acoustic-phonetic analysis is centered around pitch, particularly the use of overall pitch range for the expression of textual coherence and boundary of a dialogue. For the modelling of dialogue prosody we have been using rule synthesis. Our preliminary testing shows that variation in pitch range is a potentially important means for signalling textual aspects of a dialogue.

## INTRODUCTION

The study of prosody in spontaneous dialogue represents a new area of research at the Department of Linguistics and Phonetics in Lund. This study coincides with the beginning of a research project called CONTRASTIVE INTERACTIVE PROSODY (KIPROS), which started in 1988 and is supported by the Bank of Sweden Tercentenary Foundation [1]. The object of study is dialogue prosody in a contrastive perspective in French, Greek and Swedish. The ultimate goal of the project is to develop a model for French, Greek and Swedish interactive prosody.

Two important general questions that we hope to find an answer to in the project work are the following: 1) Do we find the same, well-known prosodic patterns in spontaneous dialogue as we have met earlier in read, laboratory speech? 2) How are the prosodic patterns observed related to dialogue structure and interactive categories? The first question relates to our "old" research tradition in prosody and the general model of prosody we have been developing in Lund [2] [3] [4] [5] [6].

Our research on prosody in a spontaneous speech framework will give us an indication of how well we have been able to simulate natural prosody in a laboratory speech environment. The second question is related to the "new" research setting for our study of prosody: spontaneous speech and dialogue. What are the factors that govern the specific choice of prosodic patterns for the speakers involved?

Our research strategy so far in the project work has been to study a fairly restricted sample of speech material in relative depth and from different angles. The choice of dialogue type for our study of prosody has been governed by a number of different criteria which we have discussed in earlier papers [1] [7].

The focus of the present report will be on the methodology that we are developing in our study of dialogue prosody, and the exemplification will be taken exclusively from Swedish.

## ANALYSIS

We have been conducting three different kinds of analysis: 1) analysis of the dialogue structure itself without specific reference to prosodic information, 2) auditory analysis in the form of a prosody-oriented transcription, and 3) acoustic-phonetic analysis centered around the examination of pitch.

### Analysis of dialogue structure

We have been considering three different aspects of dialogue structure which we have found reason to keep apart in our analysis:

*Textual aspects* pertain to the development of a dialogue as a text, which may involve one or more speakers. Specifically we are thinking of the segmentation of a dialogue into different 'speech paragraphs', each of which has a certain coherence from the point of view of topic structure.

By *interactive aspects* we refer specifically to the analysis of a dialogue as to how it is carried on in terms of the initiatives (actions) and responses

(reactions) taken and given by the speakers involved. This kind of analysis is comparable to a more traditional one into speech act categories such as questions and answers.

*Turn taking aspects* refer to the specific regulation of the speakers' turns in a dialogue, such as taking, keeping, inviting to, and yielding the turn.

In the present paper we will focus particularly on the textual aspects of dialogue and their possible connection to prosody.

## Auditory (prosodic) analysis

Our auditory (prosodic) analysis is kept distinct from the analysis of dialogue structure and interactive categories. Therefore, our prosodic transcription does not contain categories such as question intonation, continuation tone etc. It is only at a later stage, when we are relating the auditory prosodic analysis - as well as the acoustic-phonetic analysis - to the analysis of the structure of the dialogue itself, that we may establish such potential categories.

In earlier reports we have discussed some basic principles of transcription in general and specifically prosodic transcription [8]. In the following, the transcription system will be presented, and we will exemplify how the system can be used to transcribe the prosody of Swedish. Basically it is an orthographic transcription, to which are added prosodic features selected from our model of prosody. While it does not contain potentially very interesting features such as change in speech tempo, loudness and voice quality, our system does encode five prosodic features: accentual prominence, phrasing, pitch range, boundary tones and pausing. Our notation is basically phonological, and the symbolization is as far as possible in accordance with the new, current IPA system 1989 [9].

**Prominence.** The analysis of prominence levels was made in terms of three binary features: 1) The lowest level of prominence (apart from unstressed), mere stress, coded [ˌx] as in a so-called secondary stress position of a compound, 2) A higher level of prominence, word accent, coded ['x], and equivalent to the primary stress position of a word, where in Swedish either of the two word accents (accent I or accent II) is manifested, 3) The highest level of prominence at the phrase or utterance level, focal accent, coded ["x]. The symbolization of accent I is left unmarked and is implied by the use of the symbols ['x] or ["x], while accent II has the additional symbol [ˋ] above the relevant syllable nucleus.

**Phrasing.** In the analysis of prosodic phrasing in Swedish, we have so far only recognized one kind of prosodic phrase boundary [‖]. It is not unlikely, however, that we will need two kinds of prosodic group boundary for a more complete treatment of prosodic grouping in Swedish.

**Pitch range.** We have devoted special attention to the variation in the general pitch of successive prosodic phrases in a dialogue. As a starting point, we assume an auditory normalization for inter-speaker variation depending on speaker size (age, sex, voice register).

In Swedish, overall pitch range for a prosodic phrase was analyzed syntagmatically in relation to the neighbouring phrases and may assume five different values: [→] = same, [↗] = slightly raised, [↑] = markedly raised, [↘] = slightly lowered, [↓] = markedly lowered. The location of an arrow is at the beginning of each prosodic phrase.

Our notation of pitch range thus represents a fairly narrow phonetic transcription, as this has been in the focus of our attention. For a broader transcription and one that is basically phonological, we would suggest instead of five different values only two values: raised versus non-raised pitch range [10].

**Boundary tones.** Within a prosodic phrase and for a given pitch range, initial and final boundary tones are judged to be either raised (marked value = [˩ ] ) or non-raised (unmarked). This means that the range of, for example, a final pitch rise, notated as a high boundary tone, can vary considerably but still be transcribed as the same category.

**Pausing.** In our transcription system we think of prosodic phrasing and pausing as distinct categories. Boundaries between prosodic phrases can of course be marked without any real pause, although a pause may also be present there and accompany the prosodic grouping. In the transcription of Swedish dialogues, we have assumed that where a real pause is perceived, two degrees of pause length are noted: short [(.)], and long [(..)].

**Exemplification.** The prosodic transcription used here for Swedish is exemplified below. The particular Swedish dialogue that we have chosen for our study is a recording from a popular radio program "Ring så spelar vi", which is a radio listener's conversation over the telephone with the program leader:

B ↗ do dom 'fråga om du "hàde ci'vila 'kläder ||
A → jo dom frå "frågade mej 'de ||
B ↘ å du sv be"jakade 'frågan ||
A → därför att
A → 'jo ja (.) ja va 'nàmligen hade se"mester den 'här 'vèckan 'nu å ˌ || (.)
↘ och 'därför så va ja ci"vìl,klädd på "ròtarysamman,trädet i'går ||
B → 'ja så de 'blev en 'smärre "chock för 'dèl,tagarna ||
A → "ja de 'blev de || (..)
B ↑ "ja 'Bert ||(..) → "så 'e de ||
A → mm ˌ ||
B ↘ nu ska du 'vìnna en 'skìva || (.)
A → nja de e 'ìnte 'säkert ||

Simultaneous speech is marked by underlines.

### Acoustic-phonetic analysis

The acoustic-phonetic analysis that we have undertaken so far has been centered around pitch. A first part of this analysis consists in isolating relevant pitch patterns for accentuation, phrasing, boundary signalling and pitch range, where an intermediary phonological (or abstract phonetic) representation in terms of H(igh) and L(ow) turning points has proved helpful. In the present section we will focus specifically on variation and changes in overall pitch range.

A fundamental question concerns how the actual variation in F0 range is accomplished. In a study of variation in F0 range as an expression of speaker involvement in read, laboratory speech, it was shown that the 'floor' of the speaker's voice range is basically constant, while the peak values of the F0 contour are highly variable with speaker involvement (involved versus detached) [11]. That F0 peaks vary, while valleys do so much less, with variation in pitch range thus constitutes our expectation also for spontaneous dialogue.

Furthermore, variation in overall pitch range (for the domain of a phrase) can be assumed to be exploited for interactive purposes. Differing degrees of attention generally seem to correlate with variation in range. Another more specific hypothesis has been to ascribe variation in pitch range a possible connection with boundaries in the dialogue structure, for example to speech paragraphs or to the introduction of a new conversation topic [10] [12].

For Swedish we have observed the exploitation of variation in pitch range, specifically a more or less successive decrease in range, for the expression of textual coherence within a larger speech paragraph consisting of several successive prosodic phrases and a shift in pitch range (usually a widening in range) for the marking of boundaries in the dialogue structure. An apparent example of this use of variation in pitch range - a successive decrease in range during a speech paragraph produced in interaction between the two speakers and then an extra wide range for rounding off the paragraph followed by a new decrease in range for the new speech paragraph - in a section of the Swedish dialogue transcribed above is found in Figure 1.

However, our experience from studying dialogue prosody so far indicates that the connection between variation in pitch range and sectioning of a dialogue is probably best regarded as a tendency rather than as an obligatory feature.

## SYNTHESIS

An important and powerful method in our modelling of dialogue prosody and particularly the exploitation of pitch range is analysis-by-synthesis. The research tool which we have been using is the multilingual text-to-speech system developed at KTH, Stockholm [13]. The prosody rules of the Swedish text-to-speech system have recently been modified [14]. There are still, however, several limitations for its exploitation in the specific study of dialogue prosody and in simulating spontaneous speech in interaction, so that at the present stage several typical ingredients of spoken dialogue could not be implemented in the syntheses. In spite of these limitations we have found that rule synthesis can be a valuable instrument in dialogue prosody research.

The speech synthesis used here allows one to choose from a small set of speaking voices. Two different voices have been selected for participating in our simulated dialogue. In our use of rule synthesis, the starting point is a phonetic transcription of prosodic features, basically the same features as described above under auditory (prosodic) analysis.

Different versions of a dialogue section have been implemented in the rule synthesis. Two versions of our synthesis attempts are interesting for the present discussion. The first one is a neutral version of the dialogue section, where only default utterance prosody is used with no attempt to simulate interaction. Thus the same pitch range is used for the consecutive prosodic groups of the dialogue section.

The second version presents an attempt - in addition to the neutral utterance prosody - to simulate one aspect of dialogue prosody, namely the variation of pitch range for interactive purposes.

A comparison of the two synthesized versions - the neutral version and the pitch range version -

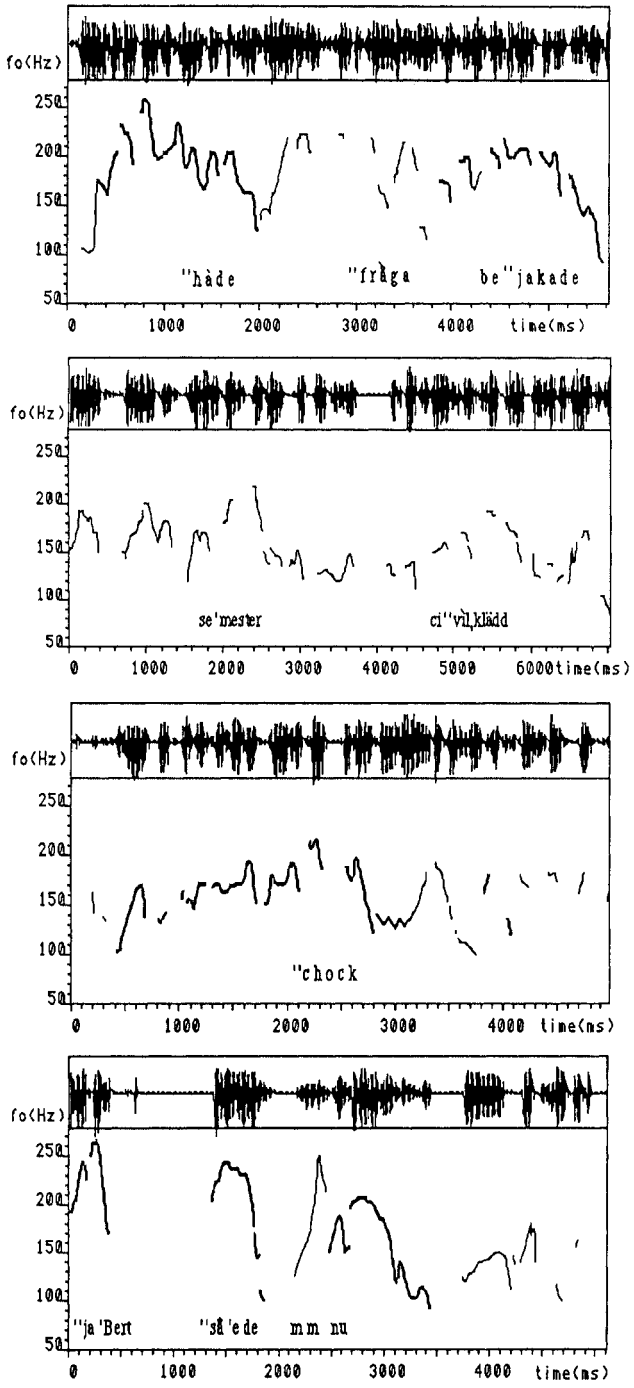


***Figure 1.*** *Pitch range variation in a section of the Swedish dialogue above; waveform (above) and pitch contour (below). Key words are aligned with important pitch events. Pitch contour of speaker A is drawn in thick lines and speaker B in thin lines.*

clearly shows that variation in overall pitch range may be considered a potentially important means for use in the development of a dialogue and its division into speech paragraphs [7].

## REFERENCES


[1] Bruce, G., Touati, P., Botinis, A., and Willstedt, U. 1988. 'Preliminary report from the KIPROS project'. *Working Papers 33,* 23-50. Lund: Dept. of Linguistics.

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

[3] Bruce, G. and Gårding, E. 1978. 'A prosodic typology for Swedish dialects'. *Nordic Prosody* , eds. E. Gårding et al., 219-228., Lund: Dept. of Linguistics.

[4] Gårding, E. 1982. 'Swedish prosody'. *Phonetica 39,* 288-301.

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

[6] Touati, P. 1987. *Structures prosodiques du suédois et du français.* Lund: Lund University Press.

[7] Bruce, G., Willstedt, U. and Touati, P. 1990. 'On Swedish interactive prosody: analysis and synthesis'. *Nordic Prosody V* , eds. K. Wiik and I. Raimo, 36-48. University of Turku: Phonetics.

[8] Bruce, G. and Touati, P. 1990. 'On the analysis of prosody in spontaneous dialogue'. *Working Papers 36,* 33-51. Lund: Dept. of Linguistics.

[9] I.P.A. 1989. 'Report on the 1989 Kiel Convention'. *Journal of the International Phonetic Association 19 (2),* 67-80.

[10] Brown, G., Currie, K. and Kenworthy, J. 1980. *Questions of intonation.* London: Croom Helm.

[11] Bruce, G. 1982. 'Developing the Swedish intonation model'. *Working Papers 22,* 51-116. Lund: Dept. of Linguistics.

[12] Hirschberg, J. and Pierrehumbert, J. 1986. 'Intonational structuring of discourse'. *Proceedings of the 24th Meeting of the Association of Computational Linguistics,* 136-144. New York.

[13] Carlsson, R. and Granström, B. 1986. 'Linguistic processing in the KTH multilingual text-to-speech system'. *In Proc. ICASSP 86, Vol.4,* 2403-2406. Tokyo.

[14] Bruce, G. and Granström, B. 1989. 'Modelling Swedish intonation in a text-to-speech system'. *STL-QPSR 1,* 17-21, Stockholm:KTH, Speech Transmission Laboratory.