

THE EXPLOITATION OF PITCH IN DIALOGUE

G. Bruce

Department of Linguistics and Phonetics, Lund, Sweden

ABSTRACT

Our study concerns the prosody of spontaneous dialogue centered around the examination of pitch and with exemplification from Swedish. Here we deal with the methodology of this research and also present some results in summary. We have undertaken four types of analysis: analysis of dialogue structure, auditory (prosodic) analysis, acoustic-phonetic analysis, and analysis-by-synthesis. Typically, the same pitch patterns that we have met in read, laboratory speech tend to occur also in spontaneous dialogue. Variation in overall pitch range and its relation to categories of dialogue structure is discussed and found to be a potentially important means for use in the sectioning and development of a dialogue.

1. INTRODUCTION

The present paper reports on and summarizes our current research on the prosody of spontaneous dialogue conducted at the Department of Linguistics and Phonetics at Lund. Our study of dialogue prosody is related to a research project called CONTRASTIVE INTERACTIVE PROSODY ('KIPROS'), which started in 1988 and is supported by the Bank of Sweden Tercentenary Foundation. 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 dialogue prosody. For recent reports from the project work see [8], [9], [10].

Two important general questions that we have been addressing 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], [4], [5], [13], [14]). 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?

In the present report we will deal with the methodology that we have been developing in our study of dialogue prosody and also present in summary some results from our work. The exemplification here will be taken exclusively from Swedish.

2. METHODOLOGY

Our research strategy in the project work has been to study a fairly restricted sample of speech material in relative depth and from different angles. We have been conducting four 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, 3) acoustic-phonetic analysis centered around the examination of pitch, and 4) analysis-by-synthesis by the use of text-to-speech.

2.1. 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 division of a dialogue into different 'speech paragraphs', each of which has a certain coherence from the point of view of topic structure. From this point of view the speakers' turns may be characterized as introducing, continuing on or terminating a certain topic.

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, receiving, keeping, and giving away the turn.

2.2. Auditory Analysis

The auditory analysis in terms of a prosodic transcription is kept distinct from the analysis of dialogue structure. 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.

Basically it is an orthographic transcription of what has been recorded. To this segmental transcription 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 with one exception fairly broad, and the symbolization is as far as possible in accordance with the new, current IPA system [16].

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 with no accent, coded [x], 2) A higher level of prominence, accented, coded ['x], 3) The highest level of prominence at the phrase or utterance level, focally accented, coded [''x].

Phrasing. In the analysis of prosodic phrasing we assume two types of boundaries: a minor phrase boundary for an accentual phrase [I] and a major phrase boundary, corresponding to a division into regular prosodic phrases [II].

Pitch range. Our notation of pitch range represents a fairly narrow phonetic transcription, as this has been in the focus of our attention. Overall pitch range for a major prosodic phrase has been analyzed syntagmatically in relation to the neighbouring phrases and may assume five different values: [→] = same [↗] = slightly raised, [↑] = markedly raised, [↘] = slightly lowered, [↓] = markedly lowered.

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 have assumed that where a real pause is perceived, two degrees of pause length are noted: short [(.)], and long [(...)].

Exemplification of our prosodically oriented transcription has been given in earlier reports (cf. [8], [9], [10]).

2.3. Acoustic-Phonetic Analysis

We consider the auditory analysis in terms of a prosody oriented transcription to be a useful basis for the acoustic-phonetic analysis of dialogue prosody: the qualitative and quantitative study of prosodic patterns from acoustic recordings of F0 and speech waveform. Our analysis has been centered around pitch. The standard procedure for us has been to have the recorded material digitalized on the VAX 11/730 at our laboratory and analyzed using the API program of the ILS package, where pitch extraction is

done with a modified cepstral technique. 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 (see e.g. [8]).

2.4. Analysis-by-Synthesis

An important and powerful method in our modelling of dialogue prosody and particularly the exploitation of pitch is analysis-by-synthesis. The research tool which we have been using is the multilingual text-to-speech system developed by Carlson and Granström [12]. The prosody rules of the Swedish text-to-speech system have recently been modified by Bruce & Granström [6], [7]. The idea is to use rule synthesis as a control instrument for checking the adequacy of our model of dialogue prosody and as a direct way of testing alternative analyses. 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, the so-called regular male voice and the deep male voice. 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 analysis.

3. RESULTS

3.1. Laboratory Speech vs. Spontaneous Dialogue

When studying the prosody of spontaneous dialogue against the background of having studied it in a laboratory speech environment, we have encountered relatively few surprises. Although we do not mean to underestimate the difference between read and spontaneous speech, it is our general

impression that the difference in prosodic patterning, particularly pitch patterns, between a specially designed, read test material and a spontaneous dialogue is less than we had expected. A typical example of the relative similarity between laboratory speech and spontaneous speech is the following.

The location of a focal accent in (Standard) Swedish represents a pivot (cf. [13]) of a prosodic phrase or utterance. The pivotal character of the focal accent in Swedish can be illustrated by its role in determining the presence or absence of a downstepping pitch contour in read speech material (cf. [3]). In a pre-focal position, up to the focal accent of a phrase (or a whole utterance) there is typically no downstepping, but instead successive non-focal accents occur on more or less the same pitch level. However, after a focal accent, the downstepping of successive non-focal accents is a characteristic pitch pattern. This downstepping seems to be the expression of equal prominence of successive post-focal accents within the phrase.

It is interesting to note that in our spontaneous dialogue speech there are several, typical examples of downstepping and non-downstepping pitch patterns, which seem to be triggered by the placement of focal accent in very much the same way as describe above. For a perspicuous example of this see [8].

3.2. Dialogue Structure and Pitch Range

A fundamental question in the study of dialogue prosody is of course how the prosodic patterns observed are related to the structure of the dialogue itself in terms of textual, interactive and turn regulating aspects. One case in point here is the variation and changes in overall pitch range, which has been in the focus of our interest. Differing degrees of attention generally seem to correlate with variation in range. A 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 (cf. [1], [15]).

In the particular Swedish dialogue that we have studied in some detail - a radio listeners' conversation over the telephone with the program leader of a popular radio

program "Ring så spelar vi" - the following regularities appear. The combined introduction of a new topic and interactive initiative is reflected by an increase in pitch range in 75 % of the cases. This can be contrasted with the combined continuation of a topic and an interactive response (which is a category four times as common in the actual dialogue), for which we find an almost equal distribution of increase, decrease and no change in pitch range.

3.3. Dialogue Prosody and Speech Synthesis

In our search for regularities of variation in overall pitch range we have also used speech synthesis. 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 of the actual dialogue section - the neutral version and the pitch range version - 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 (see further [11]).

4. REFERENCES

- [1] BROWN, G., CURRIE, K. and KENWORTHY, J. (1980), *Questions of intonation*, London: Croom Helm.
- [2] BRUCE, G. (1977), *Swedish word accents in sentence perspective*, Lund: Gleerup.
- [3] BRUCE, G. (1982), 'Developing the Swedish intonation model', *Working Papers 22*, 51-116, Lund: Dept. of Linguistics.
- [4] BRUCE, G. (1985), 'Structure and functions of prosody', *Proc. French Swedish Seminar on Speech*, eds. B. Guerin and R. Carré, 549-559, Grenoble.
- [5] 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.
- [6] 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.
- [7] BRUCE, G. and GRANSTRÖM, B. (1990), 'Modelling Swedish prosody in text-to-speech: phrasing', *Nordic Prosody V*, eds. K. Wiik and I. Raimo, 26-35, University of Turku: Phonetics.
- [8] BRUCE, G. and TOUATI, P. (1990a), 'On the analysis of prosody in spontaneous dialogue', *Working Papers 36*, 37-55, Lund: Dept. of Linguistics.
- [9] BRUCE, G. and TOUATI, P. (1990b), 'Analysis and synthesis of dialogue prosody', *Proc. ICSLP 90 Vol. 1*, 489-492, Kobe, Japan.
- [10] BRUCE, G. and TOUATI, P. (1991), 'On the analysis of prosody in spontaneous speech with exemplification from Swedish and French', To appear in *Proc. ESCA Workshop on Phonetics and Phonology of Speaking Styles*, Barcelona.
- [11] BRUCE, G., WILLSTEDT, U. & 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.
- [12] CARLSON, R and GRANSTRÖM, B. (1986), 'Linguistic processing in the KTH multilingual text-to-speech system', *In Proc. ICASSP 86, Vol. 4*, 2403-2406, Tokyo.
- [13] GÅRDING, E. (1981), 'Contrastive prosody: a model and its application', *Studia Linguistica 35*, 146-166.
- [14] GÅRDING, E. (1982), 'Swedish prosody', *Phonetica 39*, 288-301.
- [15] HIRSCHBERG, J. and PIERRE-HUMBERT, J. (1986), 'Intonational structuring of discourse', *Proc. 24th Meeting of the Association of Computational Linguistics*, 136-144, New York.
- [16] I.P.A. (1989), 'Report on the 1989 Kiel Convention', *Journal of the International Phonetic Association 19* (2), 67-80.